



Interpolation approximations for steady-state performance measures

Ane Izagirre

► To cite this version:

Ane Izagirre. Interpolation approximations for steady-state performance measures. General Mathematics [math.GM]. INSA de Toulouse; Universidad del País Vasco. Facultad de ciencias, 2015. English. NNT : 2015ISAT0019 . tel-01215869v2

HAL Id: tel-01215869

<https://theses.hal.science/tel-01215869v2>

Submitted on 26 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

l'Institut National des Sciences Appliquées de Toulouse (INSA de Toulouse)

Cotutelle internationale Université du Pays Basque

Présentée et soutenue le 21/09/2015 par :

Ane IZAGIRRE

**Interpolation approximations for steady-state performance measures
(Interpolation des mesures de performance à l'état stationnaire)**

JURY

CHRISTOPHE CHASSOT

KONSTANTIN AVRATCHENKOV

RUTESINDO NÚÑEZ - QUEIJA

DIETER FIEMS

ROB VAN DER MEI

FRANTZISKO XABIER ALBIZURI

URTZI AYESTA

INA MARIA VERLOOP

INSA Toulouse

INRIA Sophia Antipolis

Univ. of Amsterdam

Ghent Univ.

VU Univ. Amsterdam

Univ. of the Basque Country

LAAS-CNRS & Ikerbasque

IRIT-CNRS & ENSEEIHT

Président du Jury

Rapporteur

Rapporteur

Examinateur

Examinateur

Co-Directeur de Thèse

Directeur de Thèse

Directrice de Thèse

École doctorale et spécialité :

EDSYS : Systèmes embarqués 4200046

Unité de Recherche :

Laboratoire d'Analyse et d'Architecture des Systèmes (UPR 8001)

Directeur(s) de Thèse :

Frantzisko Xabier ALBIZURI, Urtzi AYESTA et Ina Maria VERLOOP

Rapporteurs :

Konstantin AVRATCHENKOV et Rudesindo NÚÑEZ - QUEIJA

Acknowledgments

I feel very lucky to have had the opportunity of carrying out the Ph.D thesis with Urtzi and Maaike, my supervisors. All my gratitude for their continuous support, patience and motivation; without their help this would not have been possible. I would also like to thank EDSYS for the Ph.D fellowship.

Xabierri, nire tesi zuzendari denari, eskerrik beroenak eman nahi dizkiot aldi oro laguntzeko erakutsi duen prestutasunagatik. Berari esker EHU-ko egonaldiak errazagoak izan dira.

The three-month visit to College Park, which was financially supported by EDSYS and INP, has been a valuable experience abroad. I am grateful to Armand Makowski for offering me this opportunity and also for the hospitality extended to me while I was in DC.

I would also like to thank all the members of the jury for accepting the invitation to attend my PhD defense.

Moreover, I wish to thank all the members of the SARA group at LAAS-CNRS laboratory and in particular, to my officemates Maialen, Samir and Tom, who have always been happy to help.

Modu berean, EHU-ko informatika fakultateko KZAA departamentuari eta hirugarren pixuko ikasleei eskerrak eman nahi dizkiet; bereziki Jonathan, Leti eta Mendi mahai inguruko tertulia atseginengatik. Mikel bulegokideari ere goizak atseginagoak egiteagatik. Eta azkenik, Martini urte hauetan zehar igarotako momentuengatik.

Bukatzeko, gurasoi eskerrak eman nahi dizkiet beraien animo eta konfindantza guztiagatik. Baita kuadrilako lagunei ere, eta bereziki Olatz eta Maddaleni krisi une horietan entzun eta laguntzeko prest egoteagatik. Eta azkenik Mikeli bihotz-bihotzez eskerrak, bera izan baita nirekin batera tesi hau hurbilen bizi izan duena eta egoerak bultzaturiko distantzia medio hurrin bezain hurbil sentitu dut.

Contents

1	Introduction	1
1.1	Queueing models	2
1.1.1	Single-server system	2
1.1.2	Parallel-server model	5
1.2	Motivation for interpolation approximation	6
1.3	Overview of the thesis	8
2	Interpolation approximation	10
2.1	Light-traffic analysis	11
2.2	Heavy-traffic regime	13
2.3	Light and heavy-traffic interpolation	14
3	Light-traffic analysis of the power-of-two policy	17
3.1	Model description	17
3.2	Mean sojourn time	19
3.3	Zeroth light-traffic derivative	21
3.4	An auxiliary result	22
3.5	First light-traffic derivative	25
3.6	Second light-traffic derivative	26
3.7	Simulation results	36
3.A	Appendix	40
3.A.1	Proof of Equation (3.18)	40
3.A.2	Two auxiliary calculations	40
3.A.3	Proof of Lemma 3.6.5	42
3.A.4	Proof of Lemma 3.6.6	43
3.A.5	Proof of Equation (3.84)	46
3.A.6	Proof of Equation (3.86)	47
4	Interpolation approximations for a discriminatory processor sharing queue	49
4.1	Model description	49
4.2	Preliminaries	50
4.2.1	Mean conditional sojourn time	50

4.2.2	Heavy-traffic results	52
4.3	Queue-length distribution	53
4.4	Waiting time distribution	56
4.5	Mean sojourn time	59
4.5.1	Mean conditional sojourn time	59
4.5.2	Mean unconditional sojourn time	63
4.5.3	Exponential service requirements	63
4.6	Numerical comparison	67
4.6.1	Queue-length distribution	69
4.6.2	Waiting time distribution	69
4.6.3	Mean sojourn time	71
4.A	Appendix	76
4.A.1	Proof of Lemma 4.3.1	76
4.A.2	Proof of Proposition 4.3.2	87
4.A.3	Proof of Lemma 4.4.1	89
4.A.4	Proof of Proposition 4.4.2	93
4.A.5	Light-traffic derivatives for the mean conditional sojourn time . . .	94
4.A.6	Proof of Corollary 4.5.3	98
5	Heavy-traffic analysis of a relative priorities queue	99
5.1	Model description	100
5.2	Queue length at departure epochs	101
5.2.1	Heavy-traffic scaling	102
5.2.2	Proof of Proposition 5.2.1	105
5.3	Queue length at arbitrary epochs	106
5.3.1	Heavy-traffic scaling	107
5.3.2	Proof of Proposition 5.3.1	108
5.4	Waiting time	110
5.5	Optimal selection of the weights	115
5.5.1	Holding cost	116
5.5.2	Moments of the waiting time	117
5.6	Numerical results	120
5.6.1	State-space collapse for the queue lengths	120
5.6.2	Moments of waiting time and queue length	120
5.6.3	Optimal values for the weights	123
5.A	Appendix	123
5.A.1	Proof of Lemma 5.2.3	123
5.A.2	Proof of Lemma 5.2.6	125
5.A.3	Proof of Lemma 5.4.4	126
5.A.4	Solution of the ODE (5.34)	127

6	Interpolation approximations for a relative priorities queue	129
6.1	Queue-length distribution	129
6.2	Waiting time distribution	132
6.3	Numerical comparison	133
6.3.1	Accuracy of interpolation approximation	134
6.3.2	RP versus DPS	134
6.A	Appendix	135
6.A.1	Proof of Lemma 6.1.2	135
6.A.2	Proof of Proposition 6.1.3	138
6.A.3	Proof of Lemma 6.2.2	140
6.A.4	Proof of Proposition 6.2.3	140
	Bibliography	142
	Annex A: Résumé de thèse en français	
	Interpolation des mesures de performance à l'état stationnaire	146
	Annex B: Tesia euskaraz	
	Errendimendu metriken interpolazio bidezko hurbilketak oreka ego- eran	167

Chapter 1

Introduction

Waiting in line is as common as unpleasant. We all wait impatiently to be served in the supermarket queue, in the hospital's emergency room or when calling a phone service provider. At a more abstract level, these queues are also encountered in telecommunication systems. For instance, every email you send or every file you download is broken up into different packets. Each packet is then sent to its destination by the best available route to avoid the queues formed by other packets.

Queueing theory is the mathematical theory that studies the aforementioned situations where queues are caused and it has two main goals. On one hand, to study the system's performance. On the other hand, to find the best policy in order to improve the performance. The number of customers in the system, also referred to as the queue length, and the waiting time, which is the time customers spend in the system minus their service requirement, are among the most frequently considered measures in the performance evaluation literature.

In this thesis we focus on the first goal and we investigate the performance of the multi-class single-server queue under the Discriminatory Processor Sharing policy and the Relative Priorities policy, and the parallel-server model under the power-of-two policy. We give further details on the models in Section 1.1.

Ideally we would like to characterise the performance metrics in their exact forms. However, due to the difficulty that this implies this is often out of reach. Therefore, queueing theory has shown a big interest in approximating the performance metrics under limiting regimes, such as, time-scale decomposition, tail asymptotics, heavy-traffic limits, fluid limits, etc. In this thesis we use the *light and heavy-traffic interpolation* approximation technique in order to derive closed-form approximations for the steady-state distribution of the queue length vector, waiting time and sojourn time. More precisely, first the performance is investigated in light traffic, that is, when the arrival rate tends to 0. Second, the performance is investigated in heavy-traffic, that is, when the system is near saturation. Then, the light-traffic and the heavy-traffic characterisations are combined in order to develop an interpolation approximation that aims at capturing the performance at any load. We motivate the technique in Section 1.2 and give further details in Chapter 2.

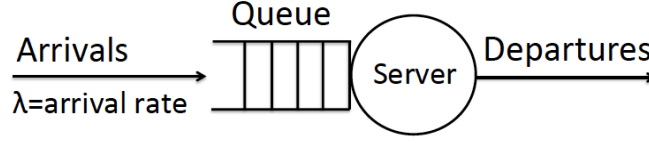


Figure 1.1: The single-server queue.

1.1 Queueing models

In this section we introduce the queueing models that we study in this thesis.

1.1.1 Single-server system

The classical single-server queue is described as in Figure 1.1. There is an arrival process of rate λ , so that λ^{-1} is the mean inter-arrival time between customers. Upon arrival, the customer will wait in the queue or is directly taken into service. This will depend on the applied scheduling policy. We assume that the capacity of the server is one. The service requirement is denoted by the random variable B . This is the time the customer will spend in the system if the server assigns its full capacity to that customer.

A common assumption is that the inter-arrival times are independent and identically distributed (i.i.d), the service requirements are i.i.d., and the sequences of inter-arrival times and service requirements are independent. This model is known as the $G/GI/1$ queue, where G stands for general distribution and GI for general and independent distributions. This notation was introduced by Kendall [29]. In this thesis we focus on the $M/GI/1$ queue, where M stands for Markovian or memoryless, that is, when customers follow a Poisson arrival process, or in other words, when the inter-arrival times are exponentially distributed.

The total workload at the system is independent of the work-conserving scheduling policy being used. A work-conserving system works at full speed whenever there is work in the system. Stability is also independent of the work-conserving scheduling policy being used. The queue will be stable as long as $\lambda\mathbb{E}[B]$ is strictly less than one. By contrast, the queue-length process depends on the policy employed.

The single-server systems considered in this thesis are multi class with K different classes of customers. Class- k customers arrive according to a Poisson process with rate λ_k and $\alpha_k := \lambda_k/\lambda$ denotes the fraction of class- k arrivals. The service requirement of a class- k customer is denoted by $B_k, k = 1, \dots, K$. Naturally, the traffic load of class- k customers is denoted as $\rho_k := \lambda_k\mathbb{E}[B_k]$, such that $\rho := \sum_{k=1}^K \rho_k$ is the total traffic load.

We will now present the work-conserving policies that play an important role in this thesis.

Processor sharing

Under the Processor Sharing (PS) policy the capacity of the server is equally shared between the customers in the system. More precisely, if at time t there are $n(t)$ customers

present in the system, under PS each customer is served at rate $1/n(t)$. We refer to the survey [57] and to [44] for a general overview of the literature.

In case of Poisson arrivals the stationary distribution of the number of customers in the system only depends on the service requirement distribution through its mean, and not through any higher-order statistics. More precisely, the steady state queue-length distribution has a geometric distribution of parameter ρ , i.e., the probability of being n customers in the system is equal to $(1 - \rho)\rho^n$, $n = 0, 1, \dots$, [52]. Because of Little's law, [39], the insensitivity of the queue-length distribution translates into insensitivity of the mean sojourn time.

In contrast to the simple geometric distribution of the queue length, the sojourn time distribution does not have any simple characterisation. Initiated by Kleinrock's analysis of the M/M/1 PS queue [35, 36], many studies in the literature have focused on the analysis of the conditional (on the service requirement) sojourn time. For results on the sojourn time distribution in the M/M/1 PS queue we refer to the summary in [6] and for results on the M/GI/1 PS queue to the survey papers [57, 58].

For a multi-class PS queue, the geometric distribution for the queue length holds as well. Under the Poisson arrival process assumption, as studied in [10, 27], the probability of having n_k class- k customers in the system, $k = 1, \dots, K$, is equal to

$$(1 - \rho) \cdot \frac{(n_1 + \dots + n_K)!}{n_1! \cdot \dots \cdot n_K!} \prod_{k=1}^K \rho_k^{n_k}.$$

The PS queue has gained a prominent role in evaluating the performance of a variety of resource allocation mechanisms (see for example [37, 27, 57]), and in recent years it has received renewed attention as a convenient abstraction for modeling the flow-level performance of bandwidth-sharing protocols in packet-switched networks, in particular TCP, see for example [16, 50].

Discriminatory processor sharing

The Discriminatory Processor Sharing queue (DPS) is a versatile queueing model providing a natural framework to model service differentiation in systems. It was introduced by Kleinrock in [36]. It is an extension of the PS policy. Again we assume there are K classes of customers, and the various classes are assigned positive weight factors, $g_1, \dots, g_K$. The service capacity is shared simultaneously among all customers present in proportion to the respective class-dependent weights. More precisely, given there are K classes of customers, if at time t there are $n_k(t)$ class- k customers present in the system, $k = 1, \dots, K$, under the DPS policy each class- k customer is served at rate

$$\frac{g_k}{\sum_{j=1}^K g_j n_j(t)}.$$

Despite the simplicity of the model description and the fact that the properties of the egalitarian PS queue are quite thoroughly understood and closed-form results exist,

the analysis of DPS has proven to be extremely difficult and no closed-form characterisation are known. We refer to the survey [1] for an extensive overview of the literature on DPS. Below we present results related to DPS that are used in Chapter 4.

In [46] Rege and Sengupta established that the generating function of the queue length vector satisfies a differential equation for *exponential service time distributions*. From this equation, the authors further show that the moments can be determined numerically as the solution of a system of equations.

The heavy-traffic regime analyses the system when it is near saturation. In [46], assuming exponential service requirement distributions, Rege and Sengupta established a state-space-collapse for the queue length distribution in the heavy-traffic regime, that is, in the limit the scaled queue length vector is distributed as the product of an exponentially distributed random variable and a deterministic vector. In [55] Verloop et al. generalised the result to *phase-type* distributions. Let $N_k^{DPS}, k = 1, \dots, K$, denote the number of class- k customers in steady state under DPS, then as $\rho \uparrow 1$

$$(1 - \lambda \mathbb{E}[B])(N_1^{DPS}, \dots, N_K^{DPS}) \xrightarrow{d} Y \cdot \left(\frac{\alpha_1 \mathbb{E}[B_1]}{g_1}, \frac{\alpha_2 \mathbb{E}[B_2]}{g_2}, \dots, \frac{\alpha_K \mathbb{E}[B_K]}{g_K} \right),$$

where $\xrightarrow{d}$ denotes convergence in distribution and Y is an exponentially distributed random variable with mean $\mathbb{E}[B^2] / (\mathbb{E}[B] \sum_{k=1}^K \alpha_k \mathbb{E}[B_k^2] / g_k)$.

In [13] Fayolle et al. studied the mean conditional (on the service requirement) and unconditional sojourn time. For general service time distributions, the authors obtained the mean conditional sojourn time as the solution of the following system of integro-differential equations:

$$\begin{aligned} \frac{\partial \overline{S}_k^{DPS}(\lambda, b)}{\partial b} &= 1 + \lambda \sum_{j=1}^K \int_0^\infty \alpha_j \frac{g_j}{g_k} \frac{\partial \overline{S}_j^{DPS}(\lambda, y)}{\partial y} \left[1 - F_j \left(y + \frac{g_j}{g_k} b \right) \right] dy \\ &\quad + \lambda \int_0^b \frac{\partial \overline{S}_k^{DPS}(\lambda, y)}{\partial y} \sum_{j=1}^K \alpha_j \frac{g_j}{g_k} \left[1 - F_j \left(\frac{g_j}{g_k} (b - y) \right) \right] dy, \end{aligned}$$

where $\overline{S}_k^{DPS}(\lambda, b)$ denotes the mean conditional (on the service requirement b) sojourn time of an arbitrary class- k customer under DPS. In addition, the authors provided a thorough analysis for the case of exponentially distributed service requirements. However, except for the case of two classes, no closed-form expression is available and numerical analysis is needed in order to calculate the mean sojourn times. We will give further details on the results of [13] in Section 4.2.1.

Relative priorities

The Relative Priorities (RP) is a multi-class queue. It provides an appropriate framework to model service differentiation in non-preemptive systems. Service is non-preemptive

and upon service completion, the probability that the next customer to be served is of class k is given by

$$\frac{n_k p_k}{\sum_{j=1}^K n_j p_j},$$

where, $p_j > 0$, $j = 1, \dots, K$, are given class-dependent weights, and n_j is the number of class- j customers at the decision epoch. Once a class is chosen to be served, an intra-class scheduling discipline determines which customer in this class will be served.

A special case of the model under study is when the intra-class scheduling discipline is uniform random, that is, within a class a customer is selected uniformly at random. This model was proposed in [20] and it is referred to as Discriminatory Random Order of Service (DROS). In recent years several interesting studies have been published on DROS, [21, 30, 31]. Expressions for the mean waiting time of a customer given its class have been obtained in [21]. In [30, 31] the authors derive differential equations that the transform of the queue lengths and the waiting time in steady-state must satisfy, respectively, and this allows the authors to find the moments of the queue lengths as a solution of linear equations. We will give further details on the results of [31] in Chapter 5.

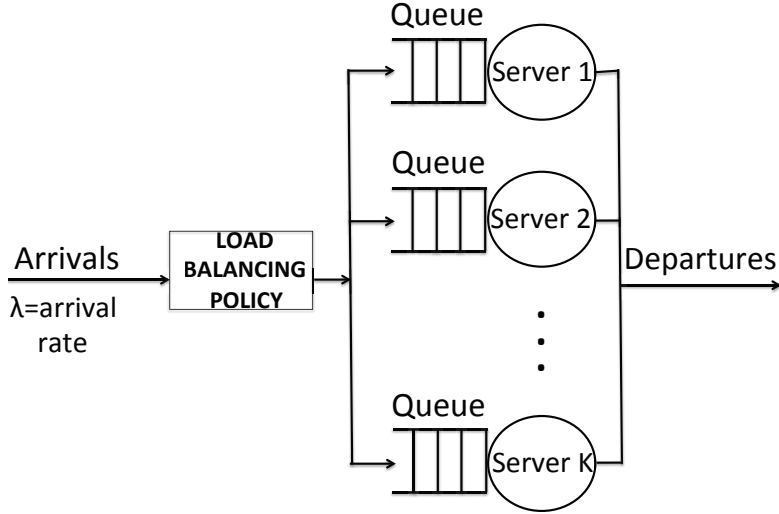
In the single-class case, $K = 1$, DROS reduces to the well-studied Random Order of Service (ROS) discipline. Classical papers on ROS are for example [33, 34, 45]. The Laplace transform for the waiting time distribution was obtained in [33]. More recently, the authors of [7] obtained the waiting time distribution in heavy traffic for certain service requirements having infinite variance. In addition, waiting-time tail asymptotics were obtained.

1.1.2 Parallel-server model

The parallel-server model consists of K *heterogeneous* (i.e., all of different speed or capacity) servers that can work simultaneously, see Figure 1.2. Upon arrival, an incoming customer is assigned to one of the K servers according to the *power-of-two* policy: The incoming customer randomly selects 2 servers from the available pool of K servers. The Join-the-shortest-queue (JSQ) policy is then used in isolation with these two servers; ties are broken randomly. The first-come-first-served (FCFS) intra server policy is then applied within each queue. The service requirement is denoted by B .

The *power-of-two* policy is often denoted as $SQ(2)$. Here $SQ(d)$, $d = 1, \dots, K$, denotes the policy that routes the customer to the shortest queue among the randomly chosen d queues. Note that $SQ(1)$ is equivalent to the random-server assignment, while $SQ(K)$ is equivalent to the JSQ among all servers in the system.

The power-of-two system has been studied for some time now; see the brief historical survey in [41, Section 1.1]. The analysis of the parallel-server model is rendered difficult because of the coupling between queues induced by local users of JSQ. This is so even when customers arrive according to a Poisson process, the servers are *homogeneous*

Figure 1.2: A parallel-server model with K servers.

(i.e., all of identical speed or capacity) and service requirements are exponentially distributed. In that setting, Mitzenmacher [41] and Vvedenskaya et al. (with $d = 2$) [43], independently, studied the limiting system obtained by letting the number of servers go to infinity. Their results point to a substantial improvement in performance of the case $d = 2$ over $d = 1$ *without* the full overhead of JSQ, whereas having $d = 3$ choices is only a constant factor better than $d = 2$. This is where the terminology “power-of-two” comes from. More recently, L. Ying et al. [38] have shown that in the case of batch arrivals pooling one server per customer is comparable to $SQ(2)$ in performance. Stolyar [51] has analysed an heterogeneous parallel-server model according to a pull-based policy. Under this policy each server sends a “pull-message” to the router when it becomes idle; the router assigns an arriving customer to a server according to a randomly chosen available pull-message, if there are any, or to a random server, otherwise. Stolyar proves that this pull-based policy provides further substantial improvements over the power-of-two policy.

1.2 Motivation for interpolation approximation

As mentioned earlier, the interpolation technique is a combination of the light-traffic and the heavy-traffic characterisations. The technique imposes the approximation to be exact in light traffic, i.e., when the load ρ is close to zero and in heavy traffic, i.e., when the load ρ of the system is close to saturation. The interpolation technique then, provides an approximation for intermediate values of the load. An important benefit of the approximation is that it provides insights into the dependency of the performance on the system parameters (weights, service time distributions, etc), and we believe this

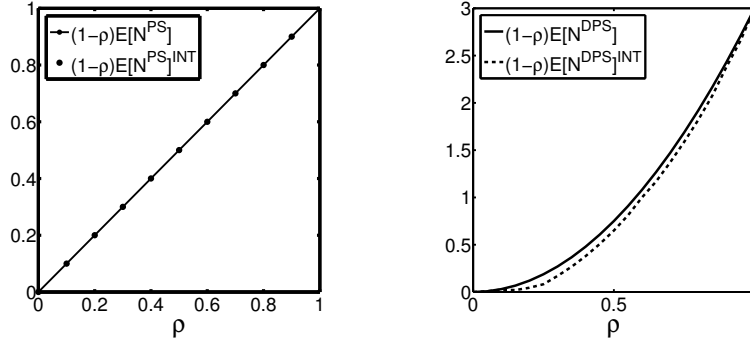


Figure 1.3: Accuracy of interpolation approximation for the total number of customers for PS (left) and DPS (right).

will be useful in their implementation. The interpolation technique, as pioneered in [49], is presented in detail in Chapter 2.

As an example, in Figure 1.3 we plot the exact form and the interpolation approximation for the total mean number of customers in a single-server system under the PS (left) and DPS (right) policies as a function of the traffic load ρ . Here N^{PS} and N^{DPS} denote the total number of customers under PS and DPS, and $\mathbb{E}[N^{PS}]^{INT}$ and $\mathbb{E}[N^{DPS}]^{INT}$ the corresponding interpolation approximations for the mean. For PS the exact expression is available and given by $\mathbb{E}[N^{PS}] = \frac{\rho}{1-\rho}$. Since by construction the interpolation approximation is exact at $\rho = 0$ and $\rho = 1$, and since for PS $(1-\rho)\mathbb{E}[N^{PS}] = \rho$ is a straight line, the interpolation approximation is exact, see Figure 1.3 left. For DPS we observe that the interpolation approximation is exact on the extreme values of the load, $\rho \downarrow 0$ and $\rho \uparrow 1$, but not for intermediate values of the load, although it remains accurate, see Figure 1.3 right.

We now describe the application of the interpolation technique to the three models studied in this thesis.

The parallel-server model under the power-of-two policy proves difficult to analyse because of dependencies among queues. In order to have asymptotically independent queues, in the literature it is often considered the number of servers to grow to infinity. Beyond the results under the infinite server assumption, explicit expressions are scarce. Motivated by the lack of results, we analysed the mean sojourn time of the power-of-two policy under the light-traffic regime. To the best of our knowledge, there is no paper in the literature that studies the power-of-two policy under the heavy-traffic regime. For this reason we did not perform the interpolation approximation for the power-of-two policy.

The analysis of DPS model has proven to be extremely difficult. For example, as stated in Section 1.1, results on an important basic metric like the mean sojourn time in the system have only been derived in a very implicit manner or under certain limiting regimes (time-scale decomposition, heavy-traffic, overload etc.). For the queue-length vector closed-form results are obtained only for exponential service requirements.

Encouraged by the difficulty in analysing the system in exact form, we derive approximations for the distribution of the queue length vector and the waiting time.

As mentioned in the previous section, for general service requirements Kim et al. [30, 31] derive for the RP and DROS models, respectively, differential equations that the transform of the queue lengths and the waiting time in steady state must satisfy. Therefore, the light-traffic approximations and the heavy-traffic results for the above mentioned metrics can be obtained directly from those differential equations, see Chapters 5 and 6, respectively. We will then combine the so-obtained results in order to develop interpolation approximations that aim at capturing the performance for any load.

1.3 Overview of the thesis

In this section we provided a general overview of the thesis, we presented the three queueing models that we study in this thesis and we motivated the approximation technique used. We now present a brief overview of the main contributions.

In Chapter 2 we provide a detailed explanation of how to obtain the light-traffic approximation, we discuss the heavy-traffic regime and we show how to build the interpolation approximation.

Chapter 3 focuses on the parallel-server model with heterogeneous servers under the power-of-two policy. We derive the light-traffic approximation for the mean sojourn time. We observe that the mean sojourn time is *decreasing* for small values of λ . This is a somewhat unexpected finding because most queueing systems are “monotone” in the sense that increasing the traffic intensity λ results in an increase in a performance metric such as the mean sojourn time of customers. The results of this chapter were published in [26].

In Chapter 4 we turn our attention to the DPS model. We obtain the light-traffic results for the probability generating function of the queue length and probability distribution of the waiting time and we combine them with the heavy-traffic results obtained in [46] and [55], respectively, in order to develop the interpolation approximations. We investigate in more detail the so-obtained approximation for the mean conditional (on the service requirement) sojourn time. We observe analytically and numerically that the approximation is accurate (if not exact). This chapter is based upon [23, 22, 25].

Chapters 5 and 6 concern the RP queue. Chapter 5 presents the heavy-traffic analysis of the RP queue. We establish a state-space collapse for the scaled queue length vector in the heavy-traffic regime, that is, in the limit the scaled queue length vector is distributed as the product of an exponentially distributed random variable and a deterministic vector. We observe that the scaled queue length reduces as classes with smaller mean service requirement obtain relatively larger weights. We finally show that the scaled waiting time of a class- k customer is distributed as the product of two exponentially distributed random variables. Chapter 6 provides the light-traffic results for the probability generating function of the queue length and Laplace Transform of the waiting time of the RP model. Combining them with the heavy-traffic results obtained

in Chapter 5, we develop the interpolation approximations. We assess the accuracy of the interpolation approximation using the first and second moments. These chapters present the results from [3, 24, 25].

The method applied throughout this thesis to develop the interpolation approximation is constructive, and it can readily be applied to other queueing systems for which no analytical results are available. We thus believe that this represents in itself one of the main contributions of the thesis.

Chapter 2

Interpolation approximation

In this chapter we describe how performance metrics can be derived for the light-traffic regime, we discuss the heavy-traffic regime and we present how the light and heavy-traffic interpolation is carried out. This approach was pioneered in a series of papers by Reiman & Simon, see for example [49], where the objective was the mean number of customers or mean sojourn time, and extended to the distribution of the sojourn time for Markovian queues in [15] and [48]. The technique has been applied with success to models like processor-sharing, fork-join, etc.; see for examples in the literature [9, 54, 23, 15, 48]. As mentioned in the previous chapter, the interpolation technique imposes the approximation to be exact in light traffic, so when the arrival rate λ is close to zero, and in heavy traffic, so when the arrival rate is close to saturation, and it aims to capture an accurate (if not exact) performance for any arbitrary value of the arrival rate λ . We will see in Chapters 4 and 6 for the DPS and RP models, respectively, that the approximation turns out to be exact for all λ for particular choices of the system parameter.

Let us denote by $G(\lambda, \vec{y})$ the performance metric we are interested in, as a function of the arrival rate λ and a vector $\vec{y}$. The interpretation of the function G and the vector $\vec{y}$ will change depending on the metric we are approximating. In this thesis the metric G will represent one of the following choices:

- (i) the generating function of the queue-length vector $(N_1, \dots, N_K)$, denoted by $\psi(\lambda, \vec{z}) := \mathbb{E} \left[z_1^{N_1} \dots z_K^{N_K} \right]$. Hence $\vec{y} = \vec{z}$,
- (ii) the complementary distribution function of the conditional waiting time of a class- k customer denoted by $W(\lambda, b, x) := \mathbb{P} [W_k(b) > x]$, where $W_k(b)$ is the conditional waiting time of a class- k customer. Hence $\vec{y} = (b, x)$,
- (iii) the Laplace-Stieltjes Transform (LST) of the waiting time of a class- k customer denoted by $\widetilde{W}_k(\lambda, u) := \mathbb{E} \left[e^{-uW_k} \right]$, where W_k is the unconditional waiting time of a class- k customer. Hence $\vec{y} = u$.

In Section 2.1 we will characterise $G(\lambda, \vec{y})$ in the light-traffic regime, $\lambda \downarrow 0$. In Section 2.2 the heavy-traffic regime is discussed. Section 2.3 presents the general setting for the interpolation approximation.

2.1 Light-traffic analysis

The light-traffic regime concerns the performance of the system when the arrival rate λ approaches zero, or in other words, when the amount of work arriving to the system per unit of time approaches zero. The light-traffic analysis has been successfully applied in several papers, see for example [54, 23, 48]. Following the discussion in [49, Appendix A] we make the next assumption on the service requirements B

$$\mathbb{E}[e^{\eta B}] = \sum_{n=0}^{\infty} \frac{\eta^n}{n!} \mathbb{E}[B^n] < \infty, \quad (2.1)$$

for some $\eta > 0$. This finite exponential moment condition requires all moments of the service requirement B to be finite.

Assuming that the first n derivatives of $G(\lambda, \vec{y})$ at $\lambda = 0$ exist, we have the following approximation for $G(\lambda, \vec{y})$ when λ is close to zero:

$$G^{LT}(\lambda, \vec{y}) := G^{(0)}(0, \vec{y}) + \lambda G^{(1)}(0, \vec{y}) + \cdots + \frac{\lambda^n}{n!} G^{(n)}(0, \vec{y}). \quad (2.2)$$

We will refer to this as the light-traffic approximation of order n . Here $G^{(0)}(0, \vec{y}) := G(0, \vec{y})$, to which we refer to as the *zeroth* light-traffic derivative. Moreover, $G^{(m)}(0, \vec{y})$, $m = 1, 2, \dots, n$, denotes the m -th derivative at $\lambda = 0$, i.e., $G^{(m)}(0, \vec{y}) := \left. \frac{\partial^m G(\lambda, \vec{y})}{\partial \lambda^m} \right|_{\lambda=0}$. The choice of the value of n will depend on the compromise between tractability and accuracy that is aimed at.

In the case of models that permit a multidimensional quasi birth-and-death representation, researchers have also used the power-series algorithm to develop light-traffic approximations of the mean queue lengths, see for example [11]. Approximations of the type (2.2) for the steady-state distribution of perturbed Markov chains have been derived in, for example, Altman et al. [12].

When a characterisation for $G(\lambda, \vec{y})$ exists, for example in terms of a differential equation that $G(\lambda, \vec{y})$ needs to satisfy, this can be exploited to carry out the light and heavy-traffic analysis (as done in Chapters 5 and 6, respectively, for the RP model). However, in general, a characterization for $G(\lambda, \vec{y})$ might not exist. In that case we use the results in [49, Section 3.] and [56, Chapter 6.3.] where it is shown how to derive the light-traffic derivatives of arbitrary order m .

The expressions as obtained in [49, 56] for the zeroth, first and second light-traffic derivatives are given in the proposition below. We want to highlight that we use these expressions to obtain the light-traffic derivatives for the parallel-server model (see Proposition 3.2.1) and the DPS queue (see Lemma 4.3.1 and 4.4.1). In the case of the RP model the light-traffic derivatives are derived from the differential equations in [31] (see Lemma 6.1.2 and 6.2.2).

Proposition 2.1.1. [49, Section 3], [56, Chapter 6.3] *Let $A(s, t)$ denote the number of arrivals in the interval $[s, t]$ in addition to the tagged customer who is assumed to arrive*

at time 0, and let $G(\lambda, \vec{y}|A)$ denote the performance metric conditioned on A . Then the zeroth, first and second light-traffic derivatives can be written as

$$G^{(0)}(0, \vec{y}) = G\left(0, \vec{y} \middle| A(-\infty, \infty) = 0\right), \quad (2.3)$$

$$G^{(1)}(0, \vec{y}) = \int_{-\infty}^{\infty} \left(G\left(0, \vec{y} \middle| A(-\infty, \infty) = 1, \tau_1 = t\right) - G\left(0, \vec{y} \middle| A(-\infty, \infty) = 0\right) \right) dt \quad (2.4)$$

and

$$\begin{aligned} G^{(2)}(0, \vec{y}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(G\left(0, \vec{y} \middle| A(-\infty, \infty) = 2, \tau_1 = t', \tau_2 = t''\right) - G\left(0, \vec{y} \middle| A(-\infty, \infty) = 1, \tau_1 = t'\right) \right. \\ &\quad \left. - G\left(0, \vec{y} \middle| A(-\infty, \infty) = 1, \tau_1 = t''\right) + G\left(0, \vec{y} \middle| A(-\infty, \infty) = 0\right) \right) dt' dt'', \end{aligned} \quad (2.5)$$

where $\tau_i, i = 1, 2$, is the arrival time of the i -th customer.

We now provide a reasoning of how to obtain the zeroth and first light-traffic derivatives. This is based on the analysis of J. Walrand in [56, Chapter 6.3]. The second and higher order light-traffic derivatives can be obtained in a similar way.

Let us consider a system that starts at time $-Z$ and that keeps going until time T , being $Z, T > 0$ given. Let $G(\lambda, \vec{y}, -Z, T)$ denote the term we are interested in approximating and note that $\lim_{Z, T \rightarrow \infty} G(\lambda, \vec{y}, -Z, T) = G(\lambda, \vec{y})$. Let $A(s, t)$ denote the number of arrivals in the interval $[s, t)$ in addition to the tagged customer who is assumed to arrive at time 0. Throughout this section we assume that the limits (with respect to Z and T) and expectations can be interchanged. We then have

$$G(\lambda, \vec{y}, -Z, T) = \sum_{a=0}^{\infty} G\left(\lambda, \vec{y}, -Z, T \middle| A(-Z, T) = a\right) \cdot \frac{(\lambda(T+Z))^a}{a!} e^{-\lambda(T+Z)}, \quad (2.6)$$

where $G\left(\lambda, \vec{y}, -Z, T \middle| A(-Z, T) = a\right)$ is conditioned on the fact that there are exactly a arrivals. Evaluating this at $\lambda = 0$ gives

$$G(\lambda, \vec{y}, -Z, T) \Big|_{\lambda=0} = G\left(0, \vec{y}, -Z, T \middle| A(-Z, T) = 0\right), \quad (2.7)$$

and now taking the limit $Z, T \rightarrow \infty$ we obtain the zeroth light-traffic derivative

$$G^{(0)}(0, \vec{y}) := \lim_{Z, T \rightarrow \infty} G(\lambda, \vec{y}, -Z, T) \Big|_{\lambda=0} = G\left(0, \vec{y} \middle| A(-\infty, \infty) = 0\right)$$

where the second equality follows from (2.7).

Next, consider the derivative with respect to λ in Equation (2.6) and evaluate it at $\lambda = 0$. This gives

$$\begin{aligned} & \frac{\partial}{\partial \lambda} G(\lambda, \vec{y}, -Z, T) \Big|_{\lambda=0} \\ &= -G(0, \vec{y}, -Z, T | A(-Z, T) = 0) \cdot (T + Z) + G(0, \vec{y}, -Z, T | A(-Z, T) = 1) \cdot (T + Z) \\ &= \int_{-Z}^T \left(G(0, \vec{y}, -Z, T | A(-Z, T) = 1, \tau_1 = t) - G(0, \vec{y}, -Z, T | A(-Z, T) = 0) \right) dt, \end{aligned} \tag{2.8}$$

where τ_1 is the arrival time of the first customer. The second equality holds because the arrivals follow a Poisson process. Hence given that the number of arrivals in $[-Z, T)$ is one ($A(-Z, T) = 1$), we have that τ is uniformly distributed on $[-Z, T)$.

Now taking $Z, T \rightarrow \infty$ we obtain the first light-traffic derivative

$$\begin{aligned} G^{(1)}(0, \vec{y}) &:= \lim_{Z, T \rightarrow \infty} \frac{\partial}{\partial \lambda} G(\lambda, \vec{y}, -Z, T) \Big|_{\lambda=0} \\ &= \int_{-\infty}^{\infty} \left(G(0, \vec{y} | A(-\infty, \infty) = 1, \tau_1 = t) - G(0, \vec{y} | A(-\infty, \infty) = 0) \right) dt, \end{aligned}$$

where the second equality follows from (2.8).

2.2 Heavy-traffic regime

The heavy-traffic regime concerns the performance of the queue when it is near saturation. In other words, the amount of work that arrives is close to (but always below) the capacity of the system. In this regime we aim at deriving an expression for our performance metric in heavy-traffic regime after applying an appropriate scaling to it. We denote this scaling by $f_\lambda(\vec{y})$ and are hence interested in obtaining an expression for $G(\lambda, f_\lambda(\vec{y}))$. In the remainder of this section we further explain this concept for the performance metrics of the single-server system as studied in this thesis.

The single-server system is near saturation when the load approaches one, $\rho \uparrow 1$. Hence, this regime can be obtained by letting

$$\lambda \uparrow \hat{\lambda} := \frac{1}{\mathbb{E}[B]}.$$

Heavy-traffic analysis has been carried out in several papers, see for example [46, 32, 55].

In Chapter 4 we provide a brief overview of the heavy-traffic results known for the DPS model and in Chapter 5 we obtain the heavy-traffic results for the RP model. The basic principle is to establish that the scaled performance metrics $(1 - \lambda \mathbb{E}[B])\vec{N}$, $(1 - \lambda \mathbb{E}[B])W_k(b)$ and $(1 - \lambda \mathbb{E}[B])W_k$, have a proper limit as $\lambda \uparrow \frac{1}{\mathbb{E}[B]}$. Hence, in the heavy-traffic regime we have expressions for the following scaled performance metrics:

- (i) $\psi(\lambda, \bar{z}^{1-\lambda\mathbb{E}[B]}) = \mathbb{E}[\bar{z}^{(1-\lambda\mathbb{E}[B])\vec{N}}]$,
- (ii) $W_k(\lambda, b, x/(1-\lambda\mathbb{E}[B])) = \mathbb{P}[(1-\lambda\mathbb{E}[B])W_k(b) > x]$,
- (iii) $\widetilde{W}_k(\lambda, u(1-\lambda\mathbb{E}[B])) = \mathbb{E}[e^{-u(1-\lambda\mathbb{E}[B])W_k}]$,

where we used the notation $\bar{z}^{\gamma\vec{N}} := (z_1^{\gamma N_1}, \dots, z_K^{\gamma N_K})$. Depending on the three metrics described above, the scaling $f_\lambda(\vec{y})$ is hence

- (i) $f_\lambda(\vec{z}) = \bar{z}^{1-\lambda\mathbb{E}[B]}$,
- (ii) $f_\lambda(b, x) = (b, x/(1-\lambda\mathbb{E}[B]))$,
- (iii) $f_\lambda(u) = u(1-\lambda\mathbb{E}[B])$.

Let $G^{HT}(\vec{y})$ be the heavy-traffic term defined as

$$G^{HT}(\vec{y}) := \lim_{\lambda \uparrow 1/\mathbb{E}[B]} G(\lambda, f_\lambda(\vec{y})). \quad (2.9)$$

2.3 Light and heavy-traffic interpolation

In the case expressions for a performance metric are known both for light traffic and heavy traffic, an approximation for an arbitrary λ can be derived following the interpolation technique. This technique was popularized by Reiman and Simon [47, 48, 49] and consists in approximating the scaled performance metric, $G(\lambda, f_\lambda(\vec{y}))$, by a polynomial $\hat{G}(\lambda, \vec{y})$ of order $n+1$:

$$\hat{G}(\lambda, \vec{y}) = h_0(\vec{y}) + h_1(\vec{y})\lambda + h_2(\vec{y})\lambda^2 + \dots + h_{n+1}(\vec{y})\lambda^{n+1}. \quad (2.10)$$

Undoing the normalisation we then obtain the interpolation approximation for the performance metric $G(\lambda, \vec{y})$, that is,

$$G^{INT}(\lambda, \vec{y}) := \hat{G}(\lambda, f_\lambda^{-1}(\vec{y})) \quad (2.11)$$

for $0 \leq \lambda < 1/\mathbb{E}[B]$. We will refer to the approximation (2.11) as the *interpolation approximation* of order $n+1$.

To determine the coefficients $h_0(\vec{y}), \dots, h_n(\vec{y})$ we take the m -th derivative, $m = 0, \dots, n$, in (2.11) at $\lambda = 0$ and set this equal to the m -th derivative of the performance metric to be approximated. Hence, we obtain the following light-traffic conditions:

$$\left. \frac{\partial^m G^{INT}(\lambda, \vec{y})}{\partial \lambda^m} \right|_{\lambda=0} = G^{(m)}(0, \vec{y}), \text{ for } m = 0, \dots, n. \quad (2.12)$$

Note that expressions for $G^{(m)}(0, \vec{y})$ are given in Proposition 2.1.1. To determine $h_{n+1}(\vec{y})$, we use the heavy-traffic condition:

$$\lim_{\lambda \uparrow 1/\mathbb{E}[B]} \hat{G}(\lambda, \vec{y}) = G^{HT}(\vec{y}), \quad (2.13)$$

where $G^{HT}(\vec{y})$ denotes the heavy-traffic result as defined in (2.9). Later on we will see how to determine the coefficients $h_0(\vec{y}), \dots, h_{n+1}(\vec{y})$ in practice.

We now state the interpolation approximation result that we will use throughout the thesis.

Proposition 2.3.1. *The interpolation approximation of order $n + 1$ can equivalently be written as*

$$G^{INT}(\lambda, \vec{y}) = \sum_{i=0}^n \lambda^i \left(1 - (\lambda \mathbb{E}[B])^{n+1-i}\right) h_i \left(f_\lambda^{-1}(\vec{y})\right) + (\lambda \mathbb{E}[B])^{n+1} G^{HT} \left(f_\lambda^{-1}(\vec{y})\right). \quad (2.14)$$

Proof: From the heavy-traffic condition (2.13) we obtain

$$h_{n+1}(\vec{y}) = \mathbb{E}[B]^{n+1} \left(G^{HT}(\vec{y}) - \sum_{i=0}^n \frac{h_i(\vec{y})}{\mathbb{E}[B]^i} \right).$$

Equation (2.14) follows after substituting this expression in (2.10) and then undoing the normalisation as in Equation (2.11). $\blacksquare$

We note that in the case $G(\lambda, \vec{y})$ denotes the unconditional sojourn time distribution, $\tilde{S}(\lambda, x) := \mathbb{P}[(1 - \lambda \mathbb{E}[B])S_k > x]$, where S_k is the sojourn time of a class- k customer, Equation (2.14) reduces to Equation (1) in [15] after undoing the normalisation.

An important observation is that the interpolation approximation obtained for the probability generating function and LST of the performance metrics might not correspond themselves to a random variable, that is, they might not be completely monotone functions as defined in [14, Section XIII.4]. However, in the thesis we will observe that they can still provide accurate approximations for the moments.

In the particular case that we are approximating the mean of a random variable V , Equation (2.10) can be further simplified in terms of the light-traffic derivatives and the heavy-traffic term, see Corollary 2.3.2 below. In this case, we define $G(\lambda) := \mathbb{E}[V]$ and the heavy-traffic term by $G^{HT} := \lim_{\lambda \uparrow 1/\mathbb{E}[B]} f_\lambda \cdot G(\lambda)$, where f_λ is the appropriate heavy-traffic scaling. Hence, we can drop all dependence on $\vec{y}$ in the notation.

Corollary 2.3.2. *The $n + 1$ order interpolation approximation for the mean can be written as*

$$\begin{aligned} G^{INT}(\lambda) &= \frac{1}{f_\lambda} \left(\sum_{i=0}^n \frac{\lambda^i}{i!} \left(1 - (\lambda \mathbb{E}[B])^{n+1-i}\right) \sum_{j=0}^i \binom{i}{j} \frac{d^{i-j} f_\lambda}{d\lambda^{i-j}} \Big|_{\lambda=0} G^{(j)}(0) + (\lambda \mathbb{E}[B])^{n+1} G^{HT} \right), \end{aligned} \quad (2.15)$$

where $G^{(j)}(\lambda), j = 0, \dots, n$ denotes the j -th light-traffic derivative.

Proof: From the light-traffic condition (2.12) and after undoing the normalisation we obtain

$$h_i(f_\lambda^{-1}) = \frac{\sum_{j=0}^i \binom{i}{j} \left. \frac{d^{i-j} f_\lambda}{d\lambda^{i-j}} \right|_{\lambda=0} G^{(j)}(0)}{i! f_\lambda}, i = 0, 1, \dots, n.$$

Equation (2.15) follows after substituting these expressions in (2.14). ■

Chapter 3

Light-traffic analysis of the power-of-two policy

In this chapter we focus on the heterogeneous parallel-server model under the power-of-two policy, described in Section 1.1.2. Due to intractability of the system in general, we apply the light-traffic approximation technique described in Section 2.1 and we point to interesting structural features associated with server heterogeneity in light traffic.

This chapter is organised as follow. In Section 3.1 we describe the model analysed and provide the needed mathematical framework to calculate the light-traffic derivatives. In Section 3.2 we gather the main results of this chapter. Sections 3.3, 3.5 and 3.6 contain the proofs of the zeroth, first and second light-traffic derivatives, respectively. In Section 3.7 we present the corresponding numerical results.

3.1 Model description

The system comprises K parallel servers labelled $k = 1, \dots, K$. Server k , which has capacity C_k , is attended by an infinite capacity buffer and operates in a FCFS manner. Customers arrive according to a Poisson process of rate $\lambda > 0$, each customer bringing a random amount of work distributed according to the probability distribution F . Upon arrival, an incoming customer is assigned to one of the K servers according to the power-of-two load balancing policy: *Two* servers are selected at random from the available pool of K servers. The JSQ policy is then used in isolation with these two servers; ties are broken randomly (but other choices are possible). As usual, the Poisson process of arrivals, the sequence of customer requirement random variables and the sequence of server selection random variables are mutually independent. Moreover, each of the two sequences is a sequence of i.i.d. random variables (across customers).

Let the random variable B be the service requirement of an arbitrary arriving customer. Let S^{SQ2} denote the stationary sojourn time of a customer at arbitrary epoch. Set the mean sojourn time

$$\overline{S^{SQ2}}(\lambda) = \mathbb{E} \left[S^{SQ2} \right].$$

We expect $\overline{S^{SQ2}}(\lambda) < \infty$ over some non-degenerate interval $[0, \lambda^*)$ for some finite $\lambda^* > 0$; see also [41, Section 2.1, Lemma 1] and [42, Lemma 6]. In what follows we shall not be concerned with this issue any further since we are mainly interested in the situation where $\lambda \simeq 0$.

We find it convenient to introduce the following terminology and notation: With t in $\mathbb{R}$, a customer arriving at time t , hereafter referred to as a t -customer, has two random variables B_t and Σ_t associated with it – The $\mathbb{R}_+$ -valued random variable B_t stipulates the amount of work requested by the t -customer from the system, while the random variable Σ_t is an (unordered) pair of servers from amongst the K available servers. The t -customer is assigned to a server ν_t selected in Σ_t according to the power-of-two policy (with a random tie-breaker). We shall refer to the random variables (B_t, Σ_t) as the characteristic pair of the t -customer.

As expected, we sometimes refer to the 0-customer with characteristics (B_0, Σ_0) as the *tagged* customer. The Reiman-Simon approach to light traffic focuses on the performance of this tagged customer under scenarios of increasing complexity. To define them, fix $n = 0, 1, \dots$. Interpret every n -uple $(t_1, \dots, t_n)$ in $\mathbb{R}^n$ as the arrival epochs of n customers into the system. For each $i = 1, \dots, n$, we lighten the notation by denoting the characteristic pair (B_{t_i}, Σ_{t_i}) of the t_i -customer arriving at time t_i simply by (B_i, Σ_i) . Throughout the following conditions are assumed to be enforced:

1. The random variables $\{B_0, B_1, \dots, B_n\}$ are i.i.d. $\mathbb{R}_+$ -valued random variables, each distributed according to the probability distribution F , namely

$$\mathbb{P}[B_i \leq x] = F(x), \quad \begin{array}{l} x \geq 0, \\ i = 0, 1, \dots, n. \end{array}$$

2. The random variables $\{\Sigma_0, \Sigma_1, \dots, \Sigma_n\}$ are $\mathcal{P}_2(K)$ -valued uniformly distributed random variables with

$$\mathbb{P}[\Sigma_i = T] = \frac{1}{\binom{K}{2}}, \quad \begin{array}{l} T \in \mathcal{P}_2(K), \\ i = 0, 1, \dots, n. \end{array}$$

3. The collections of random variables $\{B_0, B_1, \dots, B_n\}$ and $\{\Sigma_0, \Sigma_1, \dots, \Sigma_n\}$ are mutually independent.

We shall also have use for the random variables $\nu_0^*, \nu_1^*, \dots, \nu_n^*$ associated with the random pairs $\Sigma_0, \Sigma_1, \dots, \Sigma_n$, and defined in the following manner: For each $i = 0, 1, \dots, n$, the random variable ν_i^* is an Σ_i -valued random variable with

$$[\nu_i^* | \Sigma_i] \sim \mathcal{U}(\Sigma_i).$$

It is always understood that the random variables $\nu_0^*, \nu_1^*, \dots, \nu_n^*$ are conditionally mutually independent given the $2(n+1)$ random variables $B_0, B_1, \dots, B_n, \Sigma_0, \Sigma_1, \dots, \Sigma_n$. Under the enforced assumptions, we readily conclude that the random variables $\nu_0^*, \nu_1^*, \dots, \nu_n^*$ are i.i.d. random variables, each of which is uniformly distributed on $\{1, \dots, K\}$.

3.2 Mean sojourn time

When computing the light-traffic derivatives, for each $(t_1, \dots, t_n)$ in $\mathbb{R}^n$, let the random variable $S_n(t_1, \dots, t_n)$ denote the sojourn time of the tagged customer under the scenario that *in addition* to the tagged customer, only n customers are allowed to enter the system over $\mathbb{R}$, say at times $t_1, \dots, t_n$, with characteristic pairs $(B_1, \Sigma_1), \dots, (B_n, \Sigma_n)$ as defined earlier. Note that $S_n(t_1, \dots, t_n)$ depends on the random variables $\{B_0, B_1, \dots, B_n\}$, $\{\Sigma_0, \Sigma_1, \dots, \Sigma_n\}$ and $\{\nu_0^*, \nu_1^*, \dots, \nu_n^*\}$ in a complicated manner through the scheduling policy used. We shall write

$$\hat{S}_n(t_1, \dots, t_n) := \mathbb{E}[S_n(t_1, \dots, t_n)]. \quad (3.1)$$

For the power-of-two policy there is no characterisation available for the mean sojourn time with heterogeneous servers. Thus, in the next proposition we state the second order light-traffic approximation using the result given in Proposition 2.1.1. The proof method is constructive and it could be applied to other queueing systems for which no analytical results are available.

The results can be expressed compactly with the help of a random variable X which is uniformly distributed over the set of values $\frac{1}{C_1}, \dots, \frac{1}{C_K}$, i.e.,

$$\mathbb{P}\left[X = \frac{1}{C_1}\right] = \dots = \mathbb{P}\left[X = \frac{1}{C_K}\right] = \frac{1}{K}.$$

It is plain that

$$\mathbb{E}[X^p] = \frac{1}{K} \sum_{k=1}^K \frac{1}{C_k^p}, \quad p = 1, 2, \dots$$

Proposition 3.2.1. *The light-traffic approximation of the mean sojourn time under the power-of-two policy satisfies*

$$\overline{SQ2,LT}(\lambda) = \left(\overline{SQ2}\right)^{(0)}(0) + \lambda \left(\overline{SQ2}\right)^{(1)}(0) + \frac{\lambda^2}{2} \left(\overline{SQ2}\right)^{(2)}(0) \quad (3.2)$$

with

$$\left(\overline{SQ2}\right)^{(0)}(0) = \mathbb{E}[X] \cdot \mathbb{E}[B], \quad (3.3)$$

$$\left(\overline{SQ2}\right)^{(1)}(0) = -\frac{1}{K-1} \text{Var}[X] \cdot \mathbb{E}[B]^2 \quad (3.4)$$

and

$$\left(\overline{SQ2}\right)^{(2)}(0) = \frac{2}{(K-1)^2} \left(\mathbb{E}[X]^3 - 2\mathbb{E}[X]\mathbb{E}[X^2] + \mathbb{E}[X^3] \right) \cdot \mathbb{E}[B]^3. \quad (3.5)$$

Proof: The proof follows from Proposition 2.3.1, together with the expressions obtained for the zeroth, first and second light-traffic derivatives, derived in Sections 3.3, 3.5.2 and 3.6.3, respectively. ■

Several observations readily flow from Proposition 3.2.1.

Equal capacities: From (3.4) it is plain that $\left(\overline{SQ2}\right)^{(1)}(0) \leq 0$, with $\left(\overline{SQ2}\right)^{(1)}(0) = 0$ if and only if $\text{Var}[X] = 0$, or equivalently, $C_1 = \dots = C_K \equiv C$ – In that case we also have $\left(\overline{SQ2}\right)^{(2)}(0) = 0$ with

$$\overline{SQ2,LT}(\lambda) = \frac{\mathbb{E}[B]}{C} + o(\lambda^2).$$

Unequal capacities: When the capacities are different, then $\left(\overline{SQ2}\right)^{(1)}(0) < 0$ and $\overline{SQ2,LT}(\lambda)$ is *decreasing* for small values of λ . This is a somewhat unexpected finding because most queueing systems are “monotone” in the sense that increasing the traffic intensity λ results in an increase in a performance metric such as the mean sojourn time.

This fact can be explained as follows: On the average, a customer entering an empty system experiences a sojourn time given by $\left(\overline{SQ2}\right)^{(0)}(0)$ since in such a circumstance the scheduling policy will assign it to *any* of the K servers with probability $\frac{1}{K}$. However, in the presence of different server capacities, the assigned server may not have been the fastest and it is possible for subsequent customers to be served by faster servers by the luck of the draw. This will result in a decrease in the mean sojourn time if the traffic intensity increases slightly but still allows for some faster server to be starved with some non-negligible probability.

Intuitively we expect this decrease to be more pronounced if the values of $C_1, \dots, C_K$ are more “spread out”. Indeed, for a *given* average value

$$\frac{1}{K} \sum_{k=1}^K \frac{1}{C_k},$$

the decrease will be steeper the more “unbalanced” the values of $\frac{1}{C_1}, \dots, \frac{1}{C_K}$ are in the sense of majorization; this follows from (3.4) by majorization arguments and Schur-convexity (using the convexity of $t \rightarrow t^2$) [40, Prop. C.1, p. 64].

Only $\mathbb{E}[B]$ matters – The two first derivatives depend only on the first moment of B , and could be read as a form of insensitivity in light traffic. This is in sharp contrast with other systems where the first light-traffic derivative depends on $\mathbb{E}[B^2]$, e.g., $M/G/1$ -like queues [47] and the discriminatory processor sharing model [23]. This is rather unexpected because the variance of B is known to be a key factor in shaping JSQ performance with homogeneous servers under FCFS scheduling [4, Chapter 24]. See next item for a possible explanation.

FCFS vs. PS – Proposition 3.2.1 was established under the assumption that the servers operate under the FCFS discipline. However, it is easy to see that both (3.3) and (3.4) (but not (3.5)) are still valid if the servers all use the PS discipline. That the variability of B seems to play little role in light traffic is therefore consistent with the aforementioned fact that performance under the PS discipline is nearly insensitive to service variability [19].

3.3 Zeroth light-traffic derivative

The zeroth light-traffic derivative corresponds to the scenario when besides the tagged customer, no other customer enters over the entire horizon $(-\infty, \infty)$. Let S_0 denote the sojourn time of the tagged customer under these circumstances. Obviously, under the power-of-two scheduling strategy, we have

$$S_0 = \frac{B_0}{C_{\nu_0}} \quad \text{with } \nu_0 = \nu_0^*, \quad (3.6)$$

because in the absence of any other customer in the system, the tagged customer is necessarily assigned to server ν_0^* . In analogy with earlier notation we write

$$\hat{S}_0 = \mathbb{E}[S_0(t)].$$

Proposition 3.3.1. *Under the enforced assumptions, the random variable ν_0^* is uniformly distributed over $\{1, \dots, K\}$ with*

$$\mathbb{P}[\nu_0^* = k] = \frac{1}{K}, \quad k = 1, \dots, K \quad (3.7)$$

and the relation

$$\hat{S}_0 = \left(\frac{1}{K} \sum_{k=1}^K \frac{1}{C_k} \right) \cdot \mathbb{E}[B] \quad (3.8)$$

holds.

It will often be convenient to write (3.8) more compactly as

$$\hat{S}_0 = \frac{\Gamma}{K} \cdot \mathbb{E}[B], \quad (3.9)$$

with

$$\Gamma = \sum_{k=1}^K \frac{1}{C_k} \quad (3.10)$$

Proof of Proposition 3.3.1 For each $k = 1, \dots, K$, the definition of ν_0^* gives

$$\begin{aligned} \mathbb{P}[\nu_0^* = k] &= \sum_{\ell=1, \ell \neq k}^K \mathbb{P}[\Gamma_0 = \{k, \ell\}, \nu_0^* = k] \\ &= \sum_{\ell=1, \ell \neq k}^K \mathbb{P}[\nu_0^* = k | \Gamma_0 = \{k, \ell\}] \mathbb{P}[\Gamma_0 = \{k, \ell\}] \\ &= (K-1) \cdot \frac{1}{2} \cdot \frac{2}{K(K-1)} = \frac{1}{K}. \end{aligned} \quad (3.11)$$

As pointed earlier, we necessarily have $\nu_0 = \nu_0^*$. The random variables ν_0^* and B_0 being independent, we then obtain from (3.6) that

$$\hat{S}_0 = \mathbb{E} \left[\frac{B_0}{C_{\nu_0^*}} \right] = \mathbb{E}[B_0] \cdot \mathbb{E} \left[\frac{1}{C_{\nu_0^*}} \right],$$

and the conclusion (3.8) readily follows from (3.7). $\blacksquare$

We have $(\overline{S^{SQ2}})^{(0)}(0) = \hat{S}_0$ according to Equation (2.3), hence Equation (3.3) is established with the help of (3.8).

3.4 An auxiliary result

Obtaining the light-traffic derivatives of order 1 and 2 is computationally more involved. The technical result discussed next will simplify the presentation by isolating an evaluation which is repeatedly carried out during the analysis. This auxiliary result is given in a setting that mimics power-of-two scheduling with only two customers present:

Fix $y < 0$. In addition to the tagged customer arriving at time $t = 0$ with characteristic pair (B_0, Γ_0) , assume that another customer arrives at time y with (random) service requirement B . The y -customer is then assigned to the server γ , with γ being some $\{1, \dots, K\}$ -valued random variable, while the tagged customer is assigned to the server γ_0 (in Γ_0) in accordance with the power-of-two scheduling policy. Thus, if $y + \frac{B}{C_\gamma} \leq 0$, then $\gamma_0 = \nu_0^*$. On the other hand, if $y + \frac{B}{C_\gamma} > 0$, then the operational rules of the power-of-two scheduling policy preclude the tagged customer to be assigned to server γ : If γ is not in Γ_0 , then $\gamma_0 = \nu_0^*$ again, while if γ is an element of Γ_0 , then γ_0 is necessarily the other server in the pair Γ_0 , i.e., the one different from γ . In this scenario γ is given *a priori*, and should be thought as a place holder for a server assignment random variable determined via power-of-two scheduling under various circumstances. On the other hand, γ_0 depends on y , B , γ and Γ_0 (as well as ν_0^*). The explicit dependence on these quantities will dropped from the notation.

For reasons that will become apparent in subsequent developments, we also introduce an event E (to be specified later).

Lemma 3.4.1. *Given are the random variables $\mathbf{1}[E]$, B , γ , B_0 , Γ_0 and ν_0^* . We assume that (i) the random variable ν_0^* is uniformly distributed on Γ_0 conditionally on all the other random variables $\mathbf{1}[E]$, B , γ , B_0 and Γ_0 ; (ii) the collections of random variables $\{\mathbf{1}[E], B, \gamma\}$ and $\{\nu_0^*, \Gamma_0, B_0\}$ are independent; and (iii) the random variables Γ_0 and B_0 are independent. Then, for each $y < 0$ and each $k = 1, \dots, K$, we have*

$$\begin{aligned} \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \frac{B_0}{C_{\gamma_0}} \right] &= \mathbb{P} \left[E, \gamma = k, y + \frac{B}{C_k} \leq 0 \right] \cdot \hat{S}_0 \\ &\quad + \frac{1}{K-1} \mathbb{P} \left[E, \gamma = k, y + \frac{B}{C_k} > 0 \right] \left(\Gamma - \frac{1}{C_k} \right) \cdot \mathbb{E}[B_0], \end{aligned} \tag{3.12}$$

with γ_0 defined earlier.

Note that under the enforced assumptions, the random variable B_0 is independent of the collection of random variables $\{\nu_0^*, \Gamma_0\}$.

Proof of Lemma 3.4.1: Fix $k = 1, \dots$. We start with the natural decomposition

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \frac{B_0}{C_{\gamma_0}} \right] \\ &= \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_\gamma} \leq 0 \right] \frac{B_0}{C_{\gamma_0}} \right] + \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_\gamma} > 0 \right] \frac{B_0}{C_{\gamma_0}} \right]. \end{aligned} \quad (3.13)$$

For the first term, the definition of γ_0 given above leads to

$$\begin{aligned} \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_\gamma} \leq 0 \right] \frac{B_0}{C_{\gamma_0}} \right] &= \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} \leq 0 \right] \frac{B_0}{C_{\nu_0^*}} \right] \\ &= \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} \leq 0 \right] \right] \mathbb{E} \left[\frac{B_0}{C_{\nu_0^*}} \right] \\ &= \mathbb{P} \left[E, \gamma = k, y + \frac{B}{C_k} \leq 0 \right] \cdot \hat{S}_0, \end{aligned} \quad (3.14)$$

since the collections $\{\mathbf{1}[E], \gamma, B\}$ and $\{\nu_0^*, B_0\}$ are independent under the enforced assumptions.

We further decompose the second term to obtain

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_\gamma} > 0 \right] \frac{B_0}{C_{\gamma_0}} \right] \\ &= \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} > 0 \right] \mathbf{1}[k \notin \Gamma_0] \frac{B_0}{C_{\nu_0^*}} \right] \\ &\quad + \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} > 0 \right] \mathbf{1}[k \in \Gamma_0] \frac{B_0}{C_{\gamma_0}} \right]. \end{aligned} \quad (3.15)$$

It is plain that

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} > 0 \right] \mathbf{1}[k \in \Gamma_0] \frac{B_0}{C_{\gamma_0}} \right] \\ &= \sum_{\ell=1, \ell \neq k}^K \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} > 0 \right] \mathbf{1}[\Gamma_0 = \{k, \ell\}] \frac{B_0}{C_{\gamma_0}} \right] \\ &= \sum_{\ell=1, \ell \neq k}^K \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} > 0 \right] \mathbf{1}[\Gamma_0 = \{k, \ell\}] \frac{B_0}{C_\ell} \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{2}{K(K-1)} \sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[E, \gamma = k, y + \frac{B}{C_k} > 0 \right] \frac{\mathbb{E}[B_0]}{C_\ell} \\
&= \frac{2}{K(K-1)} \left(\sum_{\ell=1, \ell \neq k}^K \frac{\mathbb{E}[B_0]}{C_\ell} \right) \mathbb{P} \left[E, \gamma = k, y + \frac{B}{C_k} > 0 \right] \\
&= \frac{2}{K(K-1)} \left(\Gamma - \frac{1}{C_k} \right) \mathbb{P} \left[E, \gamma = k, y + \frac{B}{C_k} > 0 \right] \cdot \mathbb{E}[B_0], \tag{3.16}
\end{aligned}$$

since $\gamma_0 = \ell$ if $\Gamma_0 = \{k, \ell\}$ when $\gamma = k$ and $y + \frac{B}{C_k} > 0$.

On the other hand, the definition of ν_0^* implies

$$\begin{aligned}
&\mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} > 0 \right] \mathbf{1}[k \notin \Gamma_0] \frac{B_0}{C_{\nu_0^*}} \right] \\
&= \sum_{T \in \mathcal{P}_2(K)} \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} > 0 \right] \mathbf{1}[k \notin \Gamma_0] \mathbf{1}[\Gamma_0 = T] \frac{B_0}{C_{\nu_0^*}} \right] \\
&= \sum_{a=1, a \neq k}^K \left(\sum_{b=1, b \neq k}^{a-1} \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} > 0 \right] \mathbf{1}[\Gamma_0 = \{a, b\}] \frac{B_0}{C_{\nu_0^*}} \right] \right) \\
&= \sum_{a=1, a \neq k}^K \left(\sum_{b=1, b \neq k}^{a-1} \mathbb{P} \left[E, \gamma = k, y + \frac{B}{C_k} > 0 \right] \mathbb{E} \left[\mathbf{1}[\Gamma_0 = \{a, b\}] \frac{B_0}{C_{\nu_0^*}} \right] \right) \\
&= \sum_{a=1, a \neq k}^K \mathbb{P} \left[E, \gamma = k, y + \frac{B}{C_k} > 0 \right] \left(\sum_{b=1, b \neq k}^{a-1} \mathbb{P}[\Gamma_0 = \{a, b\}] \cdot \frac{1}{2} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \right) \cdot \mathbb{E}[B_0] \\
&= \frac{1}{K(K-1)} \sum_{a=1, a \neq k}^K \mathbb{P} \left[E, \gamma = k, y + \frac{B}{C_k} > 0 \right] \left(\sum_{b=1, b \neq k}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \right) \cdot \mathbb{E}[B_0]. \tag{3.17}
\end{aligned}$$

In Appendix 3.A.1 we show that

$$\sum_{a=1, a \neq k}^K \left(\sum_{b=1, b \neq k}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \right) = (K-2) \left(\Gamma - \frac{1}{C_k} \right), \tag{3.18}$$

and (3.17) can be written more compactly as

$$\begin{aligned}
&\mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} > 0 \right] \mathbf{1}[k \notin \Gamma_0] \frac{B_0}{C_{\nu_0^*}} \right] \\
&= \frac{K-2}{K(K-1)} \left(\Gamma - \frac{1}{C_k} \right) \mathbb{P} \left[E, \gamma = k, y + \frac{B}{C_k} > 0 \right] \cdot \mathbb{E}[B_0]. \tag{3.19}
\end{aligned}$$

To conclude the proof, substitute (3.16) and (3.19) into (3.15), yielding

$$\mathbb{E} \left[\mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_\gamma} > 0 \right] \frac{B_0}{C_{\gamma_0}} \right] = \frac{1}{K-1} \left(\Gamma - \frac{1}{C_k} \right) \mathbb{P} \left[\gamma = k, y + \frac{B}{C_k} > 0 \right] \cdot \mathbb{E}[B_0].$$

Combining this last expression with (3.14) we get the desired result (3.12) with the help of (3.13). ■

3.5 First light-traffic derivative

The analysis of the first derivative is associated with the following scenario: The tagged customer arrives at time $t = 0$ with characteristic pair (B_0, Γ_0) . With t in $\mathbb{R}$, in addition to the tagged customer, only a single customer arrives during the entire horizon $(-\infty, \infty)$, say at time t with characteristic pair (B_t, Γ_t) . The tagged customer and this t -customer are assigned to the servers ν_0 (in Γ_0) and ν_t (in Γ_t), respectively, in accordance with the power-of-two scheduling policy.

3.5.1 Evaluating $\hat{S}_1(t)$

For each t in $\mathbb{R}$, we have

$$\hat{S}_1(t) = \mathbb{E}[S_1(t)] \quad \text{with } S_1(t) = \frac{B_0}{C_{\nu_0}}. \quad (3.20)$$

However, with the presence of the t -customer, ν_0 does not always coincide with ν_0^* , as the determination of ν_0 may be affected by whether the t -customer completed service at the time the tagged customer arrives.

First some notation: With t arbitrary in $\mathbb{R}$, set

$$H_k(t) = \mathbb{P}[C_k t + B_t \leq 0] \cdot \hat{S}_0 + \frac{1}{K-1} \left(\Gamma - \frac{1}{C_k} \right) \mathbb{P}[C_k t + B_t > 0] \cdot \mathbb{E}[B_0] \quad (3.21)$$

for each $k = 1, \dots, K$.

Proposition 3.5.1. *Under the enforced independence assumptions, it holds that*

$$\hat{S}_1(t) = \hat{S}_0, \quad t > 0 \quad (3.22)$$

and

$$\hat{S}_1(t) = \frac{1}{K} \sum_{k=1}^K H_k(t), \quad t < 0. \quad (3.23)$$

Proof: Fix t in $\mathbb{R}$. As we seek to evaluate $\hat{S}_1(t)$ as given by (3.20), two cases need to be examined: If $0 < t$, then $\nu_0 = \nu_0^*$, whence $S_1(t) = S_0$, and the conclusion (3.22) follows.

If $t < 0$, then $\nu_t = \nu_t^*$ and we are in the setting of Lemma 3.4.1 with $y = t$, $E = \Omega$, $\tau = B_t$ and $\gamma = \nu_t^*$ (so that $\gamma_0 = \nu_0$): For each $k = 1, \dots, K$, the expression (3.12) becomes

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1}[\nu_t^* = k] \frac{B_0}{C_{\nu_0}} \right] \\ &= \mathbb{P} \left[\nu_t^* = k, t + \frac{B_t}{C_k} \leq 0 \right] \cdot \hat{S}_0 + \frac{1}{K-1} \mathbb{P} \left[\nu_t^* = k, t + \frac{B_t}{C_k} > 0 \right] \left(\Gamma - \frac{1}{C_k} \right) \cdot \mathbb{E}[B_0] \\ &= \frac{H_k(t)}{K}, \end{aligned} \quad (3.24)$$

since the random variable ν_t^* is independent of B_t and uniformly distributed on $\{1, \dots, K\}$ (as pointed out in Proposition 3.3.1). The desired result (3.23) now follows from (3.20) upon noting the decomposition

$$\widehat{S}_1(t) = \sum_{k=1}^K \mathbb{E} \left[\mathbf{1}[\nu_t^* = k] \frac{B_0}{C_{\nu_0}} \right].$$

■

3.5.2 Proof of Equation (3.4)

We can now complete the proof of Equation (3.4). Our starting point is the expression

$$\left(\overline{S^{SQ2}} \right)^{(1)}(0) = \int_{\mathbb{R}} \left(\widehat{S}_1(t) - \widehat{S}_0 \right) dt = \int_{-\infty}^0 \left(\widehat{S}_1(t) - \widehat{S}_0 \right) dt \quad (3.25)$$

as we use (3.22). Next, for $t < 0$, with the help of (3.23), we can rewrite the integrand as

$$\begin{aligned} \widehat{S}_1(t) - \widehat{S}_0 &= \frac{1}{K} \sum_{k=1}^K H_k(t) - \widehat{S}_0 \\ &= -\frac{1}{K} \sum_{k=1}^K \left(\widehat{S}_0 - \frac{1}{K-1} \left(\Gamma - \frac{1}{C_k} \right) \cdot \mathbb{E}[B] \right) \mathbb{P}[C_k t + B > 0]. \end{aligned} \quad (3.26)$$

Inserting this expression back into (3.25) we get

$$\begin{aligned} \left(\overline{S^{SQ2}} \right)^{(1)}(0) &= \int_{-\infty}^0 \left(\widehat{S}_1(t) - \widehat{S}_0 \right) dt \\ &= -\frac{1}{K} \left(\sum_{k=1}^K \frac{\mathbb{E}[B]}{C_k} \right) \widehat{S}_0 + \frac{1}{K(K-1)} \sum_{k=1}^K \frac{\mathbb{E}[B]}{C_k} \left(\Gamma - \frac{1}{C_k} \right) \cdot \mathbb{E}[B] \end{aligned} \quad (3.27)$$

upon noting that

$$\int_{-\infty}^0 \mathbb{P}[C_k t + B > 0] dt = \frac{1}{C_k} \int_0^{\infty} \mathbb{P}[B > x] dx = \frac{\mathbb{E}[B]}{C_k}, \quad k = 1, \dots, K \quad (3.28)$$

by a simple change of variable. Uninteresting algebra readily yield (3.4) with the help of (3.9), and this completes the proof of Equation (3.4). ■

3.6 Second light-traffic derivative

The computation of the second derivative is given under the following scenario: The tagged customer arrives at time $t = 0$ with characteristic pair (B_0, Γ_0) . With s and t in $\mathbb{R}$, in addition to the tagged customer, exactly two customers arrive over the entire horizon $(-\infty, \infty)$, say at times s and t with characteristic pairs (B_s, Γ_s) and (B_t, Γ_t) , respectively. The tagged customer, the s -customer and the t -customer are assigned to their respective servers ν_0 (in Γ_0), ν_s (in Γ_s) and ν_t (in Γ_t) in accordance with the power-of-two load balancing scheduling policy.

3.6.1 Evaluating $\hat{S}_2(s, t)$

For each s and t in $\mathbb{R}$, we have

$$\hat{S}_2(s, t) = \mathbb{E}[S_2(s, t)], \quad \text{with } S_2(s, t) = \frac{B_0}{C_{\nu_0}}. \quad (3.29)$$

However, the server assignment random variables ν_0 , ν_s and ν_t do not always coincide with ν_0^* , ν_s^* and ν_t^* , respectively. This is due to the fact that these random variables may be affected by whether earlier customers have completed service by the time server selection needs to be determined.

Proposition 3.6.1. *Under the enforced independence assumptions, we have*

$$\hat{S}_2(s, t) = \hat{S}_0, \quad 0 < s < t \quad (3.30)$$

and

$$\hat{S}_2(s, t) = \hat{S}_1(s) \quad s < 0 < t. \quad (3.31)$$

For $s < t < 0$, it holds that

$$\begin{aligned} \hat{S}_2(s, t) = & \left(\frac{1}{K} \sum_{k=1}^K \mathbb{P} \left[s + \frac{B_s}{C_k} \leq t \right] \right) \cdot \hat{S}_1(t) \\ & + \frac{1}{K(K-1)} \sum_{k=1}^K \left(\sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \right) \cdot H_k(t) \\ & + \frac{1}{K(K-1)^2} \sum_{k=1}^K \sum_{\ell=1, \ell \neq k}^K \left(\Gamma - \frac{1}{C_\ell} \right) \mathbb{P}[C_\ell s + B_s > 0] \mathbb{P}[C_k t + B_t \leq 0] \cdot \mathbb{E}[B_0] \\ & + \frac{1}{K^2(K-1)^2} \sum_{k=1}^K \sum_{\ell=1, \ell \neq k}^K \Gamma_{k\ell} \mathbb{P}[C_k s + B_s > 0] \mathbb{P}[C_\ell t + B_t > 0] \cdot \mathbb{E}[B_0], \end{aligned} \quad (3.32)$$

with

$$\Gamma_{k\ell} = (K+1)\Gamma - K \left(\frac{1}{C_k} + \frac{1}{C_\ell} \right), \quad k, \ell = 1, \dots, K. \quad (3.33)$$

Proof: The proof of Proposition 3.6.1 starts in Section 3.6.2. ■

We now obtain a more compact expression for (3.32).

As we focus on the last two terms in (3.32) we readily check that their sum is given

by

$$\begin{aligned}
& \frac{1}{K(K-1)^2} \sum_{k=1}^K \sum_{\ell=1, \ell \neq k}^K \left(\Gamma - \frac{1}{C_\ell} \right) \mathbb{P}[C_\ell s + B_s > 0] \mathbb{P}[C_k t + B_t \leq 0] \cdot \mathbb{E}[B_0] \\
& + \frac{1}{K^2(K-1)^2} \sum_{k=1}^K \sum_{\ell=1, \ell \neq k}^K \Gamma_{k\ell} \cdot \mathbb{P}[C_k s + B_s > 0] \mathbb{P}[C_\ell t + B_t > 0] \cdot \mathbb{E}[B_0] \\
& = \frac{1}{K^2(K-1)^2} \sum_{k=1}^K \sum_{\ell=1, \ell \neq k}^K G_{k\ell}(s, t) \cdot \mathbb{E}[B_0], \tag{3.34}
\end{aligned}$$

with

$$\begin{aligned}
G_{k\ell}(s, t) &= K \left(\Gamma - \frac{1}{C_k} \right) \mathbb{P}[C_k s + B_s > 0] \mathbb{P}[C_\ell t + B_t \leq 0] \\
&+ \left((K+1)\Gamma - K \left(\frac{1}{C_k} + \frac{1}{C_\ell} \right) \right) \cdot \mathbb{P}[C_k s + B_s > 0] \mathbb{P}[C_\ell t + B_t > 0] \\
&= K \left(\Gamma - \frac{1}{C_k} \right) \mathbb{P}[C_k s + B_s > 0] \\
&+ K \left(\frac{\Gamma}{K} - \frac{1}{C_\ell} \right) \mathbb{P}[C_k s + B_s > 0] \mathbb{P}[C_\ell t + B_t > 0], \tag{3.35}
\end{aligned}$$

for every $k, \ell = 1, \dots, K$. Upon substitution into (3.32), we then conclude that

$$\begin{aligned}
\hat{S}_2(s, t) &= \left(\frac{1}{K} \sum_{k=1}^K \mathbb{P} \left[s + \frac{B_s}{C_k} \leq t \right] \right) \cdot \hat{S}_1(t) \\
&+ \frac{1}{K(K-1)} \sum_{k=1}^K \left(\sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \right) \cdot H_k(t) \\
&+ \frac{1}{K(K-1)^2} \sum_{k=1}^K \sum_{\ell=1, \ell \neq k}^K \left(\frac{\Gamma}{K} - \frac{1}{C_\ell} \right) \mathbb{P}[C_k s + B_s > 0] \mathbb{P}[C_\ell t + B_t > 0] \cdot \mathbb{E}[B_0] \\
&+ \frac{1}{K(K-1)} \sum_{k=1}^K \left(\Gamma - \frac{1}{C_k} \right) \mathbb{P}[C_k s + B_s > 0] \cdot \mathbb{E}[B_0]. \tag{3.36}
\end{aligned}$$

Next, for each $k = 1, \dots, K$, we note that

$$\begin{aligned}
H_k(t) &= \hat{S}_0 \cdot (1 - \mathbb{P}[C_k t + B_t > 0]) + \frac{1}{K-1} \left(\Gamma - \frac{1}{C_k} \right) \mathbb{E}[B_0] \cdot \mathbb{P}[C_k t + B_t > 0] \\
&= \hat{S}_0 + \left(\frac{1}{K-1} \left(\Gamma - \frac{1}{C_k} \right) - \frac{\Gamma}{K} \right) \mathbb{E}[B_0] \cdot \mathbb{P}[C_k t + B_t > 0] \\
&= \hat{S}_0 + \frac{1}{K-1} \left(\frac{\Gamma}{K} - \frac{1}{C_k} \right) \mathbb{P}[C_k t + B_t > 0] \cdot \mathbb{E}[B_0], \tag{3.37}
\end{aligned}$$

as we make use of the expression (3.9). Therefore,

$$\begin{aligned}
& \sum_{k=1}^K \left(\sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \right) \cdot H_k(t) \\
&= \sum_{\ell=1}^K \left(\sum_{k=1, k \neq \ell}^K \mathbb{P} \left[t < s + \frac{B_s}{C_k} \leq 0 \right] \right) \cdot H_\ell(t) \\
&= \sum_{\ell=1}^K \left(\sum_{k=1, k \neq \ell}^K \mathbb{P} \left[t < s + \frac{B_s}{C_k} \leq 0 \right] \right) \left(\hat{S}_0 + \frac{1}{K-1} \left(\frac{\Gamma}{K} - \frac{1}{C_\ell} \right) \mathbb{P}[C_\ell t + B_t > 0] \cdot \mathbb{E}[B_0] \right),
\end{aligned}$$

and the second term in (3.36) becomes

$$\begin{aligned}
& \frac{1}{K(K-1)} \sum_{k=1}^K \left(\sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \right) \cdot H_k(t) \\
&= \frac{1}{K(K-1)} \sum_{\ell=1}^K \left(\sum_{k=1, k \neq \ell}^K \mathbb{P} \left[t < s + \frac{B_s}{C_k} \leq 0 \right] \right) \hat{S}_0 \\
&+ \frac{1}{K(K-1)^2} \sum_{\ell=1}^K \sum_{k=1, k \neq \ell}^K \left(\frac{\Gamma}{K} - \frac{1}{C_\ell} \right) \mathbb{P} \left[t < s + \frac{B_s}{C_k} \leq 0 \right] \mathbb{P}[C_\ell t + B_t > 0] \cdot \mathbb{E}[B_0].
\end{aligned}$$

Substituting this last expression into (3.36) we readily get the following more compact expression for the second half of Proposition 3.6.1.

Proposition 3.6.2. *Under the enforced independence assumptions, for $s < t < 0$, it holds that*

$$\begin{aligned}
\hat{S}_2(s, t) &= \left(\frac{1}{K} \sum_{\ell=1}^K \mathbb{P} \left[s + \frac{B_s}{C_\ell} \leq t \right] \right) \cdot \hat{S}_1(t) + \left(\frac{1}{K} \sum_{\ell=1}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \right) \cdot \hat{S}_0 \\
&+ H(s, t) + \frac{1}{K(K-1)} \sum_{k=1}^K \left(\frac{\Gamma}{K} - \frac{1}{C_k} \right) \mathbb{P}[C_k s + B_s > 0] \cdot \mathbb{E}[B_0]. \quad (3.38)
\end{aligned}$$

where we have set

$$\begin{aligned}
& H(s, t) \\
&= \frac{1}{K(K-1)^2} \sum_{\ell=1}^K \sum_{k=1, k \neq \ell}^K \left(\frac{\Gamma}{K} - \frac{1}{C_\ell} \right) \mathbb{P}[C_k t < C_k s + B_s] \mathbb{P}[C_\ell t + B_t > 0] \cdot \mathbb{E}[B_0].
\end{aligned} \tag{3.39}$$

3.6.2 Proof of Proposition 3.6.1

The cases $0 < s < t$ and $s < 0 < t$ are straightforward by virtue of the operational assumptions of the power-of-two load balancing policy. Indeed, when $0 < s < t$, $\nu_0 = \nu_0^*$, hence $S_2(s, t) = S_0$ and (3.30) holds. On the other hand, when $s < 0 < t$, the future t -customer does not affect the selection of ν_0 , hence has no impact on the performance of the tagged customer. As only the s -customer can possibly affect the choice of ν_0 , we get $S_2(s, t) = S_1(s)$ and (3.31) is established.

From now on we assume $s < t < 0$, in which case we have $\nu_s = \nu_s^*$. The selection of ν_t can in principle be affected by whether the s -customer has completed its service by time t , while that of ν_0 will be determined by whether the s -customer and t -customer have completed service by the time the tagged customer enters the system. Therefore, as the s -customer completes at time $s + \frac{B_s}{C_{\nu_s^*}}$, several possibilities arise; they are captured in the decomposition

$$\begin{aligned} \mathbb{E}[S_2(s, t)] &= \mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} \leq t \right] S_2(s, t) \right] \\ &\quad + \mathbb{E} \left[\mathbf{1} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0 \right] S_2(s, t) \right] \\ &\quad + \mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} > 0 \right] \mathbf{1} \left[t + \frac{B_t}{C_{\nu_t}} \leq 0 \right] S_2(s, t) \right] \\ &\quad + \mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} > 0 \right] \mathbf{1} \left[t + \frac{B_t}{C_{\nu_t}} > 0 \right] S_2(s, t) \right]. \end{aligned} \quad (3.40)$$

These four terms are evaluated separately in the next four lemmas.

Lemma 3.6.3. *With $s < t < 0$, we have*

$$\mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} \leq t \right] S_2(s, t) \right] = \left(\frac{1}{K} \sum_{k=1}^K \mathbb{P} \left[s + \frac{B_s}{C_k} \leq t \right] \right) \cdot \hat{S}_1(t). \quad (3.41)$$

Proof: When $s + \frac{B_s}{C_{\nu_s^*}} \leq t$, the s -customer will have completed service by the time the t -customer arrives. Therefore, conditionally on $s + \frac{B_s}{C_{\nu_s^*}} \leq t$, it holds that $S_2(s, t) =_{st} S_1(t)$, whence

$$\mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} \leq t \right] S_2(s, t) \right] = \mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} \leq t \right] S_1(t) \right] = \mathbb{P} \left[s + \frac{B_s}{C_{\nu_s^*}} \leq t \right] \cdot \hat{S}_1(t) \quad (3.42)$$

with

$$\mathbb{P} \left[s + \frac{B_s}{C_{\nu_s^*}} \leq t \right] = \frac{1}{K} \sum_{k=1}^K \mathbb{P} \left[s + \frac{B}{C_k} \leq t \right],$$

by the usual arguments. This completes the proof of (3.41). ■

Lemma 3.6.4. *With $s < t < 0$, we have*

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0 \right] S_2(s, t) \right] \\ &= \frac{1}{K(K-1)} \sum_{k=1}^K \left(\sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \right) \cdot H_k(t), \end{aligned} \quad (3.43)$$

with $H_k(t)$ given by (3.21) for all $k = 1, \dots, K$.

Proof: When $t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0$, the s -customer has not completed its service by time t , but will have completed it by the time the tagged customer arrives. Thus, only the t -customer can affect the definition of ν_0 (through B_t and ν_t).

With this in mind, consider the decomposition

$$\mathbb{E} \left[\mathbf{1} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0 \right] S_2(s, t) \right] = \sum_{k=1}^K \mathbb{E} \left[\mathbf{1} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0 \right] \mathbf{1} [\nu_t = k] \frac{B_0}{C_{\nu_0}} \right]. \quad (3.44)$$

Fix $k = 1, \dots, K$. We are in the setting of Lemma 3.4.1 with $y = t$, $E = [t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0]$, $\tau = B_t$ and $\gamma = \nu_t$ so that $\gamma_0 = \nu_0$: The expression (3.12) becomes

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0 \right] \mathbf{1} [\nu_t = k] \frac{B_0}{C_{\nu_0}} \right] \\ &= \mathbb{P} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0, \nu_t = k, t + \frac{B_t}{C_k} \leq 0 \right] \cdot \hat{S}_0 \\ &\quad + \frac{1}{K-1} \mathbb{P} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0, \nu_t = k, t + \frac{B_t}{C_k} > 0 \right] \left(\Gamma - \frac{1}{C_k} \right) \cdot \mathbb{E}[B_0] \\ &= \mathbb{P} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0, \nu_t = k \right] H_k(t), \end{aligned} \quad (3.45)$$

with $H_k(t)$ defined at (3.21). In the last step we used the fact that under the enforced independence assumptions, the random variable B_t is independent of the random variables $\{B_s, \nu_s^*, \nu_t\}$ when ν_t is generated by the power-of-two load balancing policy.

In Lemma 3.A.1 of Appendix 3.A.2 we show that

$$\mathbb{P} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0, \nu_t = k \right] = \frac{1}{K(K-1)} \sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right]. \quad (3.46)$$

Inserting (3.46) back into (3.45) yields

$$\mathbb{E} \left[\mathbf{1} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0 \right] \mathbf{1} [\nu_t = k] \frac{B_0}{C_{\nu_0}} \right] = \left(\frac{1}{K(K-1)} \sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \right) \cdot H_k(t),$$

and the desired result is now obtained by making use of (3.44). ■

The last two terms in the decomposition (3.40) are more cumbersome to evaluate. Their expressions are given in the next two lemmas whose proofs can be found in Appendix 3.A.3 and 3.A.4, respectively.

Lemma 3.6.5. *With $s < t < 0$, we have*

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} > 0 \right] \mathbf{1} \left[t + \frac{B_t}{C_{\nu_t}} \leq 0 \right] S_2(s, t) \right] \\ &= \frac{1}{K(K-1)^2} \sum_{k=1}^K \sum_{\ell=1, \ell \neq k}^K \left(\Gamma - \frac{1}{C_\ell} \right) \mathbb{P}[C_\ell s + B_s > 0] \mathbb{P}[C_k t + B_t \leq 0] \cdot \mathbb{E}[B_0]. \end{aligned} \quad (3.47)$$

Lemma 3.6.6. *With $s < t < 0$, we have*

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} > 0 \right] \mathbf{1} \left[t + \frac{B_t}{C_{\nu_t}} > 0 \right] S_2(s, t) \right] \\ &= \frac{1}{K^2(K-1)^2} \sum_{k=1}^K \sum_{\ell=1, \ell \neq k}^K \Gamma_{k\ell} \cdot \mathbb{P}[C_k s + B_s > 0] \mathbb{P}[C_\ell t + B_t > 0] \cdot \mathbb{E}[B_0], \end{aligned} \quad (3.48)$$

with the constants $\Gamma_{k\ell}$, $k, \ell = 1, \dots$, given by (3.33).

3.6.3 Proof of Equation (3.5)

For notational simplicity we shall write

$$R^*(s, t) = \widehat{S}_2(s, t) - \widehat{S}_1(s) - \widehat{S}_1(t) + \widehat{S}_0, \quad s, t \in \mathbb{R}.$$

We start with

$$\begin{aligned} \left(\overline{S^{SQ2}} \right)^{(2)}(0) &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} R^*(s, t) dt \right) ds \\ &= \int_{\mathbb{R}} \left(\int_{-\infty}^s R^*(s, t) dt \right) ds + \int_{\mathbb{R}} \left(\int_s^{\infty} R^*(s, t) dt \right) ds. \end{aligned} \quad (3.49)$$

Focusing on the second term in this last expression, we write

$$\int_{\mathbb{R}} \left(\int_s^{\infty} R^*(s, t) dt \right) ds = \int_{-\infty}^0 \left(\int_s^{\infty} R^*(s, t) dt \right) ds + \int_0^{\infty} \left(\int_s^{\infty} R^*(s, t) dt \right) ds$$

Now, by Propositions 3.5.1 and 3.6.1 we have

$$R^*(s, t) = \widehat{S}_0 - \widehat{S}_0 - \widehat{S}_0 + \widehat{S}_0 = 0, \quad 0 < s < t$$

and the conclusion

$$\int_0^\infty \left(\int_s^\infty R^*(s, t) ds \right) dt = 0$$

follows.

Next, consider the decomposition

$$\int_{-\infty}^0 \left(\int_s^\infty R^*(s, t) dt \right) ds = \int_{-\infty}^0 \left(\int_s^0 R^*(s, t) dt \right) ds + \int_{-\infty}^0 \left(\int_0^\infty R^*(s, t) dt \right) ds.$$

On the range $s < 0 < t$, Propositions 3.5.1 and 3.6.1 yield $\widehat{S}_2(s, t) = \widehat{S}_1(s)$ and $\widehat{S}_1(t) = \widehat{S}_0$, whence

$$R^*(s, t) = \widehat{S}_2(s, t) - \widehat{S}_1(t) - \widehat{S}_1(s) + \widehat{S}_0 = \widehat{S}_1(s) - \widehat{S}_0 - \widehat{S}_1(s) + \widehat{S}_0 = 0,$$

so that

$$\int_{-\infty}^0 \left(\int_0^\infty R^*(s, t) dt \right) ds = 0,$$

and we conclude to

$$\int_{\mathbb{R}} \left(\int_s^\infty R^*(s, t) dt \right) ds = \int_{-\infty}^0 \left(\int_s^0 R^*(s, t) dt \right) ds.$$

On the way to evaluating this last integral we consider $R^*(s, t)$ for $s < t < 0$. On that range, upon using first (3.38) and then (3.26) (with t replaced by s), we readily get

$$\begin{aligned} & R^*(s, t) \\ &= \widehat{S}_2(s, t) - \widehat{S}_1(s) - \widehat{S}_1(t) + \widehat{S}_0 \\ &= \widehat{S}_0 - \widehat{S}_1(s) - \left(\widehat{S}_1(t) - \frac{1}{K} \sum_{\ell=1}^K \mathbb{P} \left[s + \frac{B_s}{C_\ell} \leq t \right] \cdot \widehat{S}_1(t) \right) \\ &\quad + \left(\frac{1}{K} \sum_{\ell=1}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \right) \cdot \widehat{S}_0 + H(s, t) \\ &+ \frac{1}{K(K-1)} \sum_{k=1}^K \left(\Gamma - \frac{1}{C_k} \right) \mathbb{P}[C_k s + B_s > 0] \cdot \mathbb{E}[B_0] \\ &= \left(\frac{1}{K} \sum_{k=1}^K \mathbb{P}[C_k s + B_s > 0] \right) \widehat{S}_0 - \left(\frac{1}{K} \sum_{\ell=1}^K \mathbb{P} \left[s + \frac{B_s}{C_\ell} > t \right] \right) \cdot \widehat{S}_1(t) \\ &\quad + \left(\frac{1}{K} \sum_{k=1}^K \mathbb{P}[C_k t < C_k s + B_s \leq 0] \right) \cdot \widehat{S}_0 + H(s, t) \\ &= \left(\frac{1}{K} \sum_{k=1}^K \mathbb{P}[C_k t < C_k s + B_s] \right) \cdot \widehat{S}_0 - \left(\frac{1}{K} \sum_{k=1}^K \mathbb{P} \left[s + \frac{B_s}{C_k} > t \right] \right) \cdot \widehat{S}_1(t) + H(s, t) \\ &= \left(\frac{1}{K} \sum_{k=1}^K \mathbb{P}[C_k t < C_k s + B_s] \right) \cdot (\widehat{S}_0 - \widehat{S}_1(t)) + H(s, t). \end{aligned}$$

Next, after a change of order of integration and a change of variable we note that

$$\begin{aligned}
& \int_{-\infty}^0 \left(\int_s^0 \left(\frac{1}{K} \sum_{k=1}^K \mathbb{P}[C_k t < C_k s + B] \right) \cdot (\hat{S}_0 - \hat{S}_1(t)) dt \right) ds. \\
&= \frac{1}{K} \sum_{k=1}^K \int_{-\infty}^0 \left(\int_s^0 \mathbb{P}[C_k t < C_k s + B] \cdot (\hat{S}_0 - \hat{S}_1(t)) dt \right) ds \\
&= \frac{1}{K} \sum_{k=1}^K \int_{-\infty}^0 \left(\int_{-\infty}^t \mathbb{P}[C_k t < C_k s + B] \cdot (\hat{S}_0 - \hat{S}_1(t)) ds \right) dt \\
&= \frac{1}{K} \sum_{k=1}^K \int_{-\infty}^0 \left(\int_{-\infty}^t \mathbb{P}[C_k t < C_k s + B] ds \right) \cdot (\hat{S}_0 - \hat{S}_1(t)) dt
\end{aligned} \tag{3.50}$$

$$\begin{aligned}
&= \frac{1}{K} \sum_{k=1}^K \int_{-\infty}^0 \left(\int_0^\infty \mathbb{P}[C_k x < B] dx \right) \cdot (\hat{S}_0 - \hat{S}_1(t)) dt \\
&= \frac{1}{K} \sum_{k=1}^K \int_{-\infty}^0 \frac{\mathbb{E}[B]}{C_k} \cdot (\hat{S}_0 - \hat{S}_1(t)) dt \\
&= \left(\int_{-\infty}^0 (\hat{S}_0 - \hat{S}_1(t)) dt \right) \cdot \frac{\Gamma}{K} \mathbb{E}[B] \\
&= \left(\frac{1}{K} \left(\sum_{k=1}^K \frac{\mathbb{E}[B]}{C_k} \right) \hat{S}_0 - \frac{1}{K(K-1)} \sum_{k=1}^K \frac{\mathbb{E}[B]}{C_k} \left(\Gamma - \frac{1}{C_k} \right) \cdot \mathbb{E}[B] \right) \cdot \frac{\Gamma}{K} \mathbb{E}[B]
\end{aligned} \tag{3.51}$$

where the last step made used of the expression (3.27). Uninteresting calculations lead to

$$\begin{aligned}
& \left(\frac{1}{K} \left(\sum_{k=1}^K \frac{\mathbb{E}[B]}{C_k} \right) \hat{S}_0 - \frac{1}{K(K-1)} \sum_{k=1}^K \frac{\mathbb{E}[B]}{C_k} \left(\Gamma - \frac{1}{C_k} \right) \cdot \mathbb{E}[B] \right) \cdot \frac{\Gamma}{K} \mathbb{E}[B] \\
&= \left(\left(\frac{\Gamma}{K} \right)^2 - \frac{1}{K(K-1)} \left(\Gamma^2 - \sum_{k=1}^K \frac{1}{C_k^2} \right) \right) \cdot \frac{\Gamma}{K} (\mathbb{E}[B])^3 \\
&= \left(-\frac{\Gamma^2}{K^2(K-1)} + \frac{1}{K(K-1)} \sum_{k=1}^K \frac{1}{C_k^2} \right) \cdot \frac{\Gamma}{K} (\mathbb{E}[B])^3 \\
&= \left(-\left(\frac{\Gamma}{K} \right)^2 + \frac{1}{K} \sum_{k=1}^K \frac{1}{C_k^2} \right) \cdot \frac{\Gamma}{K(K-1)} (\mathbb{E}[B])^3.
\end{aligned} \tag{3.52}$$

In a similar vein, we find that

$$\int_{-\infty}^0 \left(\int_s^0 H(s, t) dt \right) ds = \frac{1}{K(K-1)^2} \sum_{\ell=1}^K \sum_{k=1, k \neq \ell}^K \left(\frac{\Gamma}{K} - \frac{1}{C_\ell} \right) I_{k\ell} \cdot \mathbb{E}[B] \tag{3.53}$$

with

$$\begin{aligned}
I_{k\ell} &= \int_{-\infty}^0 \left(\int_s^0 \mathbb{P}[C_k t < C_k s + B] \mathbb{P}[C_\ell t + B > 0] dt \right) ds \\
&= \int_{-\infty}^0 \left(\int_{-\infty}^t \mathbb{P}[C_k t < C_k s + B] ds \right) \mathbb{P}[C_\ell t + B > 0] dt \\
&= \int_{-\infty}^0 \left(\int_0^\infty \mathbb{P}[C_k x < B] dx \right) \mathbb{P}[C_\ell t + B > 0] dt \\
&= \left(\int_0^\infty \mathbb{P}[C_k x < B] dx \right) \left(\int_{-\infty}^0 \mathbb{P}[C_\ell t + B > 0] dt \right) \\
&= \frac{\mathbb{E}[B]}{C_k} \cdot \frac{\mathbb{E}[B]}{C_\ell}.
\end{aligned} \tag{3.54}$$

Therefore,

$$\int_{-\infty}^0 \left(\int_s^0 H(s, t) dt \right) ds = \frac{1}{K(K-1)^2} \sum_{\ell=1}^K \sum_{k=1, k \neq \ell}^K \left(\frac{\Gamma}{K} - \frac{1}{C_\ell} \right) \left(\frac{\mathbb{E}[B]}{C_k} \cdot \frac{\mathbb{E}[B]}{C_\ell} \right) \cdot \mathbb{E}[B]$$

with

$$\begin{aligned}
\sum_{\ell=1}^K \sum_{k=1, k \neq \ell}^K \frac{\mathbb{E}[B]}{C_k} \cdot \frac{\mathbb{E}[B]}{C_\ell} &= \sum_{\ell=1}^K \frac{\mathbb{E}[B]}{C_\ell} \left(\sum_{k=1, k \neq \ell}^K \frac{\mathbb{E}[B]}{C_k} \right) \\
&= \sum_{\ell=1}^K \frac{1}{C_\ell} \left(\Gamma - \frac{1}{C_\ell} \right) \cdot (\mathbb{E}[B])^2 \\
&= \left(\Gamma^2 - \sum_{\ell=1}^K \frac{1}{C_\ell^2} \right) \cdot (\mathbb{E}[B])^2
\end{aligned} \tag{3.55}$$

and

$$\begin{aligned}
\sum_{\ell=1}^K \sum_{k=1, k \neq \ell}^K \frac{1}{C_\ell} \left(\frac{\mathbb{E}[B]}{C_k} \cdot \frac{\mathbb{E}[B]}{C_\ell} \right) &= \sum_{\ell=1}^K \frac{\mathbb{E}[B]}{C_\ell^2} \left(\sum_{k=1, k \neq \ell}^K \frac{\mathbb{E}[B]}{C_k} \right) \\
&= \sum_{\ell=1}^K \frac{1}{C_\ell^2} \left(\Gamma - \frac{1}{C_\ell} \right) \cdot (\mathbb{E}[B])^2 \\
&= \left(\Gamma \sum_{\ell=1}^K \frac{1}{C_\ell^2} - \sum_{\ell=1}^K \frac{1}{C_\ell^3} \right) \cdot (\mathbb{E}[B])^2.
\end{aligned} \tag{3.56}$$

Collecting terms we conclude that

$$\begin{aligned}
&\int_{-\infty}^0 \left(\int_s^0 H(s, t) dt \right) ds \\
&= \frac{1}{K(K-1)^2} \left(\frac{\Gamma}{K} \left(\Gamma^2 - \sum_{\ell=1}^K \frac{1}{C_\ell^2} \right) - \left(\Gamma \sum_{\ell=1}^K \frac{1}{C_\ell^2} - \sum_{\ell=1}^K \frac{1}{C_\ell^3} \right) \right) \cdot (\mathbb{E}[B])^3 \\
&= \frac{1}{K(K-1)^2} \left(\frac{\Gamma^3}{K} - \frac{K+1}{K} \Gamma \sum_{\ell=1}^K \frac{1}{C_\ell^2} + \sum_{\ell=1}^K \frac{1}{C_\ell^3} \right) \cdot (\mathbb{E}[B])^3.
\end{aligned} \tag{3.57}$$

Finally,

$$\begin{aligned}
& \int_{-\infty}^0 \left(\int_s^0 R_2^*(s, t) dt \right) ds \\
&= \left(- \left(\frac{\Gamma}{K} \right)^2 + \frac{1}{K} \sum_{k=1}^K \frac{1}{C_k^2} \right) \cdot \frac{\Gamma}{K(K-1)} (\mathbb{E}[B])^3 \\
&\quad + \frac{1}{K(K-1)^2} \left(\frac{\Gamma^3}{K} - \frac{K+1}{K} \Gamma \sum_{\ell=1}^K \frac{1}{C_\ell^2} + \sum_{\ell=1}^K \frac{1}{C_\ell^3} \right) (\mathbb{E}[B])^3 \\
&= \left(-\frac{1}{K^3(K-1)} + \frac{1}{K^2(K-1)^2} \right) \Gamma^3 (\mathbb{E}[B])^3 \\
&\quad + \left(\frac{1}{K^2(K-1)} - \frac{K+1}{K^2(K-1)^2} \right) \left(\sum_{k=1}^K \frac{1}{C_k^2} \right) \Gamma (\mathbb{E}[B])^3 \\
&\quad + \frac{1}{K(K-1)^2} \left(\sum_{\ell=1}^K \frac{1}{C_\ell^3} \right) (\mathbb{E}[B])^3 \\
&= \frac{1}{(K-1)^2} (\dots) (\mathbb{E}[B])^3 \tag{3.58}
\end{aligned}$$

where

$$(\dots) = \left(\frac{\Gamma}{K} \right)^3 - 2 \left(\frac{1}{K} \sum_{k=1}^K \frac{1}{C_k^2} \right) \left(\frac{\Gamma}{K} \right) + \frac{1}{K} \sum_{\ell=1}^K \frac{1}{C_\ell^3} \tag{3.59}$$

In other words,

$$\begin{aligned}
& \int_{-\infty}^0 \left(\int_s^0 R_2^*(s, t) dt \right) ds \\
&= \frac{1}{(K-1)^2} \left(\left(\frac{\Gamma}{K} \right)^3 - 2 \left(\frac{1}{K} \sum_{k=1}^K \frac{1}{C_k^2} \right) \left(\frac{\Gamma}{K} \right) + \frac{1}{K} \sum_{\ell=1}^K \frac{1}{C_\ell^3} \right) (\mathbb{E}[B])^3. \tag{3.60}
\end{aligned}$$

3.7 Simulation results

In this section we analyse numerically the behaviour of the light-traffic approximation obtained in Proposition 3.2.1. As already pointed out by Reiman and Simon [47, 49], without additional information (e.g., heavy traffic information), we should not expect $\overline{SQ2,LT}(\lambda)$ to act as an accurate proxy for mean sojourn time in medium to heavy traffic. This lack of accuracy is certainly apparent in the simulation results reported below.

We have carried out simulations for different distributions of B , all with unit mean, namely hyperexponential (obtained by mixing the exponential random variables $\text{Exp}(1/2)$ and $\text{Exp}(2)$ with probability $1/3$ and $2/3$, respectively), exponential $\text{Exp}(1)$ (of parameter 1), Weibull (with shape parameter 2 and scale parameter $\Gamma(3/2)^{-1}$) and deterministic.

We next recall the distributions used:

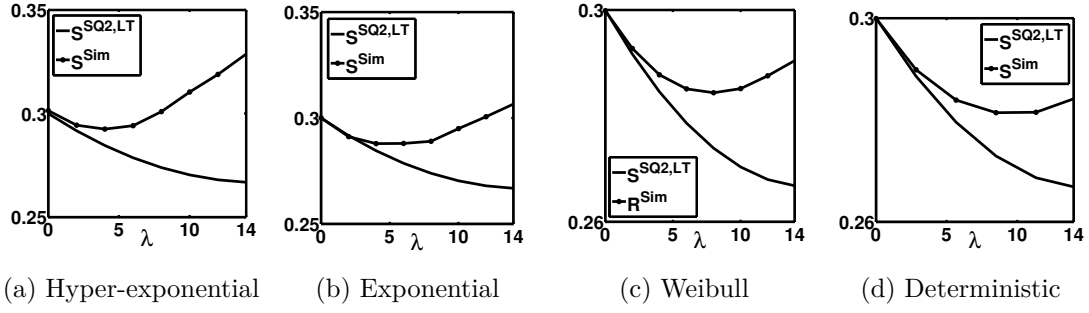


Figure 3.1: Scenario 1

- We say that B has an exponential distribution if

$$\mathbb{P}(B \leq b) = 1 - e^{-b/\mathbb{E}[B]}.$$

- We say that B has a hyper-exponential distribution with m phases if

$$\mathbb{P}(B \leq b) = 1 - \sum_{k=1}^m \beta_k e^{(-b/\mathbb{E}[B_k])},$$

with $\sum_{k=1}^m \beta_k = 1$, where β_k is the probability that a customer is exponentially distributed with mean $\mathbb{E}[B_k]$.

- We say that B has a Weibull distribution if

$$\mathbb{P}(B \leq b) = 1 - e^{(-b/\alpha)^k},$$

where $k > 0$ is the shape parameter and $\alpha > 0$ the scale parameter of the distribution.

The simulation results are based on averaging 10 runs with each run comprising 10^5 busy periods. A busy period is defined as the interval of time between two consecutive time epochs when the system becomes empty, such points being regenerative points for the stochastic process of interest. We have verified that the simulation results obtained for a system with $K = 100$ homogeneous servers and exponential service requirements agree with those given by Mitzenmacher [41, Table 1]. We are interested in the behaviour of the mean sojourn time in lightly loaded situations, and stability is therefore not a concern here as mentioned earlier. Three different scenarios were explored. In Scenarios 1 and 2 there are two types of servers, namely slow servers with capacity C_{slow} bytes/sec and fast servers with capacity C_{fast} bytes/sec. In Scenario 3 all the servers have the same capacity.

- Scenario 1: $K = 10$ servers. 5 slow servers with capacity $C_{\text{slow}} = 2$ bytes/sec and 5 fast servers with capacity $C_{\text{fast}} = 10$ bytes/sec. See Figure 3.1.

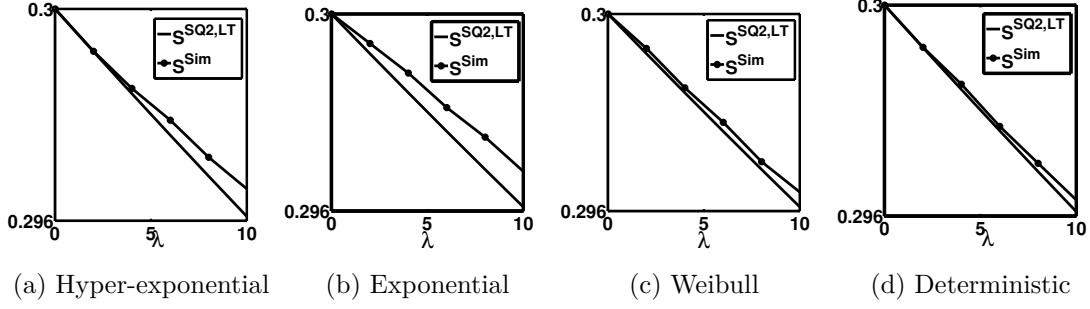


Figure 3.2: Scenario 2

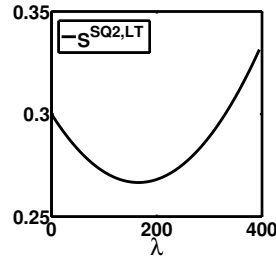


Figure 3.3: Scenario 2.

- Scenario 2: $K = 100$ servers. 50 slow servers with capacity $C_{\text{slow}} = 2$ bytes/sec and 50 fast servers with capacity $C_{\text{fast}} = 10$ bytes/sec. See Figure 3.2.
- Scenario 3: $K = 10$ servers with $C_1 = \dots = C_{10} = 10$ bytes/sec. See Figure 3.4.

In the figures we use S^{Sim} to denote the mean sojourn time obtained by simulation. Also the subscript p^* in the quantities $\left(\overline{S_{p^*}^{SQ2}}\right)^{(0)}$ and $\left(\overline{S_{p^*}^{SQ2}}\right)^{(1)}$ refers to Scenario p under $*$, where $*$ correspond with the hyper-exponential (H), exponential (E), Weibull (W) or deterministic (D) distribution. Let CV_{p^*} denote the coefficient of variation corresponding to Scenario p under $*$, with $*$ again equal to H, E, W or D distribution. Then $CV_{iH} = 1.4$, $CV_{iE} = 1$, $CV_{iW} = 0.52$ and $CV_{iD} = 0$, $\forall i = 1, 2, 3$. The simulations do confirm the aforementioned structural insights gleaned from the light traffic derivatives for non-homogeneous servers:

- (i) For all distributions, the mean sojourn time decreases as λ increases over a small neighbourhood of $\lambda = 0$,
- (ii) Over that small interval, performance seems nearly insensitive to the variability of B (as measured by its coefficient of variation).

Moreover, we observe that although in Scenario 1 and Scenario 2 there is an equal proportion of slow and fast servers, with $\left(\overline{S_{p^*}^{SQ2}}\right)^{(0)}(0) = 0.3000, \forall p = 1, 2, \forall * =$

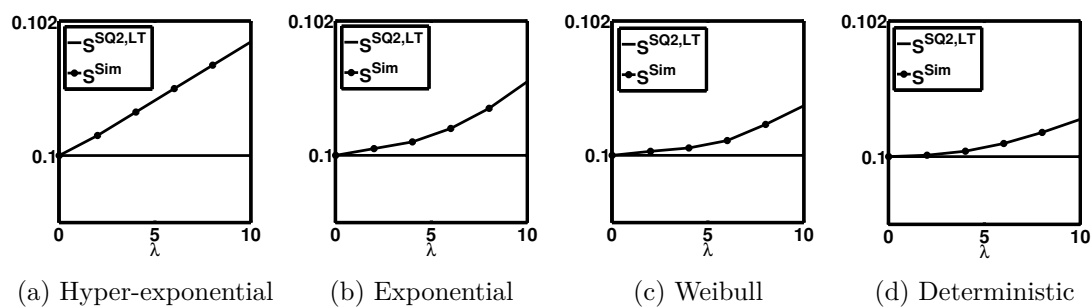


Figure 3.4: Scenario 3

H, E, W, D , the impact of the variability in server speeds is seen to diminish with increasing K since $\left(\overline{S_{1*}^{SQ2}}\right)^{(1)}(0) = -0.0044$ and $\left(\overline{S_{2*}^{SQ2}}\right)^{(1)}(0) = -4.0404 \cdot 10^{-4}, \forall * = H, E, W, D$.

Note that Figure 3.2 is a zoom of Figure 3.3. Although the sojourn time seems to be a straight line in a small interval of λ , after a while it increases. In Figure 3.4 we observe the aforementioned property for homogeneous servers; $\overline{S^{SQ2,LT}}(\lambda)$ becomes a constant line while the simulation results show that the mean sojourn time of a customer is increasing.

3.A Appendix

3.A.1 Proof of Equation (3.18)

Fix $k = 1, 2, \dots, K$. Elementary calculations give

$$\begin{aligned}
& \sum_{a=1, a \neq k}^K \left(\sum_{b=1, b \neq k}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \right) \\
&= \sum_{a=1}^{k-1} \left(\sum_{b=1, b \neq k}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \right) + \sum_{a=k+1}^K \left(\sum_{b=1, b \neq k}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \right) \\
&= \sum_{a=1}^{k-1} \left(\sum_{b=1}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \right) + \sum_{a=k+1}^K \left(-\frac{1}{C_a} - \frac{1}{C_k} + \sum_{b=1}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \right) \\
&= \sum_{a=1}^{k-1} \frac{a-1}{C_a} + \sum_{a=k+1}^K \frac{a-1}{C_a} + \sum_{a=1}^{k-1} \left(\sum_{b=1}^{a-1} \frac{1}{C_b} \right) + \sum_{a=k+1}^K \left(-\frac{1}{C_a} - \frac{1}{C_k} + \sum_{b=1}^{a-1} \frac{1}{C_b} \right) \\
&= \sum_{a=1}^K \frac{a-1}{C_a} - \frac{k-1}{C_k} + \sum_{a=1}^K \left(\sum_{b=1}^{a-1} \frac{1}{C_b} \right) - \sum_{b=1}^{k-1} \frac{1}{C_b} - \sum_{a=k+1}^K \left(\frac{1}{C_a} + \frac{1}{C_k} \right) \\
&= \sum_{a=1}^K \frac{a-1}{C_a} - \frac{k-1}{C_k} + \sum_{a=1}^K \left(\sum_{b=1}^{a-1} \frac{1}{C_b} \right) - \sum_{a=1}^K \frac{1}{C_a} - \frac{K-k-1}{C_k} \\
&= \sum_{a=1}^K \frac{a-1}{C_a} - \frac{K-2}{C_k} + \sum_{a=1}^K \left(\sum_{b=1}^{a-1} \frac{1}{C_b} \right) - \sum_{a=1}^K \frac{1}{C_a} \\
&= \sum_{a=1}^K \frac{a-1}{C_a} - \frac{K-2}{C_k} + \sum_{b=1}^{K-1} \left(\sum_{a=b+1}^K \frac{1}{C_b} \right) - \sum_{a=1}^K \frac{1}{C_a} \\
&= \sum_{a=1}^K \frac{a-1}{C_a} - \frac{K-2}{C_k} + \sum_{b=1}^{K-1} \frac{K-b}{C_b} - \sum_{a=1}^K \frac{1}{C_a} \\
&= (K-1)\Gamma - \frac{K-2}{C_k} - \Gamma
\end{aligned} \tag{3.61}$$

and the proof of (3.18) is complete. ■

3.A.2 Two auxiliary calculations

We present two results that will simplify some of the calculations.

Lemma 3.A.1. *With $s < t < 0$, we have*

$$\mathbb{P} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0, \nu_t = k \right] = \frac{1}{K(K-1)} \sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \tag{3.62}$$

for each $k = 1, \dots, K$.

Proof: Fix $k = 1, \dots, K$. Our point of departure is the obvious decomposition

$$\mathbb{P} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0, \nu_t = k \right] = \sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \nu_t = k \right]. \quad (3.63)$$

Pick $\ell = 1, \dots, K$ distinct from k , and note that

$$\begin{aligned} \mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \nu_t = k \right] &= \mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \ell \in \Gamma_t, \nu_t = k \right] \\ &\quad + \mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \ell \notin \Gamma_t, \nu_t = k \right]. \end{aligned} \quad (3.64)$$

We examine each term in turn: First, when ℓ belongs to Γ_t with $\nu_s^* = \ell$, then $\nu_t = k$ happens only if $\nu_s^* = \ell$ and $\Gamma_t = \{k, \ell\}$, whence

$$\begin{aligned} \mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \ell \in \Gamma_t, \nu_t = k \right] &= \mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \Gamma_t = \{k, \ell\}, \nu_t = k \right] \\ &= \mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \Gamma_t = \{k, \ell\} \right] \\ &= \frac{2}{K^2(K-1)} \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right]. \end{aligned} \quad (3.65)$$

Next, $\nu_t = \nu_t^*$ when ν_s^* is not in Γ_t , so that

$$\begin{aligned} &\mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \ell \notin \Gamma_t, \nu_t = k \right] \\ &= \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \ell \notin \Gamma_t, \Gamma_t = \{k, a\}, \nu_t^* = k \right] \\ &= \frac{1}{2} \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \Gamma_t = \{k, a\} \right] \\ &= \frac{1}{2} \sum_{a=1, a \neq k, a \neq \ell}^K \frac{1}{K} \frac{2}{K(K-1)} \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \\ &= \frac{1}{K^2(K-1)} \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \\ &= \frac{K-2}{K^2(K-1)} \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right]. \end{aligned} \quad (3.66)$$

Inserting (3.65) and (3.66) back into (3.64) we get

$$\mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \nu_t = k \right] = \frac{1}{K(K-1)} \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \quad (3.67)$$

and the desired conclusion (3.62) follows with the help of (3.63). $\blacksquare$

The exact same arguments lead to the following expression.

Lemma 3.A.2. *With $s < t < 0$, we have*

$$\mathbb{P} \left[s + \frac{B_s}{C_{\nu_s^*}} > 0, \nu_t = k \right] = \frac{1}{K(K-1)} \sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[s + \frac{B_s}{C_\ell} > 0 \right] \quad (3.68)$$

for each $k = 1, \dots, K$.

3.A.3 Proof of Lemma 3.6.5

If $s + \frac{B_s}{C_{\nu_s^*}} > 0$ (hence $s + \frac{B_s}{C_{\nu_s^*}} > t$), then the s -customer completes its service only after the tagged arrives, so that both the s -customer and t -customer can possibly affect the definition of ν_0 . If in addition we have $t + \frac{B_t}{C_{\nu_t}} \leq 0$, then only the s -customer can affect the selection ν_0 .

In the usual manner we have the decomposition

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} > 0 \right] \mathbf{1} \left[t + \frac{B_t}{C_{\nu_t}} \leq 0 \right] S_2(s, t) \right] \\ &= \sum_{k=1}^K \mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} > 0 \right] \mathbf{1} \left[t + \frac{B_t}{C_k} \leq 0 \right] \mathbf{1} [\nu_t = k] \frac{B_0}{C_{\nu_0}} \right] \\ &= \sum_{k=1}^K \sum_{\ell=1, \ell \neq k}^K \mathbb{E} \left[\mathbf{1} [\nu_s^* = \ell] \mathbf{1} \left[s + \frac{B_s}{C_\ell} > 0 \right] \mathbf{1} \left[t + \frac{B_t}{C_k} \leq 0 \right] \mathbf{1} [\nu_t = k] \frac{B_0}{C_{\nu_0}} \right]. \end{aligned} \quad (3.69)$$

Pick distinct $k, \ell = 1, \dots, K$. This time we apply Lemma 3.4.1 with $y = s$, $E = [s + \frac{B_s}{C_\ell} > 0, \nu_t = k, t + \frac{B_t}{C_k} \leq 0]$, $\tau = B_s$ and $\gamma = \nu_s^*$ (so that $\gamma_0 = \nu_0$). This leads to

$$\begin{aligned} & \mathbb{E} \left[\left(\mathbf{1} \left[s + \frac{B_s}{C_\ell} > 0 \right] \mathbf{1} \left[t + \frac{B_t}{C_k} \leq 0 \right] \mathbf{1} [\nu_t = k] \right) \mathbf{1} [\nu_s^* = \ell] \frac{B_0}{C_{\nu_0}} \right] \\ &= \mathbb{P} \left[s + \frac{B_s}{C_\ell} > 0, t + \frac{B_t}{C_k} \leq 0, \nu_t = k, \nu_s^* = \ell, s + \frac{B_s}{C_\ell} \leq 0 \right] \cdot \hat{S}_0 \\ &\quad + \frac{1}{K-1} \mathbb{P} \left[s + \frac{B_s}{C_\ell} > 0, t + \frac{B_t}{C_k} \leq 0, \nu_t = k, \nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0 \right] \left(\Gamma - \frac{1}{C_\ell} \right) \cdot \mathbb{E} [B_0] \\ &= \frac{1}{K-1} \mathbb{P} \left[s + \frac{B_s}{C_\ell} > 0, t + \frac{B_t}{C_k} \leq 0, \nu_t = k, \nu_s^* = \ell \right] \left(\Gamma - \frac{1}{C_\ell} \right) \cdot \mathbb{E} [B_0] \\ &= \frac{1}{K-1} \mathbb{P} \left[t + \frac{B_t}{C_k} \leq 0 \right] \mathbb{P} \left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \nu_t = k \right] \left(\Gamma - \frac{1}{C_\ell} \right) \cdot \mathbb{E} [B_0] \end{aligned} \quad (3.70)$$

since the random variables B_t is independent of the collection $\{\nu_s^*, B_s, \nu_t\}$ under the enforced independence assumptions. Next, we write

$$\begin{aligned} & \mathbb{P} \left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \nu_t = k \right] \\ &= \mathbb{P} \left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \ell \in \Gamma_t, \nu_t = k \right] + \mathbb{P} \left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \ell \notin \Gamma_t, \nu_t = k \right] \end{aligned} \quad (3.71)$$

Taking terms in turn we first get

$$\begin{aligned}
\mathbb{P}\left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \ell \in \Gamma_t, \nu_t = k\right] &= \mathbb{P}\left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \Gamma_t = \{\ell, k\}, \nu_t = k\right] \\
&= \mathbb{P}\left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \Gamma_t = \{\ell, k\}\right] \\
&= \frac{2}{K^2(K-1)} \mathbb{P}\left[s + \frac{B_s}{C_\ell} > 0\right] \tag{3.72}
\end{aligned}$$

since under the constraint $s + \frac{B_s}{C_\ell} > 0$, the fact that ν_s^* is an element of Γ_t forces ν_t to be the other element in Γ_t . In a similar way, under the constraint $s + \frac{B_s}{C_\ell} > 0$, ν_s^* not being in Γ_t implies $\nu_t = \nu_t^*$, and this leads to

$$\begin{aligned}
&\mathbb{P}\left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \ell \notin \Gamma_t, \nu_t = k\right] \\
&= \mathbb{P}\left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \ell \notin \Gamma_t, \nu_t^* = k\right] \\
&= \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{P}\left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \Gamma_t = \{a, k\}, \nu_t^* = k\right] \\
&= \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{P}[\nu_s^* = \ell] \mathbb{P}\left[s + \frac{B_s}{C_\ell} > 0\right] \frac{1}{2} \cdot \frac{2}{K(K-1)} \\
&= \frac{K-2}{K^2(K-1)} \mathbb{P}\left[s + \frac{B_s}{C_\ell} > 0\right]. \tag{3.73}
\end{aligned}$$

Collecting (3.72) and (3.73) gives

$$\mathbb{P}\left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \nu_t = k\right] = \frac{1}{K(K-1)} \mathbb{P}\left[s + \frac{B_s}{C_\ell} > 0\right], \tag{3.74}$$

and with the help of (3.70) we conclude that

$$\begin{aligned}
&\mathbb{E}\left[\left(\mathbf{1}\left[s + \frac{B_s}{C_\ell} > 0\right] \mathbf{1}\left[t + \frac{B_t}{C_k} \leq 0\right] \mathbf{1}[\nu_t = k]\right) \mathbf{1}[\nu_s^* = \ell] \frac{B_0}{C_{\nu_0}}\right] \\
&= \frac{1}{K(K-1)^2} \mathbb{P}\left[t + \frac{B_t}{C_k} \leq 0\right] \mathbb{P}\left[s + \frac{B_s}{C_\ell} > 0\right] \left(\Gamma - \frac{1}{C_\ell}\right) \cdot \mathbb{E}[B_0]. \tag{3.75}
\end{aligned}$$

Inserting this last expression into (3.69) we obtain (3.47) as desired. ■

3.A.4 Proof of Lemma 3.6.6

If $s + \frac{B_s}{C_{\nu_s^*}} > 0$ and $t + \frac{B_t}{C_{\nu_t}} > 0$, then ν_t is determined by the s -customer. At time $t = 0$, when the tagged customer arrives, both $\nu_s (= \nu_s^*)$ and ν_t would have already been

selected with both s -customer and t -customer still in service when ν_0 needs to be selected. Furthermore, $\nu_s^* \neq \nu_t$. In order to establish (3.48), we begin with the observation that

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} > 0 \right] \mathbf{1} \left[t + \frac{B_t}{C_{\nu_t}} > 0 \right] S_2(s; t) \right] \\ &= \mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} > 0 \right] \mathbf{1} \left[t + \frac{B_t}{C_{\nu_t}} > 0 \right] \frac{B_0}{C_{\nu_0}} \right] \\ &= \sum_{k=1}^K \sum_{\ell=1, \ell \neq k}^K \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \frac{B_0}{C_{\nu_0}} \right]. \quad (3.76) \end{aligned}$$

To take advantage of this decomposition, pick distinct $k, \ell = 1, \dots, K$. As we keep in mind whether ν_s^* and ν_t are in Γ_0 , we shall have to consider four possible cases: First, if both ν_s^* and ν_t are in Γ_0 , then $\nu_0 = \nu_s^*$ and we have

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [k \in \Gamma_0, \ell \in \Gamma_0] \frac{B_0}{C_{\nu_0}} \right] \\ &= \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [\Gamma_0 = \{k, \ell\}] \frac{B_0}{C_{\nu_0^*}} \right] \\ &= \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \mathbb{P} [\Gamma_0 = \{k, \ell\}] \frac{1}{2} \cdot \left(\frac{1}{C_k} + \frac{1}{C_\ell} \right) \mathbb{E} [B_0] \\ &= \frac{1}{K(K-1)} \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \left(\frac{1}{C_k} + \frac{1}{C_\ell} \right) \cdot \mathbb{E} [B_0]. \quad (3.77) \end{aligned}$$

Next, if ν_s^* is not in Γ_0 but ν_t is in Γ_0 , then ν_0 is the other element in Γ_0 , and we get

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [k \notin \Gamma_0, \ell \in \Gamma_0] \frac{B_0}{C_{\nu_0}} \right] \\ &= \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [\Gamma_0 = \{\ell, a\}] \frac{B_0}{C_{\nu_0}} \right] \\ &= \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [\Gamma_0 = \{\ell, a\}] \frac{B_0}{C_a} \right] \\ &= \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \mathbb{P} [\Gamma_0 = \{\ell, a\}] \frac{\mathbb{E} [B_0]}{C_a} \\ &= \frac{2}{K(K-1)} \left(\sum_{a=1, a \neq k, a \neq \ell}^K \frac{1}{C_a} \right) \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \mathbb{E} [B_0] \\ &= \frac{2}{K(K-1)} \left(\Gamma - \frac{1}{C_k} - \frac{1}{C_\ell} \right) \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \cdot \mathbb{E} [B_0]. \quad (3.78) \end{aligned}$$

In a similar way, if ν_s^* is in Γ_0 but ν_t is not in Γ_0 , then ν_0 is the other element in Γ_0 , and we get

$$\begin{aligned}
& \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [k \in \Gamma_0, \ell \notin \Gamma_0] \frac{B_0}{C_{\nu_0}} \right] \\
&= \sum_{b=1, b \neq k, b \neq \ell}^K \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [\Gamma_0 = \{k, b\}] \frac{B_0}{C_{\nu_0}} \right] \\
&= \sum_{b=1, b \neq k, b \neq \ell}^K \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [\Gamma_0 = \{k, b\}] \frac{B_0}{C_b} \right] \\
&= \sum_{b=1, b \neq k, b \neq \ell}^K \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \mathbb{P} [\Gamma_0 = \{b, k\}] \frac{\mathbb{E}[B_0]}{C_b} \\
&= \frac{2}{K(K-1)} \left(\sum_{b=1, b \neq k, b \neq \ell}^K \frac{1}{C_b} \right) \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \mathbb{E}[B_0] \\
&= \frac{2}{K(K-1)} \left(\Gamma - \frac{1}{C_k} - \frac{1}{C_\ell} \right) \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \cdot \mathbb{E}[B_0] \quad (3.79)
\end{aligned}$$

Finally, when neither ν_s^* and ν_t are in Γ_0 , then $\nu_0 = \nu_0^*$, whence

$$\begin{aligned}
& \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [k \notin \Gamma_0, \ell \notin \Gamma_0] \frac{B_0}{C_{\nu_0^*}} \right] \\
&= \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [k \notin \Gamma_0, \ell \notin \Gamma_0] \frac{B_0}{C_{\nu_0^*}} \right] \\
&= \sum_{a=1, a \neq k, a \neq \ell}^K \sum_{b=1, b \neq a, b \neq k, b \neq \ell}^{a-1} (\dots)_{k\ell} \quad (3.80)
\end{aligned}$$

with

$$\begin{aligned}
(\dots)_{k\ell} &= \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [\Gamma_0 = \{a, b\}] \frac{B_0}{C_{\nu_0^*}} \right] \\
&= \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \mathbb{P} [\Gamma_0 = \{a, b\}] \frac{1}{2} \cdot \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \mathbb{E}[B_0] \\
&= \frac{1}{K(K-1)} \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \cdot \mathbb{E}[B_0] \quad (3.81)
\end{aligned}$$

It then follows that

$$\begin{aligned}
& \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [k \notin \Gamma_0, \ell \notin \Gamma_0] \frac{B_0}{C_{\nu_0^*}} \right] \\
&= \frac{H_{k\ell}}{K(K-1)} \cdot \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \cdot \mathbb{E}[B_0] \quad (3.82)
\end{aligned}$$

where we have set

$$H_{k\ell} = \sum_{a=1, a \neq k, a \neq \ell}^K \left(\sum_{b=1, b \neq a, b \neq k, b \neq \ell}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \right).$$

Collecting terms (3.78)-(3.82), we conclude from (3.77) that

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \frac{B_0}{C_{\nu_0}} \right] \\ &= \frac{H_{k\ell}^*}{K(K-1)} \cdot \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \cdot \mathbb{E}[B_0] \end{aligned} \quad (3.83)$$

with

$$H_{k\ell}^* = H_{k\ell} + 4\Gamma - 3 \left(\frac{1}{C_k} + \frac{1}{C_\ell} \right).$$

In Appendix 3.A.5 we show that

$$\mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] = \frac{1}{K(K-1)} \mathbb{P} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbb{P} \left[t + \frac{B_t}{C_\ell} > 0 \right] \quad (3.84)$$

and the conclusion

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \frac{B_0}{C_{\nu_0}} \right] \\ &= \frac{H_{k\ell}^*}{K^2(K-1)^2} \cdot \mathbb{P} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbb{P} \left[t + \frac{B_t}{C_\ell} > 0 \right] \cdot \mathbb{E}[B_0] \end{aligned} \quad (3.85)$$

follows. In Appendix 3.A.6 we also show that

$$H_{k\ell} = (K-3) \left(\Gamma - \frac{1}{C_k} - \frac{1}{C_\ell} \right) \quad (3.86)$$

so that

$$\begin{aligned} H_{k\ell}^* &= H_{k\ell} + 4\Gamma - 3 \left(\frac{1}{C_k} + \frac{1}{C_\ell} \right) \\ &= (K-3) \left(\Gamma - \frac{1}{C_k} - \frac{1}{C_\ell} \right) + 4\Gamma - 3 \left(\frac{1}{C_k} + \frac{1}{C_\ell} \right) \\ &= (K+1)\Gamma - K \left(\frac{1}{C_k} + \frac{1}{C_\ell} \right). \end{aligned} \quad (3.87)$$

Inserting this last expression into (3.85) yields the desired conclusion (3.48). ■

3.A.5 Proof of Equation (3.84)

Fix distinct $k, \ell = 1, \dots, K$. We need to show that

$$\mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] = \frac{1}{K(K-1)} \mathbb{P} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbb{P} \left[t + \frac{B_t}{C_\ell} > 0 \right] \quad (3.88)$$

By arguments used earlier we get

$$\begin{aligned}
& \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, k \in \Gamma_t, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \\
&= \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \Gamma_t = \{k, \ell\}, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \\
&= \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \Gamma_t = \{k, \ell\}, t + \frac{B_t}{C_\ell} > 0 \right] \\
&= \frac{2}{K^2(K-1)} \mathbb{P} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbb{P} \left[t + \frac{B_t}{C_\ell} > 0 \right]
\end{aligned} \tag{3.89}$$

under the enforced independence assumptions.

In a similar way, we find

$$\begin{aligned}
& \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, k \notin \Gamma_t, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \\
&= \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \Gamma_t = \{a, \ell\}, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \\
&= \frac{1}{2} \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \Gamma_t = \{a, \ell\}, t + \frac{B_t}{C_\ell} > 0 \right] \\
&= \frac{1}{K^2(K-1)} \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{P} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbb{P} \left[t + \frac{B_t}{C_\ell} > 0 \right] \\
&= \frac{K-2}{K^2(K-1)} \mathbb{P} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbb{P} \left[t + \frac{B_t}{C_\ell} > 0 \right]
\end{aligned} \tag{3.90}$$

under the enforced independence assumptions. Collecting (3.89) and (3.90) we conclude to the validity of (3.88). $\blacksquare$

3.A.6 Proof of Equation (3.86)

To show (3.86) it suffices to establish this fact for $k = 1$ and $\ell = 2$. Thus,

$$\begin{aligned}
& \sum_{a=3}^K \sum_{b=3}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \\
&= \sum_{a=3}^K \left(\sum_{b=1}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) - \left(\frac{1}{C_a} + \frac{1}{C_1} + \frac{1}{C_a} + \frac{1}{C_2} \right) \right) \\
&= \sum_{a=3}^K \left(\frac{a-3}{C_a} + \sum_{b=1}^{a-1} \frac{1}{C_b} - \frac{1}{C_1} - \frac{1}{C_2} \right) \\
&= \sum_{a=3}^K \frac{a-3}{C_a} + \sum_{a=3}^K \left(\sum_{b=1}^{a-1} \frac{1}{C_b} \right) - (K-2) \left(\frac{1}{C_1} + \frac{1}{C_2} \right)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{a=3}^K \frac{a-3}{C_a} + \sum_{a=1}^K \left(\sum_{b=1}^{a-1} \frac{1}{C_b} \right) - (K-2) \left(\frac{1}{C_1} + \frac{1}{C_2} \right) - \frac{1}{C_1} \\
&= \sum_{a=1}^K \frac{a-3}{C_a} + \sum_{a=1}^K \left(\sum_{b=1}^{a-1} \frac{1}{C_b} \right) - (K-2) \left(\frac{1}{C_1} + \frac{1}{C_2} \right) - \frac{1}{C_1} - \left(-\frac{2}{C_1} - \frac{1}{C_2} \right) \\
&= \sum_{a=1}^K \frac{a-3}{C_a} + \sum_{a=1}^K \left(\sum_{b=1}^{a-1} \frac{1}{C_b} \right) - \frac{K-3}{C_1} - \frac{K-3}{C_2} \\
&= \sum_{a=1}^K \frac{a-3}{C_a} + \sum_{b=1}^{K-1} \left(\sum_{a=b+1}^K \frac{1}{C_b} \right) - \frac{K-3}{C_1} - \frac{K-3}{C_2} \\
&= \sum_{a=1}^K \frac{a-3}{C_a} + \sum_{b=1}^{K-1} \frac{K-b}{C_b} - \frac{K-3}{C_1} - \frac{K-3}{C_2} \\
&= \sum_{a=1}^{K-1} \frac{K-3}{C_a} + \frac{K-3}{C_K} - \frac{K-3}{C_1} - \frac{K-3}{C_2}, \tag{3.91}
\end{aligned}$$

whence

$$\sum_{a=3}^K \sum_{b=3}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) = (K-3) \left(\frac{1}{C_1} + \frac{1}{C_2} \right)$$

as desired.

Chapter 4

Interpolation approximations for a discriminatory processor sharing queue

In this chapter we focus on the Discriminatory Processor Sharing (DPS) policy, described in Section 1.1.1, and are interested in providing characterisations of the queue length vector and conditional (on the service requirement) waiting time in steady state. Due to intractability of the system in general, we apply the interpolation approximation technique, described in Chapter 2, and obtain closed-form approximations for the distribution of these performance metrics.

We first establish the light-traffic approximation of both the probability generating function of the stationary queue length vector and of the complementary distribution function of the conditional waiting time. Combining the so-obtained light-traffic approximations with the heavy-traffic characterisations we build the interpolation approximations, which hold for general service requirements.

This chapter is organised as follows. In Section 4.1 we provide a detailed model description. In Section 4.2 we gather results from [13], [55] and [18] that will be used in this chapter. In Section 4.3 we obtain the interpolation approximation of the probability generating function of the queue length vector. The interpolation approximation of the complementary distribution function of the waiting time is presented in Section 4.4. Section 4.5 presents the approximation results for the particular case of the mean sojourn time. In Section 4.6 we numerically test the accuracy of the obtained approximations.

4.1 Model description

We consider a multi-class single-server queue with K classes of customers. Class- k customers, $k = 1, \dots, K$, arrive according to independent Poisson processes with rate $\lambda_k \geq 0$. We denote the overall arrival rate by $\lambda = \sum_{k=1}^K \lambda_k$. A class- k customer has i.i.d generally distributed service requirements B_k and we assume $\mathbb{E}[B_k^2] < \infty$, $k = 1, \dots, K$.

The traffic intensity for class- k customers is denoted by $\rho_k := \lambda_k \mathbb{E}[B_k]$ and the total traffic intensity is denoted by

$$\rho := \sum_{k=1}^K \rho_k = \sum_{k=1}^K \lambda_k \mathbb{E}[B_k] = \lambda \sum_{k=1}^K \alpha_k \mathbb{E}[B_k] = \lambda \mathbb{E}[B],$$

where $\alpha_k = \lambda_k / \lambda$ denotes the probability that an arrival is of class k and the random variable B is the service requirement of an arbitrary arriving customer.

Under the DPS policy, the K customer classes share a common resource of capacity one. There are strictly positive weights $g_1, \dots, g_K$ associated with each of the classes. Whenever there are n_k class- k customers, $k = 1, \dots, K$, in the system, each class- k customer is served at rate

$$\frac{g_k}{\sum_{j=1}^K n_j g_j}.$$

We denote by N_k^{DPS} the steady-state number of class- k customers in the system at arbitrary epochs. The function $\psi^{DPS}(\lambda, \vec{z})$, with $\vec{z} = (z_1, \dots, z_K)$, denotes the probability generating function of $(N_1^{DPS}, \dots, N_K^{DPS})$. The waiting time of a customer is defined as its sojourn time in the system minus its service requirement. We denote the conditional (on the service requirement b) and unconditional waiting time of an arbitrary class- k customer by $W_k(b)$ and W_k , respectively, and let $W_k(\lambda, b, x) := \mathbb{P}[W_k(b) > x]$ be the complementary distribution function. The conditional (on the service requirement b) and unconditional sojourn time of an arbitrary class- k customer are denoted by $S_k^{DPS}(b)$ and S_k^{DPS} , respectively. We further define $\overline{S}_k^{DPS}(\lambda, b) := \mathbb{E}[S_k^{DPS}(b)]$ and $\overline{S}_k^{DPS}(\lambda) := \mathbb{E}[S_k^{DPS}]$, in order to reflect the dependence on λ .

4.2 Preliminaries

As mentioned earlier in the chapter exact analysis of DPS for general service requirements is out of reach. In this section we provide known results that are used later on in this chapter. As mentioned earlier, in [13] the authors obtained that the derivatives of the mean conditional sojourn times of the various classes satisfy a system of integro-differential equations. We use the results of [13] in order to numerically evaluate the accuracy of our approximation. In Section 4.2.1 we provide further details on them. In Section 4.2.2 we present heavy-traffic results available for the DPS policy that we will use in order to obtain our interpolation approximations.

4.2.1 Mean conditional sojourn time

In [13] Fayolle et al. obtained that the derivatives of the mean conditional sojourn time of a class- k customer with service requirement b , $\overline{S}_k^{DPS}(\lambda, b)$, satisfies the following system

of integro-differential equations:

$$\begin{aligned} & \frac{\partial \overline{S}_k^{DPS}(\lambda, b)}{\partial b} \\ &= 1 + \lambda \sum_{j=1}^K \int_0^\infty \alpha_j \frac{g_j}{g_k} \frac{\partial \overline{S}_j^{DPS}(\lambda, y)}{\partial y} \left[1 - F_j \left(y + \frac{g_j}{g_k} b \right) \right] dy \\ & \quad + \lambda \int_0^b \frac{\partial \overline{S}_k^{DPS}(\lambda, y)}{\partial y} \sum_{j=1}^K \alpha_j \frac{g_j}{g_k} \left[1 - F_j \left(\frac{g_j}{g_k} (b - y) \right) \right] dy, \end{aligned} \quad (4.1)$$

for $k = 1, \dots, K$. The natural boundary conditions are $\overline{S}_k^{DPS}(\lambda, 0) = 0$, $k = 1, \dots, K$.

The only known analytical solution for this system of equations has been obtained under the assumption of exponentially distributed service requirements. In this case we denote by $\mu_j := 1/\mathbb{E}[B_j]$, $\forall j$. In [13] it is proved that

$$\overline{S}_k^{DPS}(\lambda, b) = \frac{b}{1 - \rho} + \sum_{j=1}^m \frac{g_k c_j \beta_j + d_j}{\beta_j^2} \left(1 - e^{-\beta_j b / g_k} \right), \quad (4.2)$$

where $-\beta_j$, $j = 1, 2, \dots, m$, are the m distinct negative roots of

$$\sum_{j=1}^K \frac{\lambda_j g_j}{\mu_j g_j + s} = 1, \quad (4.3)$$

and where c_j and d_j , $j = 1, \dots, m$, are a function of the input parameters and β_j , $j = 1, \dots, m$.

Furthermore, for the mean unconditional sojourn time with exponentially distributed service requirements, it is shown in [13] that $\overline{S}_k^{DPS}(\lambda)$, $k = 1, \dots, K$, is the unique solution of the following system of equations:

$$\overline{S}_k^{DPS}(\lambda) \left(1 - \sum_{j=1}^K \frac{\lambda_j g_j}{\mu_j g_j + \mu_k g_k} \right) - \sum_{j=1}^K \frac{\lambda_j g_j \overline{S}_j^{DPS}(\lambda)}{\mu_j g_j + \mu_k g_k} = \frac{1}{\mu_k}. \quad (4.4)$$

A closed-form solution for this system of equations (4.4) is available only for the case of $K = 2$, and is given by

$$\overline{S}_1^{DPS}(\lambda) = \frac{1}{\mu_1(1 - \rho)} \left(1 + \frac{\mu_1 \rho_2 (g_2 - g_1)}{D} \right) \quad (4.5)$$

and

$$\overline{S}_2^{DPS}(\lambda) = \frac{1}{\mu_2(1 - \rho)} \left(1 + \frac{\mu_2 \rho_1 (g_1 - g_2)}{D} \right), \quad (4.6)$$

where $D = \mu_1 g_1 (1 - \rho_1) + \mu_2 g_2 (1 - \rho_2)$.

4.2.2 Heavy-traffic results

In this section we present the heavy-traffic results that correspond to the queue-length and conditional waiting time distributions.

As it is stated in the following proposition, a state-space collapse for the scaled queue-length vector appears in heavy-traffic. That is, in the limit the scaled queue-length vector is distributed as the product of an exponentially distributed random variable and a deterministic vector.

Proposition 4.2.1. [55, Proposition 2.1.] *The scaled probability generating function of the stationary queue length vector, $\psi^{DPS}(\lambda, \bar{z}^{(1-\lambda\mathbb{E}[B])})$, under the heavy-traffic regime satisfies*

$$\begin{aligned} \lim_{\lambda \uparrow 1/\mathbb{E}[B]} \psi^{DPS}(\lambda, \bar{z}^{(1-\lambda\mathbb{E}[B])}) &= \lim_{\lambda \uparrow 1/\mathbb{E}[B]} \mathbb{E} \left[z_1^{(1-\lambda\mathbb{E}[B])N_1^{DPS}} \cdots z_K^{(1-\lambda\mathbb{E}[B])N_K^{DPS}} \right] \\ &= \frac{\mathbb{E}[B]/\mathbb{E}[Y]}{\mathbb{E}[B]/\mathbb{E}[Y] - \sum_{i=1}^K \frac{\alpha_i \mathbb{E}[B_i]}{g_i} \ln(z_i)}, \end{aligned} \quad (4.7)$$

where

$$\mathbb{E}[Y] = \frac{\mathbb{E}[B^2]}{\mathbb{E}[B] \sum_{k=1}^K \alpha_k \mathbb{E}[B_k^2] / g_k}. \quad (4.8)$$

Or in other words, as $\lambda \uparrow 1/\mathbb{E}[B]$,

$$(1 - \lambda\mathbb{E}[B])(N_1^{DPS}, \dots, N_K^{DPS}) \xrightarrow{d} Y \cdot \left(\frac{\alpha_1 \mathbb{E}[B_1]}{g_1}, \frac{\alpha_2 \mathbb{E}[B_2]}{g_2}, \dots, \frac{\alpha_K \mathbb{E}[B_K]}{g_K} \right),$$

where $\xrightarrow{d}$ denotes convergence in distribution and Y is an exponentially distributed random variable with mean $\mathbb{E}[Y]$ as given in (4.8).

In [18] it was obtained that under the heavy-traffic regime the conditional waiting time of a tagged class- k customer, $W_k^{DPS}(b)$, is the product of an exponentially distributed random variable and a deterministic factor.

Proposition 4.2.2. [18, Theorem 4.2] *The Laplace Transform of the scaled conditional waiting time of a class- k customer under the heavy-traffic regime satisfies*

$$\lim_{\lambda \uparrow 1/\mathbb{E}[B]} \mathbb{P}[(1 - \lambda\mathbb{E}[B])W_k^{DPS}(\lambda, b) > x] = e^{-x \frac{g_k}{b\mathbb{E}[V]}}, \quad (4.9)$$

where

$$\mathbb{E}[V] = \frac{\mathbb{E}[B^2]}{\sum_{i=1}^K \alpha_i \mathbb{E}[B_i^2] / g_i}. \quad (4.10)$$

Or in other words, as $\lambda \uparrow 1/\mathbb{E}[B]$, $(1 - \lambda\mathbb{E}[B])W_k^{DPS}(\lambda, b) \xrightarrow{d} \frac{b}{g_k} V$, where $\xrightarrow{d}$ denotes convergence in distribution and V is exponentially distributed with mean $\mathbb{E}[V]$ as given in (4.10).

4.3 Queue-length distribution

In this section we are interested in approximating the probability generating function of the queue length vector under DPS and therefore, according to the notation introduced in Chapter 2, we set $\vec{y} = \vec{z}$ and let $G(\lambda, \vec{z}) = \psi^{DPS}(\lambda, \vec{z})$ be the probability generating function of $(N_1^{DPS}, \dots, N_K^{DPS})$ under DPS. Since there is no characterization available for the queue-length distribution, in the next lemma we obtain the light-traffic approximation using the result given in Proposition 2.1.1. The proof method is constructive, and it can readily be applied to other queueing systems for which no analytical results are available.

Lemma 4.3.1. *The light-traffic approximation (of order 2) of the probability generating function of $(N_1^{DPS}, \dots, N_K^{DPS})$ is given by*

$$\begin{aligned}
& \psi^{DPS,LT}(\lambda, \vec{z}) \\
&= \left(\psi^{DPS} \right)^{(0)}(\lambda, \vec{z}) \Big|_{\lambda=0} + \lambda \left(\psi^{DPS} \right)^{(1)}(\lambda, \vec{z}) \Big|_{\lambda=0} + \frac{\lambda^2}{2} \left(\psi^{DPS} \right)^{(2)}(\lambda, \vec{z}) \Big|_{\lambda=0} \\
&= 1 - \rho + \lambda \sum_{i=1}^K \alpha_i z_i \mathbb{E}[B_i] \\
&+ \frac{\lambda^2}{2} \cdot 2 \left(\sum_{i,j=1}^K \alpha_i \alpha_j (z_i - 1) \mathbb{E} \left[\left(B_i - B_j \frac{g_i}{g_j} \right) \left(B_i - \min\{B_i, B_j \frac{g_i}{g_j}\} \right) \right. \right. \\
&\quad \left. \left. - \frac{1}{2} \left(B_i - \min\{B_i, B_j \frac{g_i}{g_j}\} \right)^2 - \frac{B_i^2}{2} \right] \right. \\
&\quad \left. + \sum_{i,j=1}^K \alpha_i \alpha_j (z_i \cdot z_j - 1) \mathbb{E} \left[B_j \left(1 + \frac{g_i}{g_j} \right) \left(B_i - \min\{B_i, \frac{g_i}{g_j} B_j\} \right) \right. \right. \\
&\quad \left. \left. + \frac{1}{2} \left(1 + \frac{g_j}{g_i} \right) \min\{B_i, \frac{g_i}{g_j} B_j\}^2 \right] \right. \\
&\quad \left. + \sum_{i,j=1}^K \alpha_i \alpha_j (z_j - 1) \mathbb{E} \left[\frac{g_i}{2g_j} \min\{\frac{g_j}{g_i} B_i, B_j\}^2 - B_i \min\{\frac{g_j}{g_i} B_i, B_j\} \right] \right) \Big|_{\lambda=0} \quad (4.11)
\end{aligned}$$

Proof: To calculate the zeroth, first and second light-traffic derivatives of the probability generating function of the queue length vector we measure how many customers are in the system when the tagged customer arrives (at time 0), given that in addition 0, 1 or 2 customers arrive to the system, respectively. For instance, for the zeroth derivative the tagged customer will observe an empty system, since no additional customers are assumed to arrive to the system. Hence, the generating function will satisfy $\mathbb{E}[z_1^0 \cdots z_K^0] = 1$. For the first derivative one needs to consider the system with one other arrival, see Proposition 2.1.1 (Equation (2.4)): the customer arriving at time t might come before the tagged customer and leave before or after its arrival. In that case, the corresponding probability generating function will be $\mathbb{E}[z_{U_t}^0] = 1$ or $\mathbb{E}[z_{U_t}^1]$, respectively.

It might also arrive after the tagged customer and in that case we have $\mathbb{E}[z_{U_t}^0] = 1$. To obtain the second light-traffic derivative two arrivals to the system are considered, see Proposition 2.1.1 (Equation (2.4)). In order to get expressions for the corresponding probability generating function, one needs to consider 6 different scenarios. We refer to Appendix 4.A.1 for the detailed proof. ■

We can now derive the interpolation approximation for the probability generating function of the queue length vector.

Proposition 4.3.2. *The interpolation approximation (of order 3) of the probability generating function of $(N_1^{DPS}, \dots, N_K^{DPS})$ is given by*

$$\begin{aligned} & \psi^{DPS,INT}(\lambda, \vec{z}) \\ &= (1 - \rho^3) + \lambda (1 - \rho^2) \left(\sum_{i=1}^K \alpha_i z_i^{(1-\rho)^{-1}} \mathbb{E}[B_i] - \mathbb{E}[B_{U_t}] \right) + \lambda^2 (1 - \rho) \sum_{i,j=1}^K \alpha_i \alpha_j \left(\right. \\ & \quad \mathbb{E} \left[\left(z_i^{(1-\rho)^{-1}} - 1 \right) \left(B_j \frac{g_i}{g_j} \left(\min\{B_i, B_j \frac{g_i}{g_j}\} - B_i \right) - \frac{1}{2} \min\{B_i, B_j \frac{g_i}{g_j}\}^2 \right) \right] \\ & \quad + \mathbb{E} \left[\left(z_i^{(1-\rho)^{-1}} \cdot z_j^{(1-\rho)^{-1}} - 1 \right) \left(B_j \left(1 + \frac{g_i}{g_j} \right) \left(B_i - \min\{B_i, \frac{g_i}{g_j} B_j\} \right) \right. \right. \\ & \quad \left. \left. + \frac{1}{2} \left(1 + \frac{g_j}{g_i} \right) \min\{B_i, \frac{g_i}{g_j} B_j\}^2 \right) \right] \\ & \quad \left. + \mathbb{E} \left[\left(z_j^{(1-\rho)^{-1}} - 1 \right) \left(\frac{g_i}{2g_j} \min\{\frac{g_j}{g_i} B_i, B_j\}^2 - B_i \min\{\frac{g_j}{g_i} B_i, B_j\} \right) \right] \right] \\ & \quad - \mathbb{E}[B] \sum_{i=1}^K \alpha_i \mathbb{E}[B_i] z_i^{(1-\rho)^{-1}} \ln \left(z_i^{(1-\rho)^{-1}} \right) \Bigg) + \rho^3 \frac{\mathbb{E}[B]/\mathbb{E}[Y]}{\mathbb{E}[B]/\mathbb{E}[Y] - \sum_{i=1}^K \frac{\alpha_i \mathbb{E}[B_i]}{g_i} \ln \left(z_i^{(1-\rho)^{-1}} \right)}, \end{aligned}$$

with $\mathbb{E}[Y]$ as given in Equation (4.8).

Proof: The proof follows directly from Proposition 2.3.1, where we substitute the light-traffic derivatives as derived in Lemma 4.3.1, and the heavy-traffic term $G^{HT}(\vec{z})$ as given in (4.7). See Appendix 4.A.2 for the detailed proof. ■

We note that the interpolation approximation obtained above does not correspond itself to a random variable. Consider for example the interpolation approximation for the total number of customers, i.e., $z_i = z, \forall i$. Then it can be checked that $\left| \frac{\partial^n \psi^{DPS,INT}(\lambda, \vec{z})}{\partial z^n} \right|_{z=0} = \infty$, for n above a certain threshold. Hence the probability of having n customers explodes. However, we note that Proposition 4.3.2 does provide rather accurate approximations for the moments, as obtained below, see Section 4.6 for

numerical results. The accuracy for the moments can be explained as follows: at $z = 0$ the even probabilities tend to $+\infty$ while the odd ones tend to $-\infty$.

We now obtain the first and second moments of the number of customers in the system. The approximation for the first moment of the total number of customers is given by

$$\begin{aligned}\mathbb{E}[N^{DPS,INT}] &= \mathbb{E}[N_1^{DPS,INT} + \dots + N_K^{DPS,INT}] = \frac{\partial \left(\psi^{DPS,INT}(\lambda, \vec{z}) \Big|_{z_i=z_j=z} \right)}{\partial z} \Big|_{z=1} \\ &= \rho + \lambda^2 \sum_{i,j=1}^K \alpha_i \alpha_j \mathbb{E} \left[\left(2 + \frac{g_i}{g_j} \right) B_i B_j - \left(2 + \frac{g_i}{g_j} \right) B_j \min\{B_i, B_j \frac{g_i}{g_j}\} \right. \\ &\quad \left. + \left(\frac{1}{2} + \frac{3g_j}{2g_i} \right) \min\{B_i, B_j \frac{g_i}{g_j}\}^2 - \frac{g_j}{g_i} B_i \min\{B_i, \frac{g_i}{g_j} B_j\} \right] + \frac{\rho^3}{1-\rho} \frac{\mathbb{E}[Y]}{\mathbb{E}[B]} \sum_{i=1}^K \frac{\alpha_i \mathbb{E}[B_i]}{g_i}. \quad (4.12)\end{aligned}$$

The second derivative of $\psi^{DPS,INT}(\lambda, \vec{z})$ with respect to z , evaluated at $z = 1$, is

$$\begin{aligned}\frac{\partial^2 \left(\psi^{DPS,INT}(\lambda, \vec{z}) \Big|_{z_i=z_j=z} \right)}{\partial z^2} \Big|_{z=1} &= \mathbb{E} \left[\left(N^{DPS,INT} \right)^2 \right] - \mathbb{E} \left[N^{DPS,INT} \right] \\ &= \lambda^2 \sum_{i,j=1}^K \alpha_i \alpha_j \left(\frac{\rho}{(1-\rho)} \mathbb{E} \left[B_j \frac{g_i}{g_j} \left(-B_i + \min\{B_i, B_j \frac{g_i}{g_j}\} \right) - \frac{1}{2} \min\{B_i, B_j \frac{g_i}{g_j}\}^2 \right] \right. \\ &\quad \left. + \frac{2(1+\rho)}{(1-\rho)} \mathbb{E} \left[\left(B_j \left(1 + \frac{g_i}{g_j} \right) \left(B_i - \min\{B_i, \frac{g_i}{g_j} B_j\} \right) + \frac{1}{2} \left(1 + \frac{g_j}{g_i} \right) \min\{B_i, \frac{g_i}{g_j} B_j\}^2 \right) \right] \right. \\ &\quad \left. + \frac{\rho}{(1-\rho)} \mathbb{E} \left[\left(\frac{g_j}{2g_i} \min\{B_i, \frac{g_i}{g_j} B_j\}^2 - \frac{g_j}{g_i} B_i \min\{B_i, \frac{g_i}{g_j} B_j\} \right) \right] \right) \\ &\quad + \rho^3 \frac{\mathbb{E}[Y]}{\mathbb{E}[B]} \left(\frac{2\mathbb{E}[Y]}{\mathbb{E}[B]} \left(\sum_{i=1}^K \frac{\alpha_i \mathbb{E}[B_i]}{g_i} (1-\rho)^{-1} \right)^2 - \left(\sum_{i=1}^K \frac{\alpha_i \mathbb{E}[B_i]}{g_i} (1-\rho)^{-1} \right) \right). \quad (4.13)\end{aligned}$$

Therefore, the second moment of the total number of customers is obtained from Equations (4.12) and (4.13)

$$\mathbb{E} \left[\left(N^{DPS,INT} \right)^2 \right] = \frac{\partial^2 \left(\psi^{DPS,INT}(\lambda, \vec{z}) \Big|_{z_i=z_j=z} \right)}{\partial z^2} \Big|_{z=1} + \mathbb{E} \left[N^{DPS,INT} \right]. \quad (4.14)$$

We observe that under the assumption that the service time distributions are exponentially distributed with the same mean $1/\mu$, the first and second moments of the number of customers are equivalent to an $M/M/1$ queue and the approximation gives $\mathbb{E} \left[N^{DPS,INT} \right] = \frac{\rho}{1-\rho}$ and $\mathbb{E} \left[\left(N^{DPS,INT} \right)^2 \right] = \frac{2\rho^2}{(1-\rho)^2} + \frac{\rho}{1-\rho}$, which are exact. The

approximation of the first moment is also exact with general service time distributions if there is only one class of customers, that is, $\alpha_i = 0, \forall i \neq k$ and $\alpha_k = 1$, namely, $\mathbb{E}[N^{DPS,INT}] = \rho + \rho^2 + \frac{\rho^3}{1-\rho} = \frac{\rho}{1-\rho}$.

In Section 4.6 we use the expressions of the first and second moment, Equations (4.12) and (4.14), to numerically test the accuracy of the interpolation approximation.

4.4 Waiting time distribution

In this section we are interested in approximating the distribution of the conditional waiting time under DPS and therefore, according to the notation introduced in Chapter 2, we set $\vec{y} = (b, x)$ and let $G(\lambda, b, x) = W_k^{DPS}(\lambda, b, x) = \mathbb{P}[W_k^{DPS}(b) > x]$ be the complementary distribution function of the conditional waiting time.

As in the case of the queue-length, for DPS there is no characterisation available for the waiting time distribution with general service time distributions. Thus, in the next lemma we obtain the first order light-traffic approximation using the result given in Proposition 2.1.1. As in the case of the queue-length, the proof method is constructive and it could be applied to other queueing systems for which no analytical results are available.

We are interested in obtaining an approximation for $\mathbb{P}(W_k^{DPS}(b) > x)$. Note that the waiting time has an atom at the point $x = 0$ of size $1 - \mathbb{P}(W_k^{DPS}(b) > 0)$.

Let us denote by $W_k^{DPS}(b|A(-\infty, \infty) = 1, \tau_1 = t)$ the waiting time of the tagged class- k customer when there is exactly one arrival at time t on $\mathbb{R}$. We can write $W_k^{DPS}(b|A(-\infty, \infty) = 1, \tau_1 = t)$ as follows:

$$W_k^{DPS}(b|A(-\infty, \infty) = 1, \tau_1 = t) = \begin{cases} t + b_{u_t} & \text{if } t \leq 0 \leq t + b_{u_t} \text{ and } \frac{b}{g_k} > \frac{t+b_{u_t}}{g_{u_t}}, \\ \frac{g_{u_t}}{g_k} b & \text{if } t \leq 0 \leq t + b_{u_t} \text{ and } \frac{b}{g_k} \leq \frac{t+b_{u_t}}{g_{u_t}}, \\ 0 & \text{if } t + b_{u_t} < 0, \\ b_{u_t} & \text{if } 0 < t < b \text{ and } \frac{b-t}{g_k} > \frac{b_{u_t}}{g_{u_t}}, \\ (b-t) \frac{g_{u_t}}{g_k} & \text{if } 0 < t < b \text{ and } \frac{b-t}{g_k} \leq \frac{b_{u_t}}{g_{u_t}}, \\ 0 & \text{if } 0 < b < t, \end{cases} \quad (4.15)$$

where u_t describes the class of the customer arriving at time t and b_{u_t} denotes the service requirement of the customer arriving at time t . The first expression describes the case where the customer arrives before the tagged customer and leaves after the tagged customer arrives, but before the tagged customer leaves. Hence, by the work conserving property, the waiting time of the tagged customer is all the work present at time 0 except for its own service requirement, that is, $b_{u_t} - (-t)$. We recall that the work-conserving property states that as long as the system is non-empty, the server does not idle. The second term describes the case where the other customer is in the system

at time 0 and is still present as the tagged customer departs. Hence, the tagged class- k customer is served at rate $\frac{g_k}{g_k + g_{u_t}}$, so that its waiting time is $b \left(\frac{g_k}{g_k + g_{u_t}} \right)^{-1} - b = \frac{g_{u_t}}{g_k} b$. The fourth expression describes the case where the customer arrives after the tagged customer and leaves before the tagged customer. Hence, by the work-conserving property of the system, the waiting time of the tagged class- k customer is given by the total amount of work that needs to be done except for the service requirement of the tagged customer, that is, b_{u_t} . The fifth term describes the case where the customer arrives after the tagged customer, and departs after the tagged customer departs. Then, the waiting time of the tagged customer is composed of t , the time it was in the system until the customer arrived, plus $(b - t) \left(\frac{g_k}{g_k + g_{u_t}} \right)^{-1}$, the remaining service requirement multiplied by the inverse of the rate at which the tagged class- k customer is served, minus b , the service requirement of the tagged customer. The third and sixth case is when the tagged customer does not coincide with the other customer. Hence, the waiting time is 0.

We then obtain the following expression for the first order light-traffic approximation.

Lemma 4.4.1. *The light-traffic approximation (of order 1) of the complementary distribution function of the conditional waiting time of a tagged class- k customer with a given service requirement b is given by*

$$\begin{aligned}
 W_k^{DPS,LT}(\lambda, b, x) &= \lambda \sum_{j=1}^K \alpha_j \mathbb{E} \left[\left(1 + \frac{g_k}{g_j} \right) \left(-x + \min\{B_j, \frac{g_j}{g_k} b\} \right)^+ \right. \\
 &\quad \left. + \mathbf{1} \left[\frac{g_j}{g_k} b > x \right] \left(B_j - \min\{B_j, \frac{g_j}{g_k} b\} \right) + \mathbf{1} [B_j > x] \left(b - \frac{g_k}{g_j} \min\{B_j, \frac{g_j}{g_k} b\} \right) \right]
 \end{aligned} \tag{4.16}$$

Proof: To calculate the first light-traffic derivative of the complementary distribution function of the waiting time we use Equation (2.4) and the expression obtained in (4.15). For the first derivative six different cases might happen. See Appendix 4.A.3 for the detailed proof. ■

We consider in this chapter the light-traffic approximation of order 1. Calculating the second light-traffic derivative would imply having to consider events that either 0, 1, or 2 customers arrive in the system (besides the tagged customer). The latter would result in going through 22 different cases, while for the first derivative we only needed to go through 6 cases, see Equation (4.15). We will see in the numerical results that already the first order light-traffic approximation provides an insightful and accurate approximation of the performance. We further refer to Appendix 4.A.5 where the light-traffic approximation of the mean sojourn time of order 1 and of order 2 are numerically compared.

We can now present the interpolation approximation of the complementary distribution of the conditional waiting time in DPS.

Proposition 4.4.2. *The interpolation approximation (of order 2) of the complementary distribution of the conditional waiting time of a tagged class- k customer with a given service requirement b is given by*

$$\begin{aligned}
W_k^{DPS,INT}(\lambda, b, x) &= \lambda(1 - \rho) \sum_{j=1}^K \alpha_j \mathbb{E} \left[\left(1 + \frac{g_k}{g_j} \right) \left(-(1 - \rho)x + \min\{B_j, \frac{g_j}{g_k}b\} \right)^+ \right. \\
&\quad \left. + \mathbf{1} \left[\frac{g_j}{g_k}b > (1 - \rho)x \right] \left(B_j - \min\{B_j, \frac{g_j}{g_k}b\} \right) \right. \\
&\quad \left. + \mathbf{1} [B_j > (1 - \rho)x] \left(b - \frac{g_k}{g_j} \min\{B_j, \frac{g_j}{g_k}b\} \right) \right] + \rho^2 e^{-(1-\rho)x \frac{g_k}{b\mathbb{E}[V]}}, \tag{4.17}
\end{aligned}$$

with $\mathbb{E}[V]$ as given in (4.10).

Proof: The proof follows directly from Proposition 2.3.1, where we substitute the light-traffic derivatives as derived in Lemma 4.4.1, and the heavy-traffic term $G^{HT}(b, x)$ as given in (4.9). See Appendix 4.A.4 for the detailed proof. $\blacksquare$

We observe that the interpolation approximation of the complementary distribution function is not necessarily a distribution itself. For example, note that at the point $x = 0$, $W_k^{DPS,INT}(\lambda, b, x) = \lambda(1 - \rho)(\mathbb{E}[B] + b) + \rho^2$, which can be greater than one. However, we note that Proposition 4.4.2 does provide accurate approximations for the first moment, see Section 4.6 for numerical results.

From Proposition 4.4.2 we get as a corollary the interpolation approximation for the complementary distribution of the conditional waiting time.

Corollary 4.4.3. *The interpolation approximation (of order 2) of the complementary distribution function of the unconditional waiting time of a tagged class- k customer is given by*

$$\begin{aligned}
W_k^{DPS,INT}(\lambda, x) &:= \int_0^\infty W_k^{DPS,INT}(\lambda, b, x) dF_k(b) \\
&= \lambda(1 - \rho) \sum_{j=1}^K \alpha_j \left(\left(1 + \frac{g_k}{g_j} \right) \int_{(1-\rho)x \frac{g_k}{g_j}}^\infty \left(\int_{(1-\rho)x}^{\frac{g_j}{g_k}b} (1 - F_j(b_j)) db_j \right) dF_k(b) \right. \\
&\quad \left. + \left(\int_{(1-\rho)x \frac{g_k}{g_j}}^\infty \left(\int_{\frac{g_j}{g_k}b}^\infty (1 - F_j(b_j)) db_j \right) dF_k(b) \right) \right)
\end{aligned}$$

$$\begin{aligned}
& + \int_{(1-\rho)xg_k/g_j}^{\infty} \left(\left(b - \frac{g_k}{g_j}(1-\rho)x \right) (1 - F_j((1-\rho)x)) \right. \\
& \quad \left. - \frac{g_k}{g_j} \int_{(1-\rho)x}^{b \frac{g_j}{g_k}} (1 - F_j(b_j)) db_j \right) dF_k(b) \Bigg) + \rho^2 e^{-(1-\rho)x \frac{g_k}{\mathbb{E}[B_k]\mathbb{E}[V]}}.
\end{aligned} \tag{4.18}$$

From the complementary distribution function of the conditional waiting time, the distribution of the conditional sojourn time of the tagged customer can be obtained. Following similar steps as in Lemma 4.4.1, we can obtain the light-traffic approximation for the conditional sojourn time distribution as follows

$$\mathbb{P}[S_k^{DPS}(b) > x] = \left(\mathbb{P}[S_k^{DPS}(b) > x] \right)^{(0)} + \lambda \cdot \left(\mathbb{P}[S_k^{DPS}(b) > x] \right)^{(1)},$$

where

$$\left(\mathbb{P}[S_k^{DPS}(b) > x] \right)^{(0)} = \mathbb{P}[B_k > x]$$

and

$$\left(\mathbb{P}[S_k^{DPS}(b) > x] \right)^{(1)} = \int_0^x \left(\mathbb{P}(W_k^{DPS}(b) > x - b) \right)^{(1)} f(b) db.$$

Together with the heavy-traffic result for the waiting time presented in Proposition 4.2.2, which coincides with the heavy-traffic result for the sojourn time, the interpolation approximation for the distribution function of the sojourn time can be constructed. Unconditioning the obtained approximation, it can easily be shown that for exponentially distributed service requirements (with mean $1/\mu$) and when all the weights are the same, the interpolation approximation of the distribution function of the unconditional sojourn time becomes equivalent to the approximation for PS derived in [15, Section 2.2.], denoted by $S_k^{PS,INT}(\lambda, x)$, which satisfies

$$\begin{aligned}
& S_k^{PS,INT}(\lambda, x) \\
& = (1 - \rho^2) e^{-\mu(1-\rho)x} + \lambda(1 - \rho) e^{-\mu(1-\rho)x} \left(\frac{\mu(1-\rho)^2 x^2}{4} - \frac{(1-\rho)x}{2} \right) + \rho^2 e^{-(1-\rho)x/\mathbb{E}[B_k]}.
\end{aligned}$$

4.5 Mean sojourn time

In this section we present the interpolation approximation for the mean sojourn time and analyse it in further detail. The results of this section are a direct consequence of Proposition 4.4.2, since the mean sojourn time is the mean waiting time plus its service requirement b .

4.5.1 Mean conditional sojourn time

As a corollary of Lemma 4.4.1 and Proposition 4.4.2 we obtain the following two results for the light-traffic approximation and the interpolation approximation of the

mean conditional sojourn time of the tagged customer, denoted by $\overline{S}_k^{DPS,LT}(\lambda, b)$ and $\overline{S}_k^{DPS,INT}(\lambda, b)$, respectively.

Corollary 4.5.1. *The light-traffic approximation (of order 1) and the interpolation approximation (of order 2) of the mean conditional sojourn time for a tagged class- k customer with service requirement b is given by*

$$\begin{aligned} \overline{S}_k^{DPS,LT}(\lambda, b) \\ = b + \lambda \sum_{j=1}^K \alpha_j \mathbb{E} \left[\frac{1}{2} \left(1 + \frac{g_k}{g_j} \right) \min \left\{ B_j, b \frac{g_j}{g_k} \right\}^2 - \left(b \frac{g_j}{g_k} + \frac{g_k}{g_j} B_j \right) \min \left\{ B_j, b \frac{g_j}{g_k} \right\} + \left(1 + \frac{g_j}{g_k} \right) b B_j \right]. \end{aligned} \quad (4.19)$$

and

$$\begin{aligned} \overline{S}_k^{DPS,INT}(\lambda, b) \\ = b + \lambda \sum_{j=1}^K \alpha_j \mathbb{E} \left[\frac{1}{2} \left(1 + \frac{g_k}{g_j} \right) \min \left\{ B_j, b \frac{g_j}{g_k} \right\}^2 - \left(b \frac{g_j}{g_k} + \frac{g_k}{g_j} B_j \right) \min \left\{ B_j, b \frac{g_j}{g_k} \right\} + \left(1 + \frac{g_j}{g_k} \right) b B_j \right] \\ + \frac{(\lambda \mathbb{E}[B])^2}{(1 - \lambda \mathbb{E}[B])} \frac{b}{g_k} \frac{\mathbb{E}[B^2]}{\sum_{j=1}^K \alpha_j \mathbb{E}[B_j^2]/g_j}. \end{aligned} \quad (4.20)$$

Proof: We show how Equation (4.19) can be deduced from Lemma 4.4.1. We have

$$\begin{aligned} \overline{W}_k^{DPS,LT}(\lambda, b) &:= \int_0^\infty W_k^{DPS,LT}(\lambda, b, x) dx \\ &= \lambda \sum_{j=1}^K \alpha_j \mathbb{E} \left[\frac{1}{2} \left(1 + \frac{g_k}{g_j} \right) \min \left\{ B_j, b \frac{g_j}{g_k} \right\}^2 - \left(b \frac{g_j}{g_k} + \frac{g_k}{g_j} B_j \right) \min \left\{ B_j, b \frac{g_j}{g_k} \right\} + \left(1 + \frac{g_j}{g_k} \right) b B_j \right], \end{aligned} \quad (4.21)$$

and by the definition of the sojourn time, $\overline{S}_k^{DPS,LT}(\lambda, b) = \overline{W}_k^{DPS,LT}(\lambda, b) + b$, we obtain the desired expression for the light-traffic approximation (4.19).

The heavy-traffic result presented for the waiting time (4.9), also holds for the sojourn time. Therefore, combining the light-traffic approximation and the heavy-traffic result we obtain the interpolation approximation in (4.20). $\blacksquare$

In Appendix 4.A.5 we describe an alternative way to obtain the light-traffic derivatives for the mean sojourn times. This method makes use of Equation (4.1), and allows us to derive higher order light-traffic approximations. We will observe numerically, see Figures 4.17-4.20, that the light-traffic approximation (of order 2) gets more accurate, whereas the accuracy of the interpolation (of order 3) for intermediate loads does not necessarily get better.

Below, we make several interesting observations regarding the approximation obtained for the mean conditional sojourn time. In Section 4.6 we will numerically evaluate the accuracy of the approximation formulas derived in Equation (4.20).

Processor Sharing

For the standard Processor Sharing queue the mean conditional sojourn time is known and is given by $b/(1 - \rho)$, [28]. If either (i) there is only one class or (ii) all weights are the same, our model is equivalent to a processor-sharing queue. Below we will verify that our approximation as stated in (4.20) indeed coincides with $b/(1 - \rho)$.

We first consider the case of one class, that is, $\alpha_i = 0, \forall i \neq k$ and $\alpha_k = 1$. Then Equation (4.20) is equal to

$$\begin{aligned} & b(1 + \rho) + \lambda \alpha_k \mathbb{E} \left[\min\{B_k, b\}^2 - (b + B_k) \min\{B_k, b\} + bB_k \right] + b \frac{\rho^2}{(1 - \rho)} \\ &= b(1 + \rho + \frac{\rho^2}{(1 - \rho)}) = \frac{b}{1 - \rho}, \end{aligned}$$

where we used that $\min\{B_j, b\}^2 - (b + B_j) \min\{B_j, b\} + bB_j = 0$.

We now assume all weights are the same, i.e., $g_i = g_k, \forall i, k = 1, \dots, K$. Equation (4.20) is then equal to

$$\begin{aligned} & b(1 + \rho) + \lambda \sum_{j=1}^K \alpha_j \mathbb{E} \left[\min\{B_j, b\}^2 - (b + B_j) \min\{B_j, b\} + bB_j \right] + \frac{b\rho^2}{(1 - \rho)} \frac{\mathbb{E}[B^2]}{\sum_{j=1}^K \alpha_j \mathbb{E}[B_j^2]} \\ &= b(1 + \rho + \frac{\rho^2}{(1 - \rho)}) = \frac{b}{1 - \rho}. \end{aligned}$$

Hence, both cases coincide with the PS queue.

Priority queue

We now consider the case when the weight of the tagged customer grows large, i.e., $g_k \rightarrow \infty$. Hence, class k is prioritized in the limit. Then, the approximation simplifies to

$$\begin{aligned} & \lim_{g_k \rightarrow \infty} \left(b(1 + \rho) \right. \\ & \quad + \sum_{i=1}^K \lambda_i \mathbb{E} \left[\frac{1}{2} \left(\min\{B_i, b \frac{g_i}{g_k}\}^2 + \frac{g_k}{g_i} \min\{B_i, b \frac{g_i}{g_k}\}^2 \right) \right. \\ & \quad \left. - (b \frac{g_i}{g_k} \min\{B_i, b \frac{g_i}{g_k}\} + B_i \min\{B_i \frac{g_k}{g_i}, b\}) \right. \\ & \quad \left. + \frac{g_i}{g_k} bB_i \right] + \frac{b(\lambda \mathbb{E}[B])^2}{(1 - \lambda \mathbb{E}[B])} \frac{\mathbb{E}[B^2]}{g_k \sum_{j=1}^K \alpha_j \mathbb{E}[B_j^2]/g_j + \alpha_k \mathbb{E}[B_k^2]} \Big) \\ &= b(1 + \rho) + \mathbb{E} \left[\sum_{\substack{i=1 \\ i \neq k}}^K \lambda_i \left(\frac{1}{2} (0 + 0) - (0 + bB_i) + 0 \right) \right] \\ &= b(1 + \rho_k). \end{aligned}$$

Note that the conditional sojourn time as $g_k \rightarrow \infty$ is known and is given by $b/(1 - \rho_k)$. Since $1/(1 - \rho_k) = \sum_{i=0}^{\infty} \rho_k^i$, we directly see that the approximation is the first order approximation of the exact expression. The relative error is equal to $100\%(b/(1 - \rho_k) - b(1 + \rho_k))/b/(1 - \rho_k) = \rho_k^2 100\%$, and we thus see that the relative error increases as the load of class k increases.

Monotonicity in the weights

It can be checked that the approximation for the mean conditional sojourn time of a tagged class- k customer, $\overline{S}_k^{DPS,INT}(\lambda, b)$, is decreasing in g_k and increasing in g_i , $i \neq k$.

This can be seen as follows. From (4.20) we have

$$\begin{aligned} \overline{S}_k^{DPS,INT}(\lambda, b) &= b(1 + \rho) \\ &+ \sum_{i=1, i \neq k}^K \lambda_i \mathbb{E} \left[\frac{1}{2} \left(1 + \frac{g_k}{g_i} \right) \min\{B_i, b \frac{g_i}{g_k}\}^2 - \left(b \frac{g_i}{g_k} + \frac{g_k}{g_i} B_i \right) \min\{B_i, b \frac{g_i}{g_k}\} + b \frac{g_i}{g_k} B_i \right] \\ &+ \frac{(\lambda \mathbb{E}[B])^2}{(1 - \lambda \mathbb{E}[B])} \frac{b}{g_k} \frac{\mathbb{E}[B^2]}{\sum_{j=1}^K \alpha_j \mathbb{E}[B_j^2]/g_j}, \end{aligned}$$

where for $i = k$ we used that $\min\{B_k, b\}^2 - (b + B_k) \min\{B_k, b\} + bB_k = 0$.

Now, if $B_i \leq \frac{g_i}{g_k} b$, then

$$\frac{1}{2} \left(1 + \frac{g_k}{g_i} \right) \min\{B_i, b \frac{g_i}{g_k}\}^2 - \left(b \frac{g_i}{g_k} + \frac{g_k}{g_i} B_i \right) \min\{B_i, b \frac{g_i}{g_k}\} + b \frac{g_i}{g_k} B_i = \frac{1}{2} B_i^2 \left(1 - \frac{g_k}{g_i} \right),$$

which is decreasing in g_k and increasing in g_i . If $B_i > \frac{g_i}{g_k} b$, then

$$\frac{1}{2} \left(1 + \frac{g_k}{g_i} \right) \min\{B_i, b \frac{g_i}{g_k}\}^2 - \left(b \frac{g_i}{g_k} + \frac{g_k}{g_i} B_i \right) \min\{B_i, b \frac{g_i}{g_k}\} + b \frac{g_i}{g_k} B_i = \frac{1}{2} b^2 \frac{g_i}{g_k} \left(1 - \frac{g_i}{g_k} \right) + b B_i \left(\frac{g_i}{g_k} - 1 \right),$$

which is decreasing in g_k and increasing in g_i (can be derived by taking the derivative and the fact that $B_i > \frac{g_i}{g_k} b$). The monotonicity of $\overline{S}_k^{DPS,INT}(\lambda, b)$ in g_k and g_i now follows immediately.

Uniformly bounded in the second moment

A very relevant property of processor sharing is that the mean sojourn time depends on the service time distribution only through its mean [27]. This has been an important argument to claim the interest of time-sharing disciplines with respect to more classical scheduling policies like FCFS. Indeed, the classical Pollaczek-Khinchine formula for the mean waiting time in a FCFS queue shows that it explodes as the second moment of the service time distribution grows large. For a DPS queue, Equation (4.1) does not allow

to reach any conclusion regarding the dependence of the mean conditional sojourn time on the moments of the service time distribution.

It then becomes interesting to observe that the approximation (4.20) is uniformly bounded in the second moments of the service time distribution. To see this, we first note that $\min\{B_i, b_{g_k}^{g_i}\}^2 \leq B_i b_{g_k}^{g_i}, \forall i$, which directly implies that the first three terms in (4.20) are uniformly bounded by a function that depends on the service requirements only through its first moment. We are now left with the heavy-traffic term $\frac{\mathbb{E}[B^2]}{\sum_{j=1}^K \alpha_j \mathbb{E}[B_j^2]/g_j}$.

Let j^* be such that $\mathbb{E}[B_{j^*}^2] \geq \mathbb{E}[B_j^2], \forall j$. We then have

$$\frac{\mathbb{E}[B^2]}{\sum_{j=1}^K \alpha_j \mathbb{E}[B_j^2]/g_j} = \frac{\sum_j \alpha_j \mathbb{E}[B_j^2]}{\sum_{j=1}^K \alpha_j \mathbb{E}[B_j^2]/g_j} \leq \frac{\mathbb{E}[B_{j^*}^2]}{\alpha_{j^*} \mathbb{E}[B_{j^*}^2]/g_{j^*}} = \frac{g_{j^*}}{\alpha_{j^*}}.$$

We thus finally conclude that (4.20) can be upper bounded by an expression that depends only on the first moment of the service time distributions. This indicates that the DPS queue provides a satisfactory performance even in the presence of service time distributions with a high variability.

4.5.2 Mean unconditional sojourn time

As a corollary of Equation (4.20), we obtain the mean unconditional sojourn time of the tagged class- k customer.

Corollary 4.5.2. *The interpolation approximation (of order 2) of the mean unconditional sojourn time for a tagged class- k customer is given by*

$$\begin{aligned} \overline{S}_k^{DPS,INT}(\lambda) &:= \int_0^\infty \overline{S}_k^{DPS,INT}(\lambda, b) dF_k(b) \\ &= \mathbb{E}[B_k](1 + \rho) \\ &\quad + \lambda \sum_{j=1}^K \alpha_j \mathbb{E} \left[\frac{1}{2} \left(1 + \frac{g_k}{g_j} \right) \min\{B_j, B_k \frac{g_j}{g_k}\}^2 - \left(B_k \frac{g_j}{g_k} + \frac{g_k}{g_j} B_j \right) \min\{B_j, B_k \frac{g_j}{g_k}\} + B_k \frac{g_j}{g_k} B_j \right] \\ &\quad + \frac{(\lambda \mathbb{E}[B])^2}{(1 - \lambda \mathbb{E}[B])} \frac{\mathbb{E}[B_k]}{g_k} \frac{\mathbb{E}[B^2]}{\sum_{j=1}^K \alpha_j \mathbb{E}[B_j^2]/g_j}. \end{aligned} \quad (4.22)$$

4.5.3 Exponential service requirements

In this section we focus on the case in which the service requirements of the customers are exponentially distributed. We recall that a random variable B_i is exponentially distributed if $\mathbb{P}(B_i \leq b_i) = 1 - e^{-b_i/\mathbb{E}[B_i]}$. In the following subsection entitled *Mean conditional and unconditional sojourn time*, we further simplify the expression for the interpolation approximation of the mean conditional and unconditional sojourn time and compare the latter for two classes of customers with the exact formulas as stated in

Equations (4.5) and (4.6). In the subsection entitled *Mean unconditional sojourn time for an arbitrary customer*, we calculate the relative error (for different service requirements) and we verify that our approximation for the mean unconditional sojourn time for *an arbitrary customer* is exact when the mean service requirements of all classes are the same.

Mean conditional and unconditional sojourn time

In the case of exponentially distributed service requirements, our approximations for the mean conditional and unconditional sojourn time can be significantly simplified. This is a direct consequence of Equation (4.20) and Corollary 4.5.2, respectively.

Corollary 4.5.3. *Assume class- k customers have an exponentially distributed service requirement with mean $1/\mu_k, k = 1, \dots, K$. The interpolation approximation (of order 2) of the mean conditional sojourn time for a tagged class- k customer with service requirement b is given by*

$$\begin{aligned} \overline{S}_k^{DPS,INT}(\lambda, b) &= b + \lambda \mathbb{E}[B]b + \lambda \sum_{j=1}^K \frac{\alpha_j}{\mu_j^2} \left(1 - \frac{g_k}{g_j}\right) \left(1 - e^{-b \frac{g_j}{g_k} \mu_j}\right) \\ &\quad + \frac{(\lambda \mathbb{E}[B])^2}{(1 - \lambda \mathbb{E}[B])} \frac{b}{g_k} \frac{\sum_{j=1}^K \alpha_j / \mu_j^2}{\sum_{j=1}^K \alpha_j / (\mu_j^2 g_j)}, \end{aligned} \quad (4.23)$$

and the mean unconditional sojourn time is given by

$$\overline{S}_k^{DPS,INT}(\lambda) = \frac{1}{\mu_k} + \frac{1}{\mu_k} \lambda \mathbb{E}[B] + \lambda \sum_{j=1}^K \frac{\alpha_j}{\mu_j} \frac{(g_j - g_k)}{g_j \mu_j + g_k \mu_k} + \frac{(\lambda \mathbb{E}[B])^2}{(1 - \lambda \mathbb{E}[B])} \frac{1}{g_k \mu_k} \frac{\sum_{j=1}^K \alpha_j / \mu_j^2}{\sum_{j=1}^K \alpha_j / (\mu_j^2 g_j)}, \quad (4.24)$$

where $\mathbb{E}[B] = \sum_{j=1}^K \alpha_j / \mu_j$.

Proof: See Appendix 4.A.6 for the proof. ■

In the case of two classes of customers, Fayolle et al [13] have closed-form expressions for the mean unconditional sojourn time, see Equation (4.5) and (4.6). Rewriting (4.24) for $K = 2$ in such a way so that the similarity with Equations (4.5) and (4.6) is clear, we obtain

$$\begin{aligned} \overline{S}_k^{DPS,INT}(\lambda) &= \frac{1}{\mu_k(1 - \rho)} \left(1 + \rho^2 \left(-1 + \frac{\sum_{j=1}^K \alpha_j / \mu_j^2}{g_k \sum_{j=1}^K \alpha_j / (\mu_j^2 g_j)} \right) + \frac{\mu_k \rho_{-k} (g_{-k} - g_k)}{D} \frac{(1 - \rho)D}{\mu_1 g_1 + \mu_2 g_2} \right), \end{aligned} \quad (4.25)$$

with $k = 1, 2$, $-k = \text{mod}(k, 2) + 1$ and where $D = \mu_1 g_1 (1 - \rho_1) + \mu_2 g_2 (1 - \rho_2)$.

We directly observe that the difference with respect to the exact expression for the mean unconditional sojourn time ((4.5) and (4.6)) is in the terms

$$\rho^2 \left(-1 + \frac{\sum_{j=1}^K \alpha_j / \mu_j^2}{g_k \sum_{j=1}^K \alpha_j / (\mu_j^2 g_j)} \right) \quad \text{and} \quad \frac{(1-\rho)D}{\mu_1 g_1 + \mu_2 g_2}.$$

It can easily be seen that our approximation is exact for the two extreme values of the traffic intensity, $\rho = 0$ and $\rho = 1$. That is, the expressions

$$\overline{S_k^{DPS,INT}}(0) = 1/\mu_k \quad \text{and} \quad \lim_{\lambda \rightarrow 1/\mathbb{E}[B]} (1-\rho) \overline{S_k^{DPS,INT}}(\lambda) = \frac{1}{\mu_k} \left(1 + \frac{\mu_k \rho_{-k} (g_{-k} - g_k)}{D} \right)$$

are satisfied.

Let us denote by Rel.Error_k the relative error of a class- k customer, that is,

$$\text{Rel.Error}_k = \frac{\overline{S_k^{DPS}}(\lambda) - \overline{S_k^{DPS,INT}}(\lambda)}{\overline{S_k^{DPS}}(\lambda)}, \quad k = 1, 2.$$

Now, let us consider $g_1 + g_2 = 1$. We then obtain $\lim_{g_1 \uparrow 1} \text{Rel.Error}_1 = \rho_1^2 \cdot 100\%$ and

$$\lim_{g_1 \uparrow 1} \text{Rel.Error}_2 = \frac{\mu_2 \rho_1 - (1 - \rho_1) \left(\rho^2 \frac{\rho_1 \mu_2}{\rho_2} + \mu_2 \rho_1 (1 - \rho) \right)}{\mu_1 (1 - \rho_1) + \mu_2 \rho_1} \cdot 100\%.$$

Hence, the relative error of class-1 customers (when $g_1 \uparrow 1$) increases as the load of class 1 increases but does not depend on the parameter of class 2. The same result was obtained in the subsection entitled *Priority queue* of Section 4.5.1 for the mean conditional sojourn time for an arbitrary number of classes and general service requirements. Moreover, the absolute relative error of class-2 customers (when $g_1 \uparrow 1$) increases as the load of class 2 decreases.

In Figure 4.1 we plot the relative error of the mean unconditional sojourn time for $K = 2$ with respect to g_1 . The parameters considered are $\rho_1 = 0.2, \rho_2 = 0.4, \mu_1 = 1, \mu_2 = 1, g_2 = 1 - g_1$ and from the formulas presented above we obtain $\lim_{g_1 \uparrow 1} \text{Rel. Error}_1 = 4\%$, $\lim_{g_1 \uparrow 1} \text{Rel. Error}_2 = -0.8\%$, $\lim_{g_1 \downarrow 0} \text{Rel. Error}_1 = -12.8\%$, $\lim_{g_1 \downarrow 0} \text{Rel. Error}_2 = 16\%$, which coincide with the extreme points in the figure.

Mean unconditional sojourn time for an arbitrary customer

In this section we discuss the relative error of the mean unconditional sojourn time for an *arbitrary customer*. We first calculate the error in case $K = 2$ and when μ_1 or μ_2 take extreme values. We then show that the approximation for the mean unconditional sojourn time of an arbitrary customer is exact when the service requirements of all customers are the same, for arbitrary K .

We denote by Rel.Error the relative error of an arbitrary customer, that is,

$$\text{Rel.Error} = \left(1 - \frac{\sum_{k=1}^2 \alpha_k \overline{S_k^{DPS,INT}}(\lambda)}{\sum_{k=1}^2 \alpha_k \overline{S_k^{DPS}}(\lambda)} \right) \cdot 100\%,$$

where $\overline{S_1^{DPS}}(\lambda)$ and $\overline{S_2^{DPS}}(\lambda)$ are given in Equations (4.5) and (4.6), respectively. Then, if we keep constant ρ_1 and ρ_2 we obtain

$$\lim_{\mu_2 \downarrow 0} \text{Rel.Error} = \frac{\rho_1 \left(\frac{\rho_2(g_2 - g_1)}{g_1} \left(\frac{1}{1 - \rho_1} - 1 + \rho \right) - \rho^2 \left(-1 + \frac{g_2}{g_1} \right) \right)}{\rho_1 \left(1 + \frac{\rho_2(g_2 - g_1)}{g_1(1 - \rho_1)} \right) + \rho_2} \cdot 100\% \quad (4.26)$$

and $\lim_{\mu_1 \uparrow \infty} \text{Rel.Error} = \lim_{\mu_2 \downarrow 0} \text{Rel.Error}$. The intuition behind the latter equation can be seen as follows: having $\mu_1 \rightarrow \infty$ (and hence $\lambda_1 \rightarrow \infty$), i.e., having many class-1 arrivals of small size, is equivalent to having $\mu_2 \rightarrow 0$ (and hence $\lambda_2 \rightarrow 0$), i.e., having very few class-2 arrivals of large size.

In Figure 4.2 we plot the relative error of the mean unconditional sojourn time for *an arbitrary customer*. We fix ρ_1, ρ_2 and μ_1 and we let μ_2 and $\lambda_2 = \rho_2 \mu_2$ change. The chosen parameters are $\rho_1 = 0.2, \rho_2 = 0.4, \mu_1 = 1, g_1 = 0.2, g_2 = 1 - g_1$. We observe that the results obtained from Equation (4.26), $\lim_{\mu_2 \downarrow 0} \text{Rel.Error} = -1.3\%$ and $\lim_{\mu_2 \uparrow \infty} \text{Rel.Error} = 6.4\%$, coincide with the extreme points in the figures.

We now show that our light-traffic approximation for the mean unconditional sojourn time of *an arbitrary customer* is exact under the assumption that the mean service requirements of all customers are the same, i.e., $\mathbb{E}[B_j] = 1/\mu, \forall j = 1, \dots, K$. This result holds for an arbitrary number of classes.

As stated earlier, the mean unconditional sojourn time of *an arbitrary customer* is defined as $\overline{S^{DPS}}(\lambda) := \sum_{k=1}^K \alpha_k \overline{S_k^{DPS}}(\lambda)$. Since we assume exponentially distributed service requirements and $\mathbb{E}[B_k] = 1/\mu, \forall k = 1, \dots, K$, the total number of customers in the system is distributed as that in a processor sharing queue with arrival rate $\lambda = \sum_{k=1}^K \lambda_k$ and service rate μ . By Little's law, we therefore have that the total mean unconditional sojourn time is given by that of an $M/M/1$ queue, i.e., $\frac{1/\mu}{1 - \rho}$.

For our interpolation approximation we have

$$\begin{aligned} \overline{S^{DPS,INT}}(\lambda) &= \sum_{k=1}^K \alpha_k \overline{S_k^{DPS,INT}}(\lambda) \\ &= \frac{1 + \rho}{\mu} + \lambda \sum_{k=1}^K \alpha_k \sum_{j=1}^K \frac{\alpha_j}{\mu} \frac{(g_j - g_k)}{g_j \mu + g_k \mu} + \frac{(\lambda/\mu)^2}{(1 - \lambda/\mu)} \sum_{k=1}^K \alpha_k \frac{1}{\mu g_k} \frac{\sum_{j=1}^K \alpha_j / \mu^2}{\sum_{j=1}^K \alpha_j / (\mu^2 g_j)} \\ &= \frac{1 + \rho}{\mu} + \frac{1}{\mu} \frac{\rho^2}{1 - \rho} \sum_{k=1}^K \frac{\alpha_k}{g_k} \frac{1}{\sum_{j=1}^K \alpha_j / g_j} = \frac{1/\mu}{1 - \rho}, \end{aligned} \quad (4.27)$$

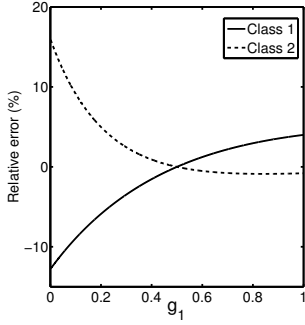


Figure 4.1: Relative error for the mean unconditional sojourn time for $K = 2$ with respect to g_1 .

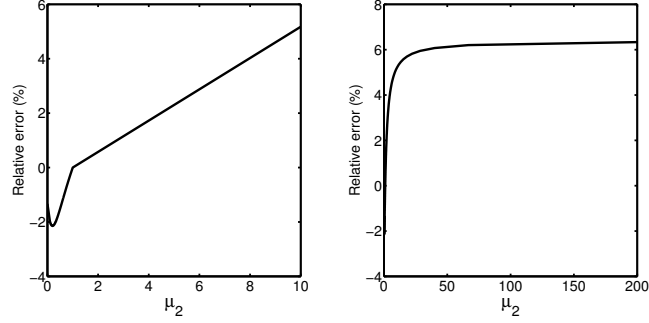


Figure 4.2: Mean unconditional sojourn time for an arbitrary customer.

where we used that

$$\sum_{k=1}^K \alpha_k \sum_{j=1}^K \alpha_j \frac{(g_j - g_k)}{g_j \mu + g_k \mu} = \frac{1}{\mu} \sum_{k=1}^K \sum_{j=1}^K \alpha_k \alpha_j \frac{g_j - g_k}{g_j + g_k} = \frac{1}{\mu} \sum_{k=1}^K \sum_{j=1}^{K-1} \alpha_k \alpha_j \left(\frac{g_j - g_k}{g_j + g_k} + \frac{g_k - g_j}{g_j + g_k} \right) = 0.$$

Hence, the obtained interpolation approximation is exact when $\mathbb{E}[B_k] = 1/\mu, \forall k = 1, \dots, K$.

In Figure 4.2 (*left*) we indeed observe that when $\mu_2 = 1$, so when the service requirements of both classes coincide, the relative error is 0, as proven in Equation (4.27).

4.6 Numerical comparison

In this section we numerically investigate the accuracy of the approximations obtained in this chapter for the distribution of the queue length vector and for the mean sojourn time under DPS.

We measure the accuracy of the queue-length distribution just checking the accuracy of the first and second moments using the algorithm proposed by Rege and Sengupta in [46, Section 1] which is only valid for exponential service times. To measure the accuracy of the mean conditional and unconditional sojourn time, for the DPS model we use Fayolle et al. results [13]. As stated earlier in the chapter, in [13] the authors obtain analytical expressions of the mean conditional and unconditional sojourn time under the assumption of exponentially distributed service requirements.

In order to obtain a more complete understanding on the accuracy of the approximation, we will also consider hyperexponential and Pareto distributions. Hyperexponential and Pareto distributions have a decreasing hazard-rate, and their second moment can be made arbitrarily large. In order to derive exact expressions for the mean unconditional sojourn time we solved numerically Equation (4.1). However, the output was not stable enough and therefore we opted to simulate the DPS queue instead (using

MATLAB). The simulation results are based on averaging 10 runs with each run comprising $5 \cdot 10^5$ busy periods. A busy period is defined as the interval of time between two consecutive time epochs when the system becomes empty, such points being regenerative points for the stochastic process of interest.

As mentioned in Section 3.7, we say that B_i has an hyper-exponential distribution with m_i phases if

$$\mathbb{P}(B_i \leq b_i) = 1 - \sum_{k=1}^{m_i} \beta_{ik} e^{(-b_i/\mathbb{E}[B_{ik}])}, \quad (4.28)$$

where β_{ik} is the probability that a class- i customer is exponentially distributed with mean $\mathbb{E}[B_{ik}]$. A particular case of the hyper-exponential distribution is the so-called *degenerate hyper-exponential*. In this case one of the phases has mean 0. For instance, let us consider the case of 2 phases, $m_i = 2$, and let $\beta_{i1} = w, \beta_{i2} = 1 - w, \mathbb{E}[B_{i1}] = 1/(\mu_i w)$ and $\mathbb{E}[B_{i2}] = 0$. It then follows that $\mathbb{E}[B_i] = 1/\mu_i$ and $\mathbb{E}[B_i^2] = \frac{2w}{(w\mu_i)^2} = \frac{2}{w\mu_i^2}$. We then easily obtain that the coefficient of variation satisfies $C_{B_i}^2 = \frac{2}{w} - 1$, and conclude that $C_{B_i} \in [1, \infty)$ as $w \in [0, 1]$. We make the observation that if classes $k = 1, \dots, m_i$ are exponentially distributed (where class k has arrival rate λ_k and mean service requirement $\mathbb{E}[B_k]$) and have the same DPS weight, $g_1 = \dots = g_{m_i}$, then they can be seen as a single (merged) class i with a hyperexponential distribution with parameters $\beta_{ik} = \lambda_k / \sum_{l=1}^{m_i} \lambda_l$ and $\mathbb{E}[B_{ik}] = \mathbb{E}[B_k]$, for each phase $k = 1, \dots, m_i$. This allows us to calculate the moments in DPS with hyperexponential distribution using the algorithm of [46].

We say that B_i has Pareto distribution with scale parameter c_i and shape parameter γ_i if

$$\mathbb{P}(B_i \leq b_i) = 1 - \left(\frac{1}{1 + c_i b_i} \right)^{\gamma_i}.$$

Note that the hyperexponential distribution satisfies the sufficient condition in Equation (2.1), whereas Pareto does not satisfy it; moments of an order higher than γ_i are unbounded. In the numerics we observe that even though condition (2.1) is not satisfied, the light-traffic interpolation approximation remains accurate.

Throughout this chapter the performance criteria will be the relative error. For the first and second moments of the number of customers, we will calculate $100\% \times \frac{\mathbb{E}[N^{DPS}] - \mathbb{E}[N^{DPS,INT}]}{\mathbb{E}[N^{DPS}]}$ and $100\% \times \frac{\mathbb{E}[(N^{DPS})^2] - \mathbb{E}[(N^{DPS,INT})^2]}{\mathbb{E}[(N^{DPS})^2]}$, respectively. And for the mean conditional and unconditional sojourn time, we will calculate

$$100\% \times \frac{\overline{S}_k^{DPS}(\lambda, b) - \overline{S}_k^{DPS,INT}(\lambda, b)}{\overline{S}_k^{DPS}(\lambda, b)} \text{ and } 100\% \times \frac{\overline{S}_k^{DPS}(\lambda) - \overline{S}_k^{DPS,INT}(\lambda)}{\overline{S}_k^{DPS}(\lambda)}, \text{ respectively.}$$

Before explaining in detail the numerical results we have obtained, we summarise our main conclusions: Our approximation is accurate over a broad range of parameter values. We observe that for a given set of parameters, the relative error for the mean conditional sojourn time increases as the service requirement of the tagged customer increases. The error increases as the disparity among the weights increases. For any given scenario, the largest relative error occurs in an intermediate load between 0 and 1. We note that the largest relative errors for the mean conditional sojourn time occur for service requirements b that are very unlikely to occur. This also explains the high

accuracy of our approximation for the mean unconditional sojourn time. We observe that although the Pareto distribution does not satisfy Equation (2.1), the light-traffic interpolation approximation remains accurate. We also compare our approximation to that obtained in [53] and conclude that our approximation outperforms that of [53]. Finally, we remark that our approximation works well across different values for the coefficient of variation.

4.6.1 Queue-length distribution

We measure the accuracy of the approximation obtained in Proposition 4.3.2 for exponentially distributed service requirements by considering the first and second moments that are given in Equations (4.12) and (4.14).

Scenario 1. In Figure 4.3 we plot the relative error of the first and second moments of the total number of customers in the system with respect to the load for exponential service time distributions. We consider two classes with $\mathbb{E}[B_1] = 11/3$, $\mathbb{E}[B_2] = 44/3$. We assume that an arriving customer is of class 1 (class 2) with probability $\alpha_1 = 8/12$ ($\alpha_2 = 4/12$). The weights are set to $g_1 = 2$, $g_2 = 5$. We observe that our approximation for the first and second moments is accurate with at most 1.2% and 3% absolute relative error, respectively.

Scenario 2. In Figure 4.4 we consider 2 classes of customers. Class-1 customers' service requirements follow an exponential distribution of rate μ_1 , while class-2 customers' service requirements follow a degenerate hyper-exponential distribution as defined in Equation (4.28) with parameters $m_2 = 2$, $\beta_{21} = w$, $\beta_{22} = 1 - w$, $\mathbb{E}[B_{21}] = 1/(\mu_2 w)$ and $\mathbb{E}[B_{22}] = 0$. We consider $g_1 = 2$, $g_2 = 5$, $\alpha_1 = 7/12$, $\alpha_2 = 5/12$, $\mathbb{E}[B_1] = 11/3$, $\mathbb{E}[B_2] = 44/3$. In Figure 4.4 we plot the relative error of the first and second moments of the total number of customers in the system with respect to w for different values of the load. Observe that as expected for $\rho \approx 0$ and $\rho \approx 1$ our interpolation approximation is exact. The absolute largest error occurs for intermediate values of the load, as w approaches 0.

4.6.2 Waiting time distribution

In Figure 4.5 we use Proposition 4.4.2 to plot the interpolation approximation of the complementary distribution of the conditional and unconditional waiting time of a class- k customer for Scenario 1. For the conditional waiting time we set the service time of the tagged customer to $b = 11/3$. Class 2 gets relatively a larger weight ($g_1 = 2$, $g_2 = 5$), and as a consequence, we see in Figure 4.5 that the conditional waiting time of class 2 is stochastically smaller than that of class 1. However, the service time of class 1 customers is smaller than that of class 2. As a result, we see that the probability that the unconditional waiting time of class 1 is bigger than x is larger than that of class 2, for x small enough.

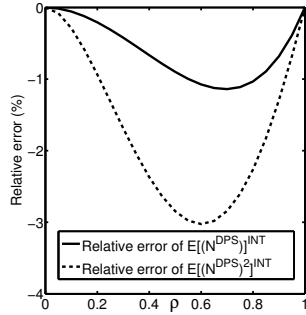


Figure 4.3: Scenario 1. Relative error in DPS of the first and second moments of the total number of customers in the system for exponential service-time distribution.

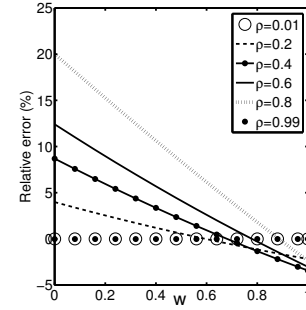
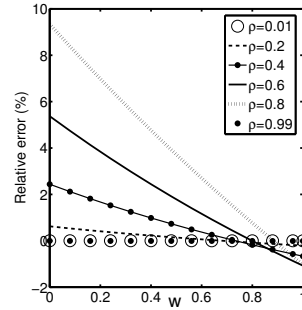


Figure 4.4: Scenario 2. Relative error in DPS of the first (left) and second moments (right) of the number of customers in the system for different values of the load.

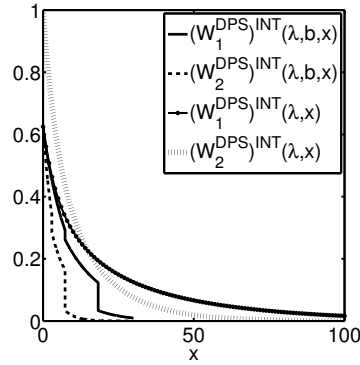


Figure 4.5: Scenario 1. Complementary distribution of the conditional and unconditional waiting time of a class- k customer.

4.6.3 Mean sojourn time

In this section we numerically investigate the accuracy of the approximations obtained for the mean conditional sojourn time and the mean unconditional sojourn time, whose approximations are stated in Equation (4.20) and Corollary 4.5.2, respectively.

As stated earlier in Section 4.2.1, in [13] the authors obtain analytical expressions of the mean conditional and unconditional sojourn time under the assumption of exponentially distributed service requirements. For exponentially distributed service requirements, we will evaluate the accuracy of the approximations by comparing the exact formulas as obtained in [13], see Equations (4.2) and (4.4), with the approximations as given in (4.20) and (4.22). The expectations in (4.20) and (4.22) are calculated numerically using MATLAB's `integral2` command, which transforms the region of integration to a rectangular shape and subdivides it into smaller rectangular regions as needed.

Mean conditional sojourn time

In this section we measure the accuracy of the mean conditional sojourn time given in Equation (4.20).

Scenario 3. In Figure 4.6 we consider four classes $K = 4$ with exponentially distributed service requirements. The parameters of the classes are fixed, and we vary the total arrival rate in order for the load to cover the range of stable values. We consider $\mathbb{E}[B_1] = 2$, $\mathbb{E}[B_2] = 5$, $\mathbb{E}[B_3] = 7$, $\mathbb{E}[B_4] = 10$, $g_1 = 30$, $g_2 = 25$, $g_3 = 20$, $g_4 = 10$, and $\alpha_1 = 10/36$, $\alpha_2 = 5/36$, $\alpha_3 = 8/36$, $\alpha_4 = 13/36$ such that $\lambda_i = \alpha_i * \lambda$, $i = 1, \dots, 4$, where λ is the total arrival rate. In Figure 4.6 we plot the relative error of our approximation for the mean conditional sojourn time of a tagged class- i customer, for $i = 1, \dots, 4$, where the size of the tagged class- i customer, b_i , is selected such that the probability of the event is $\mathbb{P}(B_i \leq b_i) = 0.01$, $\mathbb{P}(B_i \leq b_i) = 0.50$ and $\mathbb{P}(B_i \leq b_i) = 0.99$, respectively. As can be seen, the relative error for the mean conditional sojourn time remains small and always below 6%.

Scenario 4. In Figure 4.7 we consider two classes $K = 2$ with exponentially distributed service requirements. We fixed the parameters $\mathbb{E}[B_1] = 2$, $\mathbb{E}[B_2] = 1$, $g_1 = 1$, $g_2 = 3$, $\alpha_1 = 0.415$, $\alpha_2 = 0.585$ and $\lambda_i = \alpha_i * \lambda$. We let the service requirement of the class- i tagged customer span between 0 and $b_{i,max}$ where $\mathbb{P}(B_i \leq b_{i,max}) = 0.99$ and for each b we plot the largest absolute relative error that can be found for a $\rho \in [0, 1)$. We observe a largest error of at most 6%.

Scenario 5. In Figure 4.8 we consider again two classes with exponentially distributed service requirements. As parameters we fix: $\mathbb{E}[B_1] = 2$, $\mathbb{E}[B_2] = 1$, $\lambda_1 = 0.2$, $\lambda_2 = 1.5\lambda_1$ and $b = 1$. We chose $g_2 = 1 - g_1$ and let g_1 vary on the horizontal axis. In the figure we plot the mean conditional sojourn time and our approximation. We see that the property stated in the subsection entitled *Monotonicity in weights* of Section 4.5.1 is satisfied, namely as g_1 increases $S_1^{DPS,INT}(\lambda, b)$ decreases and $S_2^{DPS,INT}(\lambda, b)$ increases.

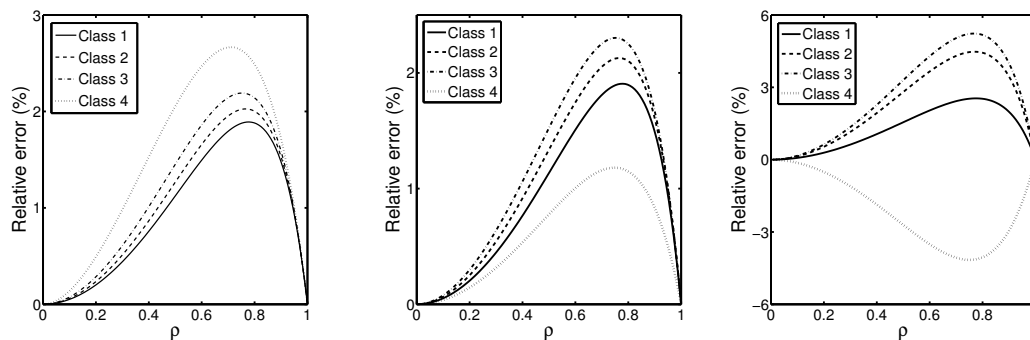


Figure 4.6: Scenario 3: Relative error for the mean conditional sojourn time for a tagged class- i customer with service requirement b_i such that $\mathbb{P}(B_i \leq b_i) = 0.01$ (left), $\mathbb{P}(B_i \leq b_i) = 0.50$ (middle), $\mathbb{P}(B_i \leq b_i) = 0.99$ (right).

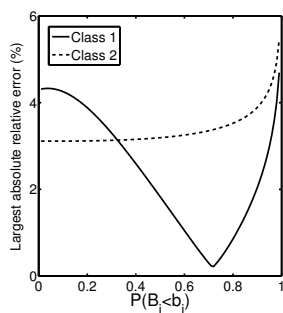


Figure 4.7: Scenario 4: Largest absolute relative error for the mean conditional sojourn time as a function of $\mathbb{P}(B_i \leq b_i)$, $i = 1, 2$.

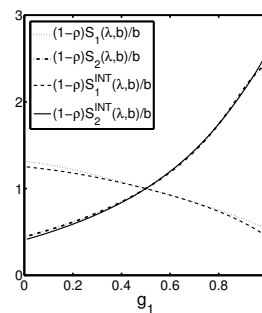


Figure 4.8: Scenario 5: Mean conditional sojourn time as a function of g_1 .

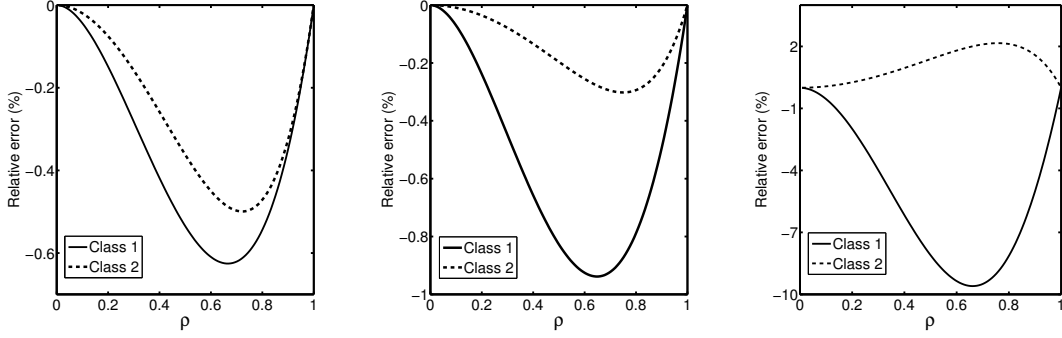


Figure 4.9: Scenario 6: Relative error for the mean conditional sojourn time for a tagged class- i customer with service requirement b_i such that $\mathbb{P}(B_i \leq b_i) = 0.01$ (left), $\mathbb{P}(B_i \leq b_i) = 0.50$ (middle), $\mathbb{P}(B_i \leq b_i) = 0.99$ (right).

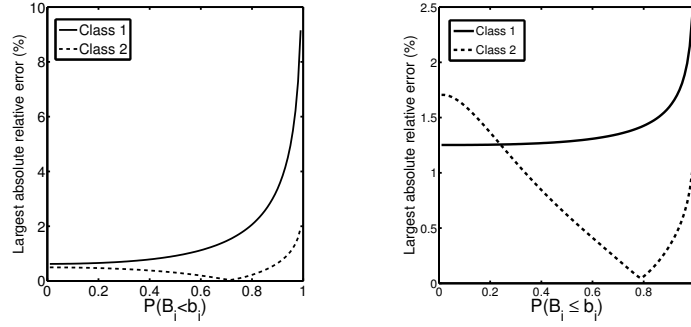


Figure 4.10: Scenario 6: Largest absolute relative error for the mean conditional sojourn time as a function of $\mathbb{P}(B_i \leq b_i)$, $i = 1, 2$ with weights $g_1 = 2, g_2 = 5$ (left) and $g_1 = 5, g_2 = 2$ (right).

Besides, it can be observed from the figure that the approximation loses accuracy as one class is given more priority, i.e., $g_1 \rightarrow 0$ or $g_1 \rightarrow 1$.

Scenario 6. In Figure 4.9 we consider two classes with hyperexponential distributed service requirements with $\mathbb{E}[B_1] = 11/3$, $\mathbb{E}[B_2] = 44/3$. Each of the hyperexponential distributions has 3 phases. The parameters are as follows: for class 1 we take $\mathbb{E}[B_{11}] = 3.5$, $\mathbb{E}[B_{12}] = 2$, $\mathbb{E}[B_{13}] = 5$, $p_{11} = 10/21$, $p_{12} = 5/21$, $p_{13} = 6/21$, and for class 2 we take $\mathbb{E}[B_{21}] = 10$, $\mathbb{E}[B_{22}] = 15$, $\mathbb{E}[B_{23}] = 20$, $p_{21} = 4/15$, $p_{22} = 8/15$, $p_{23} = 3/15$. The weights are set to $g_1 = 2$ and $g_2 = 5$. We assume that an arriving customer is of class 1 (class 2) with probability $\alpha_1 = 7/12$ ($\alpha_2 = 5/12$). As in *Scenario 3*, we select the service requirement of the tagged customer such that $\mathbb{P}(B_i \leq b_i) = 0.01, 0.5$ and 0.99 . We see that the error increases as the size of the tagged customer increases. However it is remarkable how accurate the approximation is.

In Figure 4.10 we consider *Scenario 6* with weights $g_1 = 2, g_2 = 5$ (left) and $g_1 = 5, g_2 = 2$ (right), respectively. We vary the service requirement of the class- i tagged customer between 0 and $b_{i,max}$ where $\mathbb{P}(B_i \leq b_{i,max}) = 0.99$ and for each b we plot the largest absolute relative error that can be found for a $\rho \in [0, 1)$. It can be observed that the largest absolute relative error is smaller as more priority is given to the class with small mean service requirements. In Figure 4.10 (left) the largest absolute relative error is of the order of 9% for the class with the smallest weight and of the order of 2% for the class with the highest weight. While in Figure 4.10 (right) the largest absolute relative error is of the order of 1.7% for the class with the smallest weight and of the order of 2.5% for the class with the highest weight.

Mean unconditional sojourn time

In this section we evaluate the accuracy of the mean unconditional sojourn time given in Corollary 4.5.2.

In Figure 4.11 we consider the same parameter setting as in *Scenario 3*, and we observe that the largest relative error for the mean unconditional sojourn time is less than 3.5%.

In Figure 4.12 we consider two classes with hyper-exponentially distributed service requirements. The parameters are the ones considered in *Scenario 6*. We plot the relative error of the mean unconditional sojourn time for our approximation. We conclude that our approximation works very well. The largest relative error for the mean unconditional sojourn time is around 3%. We compare our approximation to that obtained in [53]. In the latter, the unconditional sojourn time was approximated by expressions obtained when one of the classes lives on a relatively faster time scale than the other class. Under this scenario, class 1 represents the class of a vast majority of customers with small service requirements while class 2 represents the tiny fraction of customers with huge service requirements. Therefore, the approximation as given in [53] is $\mathbb{E}[S_1(\lambda)] \approx \left(\frac{g_2}{g_1} \frac{\rho_2}{1 - \rho} + 1 \right) \frac{\mathbb{E}[B_1]}{1 - \rho_1}$ and $\mathbb{E}[S_2(\lambda)] \approx \frac{\rho_2}{1 - \rho}$. In Figure 4.12 we plot the mean unconditional sojourn time of [53]. We observe that the largest relative error is 17%. We conclude that our approximation outperforms that of [53].

As pointed out in the beginning of the section, we observe that the relative error for the mean unconditional sojourn time tends to be smaller than the ones observed for the mean conditional sojourn time. This can be explained by noting that the largest errors in the mean conditional sojourn time tend to occur for service requirements that happen with a very low probability.

In Figure 4.13 we consider two classes with hyper-exponentially distributed service requirements. The parameters are the same as in *Scenario 6*. We chose $g_2 = 1 - g_1$ and let g_1 vary on the horizontal axis. For each given g_1 we calculate the largest absolute relative error for the mean unconditional sojourn time as we let ρ range from 0 to 1. We

observe that the relative error for the unconditional sojourn time is at most of 30%, and that this happens when class 2 receives full priority.

Scenario 7. In Figure 4.14 we consider 2 classes of customers. Class-1 customers' service requirements follow an exponential distribution of rate μ_1 , while class-2 customers' service requirements follow a hyper-exponential distribution as defined in Equation (4.28) with parameters $m_2 = 2, p_{21} = p, p_{22} = 1 - p, \mathbb{E}[B_{21}] = 1/(\mu_2 p)$ and $\mathbb{E}[B_{22}] = 0$. The latter distribution is referred to as a degenerate hyper-exponential distribution with $\mathbb{E}[B_2] = 1/\mu_2$ and $\mathbb{E}[B_2^2] = \frac{2p}{(p\mu_2)^2} = \frac{2}{p\mu_2^2}$. We then easily obtain that the coefficient of variation satisfies $C_{B_2}^2 = \frac{2}{p} - 1$, and conclude that $C_{B_2} \in [1, \infty)$ as $p \in [0, 1]$.

We consider $\rho = 0.7, g_1 = 1, g_2 = 5, \alpha_1 = 1/4, \alpha_2 = 3/4, \mu_1 = 1$ and $\mu_2 = 1$.

In Figure 4.14 we plot the relative error of the mean unconditional sojourn time for an *arbitrary* customer with respect to p for our approximation. We observe that our approximation works well across all values of p . In the limit $p \downarrow 0$ we obtain the following analytical expression for the relative error

$$\lim_{p \downarrow 0} \text{Rel.Error} = \left(1 - \frac{1 + \alpha_1 \left(\frac{g_2}{g_1} - 1 \right) (\rho^2 + \rho_2(1 - \rho))}{1 + \alpha_1 \left(\frac{g_2}{g_1} - 1 \right) \frac{\rho_2}{1 - \rho_1}} \right) \cdot 100\%. \quad (4.29)$$

For the parameters of Figure 4.14, the latter is equal to -0.6806% (see also the curve in Figure 4.14). This implies that for very high coefficient of variation ($p \downarrow 0$) the performance of our approximation is very good. From Equation (4.29) we note that even though the coefficient of variation explodes as $p \downarrow 0$, the relative error of our approximations remains bounded. In the limit $p \uparrow 1$, our approximation is exact since in this particular case class-2 customers follow an exponential distribution of rate μ_2 and under the assumption $\mu_1 = \mu_2$ we proved in Equation (4.27) that our approximation becomes exact. In Figure 4.14 we also plot the approximation as obtained in [53]. We note that as $p \downarrow 0$, class-2 customers arrive very rarely and are huge. Hence, the approximation of [53] becomes exact as $p \downarrow 0$. For $p > 0$ the absolute relative error is monotone increasing taking the value 60% at $p \uparrow 1$.

In the following two scenarios we consider Pareto distributed serve requirements. As mentioned in the beginning of the section Pareto does not satisfy Equation (2.1). However, we will observe that the light-traffic interpolation remains accurate.

Scenario 8. In Figure 4.15 we consider Pareto distributed service requirements. We consider four classes with $c_1 = 1/4, c_2 = 1/10, c_3 = 1/14, c_4 = 1/20$ and $\gamma_1 = 3, \gamma_2 = 3, \gamma_3 = 3, \gamma_4 = 3$, such that, $\mathbb{E}[B_1] = 2, \mathbb{E}[B_2] = 5, \mathbb{E}[B_3] = 7, \mathbb{E}[B_4] = 10$. The weights are set to $g_1 = 30, g_2 = 25, g_3 = 20, g_4 = 10$ and $\alpha_1 = 10/36, \alpha_2 = 5/36, \alpha_3 = 8/36, \alpha_4 = 13/36$ such that $\lambda_i = \alpha_i * \lambda, i = 1, \dots, 4$, where λ is the total arrival rate. Notice that $\mathbb{E}[B_i], g_i, \alpha_i, i = 1, \dots, 4$ are as in Scenario 3. We note that the point $\rho = 1$ is obtained from the heavy-traffic condition and is therefore exact. For $\rho \neq 1$, we simulated

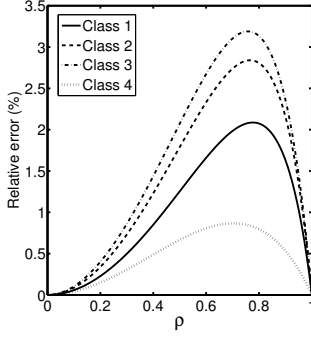


Figure 4.11: Scenario 3: Relative error for the mean unconditional sojourn time.

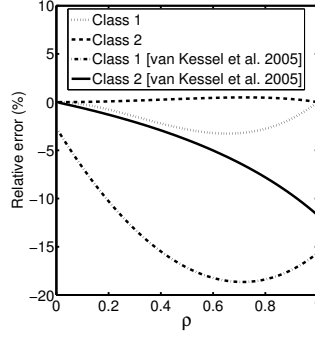


Figure 4.12: Scenario 6: Relative error for the mean unconditional sojourn time.

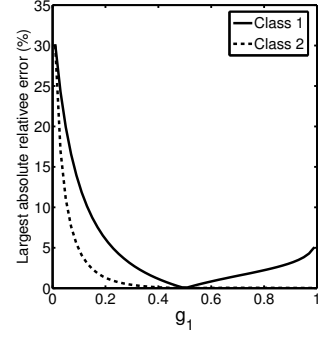


Figure 4.13: Scenario 6: Largest absolute relative error for the mean unconditional sojourn time with respect to the weight g_1 .

the system in order to compare our approximation. We conclude that the largest relative error for the mean unconditional sojourn time is around 5%.

Scenario 9. In Figure 4.16 we consider Pareto distributed service requirements. We consider two classes with $c_1 = 3/22, c_2 = 3/88$ and $\gamma_1 = 3, \gamma_2 = 3$, such that, $\mathbb{E}[B_1] = 11/3, \mathbb{E}[B_2] = 44/3$ and the weights are set to $g_1 = 2$ and $g_2 = 5$. We assume that an arriving customer is of class 1 (class 2) with probability $\alpha_1 = 7/12$ ($\alpha_2 = 5/12$). Notice that $\mathbb{E}[B_i], g_i, \alpha_i, i = 1, 2$ are as in Scenario 6. We note that the point $\rho = 1$ is obtained from the heavy-traffic condition and is therefore exact. For $\rho \neq 1$, we simulated the system in order to compare our approximation. We conclude that the largest relative error for the mean unconditional sojourn time is less than 5%.

4.A Appendix

4.A.1 Proof of Lemma 4.3.1

We obtain the zeroth, first and second light-traffic derivatives of the probability generating function of the queue length vector under DPS. The zeroth derivative satisfies

$$\left(\psi^{DPS}\right)^{(0)}(\lambda, \vec{z})\Big|_{\lambda=0} = G\left(0, \vec{z} \mid A(-\infty, \infty) = 0\right) = z_1^0 \cdots z_K^0 = 1. \quad (4.30)$$

Let t denote the time epoch in which a customer arrives to the system, and let U_t denote its class. For the first derivative there might happen two different cases:

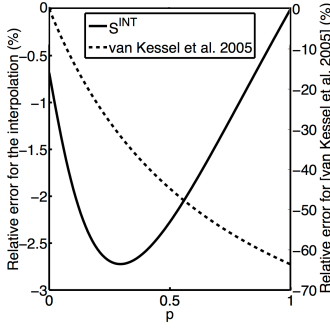


Figure 4.14: Scenario 5: Relative error for the mean unconditional sojourn time of an *arbitrary* customer for our approximation and for the approximation as obtained in [53].

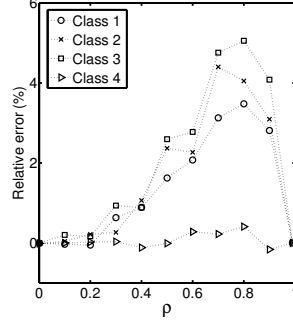


Figure 4.15: Scenario 6: Relative error for the mean unconditional sojourn time.

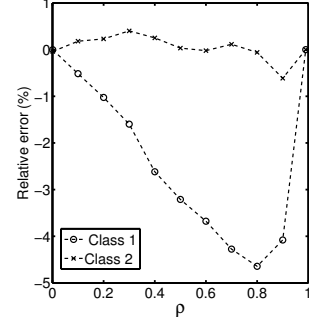


Figure 4.16: Scenario 7: Relative error for the mean unconditional sojourn time.

If $t > 0$, we have $z_{U_t}^0 = 1$ and therefore

$$G\left(0, \bar{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t\right) - G\left(0, \bar{z} \middle| A(-\infty, \infty) = 0\right) = 0.$$

If $t < 0$, we have

$$\begin{cases} z_{U_t}^0 & \text{if } t + B_{U_t} < 0 \\ z_{U_t} & \text{if } t + B_{U_t} > 0 \end{cases}$$

such that

$$G\left(0, \bar{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t\right) = \mathbb{E}[\mathbf{1}[t + B_{U_t} < 0] + \mathbf{1}[t + B_{U_t} > 0] z_{U_t}].$$

Therefore,

$$\begin{aligned} & \left(\psi^{DPS}\right)^{(1)}(\lambda, \bar{z}) \Big|_{\lambda=0} \\ &= \int_{-\infty}^0 \left(G\left(0, \bar{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t\right) - G\left(0, \bar{z} \middle| A(-\infty, \infty) = 0\right) \right) dt \\ &= \mathbb{E} \left[\int_{-\infty}^0 (\mathbf{1}[t + B_{U_t} < 0] + \mathbf{1}[t + B_{U_t} > 0] z_{U_t} - 1) dt \right] = \mathbb{E}[(z_{U_t} - 1)B_{U_t}]. \end{aligned} \tag{4.31}$$

To calculate the second derivative let us assume $t' < t''$, where t' and t'' denote the arrival epochs of two customers. At the end, because of symmetry, we multiply the final

result by 2. Then, we separate three main different cases:

If $0 < t' < t''$, the system will be empty at the tagged customer arrival and therefore we obtain

$$\begin{aligned} & G\left(0, \vec{z} \middle| A(-\infty, \infty) = 2, \tau_1 = t', \tau_2 = t''\right) - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t'\right) \\ & - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t''\right) + G\left(0, \vec{z} \middle| A(-\infty, \infty) = 0\right) = 0. \end{aligned}$$

If $t' < 0$ & $0 < t''$ two different cases are observed:

$$\begin{cases} z_{U_{t'}}^0 \cdot z_{U_{t''}}^0 & \text{if } t' + B_{U_{t'}} < 0 \\ z_{U_{t'}} & \text{if } t' + B_{U_{t'}} > 0, \end{cases}$$

such that,

$$\begin{aligned} & G\left(0, \vec{z} \middle| A(-\infty, \infty) = 2, \tau_1 = t', \tau_2 = t''\right) = \mathbb{E}\left[\mathbf{1}\left[t' + B_{U_{t'}} < 0\right] + \mathbf{1}\left[t' + B_{U_{t'}} > 0\right] z_{U_{t'}}\right] \\ & = G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t'\right). \end{aligned}$$

Therefore,

$$\begin{aligned} & G\left(0, \vec{z} \middle| A(-\infty, \infty) = 2, \tau_1 = t', \tau_2 = t''\right) - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t'\right) \\ & - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t''\right) + G\left(0, \vec{z} \middle| A(-\infty, \infty) = 0\right) = 0. \end{aligned}$$

If $t' < t'' < 0$ there might happen several cases as shown below.

First, if $t' + B_{U_{t'}} < t''$ we have

$$\begin{cases} z_{U_{t'}}^0 \cdot z_{U_{t''}}^0 & \text{if } t'' + B_{U_{t''}} < 0 \\ z_{U_{t''}} & \text{if } t'' + B_{U_{t''}} > 0. \end{cases}$$

Therefore,

$$\begin{aligned} & G\left(0, \vec{z} \middle| A(-\infty, \infty) = 2, \tau_1 = t', \tau_2 = t''\right) - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t'\right) \\ & - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t''\right) + G\left(0, \vec{z} \middle| A(-\infty, \infty) = 0\right) = 0. \end{aligned}$$

Second, if $t' + B_{U_{t'}} > t''$, the following six cases are observed

$$\left\{ \begin{array}{ll} z_{U_{t'}} & \text{if } (B_{U_{t'}} - t'' + t') \frac{g_{U_{t''}}}{g_{U_{t'}}} > B_{U_{t''}} \quad \& \quad t'' + B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} < 0 \\ & \& \quad t' + B_{U_{t'}} + B_{U_{t''}} > 0 \\ 1 & \text{if } (B_{U_{t'}} - t'' + t') \frac{g_{U_{t''}}}{g_{U_{t'}}} > B_{U_{t''}} \quad \& \quad t'' + B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} < 0 \\ & \& \quad t' + B_{U_{t'}} + B_{U_{t''}} < 0 \\ z_{U_{t'}} \cdot z_{U_{t''}} & \text{if } (B_{U_{t'}} - t'' + t') \frac{g_{U_{t''}}}{g_{U_{t'}}} > B_{U_{t''}} \quad \& \quad t'' + B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} > 0 \\ z_{U_{t''}} & \text{if } (B_{U_{t'}} - t'' + t') \frac{g_{U_{t''}}}{g_{U_{t'}}} < B_{U_{t''}} \quad \& \quad t'' + (B_{U_{t'}} - t'' + t') \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t'}}} < 0 \\ & \& \quad t' + B_{U_{t'}} + B_{U_{t''}} > 0 \\ 1 & \text{if } (B_{U_{t'}} - t'' + t') \frac{g_{U_{t''}}}{g_{U_{t'}}} < B_{U_{t''}} \quad \& \quad t'' + (B_{U_{t'}} - t'' + t') \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t'}}} < 0 \\ & \& \quad t' + B_{U_{t'}} + B_{U_{t''}} < 0 \\ z_{U_{t'}} \cdot z_{U_{t''}} & \text{if } (B_{U_{t'}} - t'' + t') \frac{g_{U_{t''}}}{g_{U_{t'}}} < B_{U_{t''}} \quad \& \quad t'' + (B_{U_{t'}} - t'' + t') \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t'}}} > 0. \end{array} \right.$$

Then,

$$\begin{aligned} & \int_{-\infty}^{\infty} \left(\int_{t'}^{\infty} \left(G(0, \vec{z} | A(-\infty, \infty) = 2, \tau_1 = t', \tau_2 = t'') - G(0, \vec{z} | A(-\infty, \infty) = 1, \tau_1 = t') \right. \right. \\ & \quad \left. \left. - G(0, \vec{z} | A(-\infty, \infty) = 1, \tau_1 = t'') + G(0, \vec{z} | A(-\infty, \infty) = 0) \right) dt'' \right) dt' \\ &= \int_{-\infty}^0 \left(\int_{t'}^0 \left(G(0, \vec{z} | A(-\infty, \infty) = 2, \tau_1 = t', \tau_2 = t'') - G(0, \vec{z} | A(-\infty, \infty) = 1, \tau_1 = t') \right. \right. \\ & \quad \left. \left. - G(0, \vec{z} | A(-\infty, \infty) = 1, \tau_1 = t'') + G(0, \vec{z} | A(-\infty, \infty) = 0) \right) dt'' \right) dt' \\ &= \int_0^{\infty} \left(\int_0^r \left(G(0, \vec{z} | A(-\infty, \infty) = 2, \tau_1 = -r, \tau_2 = -s) - G(0, \vec{z} | A(-\infty, \infty) = 1, \tau_1 = -r) \right. \right. \\ & \quad \left. \left. - G(0, \vec{z} | A(-\infty, \infty) = 1, \tau_1 = -s) + G(0, \vec{z} | A(-\infty, \infty) = 0) \right) ds \right) dr \end{aligned}$$

where

$$\begin{aligned}
& G\left(0, \vec{z} \middle| A(-\infty, \infty) = 2, \tau_1 = -r, \tau_2 = -s\right) - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = -r\right) \\
& - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = -s\right) + G\left(0, \vec{z} \middle| A(-\infty, \infty) = 0\right) \\
& = \mathbb{E} \left[\mathbf{1} \left[-r + B_{U_{t'}} > -s \right] \left(\right. \right. \\
& \quad \mathbf{1} \left[(B_{U_{t'}} + s - r) \frac{g_{U_{t''}}}{g_{U_{t'}}} > B_{U_{t''}}, -s + B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} < 0 \right] \\
& \quad \cdot \mathbf{1} \left[-r + B_{U_{t'}} + B_{U_{t''}} > 0 \right] (z_{U_{t'}} - 1) \\
& \quad + \mathbf{1} \left[(B_{U_{t'}} + s - r) \frac{g_{U_{t''}}}{g_{U_{t'}}} > B_{U_{t''}}, -s + B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} > 0 \right] (z_{U_{t'}} \cdot z_{U_{t''}} - 1) \\
& \quad + \mathbf{1} \left[(B_{U_{t'}} + s - r) \frac{g_{U_{t''}}}{g_{U_{t'}}} < B_{U_{t''}}, -s + (B_{U_{t'}} + s - r) \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t'}}} < 0 \right] \\
& \quad \cdot \mathbf{1} \left[-r + B_{U_{t'}} + B_{U_{t''}} > 0 \right] (z_{U_{t''}} - 1) \\
& \quad + \mathbf{1} \left[(B_{U_{t'}} + s - r) \frac{g_{U_{t''}}}{g_{U_{t'}}} < B_{U_{t''}}, -s + (B_{U_{t'}} + s - r) \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t'}}} > 0 \right] (z_{U_{t'}} \cdot z_{U_{t''}} - 1) \\
& \quad + \mathbb{E} \left[\mathbf{1} \left[-r + B_{U_{t'}} > 0 \right] (1 - z_{U_{t'}}) \right] \\
& \quad \left. \left. + \mathbb{E} \left[\mathbf{1} \left[-s + B_{U_{t''}} > 0 \right] (1 - z_{U_{t''}}) \right] \right) \right].
\end{aligned}$$

Let calculate the six integrals above. First,

$$\begin{aligned}
& \mathbb{E} \left[(z_{U_{t'}} - 1) \int_0^\infty \left(\int_s^\infty \mathbf{1} \left[-r + B_{U_{t'}} > -s \right] \right. \right. \\
& \quad \mathbf{1} \left[(B_{U_{t'}} + s - r) \frac{g_{U_{t''}}}{g_{U_{t'}}} > B_{U_{t''}}, -s + B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} < 0 \right] \\
& \quad \left. \left. \mathbf{1} \left[-r + B_{U_{t'}} + B_{U_{t''}} > 0 \right] dr \right) ds \right] \\
& = \mathbb{E} \left[(z_{U_{t'}} - 1) \int_0^\infty \left(\int_s^\infty \mathbf{1} \left[s + B_{U_{t'}} > r \right] \right. \right. \\
& \quad \left. \left. \mathbf{1} \left[B_{U_{t'}} + s - \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}} > r, B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} < s \right] \mathbf{1} \left[B_{U_{t'}} + B_{U_{t''}} > r \right] dr \right) ds \right]
\end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left[\left(z_{U_{t'}} - 1 \right) \int_0^\infty \mathbf{1} \left[B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} < s \right] \right. \\
&\quad \left. \left(\int_s^\infty \mathbf{1} \left[B_{U_{t'}} + s - \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}} > r, B_{U_{t'}} + B_{U_{t''}} > r \right] dr \right) ds \right] \\
&= \mathbb{E} \left[\left(z_{U_{t'}} - 1 \right) \int_0^\infty \mathbf{1} \left[B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} < s \right] \right. \\
&\quad \left. \left(\int_s^{\max\{s, \min\{B_{U_{t'}} + s - \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, B_{U_{t'}} + B_{U_{t''}}\}\}} dr \right) ds \right] \\
&= \mathbb{E} \left[\left(z_{U_{t'}} - 1 \right) \int_0^\infty \mathbf{1} \left[B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} < s \right] \right. \\
&\quad \left. \left(\max\{s, \min\{B_{U_{t'}} + s - \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, B_{U_{t'}} + B_{U_{t''}}\}\} - s \right) ds \right] \\
&= \mathbb{E} \left[\left(z_{U_{t'}} - 1 \right) \int_0^\infty \mathbf{1} \left[B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} < s \right] \right. \\
&\quad \left. \max\{0, B_{U_{t'}} + \min\{-\frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, B_{U_{t''}} - s\}\} ds \right] \\
&= \mathbb{E} \left[\left(z_{U_{t'}} - 1 \right) \int_0^\infty \mathbf{1} \left[B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} < s \right] \max\{0, B_{U_{t'}} + B_{U_{t''}} - s\} ds \right] \\
&= \mathbb{E} \left[\left(z_{U_{t'}} - 1 \right) \int_{B_{U_{t''}}(g_{U_{t'}} + g_{U_{t''}})/g_{U_{t''}}}^{\max\{B_{U_{t'}} + B_{U_{t''}}, B_{U_{t''}}(g_{U_{t'}} + g_{U_{t''}})/g_{U_{t''}}\}} (B_{U_{t'}} + B_{U_{t''}} - s) ds \right] \\
&= \mathbb{E} \left[\left(z_{U_{t'}} - 1 \right) \left((B_{U_{t'}} + B_{U_{t''}}) \max\{B_{U_{t'}} + B_{U_{t''}} - B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}}, 0\} \right. \right. \\
&\quad \left. \left. - \frac{s^2}{2} \Big|_{B_{U_{t''}}(g_{U_{t'}} + g_{U_{t''}})/g_{U_{t''}}}^{\max\{B_{U_{t'}} + B_{U_{t''}}, B_{U_{t''}}(g_{U_{t'}} + g_{U_{t''}})/g_{U_{t''}}\}} \right) \right] \\
&= \mathbb{E} \left[\left(z_{U_{t'}} - 1 \right) \left((B_{U_{t'}} + B_{U_{t''}}) \max\{B_{U_{t'}} - B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}}, 0\} \right. \right. \\
&\quad \left. \left. - \frac{1}{2} \left(\left(B_{U_{t''}} + \max\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}}\} \right)^2 - \left(B_{U_{t''}} \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \right)^2 \right) \right) \right] \\
&= \mathbb{E} \left[\left(z_{U_{t'}} - 1 \right) \left((B_{U_{t'}} + B_{U_{t''}}) \max\{B_{U_{t'}} - B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}}, 0\} \right. \right. \\
&\quad \left. \left. - \frac{1}{2} \left(\left(B_{U_{t''}} \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) + B_{U_{t'}} - \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\} \right)^2 \right. \right. \right. \\
&\quad \left. \left. \left. - \left(B_{U_{t''}} \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \right)^2 \right) \right) \right] 81
\end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left[\left(z_{U_{t'}} - 1 \right) \left((B_{U_{t'}} + B_{U_{t''}}) \left(B_{U_{t'}} - \min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \} \right) \right. \right. \\
&\quad \left. \left. - \frac{1}{2} \left(\left(B_{U_{t'}} - \min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \} \right)^2 \right. \right. \right. \\
&\quad \left. \left. + 2B_{U_{t''}} \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \left(B_{U_{t'}} - \min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \} \right) \right) \right) \right] \\
&= \mathbb{E} \left[\left(z_{U_{t'}} - 1 \right) \left(\left(B_{U_{t'}} - B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \left(B_{U_{t'}} - \min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \} \right) \right. \right. \\
&\quad \left. \left. - \frac{1}{2} \left(B_{U_{t'}} - \min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \} \right)^2 \right) \right]. \tag{4.32}
\end{aligned}$$

Second,

$$\begin{aligned}
&\mathbb{E} \left[\int_0^\infty \left(\int_s^\infty \mathbf{1}[-r + B_{U_{t'}} > -s] \mathbf{1} \left[(B_{U_{t'}} + s - r) \frac{g_{U_{t''}}}{g_{U_{t'}}} > B_{U_{t''}}, -s + B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} > 0 \right] \right. \right. \\
&\quad \left. \left. \cdot (z_{U_{t'}} \cdot z_{U_{t''}} - 1) dr \right) ds \right] \\
&= \mathbb{E} \left[\int_0^\infty \mathbf{1} \left[B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} > s \right] \right. \\
&\quad \left. \cdot \left(\int_s^\infty \mathbf{1} \left[s + B_{U_{t'}} > r, B_{U_{t'}} + s - \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}} > r \right] (z_{U_{t'}} \cdot z_{U_{t''}} - 1) dr \right) ds \right] \\
&= \mathbb{E} \left[(z_{U_{t'}} \cdot z_{U_{t''}} - 1) \int_0^\infty \mathbf{1} \left[B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} > s \right] \left(\int_s^{\max\{s, B_{U_{t'}} + s - \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\}} dr \right) ds \right] \\
&= \mathbb{E} \left[(z_{U_{t'}} \cdot z_{U_{t''}} - 1) B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} \max\{0, B_{U_{t'}} - \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\} \right] \\
&= \mathbb{E} \left[(z_{U_{t'}} \cdot z_{U_{t''}} - 1) B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} \left(B_{U_{t'}} - \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\} \right) \right]. \tag{4.33}
\end{aligned}$$

Third,

$$\begin{aligned}
& \mathbb{E} \left[\left(z_{U_{t''}} - 1 \right) \int_0^\infty \left(\int_0^r \mathbf{1} \left[-r + B_{U_{t'}} > -s \right] \right. \right. \\
& \quad \cdot \mathbf{1} \left[\left(B_{U_{t'}} + s - r \right) \frac{g_{U_{t''}}}{g_{U_{t'}}} < B_{U_{t''}}, -s + \left(B_{U_{t'}} + s - r \right) \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t'}}} < 0 \right] \\
& \quad \cdot \mathbf{1} \left[-r + B_{U_{t'}} + B_{U_{t''}} > 0 \right] \Big) ds dr \Big] \\
&= \mathbb{E} \left[\left(z_{U_{t''}} - 1 \right) \int_0^\infty \left(\mathbf{1} \left[B_{U_{t'}} + B_{U_{t''}} > r \right] \right. \right. \\
& \quad \left. \int_0^r \mathbf{1} \left[s > r - B_{U_{t'}} \right] \mathbf{1} \left[s < r - B_{U_{t'}} + \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, s < (r - B_{U_{t'}}) \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \right] \Big) ds dr \Big] \\
&= \mathbb{E} \left[\left(z_{U_{t''}} - 1 \right) \int_0^\infty \mathbf{1} \left[B_{U_{t'}} + B_{U_{t''}} > r \right] \left(\int_{(r - B_{U_{t'}})^+}^{\max\{(r - B_{U_{t'}})^+, \min\{r - B_{U_{t'}} + \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, (r - B_{U_{t'}}) \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right)\}} \right) ds dr \right] \\
&= \mathbb{E} \left[\left(z_{U_{t''}} - 1 \right) \int_0^\infty \mathbf{1} \left[B_{U_{t'}} + B_{U_{t''}} > r \right] \left(\max\{(r - B_{U_{t'}})^+, \min\{r - B_{U_{t'}} + \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, (r - B_{U_{t'}}) \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right)\}\} - (r - B_{U_{t'}})^+ \right) dr \right] \\
&= \mathbb{E} \left[\left(z_{U_{t''}} - 1 \right) \int_0^{B_{U_{t'}} + B_{U_{t''}}} \left(\max\{0, r + \min\{0, -B_{U_{t'}} + \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, r \frac{g_{U_{t'}}}{g_{U_{t''}}} - B_{U_{t'}} \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right)\} - (r - \min\{r, B_{U_{t'}}\})\} \right) dr \right] \\
&= \mathbb{E} \left[\left(z_{U_{t''}} - 1 \right) \left(\int_0^{B_{U_{t'}}} \max\{0, \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, (r - B_{U_{t'}}) \frac{g_{U_{t'}}}{g_{U_{t''}}}\} - B_{U_{t'}} + \min\{r, B_{U_{t'}}\}\} dr \right. \right. \\
& \quad \left. \left. + \int_{B_{U_{t'}}}^{B_{U_{t'}} + B_{U_{t''}}} \max\{0, \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, (r - B_{U_{t'}}) \frac{g_{U_{t'}}}{g_{U_{t''}}}\} - B_{U_{t'}} + \min\{r, B_{U_{t'}}\}\} dr \right) \right] \\
&= \mathbb{E} \left[\left(z_{U_{t''}} - 1 \right) \left(\int_0^{B_{U_{t'}}} \max\{0, (r - B_{U_{t'}}) \frac{g_{U_{t'}}}{g_{U_{t''}}} - B_{U_{t'}} + r\} dr \right. \right. \\
& \quad \left. \left. + \int_{B_{U_{t'}}}^{B_{U_{t'}} + B_{U_{t''}}} \left(\min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, (r - B_{U_{t'}}) \frac{g_{U_{t'}}}{g_{U_{t''}}}\} - B_{U_{t'}} + B_{U_{t'}} \right) dr \right) \right]
\end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left[\left(z_{U_{t''}} - 1 \right) \left(\int_0^{B_{U_{t'}}} 0 dr + \int_{B_{U_{t'}}}^{B_{U_{t'}} + B_{U_{t''}}} \min \left\{ B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, (r - B_{U_{t'}}) \frac{g_{U_{t'}}}{g_{U_{t''}}} \right\} dr \right) \right] \\
&= \mathbb{E} \left[\left(z_{U_{t''}} - 1 \right) \left(\int_{B_{U_{t'}}}^{B_{U_{t'}} + \min \left\{ \frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}} \right\}} (r - B_{U_{t'}}) \frac{g_{U_{t'}}}{g_{U_{t''}}} dr \right. \right. \\
&\quad \left. \left. + \int_{B_{U_{t'}} + \min \left\{ \frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}} \right\}}^{B_{U_{t'}} + B_{U_{t''}}} \min \left\{ B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}} \right\} dr \right) \right] \\
&= \mathbb{E} \left[\left(z_{U_{t''}} - 1 \right) \left(\int_{B_{U_{t'}}}^{B_{U_{t'}} + \min \left\{ \frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}} \right\}} (r - B_{U_{t'}}) \frac{g_{U_{t'}}}{g_{U_{t''}}} dr \right. \right. \\
&\quad \left. \left. + \int_{B_{U_{t'}} + \min \left\{ \frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}} \right\}}^{B_{U_{t'}} + B_{U_{t''}}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \min \left\{ \frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}} \right\} dr \right) \right] \\
&= \mathbb{E} \left[\left(z_{U_{t''}} - 1 \right) \left(\int_{B_{U_{t'}}}^{B_{U_{t'}} + \min \left\{ \frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}} \right\}} (r - B_{U_{t'}}) \frac{g_{U_{t'}}}{g_{U_{t''}}} dr \right. \right. \\
&\quad \left. \left. + \int_{B_{U_{t'}} + \min \left\{ \frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}} \right\}}^{B_{U_{t'}} + B_{U_{t''}}} B_{U_{t'}} dr \right) \right] \\
&= \mathbb{E} \left[\left(z_{U_{t''}} - 1 \right) \left(\frac{g_{U_{t'}}}{2g_{U_{t''}}} \min \left\{ \frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}} \right\}^2 - B_{U_{t'}} \min \left\{ \frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}} \right\} + B_{U_{t'}} B_{U_{t''}} \right) \right]. \tag{4.34}
\end{aligned}$$

Fourth,

$$\begin{aligned}
&\mathbb{E} \left[\left(z_{U_{t'}} \cdot z_{U_{t''}} - 1 \right) \int_0^\infty \left(\int_0^r \mathbf{1} \left[-r + B_{U_{t'}} > -s \right] \right. \right. \\
&\quad \left. \left. \cdot \mathbf{1} \left[(B_{U_{t'}} + s - r) \frac{g_{U_{t''}}}{g_{U_{t'}}} < B_{U_{t''}}, -s + (B_{U_{t'}} + s - r) \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t'}}} > 0 \right] ds \right) dr \right] \\
&= \mathbb{E} \left[\left(z_{U_{t'}} \cdot z_{U_{t''}} - 1 \right) \int_0^\infty \left(\int_0^r \mathbf{1} \left[s > r - B_{U_{t'}} \right] \right. \right. \\
&\quad \left. \left. \cdot \mathbf{1} \left[s < r - B_{U_{t'}} + \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, s > (r - B_{U_{t'}}) \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \right] ds \right) dr \right] \\
&= \mathbb{E} \left[\left(z_{U_{t'}} \cdot z_{U_{t''}} - 1 \right) \int_0^\infty \left(\int_0^r \mathbf{1} \left[s < r - B_{U_{t'}} + \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, s > (r - B_{U_{t'}}) \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \right] ds \right) dr \right] \\
&= \mathbb{E} \left[\left(z_{U_{t'}} \cdot z_{U_{t''}} - 1 \right) \int_0^\infty \left(\int_{\max \{ 0, r - B_{U_{t'}} \}}^{\max \{ 0, (r - B_{U_{t'}}) \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \}} \mathbf{1} \left[s < r - B_{U_{t'}} + \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}} \right] ds \right) dr \right]
\end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left[\left(z_{U_{t'}} \cdot z_{U_{t''}} - 1 \right) \right. \\
&\quad \left. \int_0^\infty \left(\max\{0, (r - B_{U_{t'}}) \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right), \min\{r, r - B_{U_{t'}} + \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\}\} \right. \right. \\
&\quad \left. \left. - \max\{0, r - B_{U_{t'}}\} \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \right) dr \right] \\
&= \mathbb{E} \left[\left(z_{U_{t'}} \cdot z_{U_{t''}} - 1 \right) \left(\int_0^{B_{U_{t'}}} \max\{0, r + \min\{0, -B_{U_{t'}} + \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\}\} dr \right. \right. \\
&\quad \left. \left. + \int_{B_{U_{t'}}}^\infty \left(\max\{(r - B_{U_{t'}}) \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right), r + \min\{0, -B_{U_{t'}} + \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\}\} \right. \right. \right. \\
&\quad \left. \left. \left. - (r - B_{U_{t'}}) \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \right) dr \right) \right] \\
&= \mathbb{E} \left[\left(z_{U_{t'}} \cdot z_{U_{t''}} - 1 \right) \right. \\
&\quad \left(\int_0^{B_{U_{t'}} - \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\}} 0 dr \right. \\
&\quad + \int_{B_{U_{t'}} - \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\}}^{B_{U_{t'}}} (r + \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\} - B_{U_{t'}}) dr \\
&\quad + \int_{B_{U_{t'}}}^{B_{U_{t'}} + \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\}} \left(\min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\} + (B_{U_{t'}} - r) \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) dr \\
&\quad \left. \left. + \int_{B_{U_{t'}} + \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\}}^\infty 0 dr \right) \right] \\
&= \mathbb{E} \left[\left(z_{U_{t'}} \cdot z_{U_{t''}} - 1 \right) \frac{1}{2} \left(1 + \frac{g_{U_{t''}}}{g_{U_{t'}}} \right) \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\}^2 \right]. \tag{4.35}
\end{aligned}$$

Fifth,

$$\begin{aligned}
& \mathbb{E} \left[(1 - z_{U_{t'}}) \int_0^\infty \left(\int_0^r \mathbf{1}[-r + B_{U_{t'}} > -s] \mathbf{1}[-r + B_{U_{t'}} > 0] ds \right) dr \right] \\
&= \mathbb{E} \left[(1 - z_{U_{t'}}) \int_0^\infty \mathbf{1}[B_{U_{t'}} > r] \left(\int_0^r \mathbf{1}[s > r - B_{U_{t'}}] ds \right) dr \right] \\
&= \mathbb{E} \left[(1 - z_{U_{t'}}) \int_0^\infty \mathbf{1}[B_{U_{t'}} > r] \left(\int_{(r - B_{U_{t'}})^+}^r ds \right) dr \right] \\
&= \mathbb{E} \left[(1 - z_{U_{t'}}) \int_0^\infty \mathbf{1}[B_{U_{t'}} > r] (r - (r - \min\{r, B_{U_{t'}}\})) dr \right] \\
&= \mathbb{E} \left[(1 - z_{U_{t'}}) \int_0^{B_{U_{t'}}} \min\{r, B_{U_{t'}}\} dr \right] \\
&= \mathbb{E} \left[(1 - z_{U_{t'}}) \int_0^{B_{U_{t'}}} r dr \right] \\
&= \mathbb{E} \left[(1 - z_{U_{t'}}) \frac{B_{U_{t'}}^2}{2} \right]. \tag{4.36}
\end{aligned}$$

Sixth,

$$\begin{aligned}
& \mathbb{E} \left[(1 - z_{U_{t''}}) \int_0^\infty \left(\int_s^\infty \mathbf{1}[-r + B_{U_{t'}} > -s] \mathbf{1}[-s + B_{U_{t''}} > 0] dr \right) ds \right] \\
&= \mathbb{E} \left[(1 - z_{U_{t''}}) \int_0^\infty \mathbf{1}[B_{U_{t'}} > s] \left(\int_s^\infty \mathbf{1}[s + B_{U_{t''}} > r] dr \right) ds \right] \\
&= \mathbb{E} \left[(1 - z_{U_{t''}}) \int_0^{B_{U_{t''}}} \left(\int_s^{s+B_{U_{t'}}} dr \right) ds \right] \\
&= \mathbb{E} \left[(1 - z_{U_{t''}}) B_{U_{t''}} B_{U_{t'}} \right]. \tag{4.37}
\end{aligned}$$

Summing Equations (4.32)-(4.37) we obtain

$$\begin{aligned}
& \mathbb{E} \left[(z_{U_{t'}} - 1) \left(\left(B_{U_{t'}} - B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \left(B_{U_{t'}} - \min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}}\} \right) \right. \right. \\
& \quad \left. \left. - \frac{1}{2} \left(B_{U_{t'}} - \min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}}\} \right)^2 \right) \right] \\
& + \mathbb{E} \left[(z_{U_{t'}} \cdot z_{U_{t''}} - 1) B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} \left(B_{U_{t'}} - \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\} \right) \right] \\
& + \mathbb{E} \left[(z_{U_{t''}} - 1) \left(\frac{g_{U_{t'}}}{2g_{U_{t''}}} \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\}^2 - B_{U_{t'}} \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\} + B_{U_{t'}} B_{U_{t''}} \right) \right]
\end{aligned}$$

$$\begin{aligned}
& + \mathbb{E} \left[\left(z_{U_{t'}} \cdot z_{U_{t''}} - 1 \right) \frac{1}{2} \left(1 + \frac{g_{U_{t''}}}{g_{U_{t'}}} \right) \min \left\{ B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}} \right\}^2 \right] + \mathbb{E} \left[(1 - z_{U_{t'}}) \frac{B_{U_{t'}}^2}{2} \right] \\
& + \mathbb{E} \left[(1 - z_{U_{t''}}) B_{U_{t''}} B_{U_{t'}} \right] \\
& = \mathbb{E} \left[\left(z_{U_{t'}} - 1 \right) \left(B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \left(\min \left\{ B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \right\} - B_{U_{t'}} \right) - \frac{1}{2} \min \left\{ B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \right\}^2 \right) \right] \\
& + \mathbb{E} \left[\left(z_{U_{t'}} \cdot z_{U_{t''}} - 1 \right) \left(B_{U_{t''}} \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \left(B_{U_{t'}} - \min \left\{ B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}} \right\} \right) \right. \right. \\
& \quad \left. \left. + \frac{1}{2} \left(1 + \frac{g_{U_{t''}}}{g_{U_{t'}}} \right) \min \left\{ B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}} \right\}^2 \right) \right] \\
& + \mathbb{E} \left[\left(z_{U_{t''}} - 1 \right) \left(\frac{g_{U_{t'}}}{2g_{U_{t''}}} \min \left\{ \frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}} \right\}^2 - B_{U_{t'}} \min \left\{ \frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}} \right\} \right) \right]
\end{aligned}$$

Realize that the case $t'' < t'$ is symmetric to the previous one, therefore,

$$\begin{aligned}
& \left(\psi^{DPS} \right)^{(2)} (\lambda, \vec{z}) \Big|_{\lambda=0} \\
& = 2 \left(\mathbb{E} \left[\left(z_{U_{t'}} - 1 \right) \left(B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \left(\min \left\{ B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \right\} - B_{U_{t'}} \right) - \frac{1}{2} \min \left\{ B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \right\}^2 \right) \right] \right. \\
& + \mathbb{E} \left[\left(z_{U_{t'}} \cdot z_{U_{t''}} - 1 \right) \left(B_{U_{t''}} \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \left(B_{U_{t'}} - \min \left\{ B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}} \right\} \right) \right. \right. \\
& \quad \left. \left. + \frac{1}{2} \left(1 + \frac{g_{U_{t''}}}{g_{U_{t'}}} \right) \min \left\{ B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}} \right\}^2 \right) \right] \\
& \left. + \mathbb{E} \left[\left(z_{U_{t''}} - 1 \right) \left(\frac{g_{U_{t'}}}{2g_{U_{t''}}} \min \left\{ \frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}} \right\}^2 - B_{U_{t'}} \min \left\{ \frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}} \right\} \right) \right] \right) \Bigg)
\end{aligned} \tag{4.38}$$

4.A.2 Proof of Proposition 4.3.2

We obtain the interpolation approximation of the probability generating function of the queue length vector under the DPS policy.

As explained in Section 2.3 we approximate $G(\lambda, \vec{z}^{(1-\rho)}) = \psi^{DPS}(\lambda, \vec{z}^{(1-\rho)})$ by the polynomial

$$\hat{G}(\lambda, \vec{z}) = h_0(\vec{z}) + \lambda h_1(\vec{z}) + \lambda^2 h_2(\vec{z}) + \lambda^3 h_3(\vec{z}).$$

Undoing the normalisation, that is, for $f_\lambda^{-1} = \bar{z}^{(1-\rho)^{-1}}$, we have

$$\begin{aligned}\psi^{DPS,INT}(\lambda, \bar{z}) &= \hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}}) \\ &= h_0(\bar{z}^{(1-\rho)^{-1}}) + \lambda h_1(\bar{z}^{(1-\rho)^{-1}}) + \lambda^2 h_2(\bar{z}^{(1-\rho)^{-1}}) + \lambda^3 h_3(\bar{z}^{(1-\rho)^{-1}}).\end{aligned}$$

Then, from the light-traffic conditions (2.12) we obtain $h_0(\bar{z}), h_1(\bar{z}), h_2(\bar{z})$. First we have, $\hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})|_{\lambda=0} = \hat{G}(0, \bar{z}) = h_0(\bar{z})$. Together with (4.30) we obtain $h_0(\bar{z}) = 1$.

Second,

$$\begin{aligned}\frac{d\hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{d\lambda}\Big|_{\lambda=0} &= \frac{d\hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{d\lambda}\Big|_{\lambda=0} + \sum_{i=1}^K \frac{d\hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{dz_i}\Big|_{\lambda=0} \cdot \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda}\Big|_{\lambda=0} \\ &= \left(h_1(\bar{z}^{(1-\rho)^{-1}}) + 2\lambda h_2(\bar{z}^{(1-\rho)^{-1}}) + 3\lambda^2 h_3(\bar{z}^{(1-\rho)^{-1}})\right)\Big|_{\lambda=0} \\ &\quad + \sum_{i=1}^K \left(\frac{dh_0(\bar{z}^{(1-\rho)^{-1}})}{dz_i} + \lambda \frac{dh_1(\bar{z}^{(1-\rho)^{-1}})}{dz_i} + \lambda^2 \frac{dh_2(\bar{z}^{(1-\rho)^{-1}})}{dz_i} + \lambda^3 \frac{dh_3(\bar{z}^{(1-\rho)^{-1}})}{dz_i}\right)\Big|_{\lambda=0} \\ &\quad \cdot \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda}\Big|_{\lambda=0} \\ &= h_1(\bar{z}) + \sum_{i=1}^K \frac{dh_0(\bar{z}^{(1-\rho)^{-1}})}{dz_i}\Big|_{\lambda=0} \cdot \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda}\Big|_{\lambda=0} \\ &= h_1(\bar{z}) + \sum_{i=1}^K \frac{d(1)}{dz_i}\Big|_{\lambda=0} \cdot \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda}\Big|_{\lambda=0} = h_1(\bar{z}).\end{aligned}$$

Together with (4.31) we obtain $h_1(\bar{z}) = -\mathbb{E}[B] + \sum_{i=1}^K \alpha_i \mathbb{E}[B_i] z_i$.

Third,

$$\begin{aligned}\frac{d^2\hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{d\lambda^2}\Big|_{\lambda=0} &= \frac{d^2\hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{d\lambda^2}\Big|_{\lambda=0} + \sum_{i=1}^K \frac{d(d\hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})/d\lambda)}{dz_i}\Big|_{\lambda=0} \cdot \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda}\Big|_{\lambda=0} \\ &\quad + \sum_{i=1}^K \left(\left(\frac{d(d\hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})/dz_i)}{d\lambda}\right)\Big|_{\lambda=0} + \frac{d(d\hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})/dz_i)}{dz_i}\Big|_{\lambda=0} \cdot \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda}\Big|_{\lambda=0}\right) \\ &\quad \cdot \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda}\Big|_{\lambda=0} + \frac{d\hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{dz_i}\Big|_{\lambda=0} \cdot \frac{d^2(z_i^{(1-\rho)^{-1}})}{d\lambda^2}\Big|_{\lambda=0} \\ &= 2h_2(\bar{z}) + 2 \sum_{i=1}^K \frac{dh_1(\bar{z})}{dz_i} \cdot \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda}\Big|_{\lambda=0} = 2h_2(\bar{z}) + 2\mathbb{E}[B] \sum_{i=1}^K \alpha_i \mathbb{E}[B_i] z_i \ln(z_i).\end{aligned}$$

Together with (4.38) we obtain

$$\begin{aligned}
& h_2(\vec{z}) \\
&= \mathbb{E} \left[\left(z_{U_{t'}}^{(1-\rho)^{-1}} - 1 \right) \left(B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \left(\min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}}\} - B_{U_{t'}} \right) - \frac{1}{2} \min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}}\}^2 \right) \right] \\
&+ \mathbb{E} \left[\left(z_{U_{t''}}^{(1-\rho)^{-1}} \cdot z_{U_{t'}}^{(1-\rho)^{-1}} - 1 \right) \left(B_{U_{t''}} \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \left(B_{U_{t'}} - \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\} \right) \right. \right. \\
&\quad \left. \left. + \frac{1}{2} \left(1 + \frac{g_{U_{t''}}}{g_{U_{t'}}} \right) \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\}^2 \right) \right] \\
&+ \mathbb{E} \left[\left(z_{U_{t''}}^{(1-\rho)^{-1}} - 1 \right) \left(\frac{g_{U_{t'}}}{2g_{U_{t''}}} \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\}^2 - B_{U_{t'}} \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\} \right) \right] \\
&- \mathbb{E}[B] \mathbb{E} \left[B_{U_{t'}} z_{U_{t'}}^{(1-\rho)^{-1}} \ln \left(z_{U_{t'}}^{(1-\rho)^{-1}} \right) \right].
\end{aligned}$$

Finally, from Proposition 2.3.1 and noting that $G^{HT}(\vec{z})$ is equal to Equation (4.7), we conclude the proof.

4.A.3 Proof of Lemma 4.4.1

The zeroth derivative satisfies

$$\left(\widetilde{W}_k^{DPS} \right)^{(0)}(0, b, x) = \mathbb{E}[\mathbf{1}[0 \geq x]] = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x > 0. \end{cases} \quad (4.39)$$

To work out the first derivative, $\left(\widetilde{W}_k^{DPS} \right)^{(1)}(0, b, x)$, we use Equation (2.4) and we need to calculate

$$\int_{-\infty}^{\infty} \left(\mathbb{E} \left[\mathbf{1} \left[W_k^{DPS} \left(b \mid A(-\infty, \infty) = 1, \tau_1 = t \right) > x \right] \right] - \mathbb{E}[\mathbf{1}[0 > x]] \right) dt,$$

where $W_k^{DPS} \left(b \mid A(-\infty, \infty) = 1, \tau_1 = t \right)$ denotes the conditional waiting time of the tagged class- k customer when there is exactly one arrival at time t on $\mathbb{R}$ and it is described by Equation (4.15).

First, we focus on the calculation corresponding to the first term of (4.15), that is, the case when $t \leq 0 \leq t + B_{U_t}$ and $t < \frac{g_{U_t}}{g_k} b - B_{U_t}$, (where the inequalities of the random

variables hold sample-path wise). We have

$$\begin{aligned}
& \int_{-\infty}^0 \mathbb{E} \left[\mathbf{1} \left[-B_{U_t} \leq t < \frac{g_{U_t} b - B_{U_t}}{g_k} \right] \mathbf{1} [t + B_{U_t} > x] \right] dt \\
&= \int_0^{\infty} \mathbb{E} \left[\mathbf{1} \left[B_{U_t} \geq t > B_{U_t} - \frac{g_{U_t} b}{g_k} \right] \mathbf{1} [-t + B_{U_t} > x] \right] dt \\
&= \mathbb{E} \left[\int_0^{\infty} \mathbf{1} \left[B_{U_t} \geq t > B_{U_t} - \frac{g_{U_t} b}{g_k} \right] \mathbf{1} [B_{U_t} - x > t] dt \right],
\end{aligned}$$

as we make use of Tonelli's Theorem. It follows that

$$\begin{aligned}
& \int_0^{\infty} \mathbf{1} \left[B_{U_t} \geq t > B_{U_t} - \frac{g_{U_t} b}{g_k} \right] \mathbf{1} [B_{U_t} - x > t] dt \\
&= \int_{\left(B_{U_t} - \frac{g_{U_t} b}{g_k}\right)^+}^{\max\left\{\left(B_{U_t} - \frac{g_{U_t} b}{g_k}\right)^+, \min\{B_{U_t}, B_{U_t} - x\}\right\}} dt \\
&= \max \left\{ \left(B_{U_t} - \frac{g_{U_t} b}{g_k}\right)^+, B_{U_t} - x \right\} - \left(B_{U_t} - \frac{g_{U_t} b}{g_k}\right)^+ \\
&= \max \left\{ 0, B_{U_t} - x - \left(B_{U_t} - \frac{g_{U_t} b}{g_k}\right)^+ \right\} \\
&= \left(B_{U_t} - x - \left(B_{U_t} - \min\{B_{U_t}, \frac{g_{U_t} b}{g_k}\}\right) \right)^+ \\
&= \left(-x + \min\{B_{U_t}, \frac{g_{U_t} b}{g_k}\} \right)^+.
\end{aligned}$$

We thus obtain

$$\begin{aligned}
& \int_0^{\infty} \mathbb{E} \left[\mathbf{1} \left[B_{U_t} \geq t > B_{U_t} - \frac{g_{U_t} b}{g_k} \right] \mathbf{1} [t + B_{U_t} > x] \right] dt \\
&= \mathbb{E} \left[\left(-x + \min\{B_{U_t}, \frac{g_{U_t} b}{g_k}\} \right)^+ \right]. \tag{4.40}
\end{aligned}$$

Second, we focus on the calculation corresponding to the second term of (4.15), that is, the case when $t \leq 0 \leq t + B_{U_t}$ and $\frac{g_{U_t} b}{g_k} - t \leq B_{U_t}$. We have

$$\begin{aligned}
& \int_{-\infty}^0 \mathbb{E} \left[\mathbf{1} \left[\frac{g_{U_t} b - B_{U_t}}{g_k} \leq t \right] \mathbf{1} \left[\frac{g_{U_t} b}{g_k} > x \right] \right] dt \\
&= \int_0^{\infty} \mathbb{E} \left[\mathbf{1} \left[\frac{g_{U_t} b - B_{U_t}}{g_k} \leq -t \right] \mathbf{1} \left[\frac{g_{U_t} b}{g_k} > x \right] \right] dt \\
&= \mathbb{E} \left[\int_0^{\infty} \mathbf{1} \left[\frac{g_{U_t} b - B_{U_t}}{g_k} \leq -t \right] \mathbf{1} \left[\frac{g_{U_t} b}{g_k} > x \right] dt \right],
\end{aligned}$$

as we make use of Tonelli's Theorem. It follows that

$$\begin{aligned} \int_0^\infty \mathbf{1} \left[t \leq B_{U_t} - \frac{g_{U_t} b}{g_k} \right] \mathbf{1} \left[\frac{g_{U_t} b}{g_k} > x \right] dt &= \mathbf{1} \left[\frac{g_{U_t} b}{g_k} > x \right] \int_0^{\left(B_{U_t} - \frac{g_{U_t} b}{g_k} \right)^+} dt \\ &= \mathbf{1} \left[\frac{g_{U_t} b}{g_k} > x \right] \left(B_{U_t} - \frac{g_{U_t} b}{g_k} \right)^+ = \mathbf{1} \left[\frac{g_{U_t} b}{g_k} > x \right] \left(B_{U_t} - \min \left\{ B_{U_t}, \frac{g_{U_t} b}{g_k} \right\} \right). \end{aligned}$$

We thus obtain

$$\begin{aligned} &\mathbb{E} \left[\int_0^\infty \mathbf{1} \left[\frac{g_{U_t} b}{g_k} - B_{U_t} \leq -t \right] \mathbf{1} \left[\frac{g_{U_t} b}{g_k} > x \right] dt \right] \\ &= \mathbb{E} \left[\mathbf{1} \left[\frac{g_{U_t} b}{g_k} > x \right] \left(B_{U_t} - \min \left\{ B_{U_t}, \frac{g_{U_t} b}{g_k} \right\} \right) \right]. \end{aligned} \quad (4.41)$$

Third, we focus on the subtraction between the third term of (4.15), that is, the case when $t + B_{U_t} < 0$, and $W_k^{DPS} (b | A(-\infty, \infty) = 0) = b$. We then have

$$\begin{aligned} &\int_{-\infty}^0 \mathbb{E} [\mathbf{1} [t < -B_{U_t}] \mathbf{1} [0 > x] - \mathbf{1} [0 > x]] dt \\ &= \int_0^\infty \mathbb{E} [\mathbf{1} [-t < -B_{U_t}] \mathbf{1} [0 > x] - \mathbf{1} [0 > x]] dt \\ &= \mathbb{E} \left[\mathbf{1} [0 > x] \int_0^\infty (\mathbf{1} [B_{U_t} < t] - 1) dt \right] \end{aligned}$$

as we make use of Tonelli's Theorem. It follows that

$$\begin{aligned} \mathbf{1} [0 > x] \int_0^\infty (\mathbf{1} [B_{U_t} < t] - 1) dt &= \mathbf{1} [0 > x] \int_0^\infty -\mathbf{1} [B_{U_t} > t] dt = -\mathbf{1} [0 > x] \int_0^{B_{U_t}} dt \\ &= -\mathbf{1} [0 > x] B_{U_t}. \end{aligned}$$

We thus obtain

$$\mathbb{E} \left[\int_0^\infty (\mathbf{1} [B_{U_t} < t] \mathbf{1} [0 > x] - \mathbf{1} [0 > x]) dt \right] = -\mathbf{1} [0 > x] \mathbb{E} [B_{U_t}]. \quad (4.42)$$

Fourth, we focus on the calculation corresponding to the fourth term of (4.15), that is, the case when $0 < t < b$ and $\frac{b-t}{g_k} > \frac{B_{U_t}}{g_{U_t}}$. We have

$$\int_0^\infty \mathbb{E} \left[\mathbf{1} \left[t < b - \frac{g_k B_{U_t}}{g_{U_t}} \right] \mathbf{1} [B_{U_t} > x] \right] dt = \mathbb{E} \left[\int_0^\infty \mathbf{1} \left[t < b - \frac{g_k B_{U_t}}{g_{U_t}} \right] \mathbf{1} [B_{U_t} > x] dt \right],$$

as we make use of Tonelli's Theorem. It follows that

$$\begin{aligned} \int_0^\infty \mathbf{1} \left[t < b - \frac{g_k B_{U_t}}{g_{U_t}} \right] \mathbf{1} [B_{U_t} > x] dt &= \int_0^{\left(b - \frac{g_k B_{U_t}}{g_{U_t}} \right)^+} \mathbf{1} [B_{U_t} > x] dt \\ &= \mathbf{1} [B_{U_t} > x] \left(b - \frac{g_k B_{U_t}}{g_{U_t}} \right)^+ = \mathbf{1} [B_{U_t} > x] \left(b - \min \left\{ b, \frac{g_k B_{U_t}}{g_{U_t}} \right\} \right) \end{aligned}$$

We thus obtain

$$\int_0^\infty \mathbb{E} \left[\mathbf{1} \left[t < b - \frac{g_k B_{U_t}}{g_{U_t}} \right] \mathbf{1} [B_{U_t} > x] \right] dt = \mathbb{E} \left[\mathbf{1} [B_{U_t} > x] \left(b - \min \left\{ b, \frac{g_k B_{U_t}}{g_{U_t}} \right\} \right) \right] \quad (4.43)$$

Fifth, we focus on the calculation corresponding to the fifth term of (4.15), that is, the case when $0 < t < b$ and $\frac{b-t}{g_k} \leq \frac{B_{U_t}}{g_{U_t}}$. We have

$$\begin{aligned} & \int_0^\infty \mathbb{E} \left[\mathbf{1} \left[b - \frac{g_k B_{U_t}}{g_{U_t}} \leq t < b \right] \mathbf{1} \left[-t \frac{g_{U_t}}{g_k} + b \frac{g_{U_t}}{g_k} > x \right] \right] dt \\ &= \mathbb{E} \left[\int_0^\infty \mathbf{1} \left[b - \frac{g_k B_{U_t}}{g_{U_t}} \leq t < b \right] \mathbf{1} \left[\left(b \frac{g_{U_t}}{g_k} - x \right) \frac{g_k}{g_{U_t}} > t \right] dt \right] \end{aligned}$$

as we make use of Tonelli's Theorem. It follows that

$$\begin{aligned} & \int_0^\infty \mathbf{1} \left[b - \frac{g_k B_{U_t}}{g_{U_t}} \leq t < b \right] \mathbf{1} \left[\left(b \frac{g_{U_t}}{g_k} - x \right) \frac{g_k}{g_{U_t}} > t \right] dt \\ &= \int_{\left(b - \frac{g_k B_{U_t}}{g_{U_t}} \right)^+}^{\max \left\{ \left(b - \frac{g_k B_{U_t}}{g_{U_t}} \right)^+, \min \left\{ b, \left(b \frac{g_{U_t}}{g_k} - x \right) \frac{g_k}{g_{U_t}} \right\} \right\}} dt \\ &= \max \left\{ \left(b - \frac{g_k B_{U_t}}{g_{U_t}} \right)^+, \min \left\{ b, \left(b \frac{g_{U_t}}{g_k} - x \right) \frac{g_k}{g_{U_t}} \right\} \right\} - \left(b - \frac{g_k B_{U_t}}{g_{U_t}} \right)^+ \\ &= \max \left\{ 0, b - x \frac{g_k}{g_{U_t}} - b + \min \left\{ b, \frac{g_k B_{U_t}}{g_{U_t}} \right\} \right\}. \end{aligned}$$

We thus obtain

$$\begin{aligned} & \int_0^\infty \mathbb{E} \left[\mathbf{1} \left[b - \frac{g_k B_{U_t}}{g_{U_t}} \leq t < b \right] \left(-t \frac{g_{U_t}}{g_k} + b \frac{g_{U_t}}{g_k} \right) \right] dt \\ &= \mathbb{E} \left[\left(-x \frac{g_k}{g_{U_t}} + \min \left\{ b, \frac{g_k B_{U_t}}{g_{U_t}} \right\} \right)^+ \right]. \end{aligned} \quad (4.44)$$

Sixth, we focus on the subtraction between the sixth term of (4.15), that is, the case when $0 < b < t$, and $W_k^{DPS} \left(b \mid A(-\infty, \infty) = 0 \right) = \mathbf{1} [0 > x]$. We then have

$$\int_0^\infty \mathbb{E} [\mathbf{1} [b < t] \mathbf{1} [0 > x] - \mathbf{1} [0 > x]] dt = -\mathbb{E} \left[\int_0^\infty \mathbf{1} [0 < t < b] \mathbf{1} [0 > x] dt \right]$$

as we make use of Tonelli's Theorem. It follows that

$$-\int_0^\infty \mathbf{1} [0 < t < b] \mathbf{1} [0 > x] dt = -\int_0^b \mathbf{1} [0 > x] dt = -b \mathbf{1} [0 > x].$$

We thus have

$$\int_0^\infty \mathbb{E} [\mathbf{1} [b < t] \mathbf{1} [0 > x] - \mathbf{1} [0 > x]] dt = -b \mathbf{1} [0 > x]. \quad (4.45)$$

In conclusion, summing Equations (4.40)-(4.45) we obtain

$$\begin{aligned}
& \left(W_k^{DPS}\right)^{(1)}(0, b, x) \\
&= \mathbb{E} \left[\left(-x + \min\{B_{U_t}, \frac{g_{U_t}}{g_k} b\} \right)^+ \right. \\
&\quad + \mathbf{1} \left[\frac{g_{U_t}}{g_k} b > x \right] \left(B_{U_t} - \min\{B_{U_t}, \frac{g_{U_t}}{g_k} b\} \right) \\
&\quad - \mathbf{1} [0 \geq x] \mathbb{E}[B_{U_t}] \\
&\quad + \mathbf{1} [B_{U_t} > x] \left(b - \min\{b, \frac{g_k B_{U_t}}{g_{U_t}}\} \right) \\
&\quad + \left(-x \frac{g_k}{g_{U_t}} + \min\{b, \frac{g_k}{g_{U_t}} B_{U_t}\} \right)^+ \\
&\quad \left. - \mathbf{1} [0 \geq x] b \right], \tag{4.46}
\end{aligned}$$

which, for $x > 0$, simplifies to

$$\begin{aligned}
& \left(W_k^{DPS}\right)^{(1)}(0, b, x) \\
&= \mathbb{E} \left[\left(1 + \frac{g_k}{g_{U_t}} \right) \left(-x + \min\{B_{U_t}, \frac{g_{U_t}}{g_k} b\} \right)^+ \right. \\
&\quad + \mathbf{1} \left[\frac{g_{U_t}}{g_k} b > x \right] \left(B_{U_t} - \min\{B_{U_t}, \frac{g_{U_t}}{g_k} b\} \right) \\
&\quad \left. + \mathbf{1} [B_{U_t} > x] \left(b - \min\{b, \frac{g_k B_{U_t}}{g_{U_t}}\} \right) \right]. \tag{4.47}
\end{aligned}$$

From Equations (2.2), (4.39) and (4.47) we end up the proof.

4.A.4 Proof of Proposition 4.4.2

We obtain the interpolation approximation of the complementary distribution function of the conditional waiting time under the DPS policy for $x > 0$. At the point $x = 0$ the waiting time has an atom.

As explained in Section 2.3 we approximate $G(\lambda, b, x(1 - \rho)^{-1}) = W_k^{DPS}(\lambda, b, x(1 - \rho)^{-1})$ by the polynomial

$$\hat{G}(\lambda, b, x) = h_0(b, x) + \lambda h_1(b, x) + \lambda^2 h_2(b, x).$$

Undoing the normalisation, that is, for $f_\lambda^{-1} = x(1 - \rho)$, we have

$$\begin{aligned}
W_k^{DPS, INT}(\lambda, b, x) &= \hat{G}(\lambda, b, x(1 - \rho)) \\
&= h_0(b, x(1 - \rho)) + \lambda h_1(b, x(1 - \rho)) + \lambda^2 h_2(b, x(1 - \rho)).
\end{aligned}$$

Then, from the light-traffic conditions (2.12) we obtain $h_0(b, x), h_1(b, x)$. First we have $\hat{G}(\lambda, b, x(1 - \rho)) \Big|_{\lambda=0} = \hat{G}(0, b, x) = h_0(b, x)$. Together with (4.39) we obtain $h_0(b, x) = \mathbf{1}[x > 0] = 0$ for $x > 0$.

Second,

$$\begin{aligned} \frac{d\hat{G}(\lambda, b, x(1 - \rho))}{d\lambda} \Big|_{\lambda=0} &= \frac{d\hat{G}(\lambda, b, x(1 - \rho))}{d\lambda} \Big|_{\lambda=0} + \frac{d\hat{G}(\lambda, b, x(1 - \rho))}{dx} \Big|_{\lambda=0} \cdot \frac{d(x(1 - \rho))}{d\lambda} \Big|_{\lambda=0} \\ &= (h_1(b, x) + 2\lambda h_2(b, x)) \\ &+ \left(\frac{dh_0(b, x(1 - \rho))}{dx} + \lambda \frac{dh_1(b, x(1 - \rho))}{dx} + \lambda^2 \frac{dh_2(b, x(1 - \rho))}{dx} \right) \Big|_{\lambda=0} \cdot \frac{d(x(1 - \rho))}{d\lambda} \Big|_{\lambda=0} \\ &= h_1(b, x) + \frac{dh_0(b, x)}{dx} \Big|_{\lambda=0} \cdot x(-\mathbb{E}[B]) = h_1(b, x). \end{aligned}$$

Together with (4.47), for $x > 0$, we obtain

$$\begin{aligned} h_1(b, x) &= \mathbb{E} \left[\left(1 + \frac{g_k}{g_{U_t}} \right) \left(-x + \min\{B_{U_t}, \frac{g_{U_t} b}{g_k}\} \right)^+ \right. \\ &\quad \left. + \mathbf{1} \left[\frac{g_{U_t} b}{g_k} > x \right] \left(B_{U_t} - \min\{B_{U_t}, \frac{g_{U_t} b}{g_k}\} \right) \right. \\ &\quad \left. + \mathbf{1}[B_{U_t} > x] \left(b - \frac{g_k}{g_{U_t}} \min\{B_{U_t}, \frac{g_{U_t} b}{g_k}\} \right) \right]. \end{aligned}$$

Finally, from Proposition 2.3.1 and noting that $G^{HT}(b, x)$ is equal to Equation (4.9) we conclude the proof.

4.A.5 Light-traffic derivatives for the mean conditional sojourn time

As explained at the beginning of Section 2.1, for λ small enough, the conditional sojourn time, $S_k^{DPS}(\lambda, b)$, can be approximated by a polynomial

$$\sum_{m=0}^{\infty} \lambda^m r_m(b), \quad (4.48)$$

where $r_m(b) = \frac{\overline{S_k^{DPS}}^{(m)}(0, b)}{m!}$, $m = 0, 1, 2, \dots$

In this chapter we used the approach initiated by Reiman and Simon [49] in order to derive expressions for the coefficients $r_m(b)$, see Appendix 4.A.3. In this section we show how the coefficients $r_m(b)$ could alternatively have been obtained for the DPS model. We do this by using the integro-differential equation for the mean conditional sojourn time as obtained by Fayolle et al., see Equation (4.1).

For the zeroth coefficient we have from Equation (4.48) $\frac{d\bar{S}_k(0, b)}{db} = \frac{dr_0(b)}{db}$ and from Equation (4.1) $\frac{d\bar{S}_k(0, b)}{db} = 1$. This immediately gives us

$$r_0(b) = b. \quad (4.49)$$

Since we assumed that for λ close to zero the function $\bar{S}_k(\lambda, b)$ can be approximated by $\sum_{m=0}^{\infty} \lambda^m r_m(b)$, we have for $m = 1, 2, \dots$

$$\begin{aligned} r_m(b) &= \lim_{\lambda \rightarrow 0} \frac{1}{\lambda^m} \left(\bar{S}_k(\lambda, b) - \sum_{i=0}^{m-1} \lambda^i r_i(b) - \sum_{i=m+1}^{\infty} \lambda^i r_i(b) \right) \\ &= \lim_{\lambda \rightarrow 0} \frac{1}{\lambda^m} \left(\bar{S}_k(\lambda, b) - \sum_{i=0}^{m-1} \lambda^i r_i(b) \right) \\ &= \lim_{\lambda \rightarrow 0} \frac{1}{\lambda^m} \int_0^b \left(\frac{\partial \bar{S}_k(\lambda, \tilde{b})}{\partial \tilde{b}} - \sum_{i=0}^{m-1} \lambda^i \frac{dr_i(\tilde{b})}{d\tilde{b}} \right) d\tilde{b}. \end{aligned}$$

Now, substituting Equation (4.1) in the above formula, one can derive the terms $r_m(b)$, $m = 1, \dots$, recursively. For the first and second order derivatives we obtain in this way

$$r_1(b) = \sum_{j=1}^K \alpha_j \frac{g_j}{g_k} \int_0^b \left(\mathbb{E}[B_j] + \left(\frac{g_k}{g_j} - 1 \right) \mathbb{E}[\min\{B_j, \frac{g_j}{g_k} \tilde{b}\}] \right) d\tilde{b} \quad (4.50)$$

and

$$\begin{aligned} r_2(b) &= \sum_{j=1}^K \alpha_j \frac{g_j}{g_k} \sum_{i=1}^K \alpha_i \frac{g_i}{g_k} \int_0^b d\tilde{b} \left(\mathbb{E}[B_i] \left(\mathbb{E}[B_j] + \left(\frac{g_k}{g_j} - 1 \right) \mathbb{E}[\min\{B_j, \frac{g_j}{g_k} \tilde{b}\}] \right) \right. \\ &\quad + \left(\frac{g_k}{g_i} - 1 \right) \int_{\frac{g_j}{g_k} \tilde{b}}^{\infty} \mathbb{E}[\min\{B_i, \frac{g_i}{g_k} (x - \frac{g_j}{g_k} \tilde{b})\}] \cdot [1 - F_j(x)] dx \\ &\quad \left. + \frac{g_k}{g_j} \left(\frac{g_k}{g_i} - 1 \right) \int_{\frac{g_j}{g_k} (\tilde{b}-b)}^{\frac{g_j}{g_k} \tilde{b}} \mathbb{E}[\min\{B_i, \frac{g_i}{g_k} (\tilde{b} - \frac{g_k}{g_j} z)\}] \cdot [1 - F_j(z)] dz \right). \quad (4.51) \end{aligned}$$

We observe that $r_0(b)$ coincides with the zeroth light-traffic derivative obtained in Equation (4.19) and we verified that $r_1(b)$, obtained in Equation (4.50), coincides with the first light-traffic derivative of the mean conditional sojourn time, which is the coefficient of λ in Equation (4.19).

We can now derive the following light-traffic approximation (of order 2) of the mean conditional sojourn time for a tagged class- k customer with service requirement b when λ is small

$$\overline{S}_k^{DPS, LT}(\lambda, b) = \sum_{m=0}^2 \lambda^m r_m(b) = r_0(b) + r_1(b)\lambda + r_2(b)\lambda^2, \quad (4.52)$$

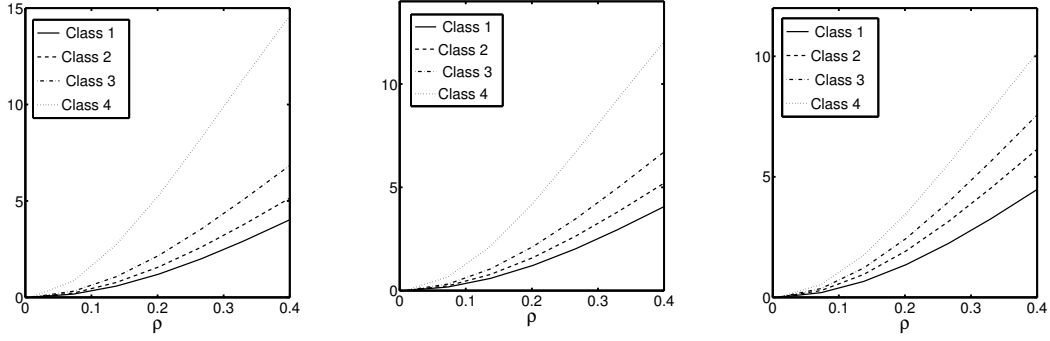


Figure 4.17: Scenario 3: Difference of the relative errors of Equations (4.19) and (4.52) for a tagged class- i customer with service requirement b_i such that $\mathbb{P}(B_i \leq b_i) = 0.01$ (left), $\mathbb{P}(B_i \leq b_i) = 0.50$ (middle), $\mathbb{P}(B_i \leq b_i) = 0.99$ (right).

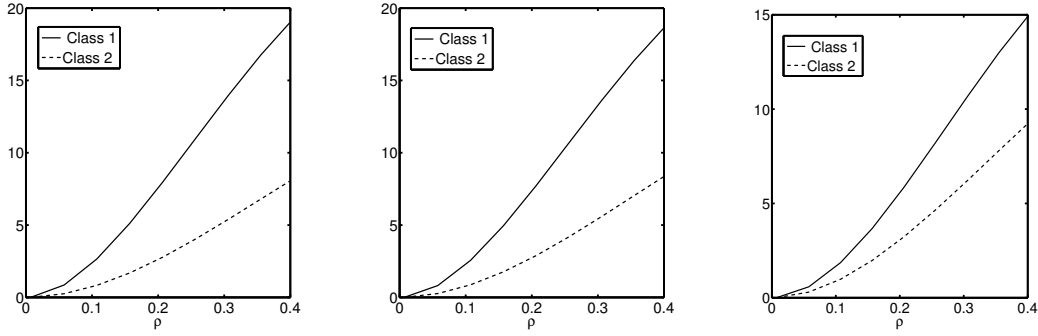


Figure 4.18: Scenario 6: Difference of the relative errors of Equations (4.19) and (4.52) for a tagged class- i customer with service requirement b_i such that $\mathbb{P}(B_i \leq b_i) = 0.01$ (left), $\mathbb{P}(B_i \leq b_i) = 0.50$ (middle), $\mathbb{P}(B_i \leq b_i) = 0.99$ (right).

where $r_0(b)$, $r_1(b)$ and $r_2(b)$ are given in Equations (4.49), (4.50) and (4.51), respectively. Using this result together with Proposition 2.3.2 and the heavy-traffic result presented in Proposition 4.2.2 we obtain the third order interpolation approximation immediately

$$\overline{S}_k^{DPS,INT}(\lambda, b) = r_0(b) + r_1(b)\lambda + r_2(b)\lambda^2 + \frac{b}{g_k} \frac{\mathbb{E}[B^2]}{\sum_{j=1}^K \alpha_j \mathbb{E}[B_j^2]/g_j} \frac{(\lambda \mathbb{E}[B])^3}{1 - \lambda \mathbb{E}[B]}. \quad (4.53)$$

We next assess the impact of adding the second light-traffic derivative to the light-traffic approximation and interpolation approximation. We numerically compare the accuracy of Equations (4.19) and (4.52), and Equations (4.20) and (4.53).

In Figures 4.17 and 4.18 we plot the difference of the relative errors of Equations (4.19) and (4.52) for Scenario 3 and Scenario 6, respectively. Since we know that all relative errors are positive and since the resulting functions are also positive, this implies that the higher order light-traffic approximation is always more accurate. In Figures 4.19 and 4.20 we plot the relative error of the mean conditional sojourn time

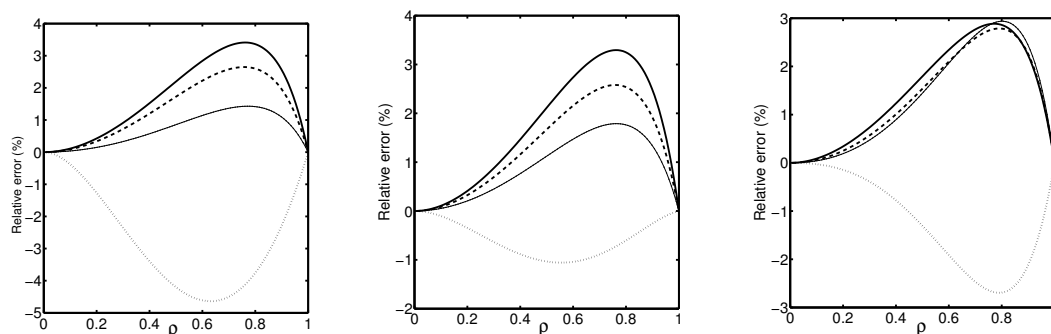


Figure 4.19: Scenario 3: Relative error for the 3rd order mean conditional sojourn time for a tagged class- i customer with service requirement b_i such that $\mathbb{P}(B_i \leq b_i) = 0.01$ (left), $\mathbb{P}(B_i \leq b_i) = 0.50$ (middle), $\mathbb{P}(B_i \leq b_i) = 0.99$ (right).

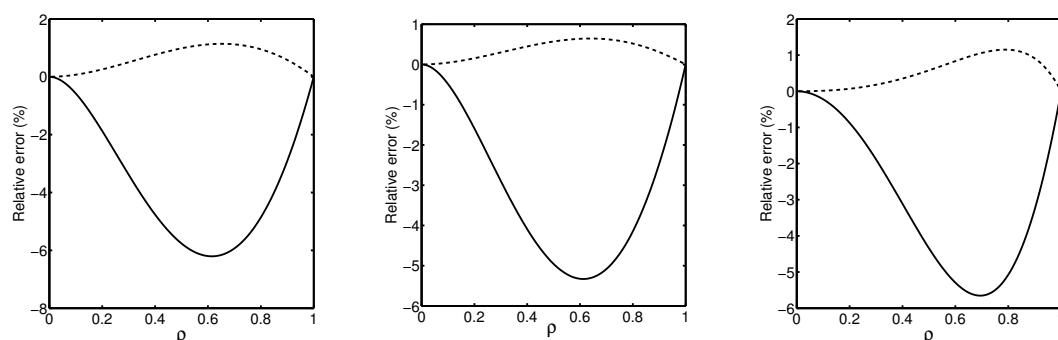


Figure 4.20: Scenario 6: Relative error for the 3rd order mean conditional sojourn time for a tagged class- i customer with service requirement b_i such that $\mathbb{P}(B_i \leq b_i) = 0.01$ (left), $\mathbb{P}(B_i \leq b_i) = 0.50$ (middle), $\mathbb{P}(B_i \leq b_i) = 0.99$ (right).

for Scenario 3 and Scenario 6, respectively. If we compare Figures 4.19 and 4.20 with Figures 4.6 and 4.9, we conclude that the accuracy of the interpolation for intermediate loads does not necessarily improve as the degree of the interpolation increases. In both cases, having a third order interpolation approximation reduces the largest relative error only for the case $\mathbb{P}(B_i \leq b_i) = 0.99$, while for $\mathbb{P}(B_i \leq b_i) = 0.01$ and $\mathbb{P}(B_i \leq b_i) = 0.5$ the third order approximation is worse than the second order approximation.

4.A.6 Proof of Corollary 4.5.3

For exponential service requirements we have the following equalities:

$$\begin{aligned}
& \mathbb{E}[\min\{B_j, b \frac{g_j}{g_k}\}] \\
&= \int_0^\infty \min\{b_j, b \frac{g_j}{g_k}\} dF_j(b_j) = \int_0^{b \frac{g_j}{g_k}} b_j dF_j(b_j) + \int_{b \frac{g_j}{g_k}}^\infty \left(b \frac{g_j}{g_k}\right) dF_j(b_j) \\
&= \frac{1}{\mu_j} \cdot F_j\left(b \frac{g_j}{g_k}\right), \\
& \mathbb{E}[\min\{B_j, b \frac{g_j}{g_k}\}^2] \\
&= \int_0^\infty \min\{b_j, b \frac{g_j}{g_k}\}^2 dF_j(b_j) = \int_0^{b \frac{g_j}{g_k}} b_j^2 dF_j(b_j) + \int_{b \frac{g_j}{g_k}}^\infty \left(b \frac{g_j}{g_k}\right)^2 dF_j(b_j) \\
&= \frac{2}{\mu_j} \left(-b \frac{g_j}{g_k} + \left(b \frac{g_j}{g_k} + \frac{1}{\mu_j}\right) F_j\left(b \frac{g_j}{g_k}\right) \right), \\
& \mathbb{E}[B_j \min\{B_j, b \frac{g_j}{g_k}\}] \\
&= \int_0^\infty b_j \min\{b_j, b \frac{g_j}{g_k}\} dF_j(b_j) = \int_0^{b \frac{g_j}{g_k}} b_j^2 dF_j(b_j) + b \frac{g_j}{g_k} \int_{b \frac{g_j}{g_k}}^\infty b_j dF_j(b_j) \\
&= -\left(b \frac{g_j}{g_k}\right)^2 \left(1 - F_j\left(b \frac{g_j}{g_k}\right)\right) + b \frac{g_j}{g_k} \frac{1}{\mu_j} + \left(1 - b \frac{g_j}{g_k} \frac{\mu_j}{2}\right) \mathbb{E}[\min\{B_j, b \frac{g_j}{g_k}\}^2].
\end{aligned}$$

Then, considering Equation (4.20) and unconditioning on U_t we obtain

$$\begin{aligned}
& \overline{S_k^{DPS,INT}}(\lambda, b) \\
&= b + \lambda b \mathbb{E}[B] + \lambda \sum_{j=1}^K \alpha_j \mathbb{E} \left[\frac{1}{2} \left(1 + \frac{g_k}{g_j}\right) \min\{B_j, b \frac{g_j}{g_k}\}^2 - \left(b \frac{g_j}{g_k} + \frac{g_k}{g_j} B_j\right) \min\{B_j, b \frac{g_j}{g_k}\} + b \frac{g_j}{g_k} B_j \right] \\
&+ \frac{(\lambda \mathbb{E}[B])^2}{(1 - \lambda \mathbb{E}[B])} \frac{b}{g_k} \frac{\mathbb{E}[B^2]}{\sum_{j=1}^K \alpha_j \mathbb{E}[B_j^2]/g_j}.
\end{aligned}$$

Together with the equalities given above, we then obtain (4.23).

The mean unconditional sojourn time, Equation (4.24), follows from Equation (4.23) together with the equality

$$\sum_{j=1}^K \frac{\alpha_j}{\mu_j^2} \left(1 - \frac{g_k}{g_j}\right) \int_0^\infty \left(1 - e^{-\mu_j b \frac{g_j}{g_k}}\right) \mu_k e^{-\mu_k b} db = \sum_{j=1}^K \frac{\alpha_j}{\mu_j} \frac{(g_j - g_k)}{g_j \mu_j + g_k \mu_k}.$$

Chapter 5

Heavy-traffic analysis of a relative priorities queue

The focus of the present and the next chapter is on the Relative Priority (RP) model described in Section 1.1.1. In this chapter we establish a state-space collapse for the scaled queue length vector in the heavy-traffic regime for a multi-class M/G/1 queue with relative priorities and non-preemptive services. In other words, in the limit the scaled queue length vector is distributed as the product of an exponentially distributed random variable and a deterministic vector. In particular, this allows to show that the scaled number of customers in the system reduces as classes with higher value of $c_k/\mathbb{E}[B_k]$ obtain a relatively larger weight, where c_k is the cost associated to class k , and $\mathbb{E}[B_k]$ is the mean service requirement of a class- k customer. This can be seen as an extension of the optimality result of the $c\mu$ -rule [17], the strict priority discipline that gives priority in decreasing order of $c_k/\mathbb{E}[B_k]$. We note that a similar state-space collapse result was observed in [55] for the discriminatory processor sharing model (presented in Proposition 4.2.1).

For DROS, i.e., under the additional assumption that the intra-class discipline is uniform random, we study the waiting time in the heavy-traffic setting. Using the state-space collapse result, we obtain the distribution of the waiting time for a customer of a given class in heavy traffic and prove that it is distributed as the product of two exponentially distributed random variables. This generalizes [34] where this result was shown for the single-class ROS queue. Moreover, we also find the value of the weights that minimizes the m -th moment of the waiting time for a customer of arbitrary class.

We note that in this chapter we consider the heavy-traffic limit of the steady-state metrics. In the literature there are state-space results available for the transient queue length processes, that is, when the heavy-traffic limit is directly taken of the queue length processes. See for example [8] for the heavy-traffic analysis of a multi-class system where all classes receive simultaneously service. In general, the heavy-traffic and steady-state limits cannot be interchanged, which explains the interest of our approach. Another important difference is that our approach allows to investigate the waiting time in the system, a metric that does not have a clear counterpart in the “process” world.

The present chapter is organised as follows. In Section 5.1 the model is introduced and the heavy-traffic scaling is defined. In Section 5.2 and 5.3 the distribution of the scaled queue length vector at departure epochs and arbitrary epochs are presented, respectively. In Section 5.4 the distribution of the scaled waiting time of a given customer is presented. In Section 5.5 it is shown how the results presented in the previous sections can be used to optimize the scaled holding cost and the moments of the scaled waiting time of an arbitrary customer. In Section 5.6 we present the numerical results.

5.1 Model description

We consider a multi-class single-server queue with K classes of customers. Class- k customers, $k = 1, \dots, K$, arrive according to independent Poisson processes with rate $\lambda_k > 0$. We denote the overall arrival rate by $\lambda = \sum_{k=1}^K \lambda_k$. We assume that class- k customers have i.i.d. generally distributed service requirements B_k , with distribution function $B_k(x)$ and Laplace-Stieltjes transform $B_k^*(s) = \int_0^\infty e^{-sx} dB_k(x)$, and we define $B_k^{*'}(s) = \frac{dB_k^*(s)}{ds}$. We assume $\mathbb{E}[B_k^2] < \infty$, for all k . The traffic intensity for class- k customers is $\rho_k = \lambda_k \mathbb{E}[B_k]$ and

$$\rho := \sum_{k=1}^K \rho_k = \sum_{k=1}^K \lambda_k \mathbb{E}[B_k] = \lambda \sum_{k=1}^K \alpha_k \mathbb{E}[B_k],$$

denotes the total traffic intensity, where $\alpha_k = \lambda_k / \lambda$ denotes the probability that an arrival is of class k . Service is non-preemptive and upon service completion, the probability that the next customer to be served is of class k is given as

$$\frac{n_k p_k}{\sum_{j=1}^K n_j p_j}, \quad (5.1)$$

where, $p_j > 0$, $j = 1, \dots, K$, are given class-dependent weights, and n_j is the number of class- j customers at the decision epoch. Once a class is chosen to be served, an intra-class scheduling discipline determines which customer in this class will be served. We assume the intra-class discipline to be non-preemptive and not to make any use of information on the actual service requirements of the customers.

We investigate the queue when it is near saturation, i.e., $\rho \uparrow 1$, which is commonly referred to as the heavy-traffic regime. This regime can be obtained by letting

$$\lambda \uparrow \hat{\lambda} := \frac{1}{\sum_{k=1}^K \alpha_k \mathbb{E}[B_k]}, \quad (5.2)$$

since then $\rho = \lambda \sum_{k=1}^K \alpha_k \mathbb{E}[B_k] \uparrow 1$. When passing to the heavy-traffic regime we keep the fraction of class- k arrivals, α_k , fixed and we define $\hat{\lambda}_k := \alpha_k \hat{\lambda}$.

We denote the steady-state number of class- k customers in the system at departure epochs by Q_k^{RP} and at arbitrary epochs by N_k^{RP} . We denote the waiting time of an arbitrary class- k customer by W_k^{DROS} . We note that, throughout the chapter, we do

not explicitly reflect the dependence of the random variables on the traffic load ρ , in order to keep notation compact. In Section 5.2, Section 5.3 and Section 5.4 we will analyze Q_k^{RP} , N_k^{RP} and W_k^{DROS} , respectively, in the heavy-traffic setting.

5.2 Queue length at departure epochs

In this section we present the state-space collapse result for the steady-state queue length distribution at departure epochs. The next proposition states the main result of this section and shows that in the limit, the queue length vector is the product of an exponentially distributed random variable and a deterministic vector. The proof is provided in Section 5.2.2.

Proposition 5.2.1. *When scaled by $1 - \rho$, the queue length vector at departure epochs has a proper limiting distribution as $(\lambda_1, \dots, \lambda_K) \rightarrow (\hat{\lambda}_1, \dots, \hat{\lambda}_K)$, such that as $\rho \uparrow 1$,*

$$(1 - \rho) (Q_1^{RP}, \dots, Q_K^{RP}) \xrightarrow{d} (\widehat{Q}_1^{RP}, \dots, \widehat{Q}_K^{RP}) \stackrel{d}{=} \left(\frac{\hat{\lambda}_1}{p_1}, \dots, \frac{\hat{\lambda}_K}{p_K} \right) \cdot Y,$$

where $\xrightarrow{d}$ denotes convergence in distribution and Y is some one-dimensional random variable.

In Remark 2 of Section 5.3 we will show that, in fact, Y is exponentially distributed. To show this we require additional results presented in Section 5.3, and thus we refer the reader to Remark 2 for more details.

Before focusing on the heavy-traffic regime, we will introduce a system of equations that is satisfied by the probability generating function of the queue length distribution at departure epochs, as obtained by Kim et al, see [31]. Define

$$\pi(q_1, \dots, q_K) := \mathbb{P} \left[(Q_1^{RP}, \dots, Q_K^{RP}) = (q_1, \dots, q_K) \right],$$

and let

$$p(\vec{z}) = \mathbb{E} \left[z_1^{Q_1^{RP}} \dots z_K^{Q_K^{RP}} \right] = \sum_{q_1=0}^{\infty} \dots \sum_{q_K=0}^{\infty} z_1^{q_1} \dots z_K^{q_K} \pi(\vec{q})$$

be its joint probability generating function. We define

$$r(\vec{z}) := \mathbb{E} \left[\frac{z_1^{Q_1^{RP}} \dots z_K^{Q_K^{RP}}}{\sum_{k=1}^K Q_k^{RP} p_k} \cdot \mathbf{1}_{(\sum_{k=1}^K Q_k^{RP} > 0)} \right] = \sum_{(q_1, \dots, q_K) \neq (0, \dots, 0)} \frac{\pi(\vec{q})}{q_1 p_1 + \dots + q_K p_K} z_1^{q_1} \dots z_K^{q_K}.$$

In [31] the distribution of the queue length was studied assuming that the intra-class scheduling is uniform random. However, since the service discipline is non-preemptive, non-anticipating and all class- k customers in the queue are stochastically equivalent, the distribution of the queue length vector does not depend on the particular choice of the intra-class policy. Hence, for any arbitrary work-conserving intra-class policy we have the following result from [31].

Theorem 5.2.2. [31, Theorem 1 and 2] (a) *The probability generating function $p(z_1, \dots, z_K)$ of the joint stationary queue lengths at departure epochs satisfies*

$$p(z_1, \dots, z_K) = 1 - \rho + \sum_{i=1}^K p_i z_i \frac{\partial}{\partial z_i} r(z_1, \dots, z_K). \quad (5.3)$$

(b) *The function $r(z_1, \dots, z_K)$ satisfies*

$$\sum_{i=1}^K p_i \left(z_i - B_i^* \left(\lambda - \sum_{j=1}^K \lambda_j z_j \right) \right) \frac{\partial}{\partial z_i} r(z_1, \dots, z_K) = (\rho - 1) \left(1 - \sum_{i=1}^K \frac{\lambda_i}{\lambda} B_i^* \left(\lambda - \sum_{j=1}^K \lambda_j z_j \right) \right). \quad (5.4)$$

In Section 5.2.1 we will show that Equations (5.3) and (5.4) simplify under the heavy-traffic scaling, which we will use in Section 5.2.2 to prove Proposition 5.2.1.

5.2.1 Heavy-traffic scaling

In this section we present three lemmas needed for the proof of Proposition 5.2.1. In the first lemma we show that the scaled queue length at departure epochs is tight. The proof may be found in Appendix 5.A.1.

Lemma 5.2.3. *The random vector $(1 - \rho)(Q_1^{RP}, \dots, Q_K^{RP})$ is tight for ρ close enough to 1, that is, for all ϵ there is a $\bar{\rho} \in (0, 1)$ and $M > 0$ such that $\mathbb{P}((1 - \rho)Q_k^{RP} \geq M) < \epsilon$, for all $k = 1, \dots, K$, and $\rho > \bar{\rho}$.*

It will be convenient to use the change of variables $z_i = e^{-s_i}$ with $s_i > 0, i = 1, \dots, K$. Denote $\vec{s} = (s_1, \dots, s_K)$ and $e^{-(1-\rho)\vec{s}} = (e^{-(1-\rho)s_1}, \dots, e^{-(1-\rho)s_K})$. If

$$\lim_{\rho \uparrow 1} p(e^{-(1-\rho)\vec{s}}) = \lim_{\rho \uparrow 1} \mathbb{E} \left[e^{-(1-\rho)s_1} Q_1^{RP} \dots e^{-(1-\rho)s_K} Q_K^{RP} \right] \quad (5.5)$$

exists, then there is a (possibly defective) random vector $(\widehat{Q}_1^{RP}, \widehat{Q}_2^{RP}, \dots, \widehat{Q}_K^{RP})$ such that $(1 - \rho)(Q_1^{RP}, Q_2^{RP}, \dots, Q_K^{RP})$ converges in distribution to $(\widehat{Q}_1^{RP}, \widehat{Q}_2^{RP}, \dots, \widehat{Q}_K^{RP})$, and the distribution of $(\widehat{Q}_1^{RP}, \widehat{Q}_2^{RP}, \dots, \widehat{Q}_K^{RP})$ is uniquely determined by the limit in (5.5) (cf. the Continuity theorem, see Feller (1971) [14]). For now, we assume that the limit exists; we come back to this assumption in the last part of the proof of Proposition 5.2.1.

Below we give two lemmas that describe properties of $\lim_{\rho \uparrow 1} p(e^{-(1-\rho)\vec{s}})$. In particular, in Lemma 5.2.5 we obtain a partial differential equation which will be the key element in the proof of Proposition 5.2.1.

In order to describe the behaviour of the generating function, we define

$$\hat{r}(\vec{s}) = \mathbb{E} \left[\frac{1 - e^{-s_1} \widehat{Q}_1^{RP} \dots e^{-s_K} \widehat{Q}_K^{RP}}{\sum_{k=1}^K \widehat{Q}_k^{RP} p_k} 1_{\left(\sum_{k=1}^K \widehat{Q}_k^{RP} > 0 \right)} \right]. \quad (5.6)$$

The “1” in the numerator is to ensure that the expression between brackets remains bounded when the $\widehat{Q}_j^{RP}$ ’s are all near zero.

Lemma 5.2.4. *If $\lim_{\rho \uparrow 1} p(e^{-(1-\rho)\vec{s}})$ exists, then it satisfies*

$$\lim_{\rho \uparrow 1} p(e^{-(1-\rho)\vec{s}}) = \sum_{i=1}^K p_i \frac{\partial \hat{r}(\vec{s})}{\partial s_i}. \quad (5.7)$$

Proof: From (5.3) we have

$$\lim_{\rho \uparrow 1} p(e^{-(1-\rho)\vec{s}}) = \lim_{\rho \uparrow 1} \sum_{i=1}^K p_i \frac{\partial r(\vec{z})}{\partial z_i} \Big|_{\vec{z}=e^{-(1-\rho)\vec{s}}}. \quad (5.8)$$

By definition of $r(\vec{z})$ we can write

$$\begin{aligned} & \lim_{\rho \uparrow 1} \frac{\partial r(\vec{z})}{\partial z_i} \Big|_{\vec{z}=e^{-(1-\rho)\vec{s}}} \\ &= \lim_{\rho \uparrow 1} \frac{\partial \mathbb{E} \left[\frac{z_1^{Q_1^{RP}} \cdots z_K^{Q_K^{RP}}}{\sum_{k=1}^K Q_k^{RP} p_k} \cdot \mathbf{1}_{(\sum_{k=1}^K Q_k^{RP} > 0)} \right]}{\partial z_i} \Big|_{\vec{z}=e^{-(1-\rho)\vec{s}}} \\ &= \lim_{\rho \uparrow 1} \mathbb{E} \left[\frac{Q_i^{RP}}{\sum_{k=1}^K Q_k^{RP} p_k} \cdot \frac{e^{-(1-\rho)s_1 Q_1^{RP}} \cdots e^{-(1-\rho)s_K Q_K^{RP}}}{e^{-(1-\rho)s_i}} \cdot \mathbf{1}_{(\sum_{k=1}^K Q_k^{RP} > 0)} \right] \\ &= \mathbb{E} \left[\frac{\widehat{Q}_i^{RP}}{\sum_{k=1}^K \widehat{Q}_k^{RP} p_k} \cdot e^{-s_1 \widehat{Q}_1^{RP}} \cdots e^{-s_K \widehat{Q}_K^{RP}} \cdot \mathbf{1}_{(\sum_{k=1}^K \widehat{Q}_k^{RP} > 0)} \right] \\ &= \frac{\partial \hat{r}(\vec{s})}{\partial s_i}. \end{aligned} \quad (5.9)$$

In the third step we used that $\frac{Q_i^{RP}}{\sum_{k=1}^K Q_k^{RP} p_k} \cdot e^{-(1-\rho)s_1 Q_1^{RP}} \cdots e^{-(1-\rho)s_K Q_K^{RP}} \cdot \mathbf{1}_{(\sum_{k=1}^K Q_k^{RP} > 0)}$ is upper bounded by $\frac{1}{\min_j(p_j)}$, and, cf. the continuous mapping theorem, converges in distribution to $\frac{\widehat{Q}_i^{RP}}{\sum_{k=1}^K \widehat{Q}_k^{RP} p_k} \cdot e^{-s_1 \widehat{Q}_1^{RP}} \cdots e^{-s_K \widehat{Q}_K^{RP}} \cdot \mathbf{1}_{(\sum_{k=1}^K \widehat{Q}_k^{RP} > 0)}$. From (5.8) and (5.9) we obtain (5.7). $\blacksquare$

In the following lemma we show that the partial differential equation as given in (5.4) simplifies considerably in the heavy-traffic regime.

Lemma 5.2.5. *If $\lim_{\rho \uparrow 1} p(e^{-(1-\rho)\vec{s}})$ exists, then the function $\hat{r}(\vec{s})$ satisfies the following partial differential equation:*

$$0 = \sum_{i=1}^K F_i(\vec{s}) \frac{\partial \hat{r}(\vec{s})}{\partial s_i} = \vec{F}(\vec{s}) \cdot \nabla \hat{r}(\vec{s}), \quad \forall \vec{s} \geq \vec{0},$$

where $\vec{F}(\vec{s}) = (F_1(\vec{s}), \dots, F_K(\vec{s}))$, and

$$F_i(\vec{s}) = p_i \left(-s_i + \mathbb{E}[B_i] \sum_{k=1}^K \hat{\lambda}_k s_k \right) \quad i = 1, \dots, K, \quad (5.10)$$

with $\hat{\lambda}_j = \alpha_j \hat{\lambda}$ and $\hat{\lambda}$ as defined in (5.2).

Proof: Taking $\vec{z}$ equal to $e^{-(1-\rho)\vec{s}}$ in (5.4), dividing both sides by $(1-\rho)$ and taking the limit of $\rho \uparrow 1$, we obtain

$$\begin{aligned} & \lim_{\rho \uparrow 1} \frac{\sum_{i=1}^K p_i \left(e^{-(1-\rho)s_i} - B_i^*(\lambda - \sum_{j=1}^K \lambda_j e^{-(1-\rho)s_j}) \right)}{1-\rho} \frac{\partial}{\partial z_i} r(z_1, \dots, z_K) \Big|_{z_i = e^{-(1-\rho)s_i}} \\ &= \lim_{\rho \uparrow 1} - \left(1 - \sum_{i=1}^K \frac{\lambda_i}{\lambda} B_i^* \left(\lambda - \sum_{j=1}^K \lambda_j e^{-(1-\rho)s_j} \right) \right) = 0. \end{aligned} \quad (5.11)$$

where the last equality follows by noting that $B_i^*(0) = 1, \forall i$. Making the change of variable $x_i = e^{-s_i}$ we obtain

$$\begin{aligned} & \lim_{\rho \uparrow 1} \frac{\sum_{i=1}^K p_i \left(x_i^{1-\rho} - B_i^*(\lambda - \sum_{j=1}^K \lambda_j x_j^{1-\rho}) \right)}{1-\rho} \frac{\partial}{\partial z_i} r(z_1, \dots, z_K) \Big|_{z_i = x_i^{1-\rho}} \\ &= \lim_{\rho \uparrow 1} \frac{\sum_{i=1}^K p_i \left(x_i^{1-\rho} - B_i^*(\lambda - \sum_{j=1}^K \lambda_j x_j^{1-\rho}) \right)}{1-\rho} \frac{\partial}{\partial z_i} r(z_1, \dots, z_K) \Big|_{z_i = x_i^{1-\rho}} \\ &= \lim_{\rho \uparrow 1} \sum_{i=1}^K p_i \left(x_i^{1-\rho} \ln x_i \right. \\ & \quad \left. + \left(\frac{1}{\mathbb{E}[B]} - \left(\frac{1}{\mathbb{E}[B]} \sum_{j=1}^K \alpha_j x_j^{1-\rho} - \sum_{j=1}^K \lambda_j x_j^{1-\rho} \ln x_j \right) \right) \right. \\ & \quad \left. \cdot \left(B_i^{*'} \left(\lambda - \sum_{j=1}^K \lambda_j x_j^{1-\rho} \right) \right) \right) \frac{\partial}{\partial z_i} r(z_1, \dots, z_K) \Big|_{z_i = x_i^{1-\rho}} \\ &= \sum_{i=1}^K p_i \left(-s_i + \mathbb{E}[B_i] \sum_{j=1}^K \hat{\lambda}_j s_j \right) \frac{\partial \hat{r}(\vec{s})}{\partial s_i}, \end{aligned}$$

where in the second step we used l'Hopital's rule and in the third step we used (5.9) and that $B_i^{*'}(0) := \frac{dB_i^*(s)}{ds} \Big|_{s=0} = -\mathbb{E}[B_i]$ for all i .

Together with (5.11), we then obtain that

$$\sum_{i=1}^K p_i \left(-s_i + \mathbb{E}[B_i] \sum_{j=1}^K \hat{\lambda}_j s_j \right) \frac{\partial \hat{r}(\vec{s})}{\partial s_i} = 0.$$

■

5.2.2 Proof of Proposition 5.2.1

This subsection contains the proof of Proposition 5.2.1. The proof is based on the fact that the function $\hat{r}(\vec{s})$ satisfies the partial differential equation as described in Lemma 5.2.5. From this partial differential equation the following property for the function $\hat{r}(\cdot)$ can be derived:

Lemma 5.2.6. *If $\lim_{\rho \uparrow 1} p(e^{-(1-\rho)\vec{s}})$ exists, then the function $\hat{r}(s)$ is constant on the $(K-1)$ -dimensional hyperplane*

$$H_c := \{\vec{s} \geq \vec{0} : \sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k = c\}, c > 0.$$

The proof of Lemma 5.2.6 may be found in Appendix 5.A.2. We can now give the proof of Proposition 5.2.1.

Proof of Proposition 5.2.1: Assume $\lim_{\rho \uparrow 1} p(e^{-(1-\rho)\vec{s}})$ exists. We come back to this assumption at the end of the proof. As $\hat{r}(\vec{s})$ is constant on H_c , see Lemma 5.2.6, $\hat{r}(\cdot)$ depends on $\vec{s}$ only through $\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k$, so there exists a function $\hat{r}^* : \mathbb{R} \rightarrow \mathbb{R}$ such that $\hat{r}(\vec{s}) = \hat{r}^*\left(\sum_{k=1}^K \frac{\lambda_k}{p_k} s_k\right)$. From Lemma 5.2.4 and $\frac{\partial \hat{r}(s)}{\partial s_i} = \frac{\hat{\lambda}_i}{p_i} \frac{d\hat{r}^*(v)}{dv} \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k}$ we obtain

$$\begin{aligned} \mathbb{E} \left[e^{-\sum_{i=1}^K s_i \widehat{Q_i^{RP}}} \right] &= \lim_{\rho \rightarrow 1} p(e^{-(1-\rho)\vec{s}}) = \sum_{i=1}^K p_i \frac{\partial \hat{r}(s)}{\partial s_i} = \sum_{i=1}^K \hat{\lambda}_i \frac{d\hat{r}^*(v)}{dv} \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k} \\ &= \hat{\lambda} \frac{d\hat{r}^*(v)}{dv} \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k}, \end{aligned} \quad (5.12)$$

which again depends on $\vec{s}$ only through $\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k$. Equivalently, we can write

$$\mathbb{E} \left[e^{-\sum_{i=1}^K s_i \widehat{Q_i^{RP}}} \right] = \mathbb{E} \left[e^{-\frac{p_1}{\hat{\lambda}_1} \widehat{Q_1^{RP}} \sum_{i=1}^K \frac{\hat{\lambda}_i}{p_i} s_i - s_2 \frac{\hat{\lambda}_2}{p_2} \left(\frac{p_2}{\hat{\lambda}_2} \widehat{Q_2^{RP}} - \frac{p_1}{\hat{\lambda}_1} \widehat{Q_1^{RP}} \right) - \dots - s_K \frac{\hat{\lambda}_K}{p_K} \left(\frac{p_K}{\hat{\lambda}_K} \widehat{Q_K^{RP}} - \frac{p_1}{\hat{\lambda}_1} \widehat{Q_1^{RP}} \right)} \right].$$

Since (by (5.12)) this only depends on $\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k$, it implies $\frac{p_i}{\hat{\lambda}_i} \widehat{Q_i^{RP}} = \frac{p_j}{\hat{\lambda}_j} \widehat{Q_j^{RP}}$ almost surely for all i, j , and we obtain that

$$\left(\widehat{Q_1^{RP}}, \dots, \widehat{Q_K^{RP}} \right) = \left(\frac{\hat{\lambda}_1}{p_1}, \frac{\hat{\lambda}_2}{p_2}, \dots, \frac{\hat{\lambda}_K}{p_K} \right) \frac{p_1}{\hat{\lambda}_1} \widehat{Q_1^{RP}},$$

almost surely. Writing $Y \stackrel{d}{=} \frac{p_1}{\hat{\lambda}_1} \widehat{Q_1^{RP}}$ we get

$$\left(\widehat{Q_1^{RP}}, \dots, \widehat{Q_K^{RP}} \right) \stackrel{d}{=} \left(\frac{\hat{\lambda}_1}{p_1}, \frac{\hat{\lambda}_2}{p_2}, \dots, \frac{\hat{\lambda}_K}{p_K} \right) Y. \quad (5.13)$$

Recall that we assumed that, for the sequence ρ , $\lim_{\rho \uparrow 1} p \left(e^{-(1-\rho)\vec{s}} \right)$ exists, thereby showing that there is a unique limit (5.13). Since $(1-\rho)(Q_1^{RP}, \dots, Q_K^{RP})$ is tight, see Lemma 5.2.3, and since for any converging subsequence of ρ we obtain the same limit, we obtain that the limit itself exists (see corollary on page 59, Billingsley 1999). This concludes the proof. $\blacksquare$

5.3 Queue length at arbitrary epochs

In this section we focus on the number of customers in the system at arbitrary epochs, $(N_1^{RP}, \dots, N_K^{RP})$. The following result shows that in the limit the queue length vector at arbitrary epochs is the product of an exponentially distributed random variable and a deterministic vector. We refer to the latter as a state-space collapse. The proof is presented in Section 5.3.2

Remark 1. *We note that a similar state-space collapse result was observed in [55, Proposition 2.1] for the discriminatory processor sharing model. In fact, the proof technique is similar to that of [55].*

Proposition 5.3.1. *When scaled by $1-\rho$, the queue length vector at arbitrary epochs has a proper limiting distribution as $(\lambda_1, \dots, \lambda_K) \rightarrow (\hat{\lambda}_1, \dots, \hat{\lambda}_K)$, such that $\rho \uparrow 1$,*

$$(1-\rho) \left(N_1^{RP}, \dots, N_K^{RP} \right) \xrightarrow{d} \left(\widehat{N}_1^{RP}, \dots, \widehat{N}_K^{RP} \right) \stackrel{d}{=} \left(\frac{\hat{\lambda}_1}{p_1}, \frac{\hat{\lambda}_2}{p_2}, \dots, \frac{\hat{\lambda}_K}{p_K} \right) X, \quad (5.14)$$

where $\xrightarrow{d}$ denotes convergence in distribution and X is an exponentially distributed random variable with mean $1/\nu(\vec{p})$, where

$$\nu(\vec{p}) := \frac{2 \sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k]}{\sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k^2]}. \quad (5.15)$$

Before focusing on the heavy-traffic regime, we will introduce a system of equations that is satisfied by the probability generating function of the queue length distribution, as obtained by Kim et al, see [31]. Let $\psi^{RP}(z_1, \dots, z_K)$ be the joint probability generating function of $(N_1^{RP}, \dots, N_K^{RP})$, i.e.,

$$\psi^{RP}(z_1, \dots, z_K) := \mathbb{E} \left[z_1^{N_1^{RP}} \dots z_K^{N_K^{RP}} \right].$$

As mentioned in Section 5.2, the distribution of the queue length vector is independent of the particular choice of the intra-class scheduling discipline. We can therefore use the following result from [31].

Theorem 5.3.2. [31, Theorem 3 and Theorem 4] *The joint probability generating function $\psi^{RP}(z_1, \dots, z_K)$ of the joint stationary queue lengths at arbitrary time epochs is*

given by

$$\psi^{RP}(z_1, \dots, z_K) = 1 - \rho + \sum_{i=1}^K \lambda_i z_i \phi_i(z_1, \dots, z_K) \frac{1 - B_i^* \left(\lambda - \sum_{k=1}^K \lambda_k z_k \right)}{\lambda - \sum_{k=1}^K \lambda_k z_k}, \quad (5.16)$$

where $\phi_i(z_1, \dots, z_K)$ (representing the joint probability generating function of the stationary queue lengths excluding the customer who has already started service, at service initiation epochs of class- i customers) is given by

$$\phi_i(z_1, \dots, z_K) = 1 - \rho + \frac{\lambda p_i}{\lambda_i} \frac{\partial}{\partial z_i} r(z_1, \dots, z_K). \quad (5.17)$$

In Section 5.3.1 we will show that Equation (5.16) simplifies under the heavy-traffic scaling, and in Section 5.3.2 we will use this to characterize the distribution of the scaled queue length vector at arbitrary epochs, that is, to prove Proposition 5.3.1.

5.3.1 Heavy-traffic scaling

In the next lemma we characterize Equation (5.16) in heavy traffic.

Lemma 5.3.3. *The limit of $\psi^{RP}(e^{-(1-\rho)\vec{s}})$ as $\rho \uparrow 1$ exists and satisfies*

$$\lim_{\rho \uparrow 1} \psi^{RP}(e^{-(1-\rho)\vec{s}}) = \sum_{i=1}^K p_i \frac{\partial \hat{r}(s)}{\partial s_i} = \hat{\lambda} \frac{d\hat{r}^*(v)}{dv} \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k},$$

with $\hat{r}^*(\cdot)$ some function $\hat{r}^* : \mathbb{R} \rightarrow \mathbb{R}$.

Proof: Since $(1 - \rho)(Q_1^{RP}, \dots, Q_K^{RP})$ converges in distribution to $(\widehat{Q}_1^{RP}, \dots, \widehat{Q}_K^{RP})$, we know that the limit of $p(e^{-(1-\rho)\vec{s}})$ exists, and hence, by Equation (5.8), the limit of $\frac{\partial r(\vec{z})}{\partial z_i} \Big|_{\vec{z}=e^{-(1-\rho)\vec{s}}}$ exists. It now follows directly from (5.17) that $\lim_{\rho \uparrow 1} \phi_i(e^{-(1-\rho)\vec{s}})$ exists and it is given by $\frac{\hat{\lambda} p_i}{\lambda_i} \frac{\partial \hat{r}(s)}{\partial s_i}$.

As we have seen in Lemma 5.2.6, $\hat{r}(\vec{s})$ is constant on H_c . Therefore, it depends on $\vec{s}$ only through $\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k$, so there exists a function $\hat{r}^* : \mathbb{R} \rightarrow \mathbb{R}$ such that $\hat{r}(\vec{s}) = \hat{r}^*\left(\sum_{k=1}^K \frac{\lambda_k}{p_k} s_k\right)$ and $\frac{\partial \hat{r}(s)}{\partial s_i} = \frac{\hat{\lambda}_i}{p_i} \frac{d\hat{r}^*(v)}{dv} \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k}$.

This, together with (5.16) gives that

$$\begin{aligned}
& \lim_{\rho \uparrow 1} \psi^{RP} \left(e^{-(1-\rho)\vec{s}} \right) \\
&= \lim_{\rho \uparrow 1} \left(1 - \rho + \sum_{i=1}^K \lambda_i e^{-(1-\rho)s_i} \phi_i(e^{-(1-\rho)\vec{s}}) \frac{1 - B_i^* \left(\lambda - \sum_{k=1}^K \lambda_k e^{-(1-\rho)s_k} \right)}{\lambda - \sum_{k=1}^K \lambda_k e^{-(1-\rho)s_k}} \right) \\
&= \sum_{i=1}^K \hat{\lambda}_i \frac{\hat{\lambda}_i p_i}{\hat{\lambda}_i} \frac{\partial \hat{r}(\vec{s})}{\partial s_i} \left(-B_i^{*'}(0) \right) = \sum_{i=1}^K \hat{\lambda}_i \mathbb{E}[B_i] \frac{d\hat{r}^*(v)}{dv} \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k} \\
&= \hat{\lambda} \frac{d\hat{r}^*(v)}{dv} \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k} = \hat{\lambda} \frac{d\hat{r}^*(v)}{dv} \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k},
\end{aligned}$$

where in the first step we used l'Hopital's rule and $B_i^{*'}(0) := \frac{dB_i^*(s)}{ds} \Big|_{s=0} = -\mathbb{E}[B_i]$ for all i . $\square$

In particular, Lemma 5.3.3 implies that there exists a vector $(\widehat{N}_1^{RP}, \dots, \widehat{N}_K^{RP})$ such that the scaled queue length vector at arbitrary epochs converges in distribution to it.

5.3.2 Proof of Proposition 5.3.1

This subsection contains the proof of Proposition 5.3.1. It consists of two steps. Firstly, we show that the queue length vector is the product of a random variable and a deterministic vector, and secondly, we determine the distribution of the random variable X , concluding that it is exponentially distributed with mean as given in (5.15).

Proof of Proposition 5.3.1: Since $\lim_{\rho \uparrow 1} \psi^{RP} \left(e^{-(1-\rho)\vec{s}} \right)$ exists, see Lemma 5.3.3 we know there exists a random vector $(\widehat{N}_1^{RP}, \dots, \widehat{N}_K^{RP})$ such that

$$\mathbb{E} \left[e^{-\sum_{k=1}^K s_k \widehat{N}_k^{RP}} \right] = \lim_{\rho \uparrow 1} \psi^{RP} \left(e^{-(1-\rho)\vec{s}} \right) = \hat{\lambda} \frac{d\hat{r}^*(v)}{dv} \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k}. \quad (5.18)$$

Using the same steps as in the proof of Proposition 5.2.1 we obtain that

$$(\widehat{N}_1^{RP}, \dots, \widehat{N}_K^{RP}) \stackrel{d}{=} \left(\frac{\hat{\lambda}_1}{p_1}, \frac{\hat{\lambda}_2}{p_2}, \dots, \frac{\hat{\lambda}_K}{p_K} \right) X, \quad (5.19)$$

with X distributed as $\frac{p_1}{\hat{\lambda}_1} \widehat{N}_1^{RP}$.

In order to determine the distribution of X , we consider the total workload in the queue at arbitrary epochs, denoted by V^{arb} . We first note that the total workload at the system is independent of the work-conserving scheduling discipline being used. In

[32], Kingman considered a FCFS queue and showed that the scaled total workload in a $M/G/1$ queue has a proper distribution as $\rho \uparrow 1$:

$$(1 - \rho)V^{arb} \xrightarrow{d} \hat{V}^{arb},$$

where $\hat{V}^{arb}$ is exponentially distributed with mean

$$\mathbb{E}[\hat{V}^{arb}] = \frac{\sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k^2]}{2}. \quad (5.20)$$

Under the discipline DROS, the total workload at arbitrary epochs can equivalently be represented as

$$V^{arb} = \sum_{k=1}^K \sum_{h=1}^{N_k^{RP}-1} B_{k,h} + \sum_{k=1}^K \tilde{B}_k,$$

with $B_{k,h}$ the service requirement of the h -th class- k customer and $\tilde{B}_k$ the remaining service requirement of the first class- k customer in line. On one hand, note that the service requirements of all class- k customers are i.i.d., more precisely, $B_{k,h} \stackrel{d}{=} B_k$ for all h . On the other hand, $\tilde{B}_k$ is distributed as B_k if the N_k^{RP} -th class- k customer is not being served, and otherwise is given by the forward-recurrence time of B_k . Hence, for the scaled workload at arbitrary epochs we can write

$$\begin{aligned} \mathbb{E} \left[e^{-s \hat{V}^{arb}} \right] &= \lim_{\rho \uparrow 1} \mathbb{E} \left[e^{-(1-\rho)s V^{arb}} \right] \\ &= \lim_{\rho \uparrow 1} \mathbb{E} \left[e^{-(1-\rho)s \left(\sum_{k=1}^K \sum_{h=1}^{N_k^{RP}-1} B_{k,h} + \sum_{k=1}^K \tilde{B}_k \right)} \right] \\ &= \lim_{\rho \uparrow 1} \mathbb{E} \left[e^{-s \sum_{k=1}^K (1-\rho)(N_k^{RP}-1) \frac{\sum_{h=1}^{N_k^{RP}-1} B_{k,h}}{(N_k^{RP}-1)}} e^{-(1-\rho)s \sum_{k=1}^K \tilde{B}_k} \right] \\ &= \mathbb{E} \left[e^{-s \sum_{k=1}^K \mathbb{E}[B_k] \hat{N}_k^{RP}} \right], \end{aligned} \quad (5.21)$$

where in the last step we used that $e^{-s \sum_{k=1}^K (1-\rho)(N_k^{RP}-1) \frac{\sum_{h=1}^{N_k^{RP}-1} B_{k,h}}{(N_k^{RP}-1)}}$ is bounded by 1 and converges in distribution to $e^{-s \sum_{k=1}^K \mathbb{E}[B_k] \hat{N}_k^{RP}}$. From (5.21) we obtain that

$$\hat{V}^{arb} \stackrel{d}{=} \sum_{k=1}^K \mathbb{E}[B_k] \hat{N}_k^{RP}, \quad (5.22)$$

and together with (5.19) this gives

$$\hat{V}^{arb} \stackrel{d}{=} X \sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k]. \quad (5.23)$$

Since $\hat{V}^{arb}$ is exponentially distributed, the same is true for X . Hence, taking expectations in (5.23) and applying (5.20) we obtain

$$\mathbb{E}[X] = \frac{\sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k^2]}{2 \sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k]},$$

which concludes the proof of Proposition 5.3.1. $\blacksquare$

Remark 2. *In this remark we show that the two random variables that characterize the heavy traffic at departure and arbitrary epochs, Y and X , respectively, are equal in distribution. Let us consider the arbitrary arrival of a class- k customer. By the PASTA property, the number of class- k customers in the system at this time is equal to N_k^{RP} . The number of customers in the system after the first departure epoch is distributed as $(Q_1^{RP}, \dots, Q_K^{RP})$. The number of customers that arrive in the time it takes for the customer in service to depart is of the order ρ , since it is distributed as the number of arrivals in a residual service requirement. It then follows that*

$$Q_k^{RP} \stackrel{d}{=} N_k^{RP} + \mathbf{O}(\rho).$$

Multiplying the above equation by $(1 - \rho)$ and taking the limit $\rho \uparrow 1$ we get that $\widehat{Q_k^{RP}} \stackrel{d}{=} \widehat{N_k^{RP}}$ and hence $X \stackrel{d}{=} Y$.

5.4 Waiting time

In this section we investigate the waiting time in the heavy-traffic setting. We focus on the random intra-class scheduling discipline, that is, we consider the specific model DROS.

Let W_l^{DROS} denote a generic random variable for the waiting time of an arbitrary class- l customer. We refer to this customer as the tagged class- l customer. Let Q_k^* denote the number of class- k customers in the system (excluding the tagged customer) immediately after service initiation of the tagged customer in case the tagged customer arrives while the server is busy, i.e., $W_l^{DROS} > 0$. We now define the following joint transform:

$$T_l^{DROS}(u, z_1, \dots, z_K) := \mathbb{E} \left[e^{-u W_l^{DROS}} z_1^{Q_1^*} \dots z_K^{Q_K^*} \mathbf{1}_{\{W_l^{DROS} > 0\}} \right]. \quad (5.24)$$

Note that the transform of the waiting time W_l^{DROS} of the tagged class- l customer is given by

$$\mathbb{E} \left[e^{-u W_l^{DROS}} \right] = \mathbb{E} \left[e^{-u \cdot 0} \mathbf{1}_{\{W_l^{DROS} = 0\}} + e^{-u W_l^{DROS}} \mathbf{1}_{\{W_l^{DROS} > 0\}} \right] = 1 - \rho + T_l^{DROS}(u, \vec{0}), \quad (5.25)$$

since $1 - \rho$ is the probability that the tagged class- l customer arrives in an idle period. For the random intra-class scheduling discipline we have from [31] the following result for the transform $T_l^{DROS}(u, \vec{z})$.

Theorem 5.4.1. [31, Theorem 8] *For the random intra-class scheduling discipline, the joint transform $T_l^{DROS}(u, \vec{z})$, where $\vec{z} = (z_1, \dots, z_K)$, satisfies*

$$\begin{aligned} & \sum_{i=1}^K \frac{p_i}{p_l} \left(\frac{\partial}{\partial z_i} T_l^{DROS}(u, \vec{z}) \right) \left(z_i - B_i^* \left(u + \lambda - \sum_{k=1}^K \lambda_k z_k \right) \right) + T_l^{DROS}(u, \vec{z}) \\ &= W_l^1(u, \vec{z}), \end{aligned} \quad (5.26)$$

where $W_l^1(u, \vec{z})$ satisfies

$$\begin{aligned} & W_l^1(u, \vec{z}) \\ &= \sum_{i=1}^K \left((1 - \rho) \lambda_i + \lambda p_i \frac{\partial}{\partial z_i} r(z_1, \dots, z_K) \right) \frac{B_i^* \left(\lambda - \sum_{k=1}^K \lambda_k z_k \right) - B_i^* \left(u + \lambda - \sum_{k=1}^K \lambda_k z_k \right)}{u}. \end{aligned} \quad (5.27)$$

In order to study the scaled waiting time, we will need to assume throughout this section that $(1 - \rho)Q_k^*$ is uniform integrable, for all k . As we mention in Section 5.6.2, numerics show arguments to believe that this is indeed satisfied.

Assumption 1. *For a random intra-class scheduling discipline, the family of random variables $\{(1 - \rho)Q_k^*\}$ is uniform integrable for all k .*

We can now state our result that shows that in the limit the waiting time of a tagged class- l customer, W_l^{DROS} , is the product of two exponentially distributed independent random variables.

Proposition 5.4.2. *Let Assumption 1 be satisfied and consider the random intra-class scheduling discipline (i.e., DROS). Then, as $\rho \uparrow 1$,*

$$(1 - \rho) \left(W_l^{DROS}, Q_1^*, \dots, Q_K^* \right) \xrightarrow{d} \left(\widehat{W_l^{DROS}}, \widehat{Q_1^*}, \dots, \widehat{Q_K^*} \right) \stackrel{d}{=} \left(Z_l, \frac{\hat{\lambda}_1}{p_1}, \frac{\hat{\lambda}_2}{p_2}, \dots, \frac{\hat{\lambda}_K}{p_K} \right) X,$$

where $\xrightarrow{d}$ denotes convergence in distribution and X and Z_l are exponentially distributed independent random variables with $\mathbb{E}[X] = 1/\nu(\vec{p})$ and $\mathbb{E}[Z_l] = 1/p_l$.

Remark 3. *Proposition 5.4.2 is a generalization of Kingman's result, see [34], where the asymptotic waiting time distribution is obtained for the single-class DROS queue (i.e., ROS).*

In order to prove Proposition 5.4.2, we will need the following three technical lemmas. The first lemma states that the scaled vector $(Q_1^*, \dots, Q_K^*)$ has a proper limit.

Lemma 5.4.3. *When scaled by $1 - \rho$, the queue length vector $(Q_1^*, \dots, Q_K^*)$ has a proper limiting distribution as $(\lambda_1, \dots, \lambda_K) \rightarrow (\hat{\lambda}_1, \dots, \hat{\lambda}_K)$, such that as $\rho \uparrow 1$,*

$$(1 - \rho)(Q_1^*, \dots, Q_K^*) \xrightarrow{d} (\widehat{Q}_1^*, \dots, \widehat{Q}_K^*) \stackrel{d}{=} \left(\frac{\hat{\lambda}_1}{p_1}, \dots, \frac{\hat{\lambda}_K}{p_K} \right) \cdot X,$$

where $\xrightarrow{d}$ denotes convergence in distribution and X is an exponentially distributed random variable with mean $1/\nu(\vec{p})$.

Proof: Denote by $\tilde{Q}_i$ the class- i queue length at a service initiation epoch of a tagged class- i customers (excluding the tagged customer). By definition the following equality is satisfied:

$$\begin{aligned} \phi_l(e^{-s_1}, \dots, e^{-s_K}) &= \mathbb{E} \left[e^{-\sum_{i=1}^K s_i \tilde{Q}_i} \right] \\ &= \mathbb{E} \left[e^{-\sum_{i=1}^K s_i \tilde{Q}_i} \mathbf{1}_{\{W_i=0\}} \right] + \mathbb{E} \left[e^{-\sum_{i=1}^K s_i \tilde{Q}_i} \mathbf{1}_{\{W_i>0\}} \right] \\ &= 1 - \rho + T_l^{DROS}(0, e^{-s_1}, \dots, e^{-s_K}). \end{aligned}$$

Hence, from Equation (5.17) we obtain that $T_l^{DROS}(0, e^{-s_1}, \dots, e^{-s_K}) = \frac{\lambda_{p_i}}{\lambda_i} \frac{\partial}{\partial z_i} r(z_1, \dots, z_K) \big|_{\vec{z}=e^{-\vec{s}}}$.

We have $\lim_{\rho \uparrow 1} \frac{\partial}{\partial z_i} r(z_1, \dots, z_K) \big|_{\vec{z}=e^{-(1-\rho)\vec{s}}} = \frac{\hat{\lambda}_i}{p_i} \frac{d\hat{r}^*(v)}{dv} \big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k}$ (see proof of Lemma 5.3.3),

hence

$$\lim_{\rho \uparrow 1} T_l^{DROS}(0, e^{-s_1}, \dots, e^{-s_K}) = \hat{\lambda} \frac{d\hat{r}^*(v)}{dv} \big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k}. \quad (5.28)$$

From (5.18) we obtain $\mathbb{E} \left[e^{-\sum_{k=1}^K s_k \widehat{N}_k^{R\bar{P}}} \right] = \hat{\lambda} \frac{d\hat{r}^*(v)}{dv} \big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k}$. Together with Equation (5.28) and Proposition 5.3.1 this concludes the proof. $\blacksquare$

The following technical lemma characterizes the value that the function $W_l^1(u, z_1, \dots, z_K)$, as defined in (5.27), takes in heavy traffic.

Lemma 5.4.4. *We consider the random intra-class scheduling discipline (i.e., DROS). Then, as $\rho \uparrow 1$, the limit $W_l^1((1 - \rho)u, e^{-(1-\rho)s_1}, \dots, e^{-(1-\rho)s_K})$ exists and satisfies*

$$\lim_{\rho \uparrow 1} W_l^1 \left((1 - \rho)u, e^{-(1-\rho)s_1}, \dots, e^{-(1-\rho)s_K} \right) = \frac{\nu(\vec{p})}{\nu(\vec{p}) + \sum_{k=1}^K s_k \frac{\hat{\lambda}_k}{p_k}},$$

with $1/\nu(\vec{p})$ as given in (5.15).

The result of Lemma 5.4.4 implies that in heavy traffic the function $\lim_{\rho \uparrow 1} W_l^1((1 - \rho)u, e^{-(1-\rho)s_1}, \dots, e^{-(1-\rho)s_K})$ depends on s only through a linear combination of its components. The proof of Lemma 5.4.4 may be found in Appendix 5.A.3. In the following lemma we show that the scaled waiting time of a class- l customer has a proper limit.

Lemma 5.4.5. *Let Assumption 1 be satisfied and consider the random intra-class scheduling discipline (i.e., DROS). Then, there exists a $\widehat{W}_l^{DROS}$ such that $(1 - \rho)W_l^{DROS}$ converges in distribution to $\widehat{W}_l^{DROS}$ as $\rho \uparrow 1$.*

Proof: By definition, the following two equalities are satisfied:

$$T_l^{DROS}(u, 1, \dots, 1) = \mathbb{E}[e^{-uW_l^{DROS}} \mathbf{1}_{\{W_l^{DROS} > 0\}}], \quad (5.29)$$

and

$$\frac{\partial}{\partial z_i} T_l^{DROS}(u, z_1, \dots, z_K) \Big|_{\vec{z}=1} = \mathbb{E}[Q_i^* e^{-uW_l^{DROS}} \mathbf{1}_{\{W_l^{DROS} > 0\}}]. \quad (5.30)$$

Now, considering Equation (5.26) with $\vec{z} = 1$ in heavy traffic we get:

$$\begin{aligned} & \lim_{\rho \uparrow 1} W_l^1((1 - \rho)u, 1, \dots, 1) \\ &= \lim_{\rho \uparrow 1} \sum_{i=1}^K \frac{p_i}{p_l} (1 - \rho) \frac{\partial}{\partial z_i} T_l^{DROS}((1 - \rho)u, z_1, \dots, z_K) \Big|_{\vec{z}=1} \frac{1 - B_i^*((1 - \rho)u)}{(1 - \rho)} \\ & \quad + \lim_{\rho \uparrow 1} T_l^{DROS}((1 - \rho)u, 1, \dots, 1) \\ &= \lim_{\rho \uparrow 1} \sum_{i=1}^K \frac{p_i}{p_l} u \mathbb{E}[B_i] \mathbb{E}[(1 - \rho)Q_i^* e^{-(1 - \rho)uW_l^{DROS}} \mathbf{1}_{\{W_l^{DROS} > 0\}}] \\ & \quad + \lim_{\rho \uparrow 1} \mathbb{E}[e^{-(1 - \rho)uW_l^{DROS}} \mathbf{1}_{\{W_l^{DROS} > 0\}}] \\ &= \lim_{\rho \uparrow 1} \mathbb{E} \left[\left(\sum_{i=1}^K \frac{p_i}{p_l} u \mathbb{E}[B_i] (1 - \rho)Q_i^* + 1 \right) e^{-(1 - \rho)uW_l^{DROS}} \mathbf{1}_{\{W_l^{DROS} > 0\}} \right] \\ &= \mathbb{E} \left[\lim_{\rho \uparrow 1} \left(\sum_{i=1}^K \frac{p_i}{p_l} u \mathbb{E}[B_i] (1 - \rho)Q_i^* + 1 \right) e^{-(1 - \rho)uW_l^{DROS}} \mathbf{1}_{\{W_l^{DROS} > 0\}} \right], \quad (5.31) \end{aligned}$$

where in the second step we used (5.29) and (5.30) and in the fourth step we used the hypothesis that $(1 - \rho)Q_i^*$ is uniformly integrable (Assumption 1), [5, Theorem 3.5]. Note that $W_l^1((1 - \rho)u, e^{-(1 - \rho)\vec{s}})$, which is defined in Equation (5.27), has a proper limit when $\rho \uparrow 1$, see Lemma 5.4.4. Since (5.31) converges, the same must hold for (5.31).

Besides, $\sum_{i=1}^K \frac{p_i}{p_l} u \mathbb{E}[B_i] (1 - \rho)Q_i^*$ converges in distribution to $\sum_{i=1}^K \frac{p_i}{p_l} u \mathbb{E}[B_i] \frac{\hat{\lambda}_i}{p_i} X$ (see Lemma 5.4.3) and therefore, we conclude that the waiting time of an arbitrary class- l customer in heavy traffic converges in distribution to some random variable $\widehat{W}_l^{DROS}$. ■

From Lemma 5.4.4, we note that (5.31) should in fact be independent of u . It can be checked that in case $(1 - \rho)(W_l^{DROS}, Q_1^*, \dots, Q_K^*)$ is distributed as $X(Z_l, \frac{\hat{\lambda}_1}{p_1}, \dots, \frac{\hat{\lambda}_K}{p_K})$, as we want to show, see Proposition 5.4.2, this is indeed satisfied.

We can now prove Proposition 5.4.2 which consists in finding $T_l^{DROS}(\cdot)$ by solving Equation (5.26) after the heavy-traffic scaling.

Proof of Proposition 5.4.2: We know by Lemma 5.4.5 that there is a random variable $\widehat{W}_l^{DROS}$ such that $(1 - \rho)W_l^{DROS}$ converges in distribution to $\widehat{W}_l^{DROS}$. Hence, we can define the function $\widehat{T}_l^{DROS}(u, \vec{s})$ as follows:

$$\begin{aligned}\widehat{T}_l^{DROS}(u, \vec{s}) &:= \mathbb{E} \left[e^{-u\widehat{W}_l^{DROS}} e^{-\sum_{i=1}^K s_i \widehat{Q}_i^*} \right] \\ &= \lim_{\rho \uparrow 1} \mathbb{E} \left[e^{-(1-\rho)uW_l^{DROS}} e^{-(1-\rho)s_1 Q_1^*} \dots e^{-(1-\rho)s_K Q_K^*} \right] \\ &= \lim_{\rho \uparrow 1} T_l^{DROS}((1 - \rho)u, e^{-(1-\rho)s_1}, \dots, e^{-(1-\rho)s_K}).\end{aligned}$$

We will evaluate Equation (5.26) in the point $(u, \vec{z}) = (u(1 - \rho), e^{-(1-\rho)\vec{s}})$ as $\rho \uparrow 1$. We first focus on the first term. We have

$$\begin{aligned}&\lim_{\rho \uparrow 1} (1 - \rho) \frac{\partial}{\partial z_i} T_l^{DROS}(u, \vec{z}) \Big|_{u=(1-\rho)u, \vec{z}=e^{-(1-\rho)\vec{s}}} \\ &= \lim_{\rho \uparrow 1} (1 - \rho) \mathbb{E} \left[\frac{Q_i^* e^{-(1-\rho)uW_l^{DROS}} e^{-(1-\rho)s_1 Q_1^*} \dots e^{-(1-\rho)s_i (Q_i^* - 1)} \dots e^{-(1-\rho)s_K Q_K^*}}{e^{-(1-\rho)s_i}} \right] \\ &= \mathbb{E} \left[\widehat{Q}_i^* e^{-u\widehat{W}_l^{DROS}} e^{-s_1 \widehat{Q}_1^*} \dots e^{-s_K \widehat{Q}_K^*} \right] = -\frac{\partial}{\partial s_i} \widehat{T}_l^{DROS}(u, \vec{s}),\end{aligned}\tag{5.32}$$

where in the second step we used Assumption 1 and [5, Theorem 3.5].

Moreover, realize that $\widehat{T}_l^{DROS}(u, \vec{s}) = \mathbb{E} \left[e^{-u\widehat{W}_l^{DROS}} e^{-\sum_{i=1}^K s_i \widehat{Q}_i^*} \right]$ depends on $\vec{s} = (s_1, \dots, s_K)$ only through $y = \sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k$ (see Lemma 5.4.3). Thus, we will write $\widehat{T}_l^{DROS}(u, \vec{s}) = \widehat{T}_l^{DROS}(u, y)$ and by the chain rule:

$$\begin{aligned}\frac{\partial}{\partial s_i} \widehat{T}_l^{DROS}(u, \vec{s}) &= \frac{\partial}{\partial s_i} \widehat{T}_l^{DROS}(u, y) = \frac{\partial}{\partial y} \widehat{T}_l^{DROS}(u, y) \frac{\partial y}{\partial s_i} \\ &= \frac{\hat{\lambda}_i}{p_i} \frac{\partial}{\partial y} \widehat{T}_l^{DROS}(u, y) \Big|_{y=\sum_{k=1}^K s_k \frac{\hat{\lambda}_k}{p_k}}.\end{aligned}\tag{5.33}$$

Then, taking the heavy-traffic limit in Equation (5.26) and using Equations (5.32),

$$\begin{aligned}&\left(e^{-(1-\rho)s_i} - B_i^*((1 - \rho)u + \lambda - \sum_{k=1}^K \lambda_k e^{-(1-\rho)s_k}) \right) \\ (5.33), \text{ Lemma 5.4.4 and the relation } \lim_{\rho \uparrow 1} &\frac{\left(e^{-(1-\rho)s_i} - B_i^*((1 - \rho)u + \lambda - \sum_{k=1}^K \lambda_k e^{-(1-\rho)s_k}) \right)}{1 - \rho} = \\ &-s_i + \mathbb{E}[B_i] \left(u + \sum_{k=1}^K \hat{\lambda}_k s_k \right), \text{ which follows from l'Hopital's rule, we arrive to the following} \\ &\text{ordinary differential equation (ODE):}\end{aligned}$$

$$-\sum_{i=1}^K \frac{p_i}{p_i} \frac{\hat{\lambda}_i}{p_i} \frac{\partial \widehat{T}_l^{DROS}(u, y)}{\partial y} \Big|_{y=\sum_{i=1}^K s_i \frac{\hat{\lambda}_i}{p_i}} \left(-s_i + \mathbb{E}[B_i] \left(u + \sum_{k=1}^K \hat{\lambda}_k s_k \right) \right) + \widehat{T}_l^{DROS}(u, y) = \frac{\nu(\vec{p})}{\nu(\vec{p}) + y}.$$

Since $\sum_{i=1}^K \hat{\lambda}_i \left(-s_i + \mathbb{E}[B_i] \left(u + \sum_{k=1}^K \hat{\lambda}_k s_k \right) \right) = u$, the latter can be written as

$$-\frac{u}{p_l} \frac{\partial \widehat{T_l^{DROS}}(u, y)}{\partial y} \Big|_{y=\sum_{k=1}^K s_k \frac{\hat{\lambda}_k}{p_k}} + \widehat{T_l^{DROS}}(u, y) = \frac{\nu(\vec{p})}{\nu(\vec{p}) + y}. \quad (5.34)$$

The solution of the ODE (5.34) is

$$\widehat{T_l^{DROS}}(u, y) = \frac{p_l}{u} e^{\frac{p_l}{u} y} \int_y^\infty e^{-\frac{p_l}{u} x} \frac{\nu(\vec{p})}{\nu(\vec{p}) + x} dx, \quad (5.35)$$

see Appendix 5.A.4 for the details.

Let Z_l and X be two exponentially distributed random variables with $\mathbb{E}[Z_l] = 1/\eta$ and $\mathbb{E}[X] = 1/\nu(\vec{p})$. Then, the Laplace Transform of $(Z_l \cdot X, \frac{\lambda_1}{p_1} X, \dots, \frac{\lambda_K}{p_K} X)$ is given by:

$$\begin{aligned} \mathbb{E}[e^{-u Z_l \cdot X - s_1 \frac{\lambda_1}{p_1} X - \dots - s_K \frac{\lambda_K}{p_K} X}] &= \mathbb{E}[e^{-u Z_l \cdot X - \sum_{k=1}^K s_k \frac{\lambda_k}{p_k} X}] = \mathbb{E}[e^{-u Z_l \cdot X - y X}] \\ &= \mathbb{E}[\mathbb{E}[e^{-u z_l X - y X} | z_l = Z_l]] = \mathbb{E}\left[\frac{\nu(\vec{p})}{\nu(\vec{p}) + u Z_l + y}\right] = \int_0^\infty \eta e^{-\eta z_l} \frac{\nu(\vec{p})}{\nu(\vec{p}) + u z_l + y} dz_l \\ &= \frac{1}{u} \int_{\nu(\vec{p})+y}^\infty \eta e^{-\eta \frac{x-\nu(\vec{p})-y}{u}} \frac{\nu(\vec{p})}{x} dx = \frac{\nu(\vec{p})\eta}{u} e^{\eta \frac{\nu(\vec{p})+y}{u}} \int_{\nu(\vec{p})+y}^\infty e^{-\eta \frac{x}{u}} \frac{1}{x} dx \\ &= \frac{\nu(\vec{p})\eta}{u} e^{\eta \frac{\nu(\vec{p})+y}{u}} \int_{\eta \frac{\nu(\vec{p})+y}{u}}^\infty \frac{e^{-l}}{l} dl. \end{aligned} \quad (5.36)$$

Making the change of variable $z = p_l \frac{\nu(\vec{p})+x}{u}$ in Equation (5.35) we obtain

$$\hat{T}_l(u, y) = \nu(\vec{p}) \frac{p_l}{u} e^{p_l \frac{y}{u}} \int_y^\infty e^{-p_l \frac{x}{u}} \frac{1}{\nu(\vec{p}) + x} dx = p_l \frac{\nu(\vec{p})}{u} e^{p_l \frac{\nu(\vec{p})+y}{u}} \int_{p_l \frac{\nu(\vec{p})+y}{u}}^\infty e^{-z} \frac{1}{z} dz.$$

Hence, it coincides with the Laplace Transform of $(Z_l \cdot X, \frac{\lambda_1}{p_1} X, \dots, \frac{\lambda_K}{p_K} X)$ obtained in (5.36) with $p_l = \eta$. Since the Laplace transform of a probability distribution is unique, (uniqueness theorem, [14]), we conclude that $(1 - \rho)(W_l^{DROS}, Q_1^*, \dots, Q_K^*)$ converges in distribution to $(Z_l \cdot X, \frac{\lambda_1}{p_1} X, \dots, \frac{\lambda_K}{p_K} X)$, where, as we have previously mentioned, Z_l and X are exponentially distributed independent random variables with $\mathbb{E}(Z_l) = 1/p_l$ and $\mathbb{E}(X) = 1/\nu(\vec{p})$. $\blacksquare$

5.5 Optimal selection of the weights

In this section we show how the results of Proposition 5.3.1 and Proposition 5.4.2 can be used in order to optimize the performance. In particular, in Section 5.5.1 we focus on the holding cost and in Section 5.5.2 we find the weights that minimize the moments of the waiting time of an arbitrary customer.

5.5.1 Holding cost

With each class of customers we associate a cost $c_k \geq 0, k = 1, \dots, K$. As performance measure we take the holding cost $\sum_{k=1}^K c_k N_k^{RP}$. In this section we will write $N_k^{RP}(\vec{p}), \widehat{N}_k^{RP}(\vec{p})$ instead of $N_k^{RP}, \widehat{N}_k^{RP}$ to emphasize the dependence on the weights $\vec{p} := (p_1, \dots, p_K)$. From Proposition 5.3.1 we obtain that the scaled holding cost, $(1-\rho) \sum_{k=1}^K c_k N_k^{RP}(\vec{p})$, converges in distribution to an exponentially distributed random variable with mean

$$\sum_{k=1}^K c_k \mathbb{E} \left[\widehat{N}_k^{RP}(\vec{p}) \right] = \frac{\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} c_k}{2 \sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k]} \sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k^2], \quad (5.37)$$

as $\rho \uparrow 1$. Using this expression, we obtain the following monotonicity result in the heavy-traffic regime: The holding cost decreases “stochastically” as more preference is given to customers with a large value of $\frac{c_i}{\mathbb{E}[B_i]}$. This can be seen as an extension of the $c\mu$ -rule for the heavy-traffic setting [17].

Proposition 5.5.1. *Consider two policies with weights $(p_1, \dots, p_K)$ and $(q_1, \dots, q_K)$, respectively. Let $c_k \geq 0, k = 1, \dots, K$. Without loss of generality we assume that the classes are ordered such that $\frac{c_1}{\mathbb{E}[B_1]} \geq \frac{c_2}{\mathbb{E}[B_2]} \geq \dots \geq \frac{c_K}{\mathbb{E}[B_K]}$. If $\frac{p_k}{p_{k+1}} \leq \frac{q_k}{q_{k+1}}$, for all $k = 1, \dots, K-1$, then*

$$\lim_{\rho \uparrow 1} (1-\rho) \sum_{k=1}^K c_k N_k^{RP}(\vec{p}) \geq_{st} \lim_{\rho \uparrow 1} (1-\rho) \sum_{k=1}^K c_k N_k^{RP}(\vec{q}),$$

where $\geq_{st}$ denotes the usual stochastic ordering, i.e., $X \geq_{st} Y$ if and only if $\mathbb{P}(X \geq z) \geq \mathbb{P}(Y \geq z)$ for all z .

Proof: We have that $(1-\rho) \sum_{k=1}^K c_k N_k^{RP}(\vec{p})$ converges in distribution to an exponentially distributed random variable with mean as stated in (5.37). Since exponentially distributed random variables are stochastically ordered according to their means, it only remains to check that

$$\frac{\sum_{k=1}^K \frac{c_k \hat{\lambda}_k}{p_k}}{\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k]} \geq \frac{\sum_{k=1}^K \frac{c_k \hat{\lambda}_k}{q_k}}{\sum_{k=1}^K \frac{\hat{\lambda}_k}{q_k} \mathbb{E}[B_k]}.$$

This holds since

$$\begin{aligned}
& \left(\sum_{k=1}^K \frac{c_k \hat{\lambda}_k}{p_k} \right) \left(\sum_{k=1}^K \frac{\hat{\lambda}_k}{q_k} \mathbb{E}[B_k] \right) \\
&= \sum_{k,i:k \neq i} \hat{\lambda}_k \hat{\lambda}_i \left(\frac{1}{p_k q_i} c_k \mathbb{E}[B_i] + \frac{1}{p_i q_k} c_i \mathbb{E}[B_k] \right) + \sum_{k=1}^K \hat{\lambda}_k^2 \frac{1}{p_k q_k} c_k \mathbb{E}[B_k] \\
&\geq \sum_{k,i:k \neq i} \hat{\lambda}_k \hat{\lambda}_i \left(\frac{1}{p_i q_k} c_k \mathbb{E}[B_i] + \frac{1}{p_k q_i} c_i \mathbb{E}[B_k] \right) + \sum_{k=1}^K \hat{\lambda}_k^2 \frac{1}{p_k q_k} c_k \mathbb{E}[B_k] \\
&= \left(\sum_{k=1}^K \frac{c_k \hat{\lambda}_k}{q_k} \right) \left(\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k] \right).
\end{aligned}$$

Here we used that $c_i \mathbb{E}[B_k] \left(\frac{1}{p_i q_k} - \frac{1}{p_k q_i} \right) \geq c_k \mathbb{E}[B_i] \left(\frac{1}{p_i q_k} - \frac{1}{p_k q_i} \right)$, which follows from the fact that $\frac{p_i}{p_k} \leq \frac{q_i}{q_k}$ and $\frac{c_i}{\mathbb{E}[B_i]} \geq \frac{c_k}{\mathbb{E}[B_k]}$, for $i \leq k$. $\blacksquare$

5.5.2 Moments of the waiting time

In this section we will give the optimal values for the weights that minimize the m -th moment of the limit of the scaled waiting time of a tagged class- k customer, $\widehat{W}_k^{DROS}$. From Proposition 5.4.2 we know that

$$\widehat{W}_k^{DROS} \stackrel{d}{=} X \cdot Z_k, \quad (5.38)$$

where X and Z_k are exponentially distributed independent random variables with $\mathbb{E}(Z_k) = 1/p_k$ and $\mathbb{E}(X) = 1/\nu(\vec{p})$. Now taking the expression in (5.38) and using that X and Z_k are independent random variables we observe that the m -th moment of $\widehat{W}_k^{DROS}$ is given by

$$\mathbb{E} \left[\left(\widehat{W}_k^{DROS} \right)^m \right] = \mathbb{E}[X^m Z_k^m] = \mathbb{E}[X^m] \mathbb{E}[Z_k^m] = \frac{m!}{\nu(\vec{p})^m} \frac{m!}{p_k^m} = (m!)^2 \frac{1}{p_k^m} \left(\frac{\sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k^2]}{2 \sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k]} \right)^m.$$

Hence the m -th moment of the waiting time for an arbitrary customer is given by

$$\mathbb{E} \left[\left(\widehat{W}^{DROS} \right)^m \right] = \sum_{k=1}^K \frac{\hat{\lambda}_k}{\hat{\lambda}} \mathbb{E} \left[\left(\widehat{W}_k^{DROS} \right)^m \right] = (m!)^2 \left(\sum_{k=1}^K \frac{\hat{\lambda}_k}{\hat{\lambda} p_k^m} \right) \left(\frac{\sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k^2]}{2 \sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k]} \right)^m. \quad (5.39)$$

In what follows we will write $\widehat{W}^{DROS}(\vec{p})$ instead of $\widehat{W}^{DROS}$ to emphasize the dependence on the weights $\vec{p}$.

Note that $\mathbb{E} \left[\widehat{W}^{DROS}(\vec{p}) \right] = \frac{1}{\hat{\lambda}} \sum_{k=1}^K \mathbb{E} \left[\widehat{N}_k^{RP}(\vec{p}) \right]$. Hence, by applying Little's law to the result obtained in Proposition 5.5.1, we obtain the following corollary, which means that the mean waiting time decreases as more preference is given to customers with a small value of $\mathbb{E}[B_i]$, $i = 1, \dots, K$.

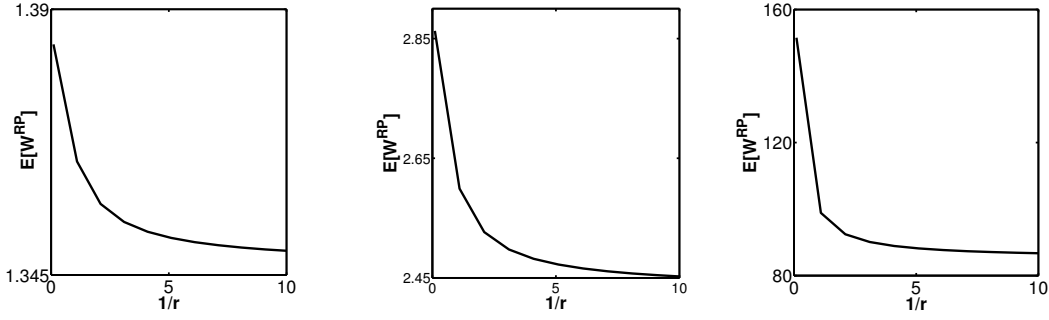


Figure 5.1: The mean waiting time for three classes of customers in the system, $K=3$, under DROS for the loads $\rho = 0.2610, \rho = 0.6083$ and $\rho = 0.99$, respectively. The horizontal axis corresponds to $\frac{1}{r} = \frac{p_j}{p_{j+1}}, j = 1, \dots, K-1$.

Corollary 5.5.2. *Without loss of generality we assume that the classes are ordered such that $\mathbb{E}[B_1] \leq \dots \leq \mathbb{E}[B_K]$. If $\frac{p_j}{p_{j+1}} \leq \frac{q_j}{q_{j+1}}$, for all $j = 1, \dots, K-1$, then $\mathbb{E}[\widehat{W^{DROS}}(\vec{p})] \geq \mathbb{E}[\widehat{W^{DROS}}(\vec{q})]$.*

Remark 4. *The monotonicity result for the waiting time holds in the heavy-traffic setting. In the case of two classes, $K=2$, Corollary 5.5.2 is true for any stable system, i.e., for any value of ρ , not necessarily close to one. This can be seen as follows. The expression for the mean waiting time for $K=2$ is the following:*

$$\begin{aligned} \mathbb{E}[W^{DROS}(\vec{p})] &= \sum_{i=1}^2 \frac{\lambda_i}{\lambda} \mathbb{E}[W_i^{DROS}] \\ &= \frac{\lambda_1 \mathbb{E}[B_1^2] + \lambda_2 \mathbb{E}[B_2^2]}{2\lambda} \frac{\lambda_1(1 - \rho p_1) + \lambda_2(1 - \rho p_2)}{(1 - \rho_1 - p_2 \rho_2)(1 - \rho_2 - p_1 \rho_1) - p_1 p_2 \rho_1 \rho_2} \end{aligned} \quad (5.40)$$

where the expression of $\mathbb{E}[W_i^{DROS}]$, $i = 1, 2$, was obtained in [31], Equation (38). Without loss of generality we assume that $p_1 + p_2 = 1$. Then, taking the derivative of (5.40) with respect to p_1 we obtain the monotonicity result as stated in Corollary 5.5.2.

Moreover, we have written a code to calculate the mean waiting time as given in [31] for any value of K , i.e., for any number of classes of customers. We choose the weights such that $\frac{p_j}{p_{j+1}} = \frac{1}{r}, \forall j$. In the figures we chose exponentially distributed service requirements, however, the monotonicity observed holds for any service requirement distribution (with the same first moment). The results obtained are shown in Figure 5.1 and Figure 5.2, for $K=3$ and $K=4$, respectively, for different values of the load. It can be seen for these examples that as more priority is given to customers with small mean service requirement (i.e., as $\frac{1}{r}$ becomes large), the mean waiting time decreases for any value of the load.

In Corollary 5.5.2 we considered the first moment of the scaled waiting time. In

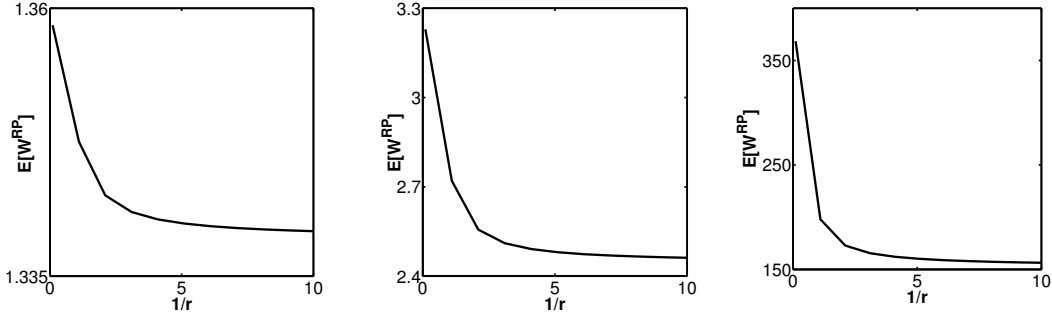


Figure 5.2: The mean waiting time for four classes of customers in the system, $K=4$, under DROS for the loads $\rho = 0.2535, \rho = 0.6304$ and $\rho = 0.9951$, respectively. The horizontal axis corresponds to $\frac{1}{r} = \frac{p_j}{p_{j+1}}, j = 1, \dots, K-1$.

Proposition 5.5.3 we will investigate the m -th moment of the scaled waiting time and find the optimal value for the weights, which is non-trivial.

Proposition 5.5.3. *The m -th moment of the limit of the scaled waiting time, $\mathbb{E} \left[\left(\widehat{W^{DROS}}(\vec{p}) \right)^m \right]$, is minimized in $\vec{p}^* = (p_1^*, \dots, p_K^*)$, with*

$$p_k^* := \frac{1/\mathbb{E}[B_k]^{1/m-1}}{\sum_{i=1}^K 1/\mathbb{E}[B_i]^{1/m-1}}, \quad (5.41)$$

for each $k \in \{1, \dots, K\}$, $m = 2, 3, \dots$

Proof: We need to show that $\mathbb{E} \left[\left(\widehat{W^{DROS}}(p^*) \right)^m \right] \leq \mathbb{E} \left[\left(\widehat{W^{DROS}}(p) \right)^m \right]$. This holds if and only if

$$\frac{\sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k]^{m/m-1}}{\hat{\lambda}(\sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k]^{m/m-1})^m} \cdot \left(\frac{\sum_{j=1}^K \hat{\lambda}_j \mathbb{E}[B_j^2]}{2} \right)^m \leq \frac{\sum_{k=1}^K \frac{\hat{\lambda}_k}{\hat{\lambda} p_k^m}}{(\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k])^m} \cdot \left(\frac{\sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k^2]}{2} \right)^m,$$

which follows by definition. This is equivalent to

$$\frac{1}{(\sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k]^{m/m-1})^{m-1}} \leq \frac{\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k^m}}{(\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k])^m}$$

and rewriting it we obtain

$$\left(\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k] \right)^m \leq \sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k^m} \left(\sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k]^{m/m-1} \right)^{m-1}. \quad (5.42)$$

The latter holds by Hölder's inequality. ■

Remark 5. By Equation (5.41) we get that the ratio of two optimal weights is the following:

$$\frac{p_k^*}{p_j^*} = \left(\frac{\mathbb{E}[B_j]}{\mathbb{E}[B_k]} \right)^{1/(m-1)}. \quad (5.43)$$

In general, the optimal choice for the weights is non trivial. However, note that when $m \rightarrow 1$ we deduce that under the optimal weights (for the m -th moment) a class- k customer has strict priority over a class- j customer if $\mathbb{E}[B_k] < \mathbb{E}[B_j]$. This is exactly the result that the $c\mu$ -rule states. In addition, when $m \rightarrow \infty$, from (5.43) we see that the ratio of the optimal weights converges to 1. This implies that as m gets larger, it becomes optimal (for the m -th moment) to treat all classes equal.

5.6 Numerical results

In this section we present numerical experiments related to the results obtained in this chapter. We consider a system under the discipline DROS with two classes of customers ($K = 2$) and assume exponentially distributed service requirements. For each experiment in the order of 10^5 busy periods are simulated. A busy period refers to the period of time between two consecutive time epochs in which the system is empty, and every busy period is a regenerative point of the stochastic process. In Section 5.6.1 we present the numerical results corresponding to the distribution of the number of customers in the queue, in Sections 5.6.2 we focus on the moments of the queue length and waiting time and in Section 5.6.3 we investigate the optimal weights.

5.6.1 State-space collapse for the queue lengths

In this section we simulate the distribution of the joint queue length vector. As parameters we chose $\lambda_1 = 2.15$, $\lambda_2 = 2.85$, $\mathbb{E}[B_1] = 1/4$ and $\mathbb{E}[B_2] = 1/6$, so that $\rho = 0.9994$. In Figure 5.3 we plot the joint queue length probabilities (obtained by simulation) for the weights $p_1 = 0.7$, $p_2 = 0.3$. The horizontal and vertical axis correspond to N_1^{RP} and N_2^{RP} , respectively. As a consequence of the state-space collapse stated in Proposition 5.3.1, in heavy traffic the probabilities will lie on a straight line with slope $\frac{\widehat{N_2^{RP}}}{\widehat{N_1^{RP}}} = \frac{p_1}{\hat{\lambda}_1} \frac{\hat{\lambda}_2}{p_2} \approx 3.1$, starting from the origin. This result coincides with the slope of the figure obtained.

5.6.2 Moments of waiting time and queue length

In Figure 5.4 we plot $(1 - \rho)\mathbb{E}[N^{RP}]$ (using Little's Law and Equation (5.40)) and $(1 - \rho)^2\mathbb{E}\left[\left(N^{RP}\right)^2\right]$ (obtained by simulation) for different values of the load ρ . When doing so, we keep the mean service requirements fixed, $\mathbb{E}[B_1] = 1/4$ and $\mathbb{E}[B_2] = 1/6$, and take $\lambda_2 = 1.5\lambda_1$. Moreover, we calculate the first and second moment of the limit of the scaled queue length, i.e., $\mathbb{E}\left[\lim_{\rho \uparrow 1} \left((1 - \rho)N^{RP}\right)^m\right] = \mathbb{E}\left[\left(\widehat{N^{RP}}\right)^m\right]$, $m = 1, 2$,

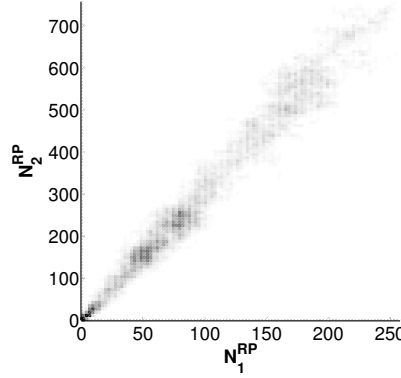


Figure 5.3: Joint queue length probability. The darkness of the points specifies the probability of being into a particular state. The darker the point is, the higher the probability of being in that state.

using (5.14), giving as a result the values indicated with a dot in Figure 5.4, which are $\mathbb{E}[\widehat{N^{RP}}] = 0.9589$ and $\mathbb{E}[(\widehat{N^{RP}})^2] = 1.9510$.

As it can be seen in Figure 5.4, in both cases, as the load gets close to one the functions $\mathbb{E}[(1-\rho)N_k^{RP}]^m$, $m = 1, 2$, converge to the values indicated with the dot. This would imply that an interchange of the limit and expectation holds for the random variable $(1-\rho)N_k^{RP}$, i.e., $\lim_{\rho \uparrow 1} \mathbb{E}[(1-\rho)N_k^{RP}]^m = \mathbb{E}[\lim_{\rho \uparrow 1} ((1-\rho)N_k^{RP})^m]$, $m = 1, 2$.

We note that if the limits are indeed interchangeable, together with the convergence in distribution of the scaled queue lengths this would imply the uniform integrability of the scaled queue length (see [5, Theorem 3.5]), as assumed in Assumption 1.

In Figure 5.5 we plot $(1-\rho)\mathbb{E}[W^{DROS}]$ (using Equation (5.40)) and $(1-\rho)^2\mathbb{E}[(W^{DROS})^2]$ (obtained by simulation) for different values of the load ρ . The simulation setting is the same as the one used for the queue length. We calculate the value of Equation (5.39) for the cases $m = 1$ and $m = 2$ giving as a result the values indicated with a dot in Figure 5.5, which are $\mathbb{E}[\widehat{W^{DROS}}] = 0.1906$ and $\mathbb{E}[(\widehat{W^{DROS}})^2] = 0.1713$. In both cases, as the load gets close to one the functions converge to the value obtained in Equation (5.39), which would imply again that an interchange of the limit and expectation holds for the random variable $(1-\rho)W_k^{DROS}$, i.e., $\lim_{\rho \uparrow 1} \mathbb{E}[(1-\rho)W_k^{DROS}]^m = \mathbb{E}[\lim_{\rho \uparrow 1} ((1-\rho)W_k^{DROS})^m]$, $m = 1, 2$. In fact, for the first moment, taking the limit as $\rho \uparrow 1$ in Equation (5.40) it is easy to see that it indeed converges to the heavy-traffic limit as characterized by Equation (5.39) when $K = 2$.

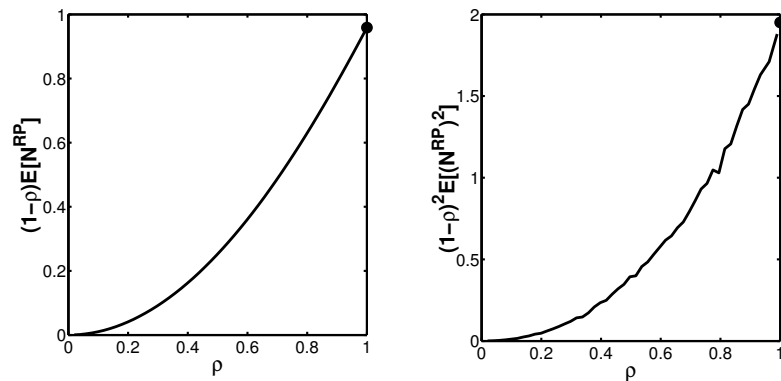


Figure 5.4: First and second moment of the scaled queue length obtained for different values of the load ρ . The dots in both pictures are calculated by using (5.14) giving as a result $\mathbb{E}[\widehat{N^{RP}}] = 0.9589$ and $\mathbb{E}[(\widehat{N^{RP}})^2] = 1.9510$.

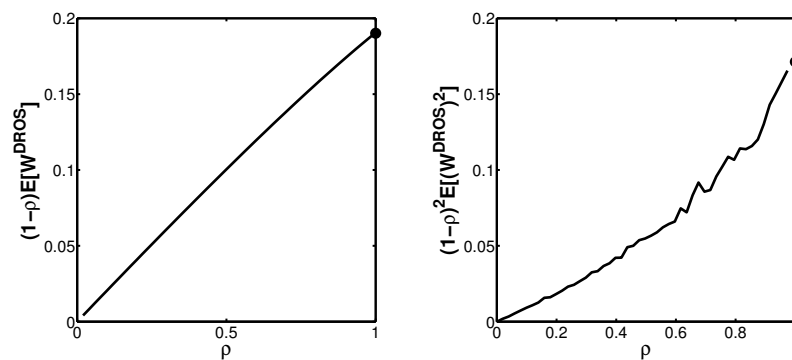


Figure 5.5: First and second moment of the scaled waiting time obtained for different values of the load ρ . The dots in both pictures are calculated by using Equation (5.39) giving as a result $\mathbb{E}[\widehat{W^{DROS}}] = 0.1906$ and $\mathbb{E}[(\widehat{W^{DROS}})^2] = 0.1713$.

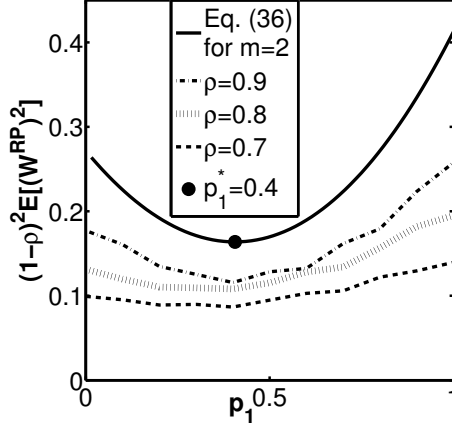


Figure 5.6: The second moment $\mathbb{E} \left[\left((1 - \rho) W^{DROS} \right)^2 \right]$ for different values of the weights $(p_1, p_2) = (p_1, 1 - p_1)$.

5.6.3 Optimal values for the weights

In Proposition 5.5.3 we presented the optimal choices for the weights p^* in order to minimize the moments of the scaled waiting time $\widehat{W^{DROS}}$. In this section we numerically evaluate the validity of the optimal weights outside the heavy-traffic regime. We set $\mathbb{E}[B_1] = 0.2439$ and $\mathbb{E}[B_2] = 0.1667$ and plot $(1 - \rho)^2 \mathbb{E} \left[\left(W^{DROS} \right)^2 (p_1, 1 - p_1) \right]$ for three different values of the load, $\rho = 0.7, \rho = 0.8$ and $\rho = 0.9$, see Figure 5.6. The value of p_1^* is in this particular case equal to $p_1^* = 0.4059$ (see (5.41)). It can be seen that the weight $p_1^* = 0.4059$ is a good approximation for the minimizer of $(1 - \rho)^2 \mathbb{E} \left[\left(W^{DROS} \right)^2 (p_1, 1 - p_1) \right]$ for load equal to $\rho = 0.9$. As the load decreases the approximation becomes worse, but it is still close to the minimum of the function. We also plot $\mathbb{E} \left[\lim_{\rho \uparrow 1} (1 - \rho)^2 \left(W^{DROS} \right)^2 (p_1, 1 - p_1) \right] = \mathbb{E} \left[\left(\widehat{W^{DROS}} \right)^2 (p_1, 1 - p_1) \right]$, which is seen to be a good approximation for $(1 - \rho)^2 \mathbb{E} \left[\left(W^{DROS} \right)^2 (p_1, 1 - p_1) \right]$ as the load gets close to 1.

5.A Appendix

5.A.1 Proof of Lemma 5.2.3

The total workload at departure epochs can be represented as

$$V^{\text{dep}} = \sum_{k=1}^K \sum_{h=1}^{Q_k^{RP}} B_{k,h},$$

with $B_{k,h}$ the service requirement of the h -th class- k customer. Note that the service requirements of all class- k customers are i.i.d., and $B_{k,h} \stackrel{d}{=} B_k$ for all h .

For $\epsilon' > 0$ we have

$$\begin{aligned}
\mathbb{P}((1-\rho)Q_k^{RP} \geq M) &= \mathbb{P}\left(Q_k^{RP} \geq \frac{M}{(1-\rho)}\right) \\
&\leq \mathbb{P}\left(\sum_{h=1}^{Q_k^{RP}} B_{k,h} \geq \sum_{h=1}^{\lfloor M/(1-\rho) \rfloor} B_{k,h}\right) \\
&\leq \mathbb{P}\left((1-\rho) \sum_{k=1}^K \sum_{h=1}^{Q_k^{RP}} B_{k,h} \geq M \frac{(1-\rho)}{M} \sum_{h=1}^{\lfloor M/(1-\rho) \rfloor} B_{k,h}\right) \\
&= \mathbb{P}\left(\frac{(1-\rho)V^{\text{dep}}}{M} - \mathbb{E}[B_k] \geq \frac{(1-\rho)}{M} \sum_{h=1}^{\lfloor M/(1-\rho) \rfloor} B_{k,h} - \mathbb{E}[B_k]\right) \\
&= \mathbb{P}\left(\frac{(1-\rho)V^{\text{dep}}}{M} - \mathbb{E}[B_k] \right. \\
&\quad \left. \geq \frac{(1-\rho)}{M} \sum_{h=1}^{\lfloor M/(1-\rho) \rfloor} B_{k,h} - \mathbb{E}[B_k], \frac{(1-\rho)}{M} \sum_{h=1}^{\lfloor M/(1-\rho) \rfloor} B_{k,h} - \mathbb{E}[B_k] > -\epsilon\right) \\
&\quad + \mathbb{P}\left(\frac{(1-\rho)V^{\text{dep}}}{M} - \mathbb{E}[B_k] \right. \\
&\quad \left. \geq \frac{(1-\rho)}{M} \sum_{h=1}^{\lfloor M/(1-\rho) \rfloor} B_{k,h} - \mathbb{E}[B_k] \middle| \frac{(1-\rho)}{M} \sum_{h=1}^{\lfloor M/(1-\rho) \rfloor} B_{k,h} - \mathbb{E}[B_k] \leq -\epsilon\right) \\
&\quad \cdot \mathbb{P}\left(\frac{(1-\rho)}{M} \sum_{h=1}^{\lfloor M/(1-\rho) \rfloor} B_{k,h} - \mathbb{E}[B_k] \leq -\epsilon\right) \\
&\leq \mathbb{P}\left(\frac{(1-\rho)V^{\text{dep}}}{M} - \mathbb{E}[B_k] > -\epsilon\right) + \tilde{\epsilon} = \mathbb{P}\left((1-\rho)V^{\text{dep}} \geq M(\mathbb{E}[B_k] - \epsilon)\right) + \tilde{\epsilon} \\
&< \bar{\epsilon} + \tilde{\epsilon} = \epsilon', \text{ for } \rho \text{ close enough to 1 and } M \text{ large enough.} \tag{5.44}
\end{aligned}$$

In the fifth step we used that $\frac{(1-\rho)}{M} \sum_{h=1}^{\lfloor M/(1-\rho) \rfloor} B_{k,h}$ converges in distribution to $\mathbb{E}[B_k]$ as $\rho \uparrow 1$, hence $\mathbb{P}\left(\frac{(1-\rho)}{M} \sum_{h=1}^{\lfloor M/(1-\rho) \rfloor} B_{k,h} - \mathbb{E}[B_k] \leq -\epsilon\right) \leq \bar{\epsilon}$, for ρ close enough to 1. In the last step we used the fact that the workload, independently of the work-conserving scheduling discipline being used, is tight in heavy traffic, see Kingman [32], that is $\forall \bar{\epsilon} \exists M'$ such that $\mathbb{P}((1-\rho)V^{\text{dep}} \geq M') < \bar{\epsilon}$. From (5.44) we conclude that $(1-\rho)(Q_1^{RP}, Q_2^{RP}, \dots, Q_K^{RP})$ is tight.

5.A.2 Proof of Lemma 5.2.6

The proof of Lemma 5.2.6 is based on the proof of Lemma 3 in [55]. We have

$$\begin{aligned}
 \sum_{i=1}^K \frac{\hat{\lambda}_i}{p_i} F_i(\vec{s}) &= \sum_{i=1}^K \frac{\hat{\lambda}_i}{p_i} p_i(-s_i + \mathbb{E}[B_i]) \sum_{k=1}^K \hat{\lambda}_k s_k \\
 &= -\sum_{i=1}^K \hat{\lambda}_i s_i + \sum_{i=1}^K \hat{\lambda}_i \mathbb{E}[B_i] \sum_{k=1}^K \hat{\lambda}_k s_k \\
 &= -\sum_{i=1}^K \hat{\lambda}_i s_i + \sum_{k=1}^K \hat{\lambda}_k s_k \\
 &= 0.
 \end{aligned}$$

This implies that for all $\vec{s} \in H_c$, the vector $\vec{F}(\vec{s})$ is parallel to the hyperplane H_c . Since $\vec{F}$ is C^1 , for each state $\vec{s} \geq \vec{0}$ there exists a unique flow $\vec{f}(u) = (f_1(u), \dots, f_K(u))$, parametrized by $u \geq 0$, such that

$$\vec{f}(0) = \vec{s} \quad \text{and} \quad \frac{\partial f_i(u)}{\partial u} = F_i(\vec{f}(u)), \quad \text{for all } i \text{ and } u \geq 0. \quad (5.45)$$

Since $\vec{F}(\vec{s})$ is parallel to H_c for all $\vec{s} \in H_c$, when started in H_c , the flow $\vec{f}(u)$ will stay in H_c . Another important property of this flow $\vec{f}(u)$ is that

$$\frac{\partial \hat{r}(\vec{f}(u))}{\partial u} = \sum_{i=1}^K \frac{\partial f_i(u)}{\partial u} \cdot \frac{\partial \hat{r}(\vec{s})}{\partial s_i} \Big|_{\vec{s}=\vec{f}(u)} = 0,$$

which follows from the chain rule, Lemma 5.2.5, and Equation (5.45). Hence, along each flow $\vec{f}(u)$, which lies in H_c , the function $\hat{r}(\vec{f}(u))$ is constant. We will now show that each flow in H_c converges to a certain point $c \cdot \vec{s}^* \geq \vec{0}$ as $u \rightarrow \infty$.

From (5.10) we get that (5.45) can be written as $\vec{f}'(u) = \vec{s}$ and $\vec{f}'(u)^T = A\vec{f}(u)^T$ with

$$A = \begin{pmatrix} p_1(-1 + \mathbb{E}[B_1]\hat{\lambda}_1) & p_1\mathbb{E}[B_1]\hat{\lambda}_2 & \cdots & p_1\mathbb{E}[B_1]\hat{\lambda}_K \\ p_2\mathbb{E}[B_2]\hat{\lambda}_1 & p_2(-1 + \mathbb{E}[B_2]\hat{\lambda}_2) & \cdots & p_2\mathbb{E}[B_2]\hat{\lambda}_K \\ \vdots & \vdots & \ddots & \vdots \\ p_K\mathbb{E}[B_K]\hat{\lambda}_1 & p_K\mathbb{E}[B_K]\hat{\lambda}_2 & \cdots & p_K(-1 + \mathbb{E}[B_K]\hat{\lambda}_K) \end{pmatrix}. \quad (5.46)$$

In Lemma 5.A.1 below it is proved that one eigenvalue of A is 0 with eigenvector $\vec{s}^* \geq \vec{0}$, $\vec{s}^* \in H_1$, and all the other eigenvalues have a strictly negative real part. Hence, the solution of $\vec{f}'(u)^T = A\vec{f}(u)^T$ with $\vec{f}(0) \in H_c$ can be written as $\vec{f}(u) = c \cdot \vec{s}^* + \vec{g}(u)$, where $\lim_{u \rightarrow \infty} \vec{g}(u) = \vec{0}$ and $\vec{s}^* \geq \vec{0}$. This implies that all the flows in the hyperplane H_c converge to one common point $c \cdot \vec{s}^* \geq \vec{0}$.

Since the continuous function $\hat{r}(\vec{s})$ is constant along each flow, and all flows in the hyperplane H_c converge to $c \cdot \vec{s}^* \in H_c$, we obtain that the function $\hat{r}(\vec{s})$ is constant on H_c .

The following technical lemma is used in the proof of Lemma 5.2.6.

Lemma 5.A.1. *Consider the matrix A as defined in (5.46). One eigenvalue of A is 0 (with multiplicity 1), and all the other eigenvalues have a strictly negative real part. In addition, there exists a vector $\vec{\eta} = (\eta_1, \dots, \eta_K) \geq \vec{0}$ with $\sum_{j=1}^K \eta_j = 1$ such that $\vec{s}^* = (s_1^*, \dots, s_K^*)$ with $s_j^* := \frac{p_j}{\lambda_j} \eta_j$ is an eigenvector of A corresponding to the eigenvalue 0, and $\vec{s}^* \in H_1$.*

Proof: Define D as the diagonal matrix $\text{diag}[d_1, d_2, \dots, d_K]$ with $d_i = \frac{\hat{\lambda}_i}{p_i}$, and let S be the matrix

$$S = DAD^{-1} = \begin{pmatrix} p_1(-1 + \mathbb{E}[B_1]\hat{\lambda}_1) & p_2\hat{\lambda}_1\mathbb{E}[B_1] & \cdots & p_K\hat{\lambda}_1\mathbb{E}[B_1] \\ p_1\hat{\lambda}_2\mathbb{E}[B_2] & p_2(-1 + \mathbb{E}[B_2]\hat{\lambda}_2) & \cdots & p_K\hat{\lambda}_2\mathbb{E}[B_2] \\ \vdots & \vdots & \ddots & \vdots \\ p_1\hat{\lambda}_K\mathbb{E}[B_K] & p_2\hat{\lambda}_K\mathbb{E}[B_K] & \cdots & p_K(-1 + \mathbb{E}[B_K]\hat{\lambda}_K) \end{pmatrix}. \quad (5.47)$$

The matrix A is similar to S and therefore A, S and S^T have the same eigenvalues. The sum of each row of S^T is 0 because $\sum_{i=1}^K \mathbb{E}[B_i]\hat{\lambda}_i = 1$, and the off-diagonal elements in S^T are all strictly positive. This implies that the matrix S^T is a generator corresponding to a finite-state continuous-time irreducible Markov chain. Hence, it has a unique equilibrium distribution $\vec{\eta}$, i.e., $\vec{\eta}S^T = \vec{0}$ and $\sum_{k=1}^K \eta_k = 1$. In particular, 0 is an eigenvalue of the matrix S^T , with multiplicity 1 and corresponding to the left eigenvector $\vec{\eta}$, and, cf.(Proposition 6.2,[2]), the real parts of all other eigenvalues are strictly negative. Since the eigenvalues of A and S^T coincide, the same holds for the matrix A . The eigenvector of A corresponding to the eigenvalue 0 is given by $\vec{s}^{*T} = D^{-1}\vec{\eta}^T$, since $A\vec{s}^{*T} = D^{-1}DAD^{-1}\vec{\eta}^T = D^{-1}S\vec{\eta}^T = \vec{0}^T$. ■

5.A.3 Proof of Lemma 5.4.4

Taking $(u, z_1, \dots, z_K) = ((1 - \rho)u, e^{-(1-\rho)s_1}, \dots, e^{-(1-\rho)s_K})$ in (5.27) we get

$$\begin{aligned} W_l^1((1 - \rho)u, e^{-(1-\rho)s_1}, \dots, e^{-(1-\rho)s_K}) = \\ \sum_{i=1}^K ((1 - \rho)\lambda_i + \lambda p_i \frac{\partial}{\partial z_i} r(z_1, \dots, z_K) \Big|_{z_i = e^{-(1-\rho)s_i}}) \cdot \\ \frac{B_i^*(\lambda - \sum_{k=1}^K \lambda_k e^{-(1-\rho)s_k}) - B_i^*((1 - \rho)u + \lambda - \sum_{k=1}^K \lambda_k e^{-(1-\rho)s_k})}{(1 - \rho)u}. \end{aligned}$$

By applying l'Hopital's rule we get the expression below:

$$\begin{aligned}
& \lim_{\rho \uparrow 1} W_l^1((1-\rho)u, e^{-(1-\rho)s_1}, \dots, e^{-(1-\rho)s_K}) \\
&= -\frac{1}{u} \sum_{i=1}^K \left(\hat{\lambda}_i \cdot 0 + \hat{\lambda}_i p_i \frac{\partial}{\partial s_i} \hat{r}(s_1, \dots, s_K) \left(B_i^{*'}(0) \left(-\sum_{k=1}^K \hat{\lambda}_k s_k \right) \right. \right. \\
&\quad \left. \left. - B_i^{*'}(0) \left(-u - \sum_{k=1}^K \hat{\lambda}_k s_k \right) \right) \right) \\
&= \hat{\lambda} \sum_{i=1}^K \mathbb{E}[B_i] p_i \frac{\hat{\lambda}_i}{p_i} \frac{d}{dv} \hat{r}^*(v) \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k} = \hat{\lambda} \frac{d}{dv} \hat{r}^*(v) \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k},
\end{aligned}$$

with $\hat{r}(\vec{s})$ as defined in (5.6) and $\hat{r}^*(\vec{s})$ as defined in the proof of Lemma 5.3.3. The result now follows from Equation (5.18), together with the fact that the latter is equal to $\frac{\nu(\vec{p})}{\nu(\vec{p}) + \sum_{k=1}^K s_k \frac{\hat{\lambda}_k}{p_k}}$ since $\left(\widehat{N_1^{RP}}, \dots, \widehat{N_K^{RP}} \right) \stackrel{d}{=} X \cdot \left(\frac{\lambda_1}{p_1}, \dots, \frac{\lambda_K}{p_K} \right)$, with X exponentially distributed with mean $1/\nu(\vec{p})$.

5.A.4 Solution of the ODE (5.34)

The solution of (5.34) is given by the sum of the solution of the homogeneous case, $T_l^{\widehat{DROS},H}(u, y)$, and a particular solution, $T_l^{\widehat{DROS},P}(u, y)$. The homogeneous solution is given by:

$$T_l^{\widehat{DROS},H}(u, y) = C(u) e^{\frac{p_l}{u} y}, \quad (5.48)$$

where $C(u)$ is an arbitrary function of u . In order to find the particular solution we rewrite (5.34) as

$$\frac{\partial T_l^{\widehat{DROS},P}(u, y)}{\partial y} \Big|_{y=\sum_{k=1}^K s_k \frac{\hat{\lambda}_k}{p_k}} - \frac{p_l}{u} T_l^{\widehat{DROS},P}(u, y) = -\frac{p_l}{u} \frac{\nu(\vec{p})}{\nu(\vec{p}) + y}. \quad (5.49)$$

Let us solve the new equation using the integrating factor technique. In order to do so, we define the function $\mu(y) = e^{-\frac{p_l}{u} y}$ and multiply (5.49) by it. The derivative of $\mu(y)$ satisfies $\frac{d\mu(y)}{dy} = -\mu(y) \frac{p_l}{u}$.

Then, our equation becomes

$$\begin{aligned}
-\mu(y) \frac{p_l}{u} \frac{\nu(\vec{p})}{\nu(\vec{p}) + y} &= \mu(y) \frac{\partial T_l^{\widehat{DROS},P}(u, y)}{\partial y} \Big|_{y=\sum_{k=1}^K s_k \frac{\hat{\lambda}_k}{p_k}} - \mu(y) \frac{p_l}{u} T_l^{\widehat{DROS},P}(u, y) \\
&= \mu(y) \frac{\partial T_l^{\widehat{DROS},P}(u, y)}{\partial y} \Big|_{y=\sum_{k=1}^K s_k \frac{\hat{\lambda}_k}{p_k}} + T_l^{\widehat{DROS},P}(u, y) \frac{d\mu(y)}{dy} \\
&= \frac{\partial}{\partial y} (\mu(y) T_l^{\widehat{DROS},P}(u, y)),
\end{aligned}$$

which can be solved by integration. Integrating each side with respect to y gives us a particular solution for (5.34), which is,

$$T_l^{\widehat{DROS},P}(u, y) = -\frac{p_l}{u} e^{\frac{p_l}{u}y} \int_0^y e^{-\frac{p_l}{u}x} \frac{\nu(\vec{p})}{\nu(\vec{p}) + x} dx,$$

which is identically written as

$$T_l^{\widehat{DROS},P}(u, y) = \frac{p_l}{u} e^{\frac{p_l}{u}y} \int_y^\infty e^{-\frac{p_l}{u}x} \frac{\nu(\vec{p})}{\nu(\vec{p}) + x} dx. \quad (5.50)$$

In conclusion, the general solution of the ODE (5.34) is given by

$$T_l^{\widehat{DROS}}(u, y) = T_l^{\widehat{DROS},H}(u, y) + T_k^{\widehat{DROS},P}(u, y) = C(u) e^{\frac{p_l}{u}y} + \frac{p_l}{u} e^{\frac{p_l}{u}y} \int_y^\infty e^{-\frac{p_l}{u}x} \frac{\nu(\vec{p})}{\nu(\vec{p}) + x} dx. \quad (5.51)$$

We will now show that the constant $C(u)$ is equal to zero. First, note that $T_l^{\widehat{DROS},H}(u, y) \rightarrow \infty$ as $y \rightarrow \infty$. Second, since $T_l^{\widehat{DROS}}(0, y) = \frac{\nu(\vec{p})}{\nu(\vec{p}) + y}$ (from Equation (5.24) and Lemma 5.4.3) it is immediate that $T_l^{\widehat{DROS}}(u, y)$ converges to 0 when $y = \sum_{k=1}^K s_k \frac{\hat{\lambda}_k}{p_k} \rightarrow \infty$. Moreover, if we take the particular solution (5.50), applying l'Hopital's rule for $y \rightarrow \infty$ we obtain that it also converges to zero, namely:

$$\begin{aligned}
\lim_{y \rightarrow \infty} T_l^{\widehat{DROS}}(u, y) &= \lim_{y \rightarrow \infty} \frac{\nu(\vec{p}) p_l}{u} \frac{\int_y^\infty e^{-\frac{p_l}{u}x} \frac{1}{\nu(\vec{p}) + x} dx}{e^{-\frac{p_l}{u}y}} = \lim_{y \rightarrow \infty} \frac{\nu(\vec{p}) p_l}{u} \frac{-e^{-\frac{p_l}{u}y} \frac{1}{\nu(\vec{p}) + y}}{-\frac{p_l}{u} e^{-\frac{p_l}{u}y}} \\
&= \lim_{y \rightarrow \infty} \frac{\nu(\vec{p})}{\nu(\vec{p}) + y} = 0.
\end{aligned}$$

Then, the necessary condition for $C(u) e^{\frac{p_l}{u}y}$ to converge to zero as $y \rightarrow \infty$ is $C(u) = 0$. As a consequence, we conclude that the solution of (5.34) is

$$T_l^{\widehat{DROS}}(u, y) = \frac{p_l}{u} e^{\frac{p_l}{u}y} \int_y^\infty e^{-\frac{p_l}{u}x} \frac{\nu(\vec{p})}{\nu(\vec{p}) + x} dx.$$

Chapter 6

Interpolation approximations for a relative priorities queue

This chapter is devoted to the light and heavy-traffic interpolation approximations of the RP model, see Section 1.1 for the model description. In order to build the interpolation approximations we will make use of the heavy-traffic characterisations obtained in Chapter 5.

Kim et al.'s results, Theorems 5.3.2 and 5.4.1 presented in Chapter 5, cannot be solved in general, however, they are very valuable in obtaining insights into the performance of the system. In particular, in Chapter 5, they were key in carrying out a heavy-traffic analysis of the queue length and waiting time under RP and DROS, respectively, and in the current chapter they will be key in obtaining the light-traffic approximation for the distribution of the stationary queue length vector and the waiting time. Combining the so-obtained light-traffic approximations and the heavy-traffic characterisations we build the interpolation approximations for the aforementioned performance metrics. An important benefit of the approximations is that they provide insights into the dependency of the performance on the system parameters (weights, service time distributions, etc), and we thus believe it will provide an interesting tool.

The remainder of the chapter is organised as follows. In Section 6.1 we obtain the interpolation approximation for the steady-state queue-length, and in Section 6.2 the interpolation approximation for the waiting time. In Section 6.3 we present the corresponding numerical results.

6.1 Queue-length distribution

We recall from Chapter 5 that the steady-state number of class- k customers in the system at arbitrary epochs is denoted by N_k^{RP} . We define the vector $\vec{N}^{RP} = (N_1^{RP}, \dots, N_K^{RP})$ and $N^{RP} := \sum_{k=1}^K N_k^{RP}$.

In this section we are interested in approximating the probability generating function of the queue length vector under RP and therefore, according to the notation in-

roduced in Chapter 2, we set $\vec{y} = \vec{z}$ and let $G(\lambda, \vec{z}) = \psi^{RP}(\lambda, \vec{z})$ be the probability generating function of $(N_1^{RP}, \dots, N_K^{RP})$ under RP.

Now, using the result of Proposition 5.3.1 obtained in Chapter 5, we obtain the following heavy-traffic result. As stated in Chapter 5, a state-space collapse for the scaled queue length vector in the heavy-traffic regime was established, that is, in the limit the scaled queue length vector is distributed as the product of an exponentially distributed random variable and a deterministic vector.

Lemma 6.1.1. *The scaled probability generating function of the stationary queue length vector, $\psi^{RP}(\lambda, \vec{z}^{(1-\lambda\mathbb{E}[B])})$, under the heavy-traffic regime satisfies*

$$\lim_{\lambda \rightarrow 1/\mathbb{E}[B]} \psi^{RP}(\lambda, \vec{z}^{(1-\lambda\mathbb{E}[B])}) = \frac{\mathbb{E}[B]\nu(\vec{p})}{\mathbb{E}[B]\nu(\vec{p}) - \sum_{i=1}^K \frac{\hat{\alpha}_i}{p_i} \ln(z_i)}. \quad (6.1)$$

Proof:

$$\begin{aligned} \lim_{\lambda \rightarrow 1/\mathbb{E}[B]} \psi^{RP}(\lambda, \vec{z}^{(1-\lambda\mathbb{E}[B])}) &= \lim_{\lambda \rightarrow 1/\mathbb{E}[B]} \mathbb{E}[z_1^{(1-\lambda\mathbb{E}[B])N_1^{RP}} \dots z_K^{(1-\lambda\mathbb{E}[B])N_K^{RP}}] \\ &= \lim_{\lambda \rightarrow 1/\mathbb{E}[B]} \mathbb{E} \left[e^{\ln(z_1)^{(1-\lambda\mathbb{E}[B])N_1^{RP}}} \dots e^{\ln(z_K)^{(1-\lambda\mathbb{E}[B])N_K^{RP}}} \right] = \mathbb{E} \left[e^{\widehat{N_1^{RP} \ln(z_1)}} \dots e^{\widehat{N_K^{RP} \ln(z_K)}} \right] \\ &= \mathbb{E} \left[e^{X \sum_{i=1}^K \frac{\hat{\lambda}_i}{p_i} \ln(z_i)} \right] = \frac{\mathbb{E}[B]\nu(\vec{p})}{\mathbb{E}[B]\nu(\vec{p}) - \sum_{i=1}^K \frac{\hat{\alpha}_i}{p_i} \ln(z_i)}, \end{aligned} \quad (6.2)$$

where in the fourth step we used Proposition 5.3.1 and $\nu(\vec{p})$ as given in Equation (5.15). ■

From Theorem 5.3.2, where it is given an expression that the joint probability generating function of the joint stationary queue lengths at arbitrary time epochs satisfies, we obtain the following light-traffic approximation (when $\lambda = 0$).

Lemma 6.1.2. *The light-traffic approximation (of order 2) of the probability generating function of the queue-length distribution is given by*

$$\begin{aligned} \psi^{RP,LT}(\lambda, \vec{z}) &= (\psi^{RP})^{(0)}(0, \vec{z}) + \lambda (\psi^{RP})^{(1)}(0, \vec{z}) + \frac{\lambda^2}{2} (\psi^{RP})^{(2)}(0, \vec{z}) \\ &= 1 - \rho + \lambda \sum_{i=1}^K \alpha_i \mathbb{E}[B_i] z_i + \frac{\lambda^2}{2} \sum_{i=1}^K \alpha_i z_i \mathbb{E}[B_i^2] \left(\sum_{k=1}^K \alpha_k z_k - 1 \right). \end{aligned}$$

Proof: See Appendix 6.A.1 for the proof. ■

We now present the interpolation approximation for the queue length.

Proposition 6.1.3. *The interpolation approximation (of order 3) of the probability generating function of $(N_1^{RP}, \dots, N_K^{RP})$ is given by*

$$\begin{aligned}
& \psi^{RP,INT}(\lambda, \vec{z}) \\
&= (1 - \rho^3) + \lambda (1 - \rho^2) \left(-\mathbb{E}[B] + \sum_{i=1}^K \alpha_i \mathbb{E}[B_i] z_i^{(1-\rho)^{-1}} \right) \\
&+ \frac{\lambda^2 (1 - \rho)}{2} \left(-2\mathbb{E}[B] \sum_{i=1}^K \alpha_i z_i^{(1-\rho)^{-1}} \mathbb{E}[B_i] \frac{\ln(z_i)}{1 - \rho} \right. \\
&\quad \left. + \sum_{i=1}^K \alpha_i z_i^{(1-\rho)^{-1}} \mathbb{E}[B_i^2] \left(\sum_{k=1}^K \alpha_k z_k^{(1-\rho)^{-1}} - 1 \right) \right) \\
&+ \rho^3 \frac{\mathbb{E}[B] \nu(\vec{p})}{\mathbb{E}[B] \nu(\vec{p}) - \sum_{i=1}^K \frac{\alpha_i}{p_i} \ln(z_i^{(1-\rho)^{-1}})}. \tag{6.3}
\end{aligned}$$

Proof: The proof follows directly from Proposition 2.3.1, where we substitute the light-traffic derivatives as derived in Lemma 6.1.2, and the heavy-traffic term $G^{HT}(\vec{z})$ as given in (6.1). A detailed proof is provided in Appendix 6.A.2. $\blacksquare$

Equation (6.3) can be readily used to obtain the first and second moments of the number of customers in the system. As an example, we obtain now the approximation for the first moment of the total number of customers. For the mean we have

$$\begin{aligned}
\mathbb{E}[N^{RP,INT}] &= \mathbb{E}[N_1^{RP,INT} + \dots + N_K^{RP,INT}] = \frac{\partial \left(\psi^{RP,INT}(\lambda, \vec{z}) \Big|_{z_i=z_j=z} \right)}{\partial z} \Big|_{z=1} \\
&= \rho + \frac{\lambda^2 \mathbb{E}[B^2]}{2} + \frac{\rho^3}{(1 - \rho)} \frac{\mathbb{E}[B^2]}{2\mathbb{E}[B] \sum_{k=1}^K \frac{\alpha_k}{p_k} \mathbb{E}[B_k]} \cdot \sum_{i=1}^K \frac{\alpha_i}{p_i}. \tag{6.4}
\end{aligned}$$

Under the assumption that there is one class in the system, that is, $\alpha_i = 0, \forall i \neq k$ and $\alpha_k = 1$, Equation (6.4) is exact, that is, it coincides with the well known Pollaczek-Khinchine formula for the M/G/1 queue. We have

$$\mathbb{E}[N^{RP,INT}] = \rho + \frac{\lambda^2 \mathbb{E}[B^2]}{2} \left(1 + \frac{\rho}{1 - \rho} \right) = \rho + \frac{\lambda^2 \mathbb{E}[B^2]}{2(1 - \rho)}.$$

The second derivative of $\psi^{RP,INT}(\lambda, \vec{z})$ with respect to z , evaluated at $z = 1$, is

$$\begin{aligned}
& \frac{\partial^2 \left(\psi^{RP,INT}(\lambda, \vec{z}) \Big|_{z_i=z_j=z} \right)}{\partial z^2} \Big|_{z=1} = \mathbb{E} \left[\left(N^{RP,INT} \right)^2 \right] - \mathbb{E}[N^{RP,INT}] = \frac{\lambda^2 \mathbb{E}[B^2]}{1 - \rho} \frac{2 + \rho}{2} \\
&+ \frac{\rho^3}{1 - \rho} \frac{2\mathbb{E}[B]}{\mathbb{E}[B^2]} \sum_{k=1}^K \frac{\alpha_k}{p_k} \mathbb{E}[B_k] \left(\frac{\mathbb{E}[B^2]}{2\mathbb{E}[B] \sum_{k=1}^K \frac{\alpha_k}{p_k} \mathbb{E}[B_k]} \right)^2 \\
&\cdot \sum_{i=1}^K \frac{\alpha_i}{p_i} \left(\frac{2}{(1 - \rho)} \frac{\mathbb{E}[B^2]}{2\mathbb{E}[B] \sum_{k=1}^K \frac{\alpha_k}{p_k} \mathbb{E}[B_k]} \sum_{i=1}^K \frac{\alpha_i}{p_i} - 1 \right). \tag{6.5}
\end{aligned}$$

Therefore, the second moment is given by the sum of Equations (6.4) and (6.5)

$$\mathbb{E} \left[\left(N^{RP,INT} \right)^2 \right] = \frac{\partial^2 \left(\psi^{RP,INT}(\lambda, \vec{z}) \right) \Big|_{z_i=z_j=z}}{\partial z^2} \Big|_{z=1} + \mathbb{E}[N^{RP,INT}]. \quad (6.6)$$

In Section 6.3 we use the expression for the first and second moment, Equations (6.4) and (6.6) to numerically test the accuracy of the interpolation approximation.

6.2 Waiting time distribution

The waiting time in RP depends on the intra-class scheduling discipline being implemented. As in Chapter 5 we consider the particular case in which the intra-class scheduling discipline is random and we refer to it as DROS. We recall from Chapter 5 that the waiting time of an arbitrary class- k customer is denoted by W_k^{DROS} . We refer to this customer as the tagged class- k customer. Let Q_k^* denote the number of class- k customers in the system (excluding the tagged customer) immediately after service initiation of the tagged customer in case the tagged customer arrives while the server is busy, i.e., $W_k^{DROS} > 0$.

In this section we are interested in approximating the LST of the waiting time under DROS and therefore, according to the notation introduced in Chapter 2, we set $\vec{y} = u$ and let $G(\lambda, u) = \widetilde{W}_k^{DROS}(\lambda, u) = \mathbb{E}[e^{-uW_k^{DROS}}]$ be the LST of a class- k customer's waiting time under DROS. Then, using Proposition 5.4.2 we obtain the following heavy-traffic result.

Lemma 6.2.1. *The Laplace Transform of the scaled waiting time of a class- k customer under the heavy-traffic regime satisfies*

$$\lim_{\lambda \rightarrow 1/\mathbb{E}[B]} \widetilde{W}_k^{DROS}(\lambda, (1 - \lambda\mathbb{E}[B])u) = \frac{\nu(\vec{p})p_k}{u} e^{p_k \frac{\nu(\vec{p})}{u}} \int_{p_k \frac{\nu(\vec{p})}{u}}^{\infty} \frac{e^{-l}}{l} dl. \quad (6.7)$$

Now, taking the derivatives of Equation (5.26) with respect to λ we obtain the following light-traffic approximation. Equation (5.26) is a partial differential equation that satisfies the joint transform of the waiting time and queue length.

Lemma 6.2.2. *The light-traffic approximation (of order 1) of the Laplace Transform of the waiting time under DROS is given by*

$$\widetilde{W}_k^{DROS,LT}(\lambda, u) = \left(\widetilde{W}_k^{DROS} \right)^{(0)}(0, u) + \lambda \left(\widetilde{W}_k^{DROS} \right)^{(1)}(0, u) = 1 - \rho + \lambda \left(\sum_{i=1}^K \alpha_i \frac{1 - B_i^*(u)}{u} \right)$$

Proof: See Appendix 6.A.3 for the proof. ■

We note that the light traffic approximation is independent of the class. Indeed, from Proposition 2.1.1 we know that the 1st order approximation is calculated when

there is only one arrival to the system, and thus, the scheduling policy does not play any role. The 2nd order approximation can be calculated, however, the final expression is much more cumbersome, and yet the numerical accuracy does not significantly improve. In the next proposition we present the interpolation approximation which does depend on the weights due to the heavy-traffic term.

Proposition 6.2.3. *The interpolation approximation (of order 2) of the LST of the waiting time under DROS is given by*

$$\begin{aligned} \widetilde{W}_k^{DROS,INT}(\lambda, u) &= (1 - \rho)^2 + \lambda(1 - \rho) \left(-\mathbb{E}[B] + \sum_{i=1}^K \alpha_i \frac{1 - B_i^*((1 - \rho)^{-1}u)}{(1 - \rho)^{-1}u} \right) \\ &\quad + \rho^2 \frac{\nu(\vec{p})p_k}{(1 - \rho)^{-1}u} e^{p_k \frac{\nu(\vec{p})}{(1 - \rho)^{-1}u}} \int_{p_k \frac{\nu(\vec{p})}{(1 - \rho)^{-1}u}}^{\infty} \frac{e^{-l}}{l} dl, \end{aligned}$$

with $\nu(\vec{p})$ given as in Equation (5.15).

Proof: The proof follows directly from Proposition 2.3.1, where we substitute the light-traffic derivatives as derived in Lemma 6.2.2, and the heavy-traffic term $G^{HT}(u)$ as given in (6.7). See Appendix 6.A.4 for the proof. $\blacksquare$

6.3 Numerical comparison

In this section we numerically investigate the accuracy of the approximation obtained in Proposition 6.1.3 for the probability generating function of the queue length vector under RP.

To measure the accuracy, we use as reference the result of Kim et al. given in Equation (5.16) that unfortunately cannot be solved analytically for arbitrary λ , but in [31, Section 3.2] the authors present a numerical scheme to obtain the moments numerically for any service-time distribution. We denote their results by $(*)^{KIM}$, where $*$ refers to the metric studied.

We refer to Sections 3.7 and 4.6 for a review of the service time distributions that we use throughout this section. The performance criteria will again be the relative error. For the first and second moments of the number of customers, we will calculate $100\% \times \frac{\mathbb{E}[N^{RP}] - \mathbb{E}[N^{RP,INT}]}{\mathbb{E}[N^{RP}]}$ and $100\% \times \frac{\mathbb{E}[(N^{RP})^2] - \mathbb{E}[(N^{RP,INT})^2]}{\mathbb{E}[(N^{RP})^2]}$, respectively. We have decided to focus on the queue length because we have observed numerically that the error of the interpolation approximation is larger for the queue length than for the sojourn time or waiting time.

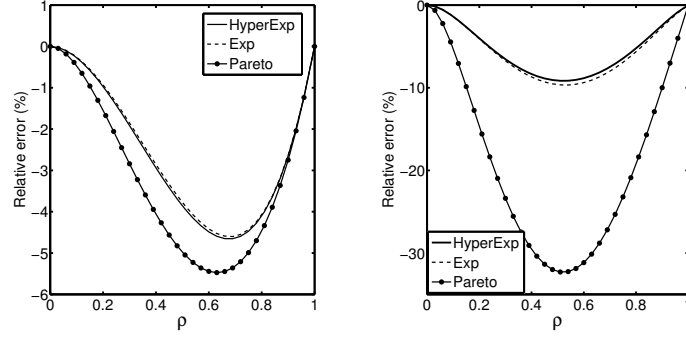


Figure 6.1: Scenario 1. Relative error with RP of the first (left) and second moments (right) of the total number of customers in the system for hyper-exponential, exponential and Pareto service-time distributions.

6.3.1 Accuracy of interpolation approximation

We now measure the accuracy of the approximation obtained in Proposition 6.1.3 by considering the first and second moments that are given in Equations (6.4) and (6.6).

Scenario 1. In Figure 6.1 we plot the relative error of the first and second moments of the total number of customers in the system with respect to the load for exponential, hyper-exponential and Pareto service time distributions. We consider two classes with $\mathbb{E}[B_1] = 11/3$, $\mathbb{E}[B_2] = 44/3$. We assume that an arriving customer is of class 1 (class 2) with probability $\alpha_1 = 8/12$ ($\alpha_2 = 4/12$). The weights are set to $p_1 = 2$, $p_2 = 5$. We observe that the first moment remains accurate for the three distributions while the absolute relative error of the second moment for Pareto distribution reached the value of 30%. The fact that Pareto does not satisfy Equation (2.1) might explain the large relative error.

Scenario 2. In Figure 6.2 we consider 2 classes of customers. Class-1 customers' service requirements follow an exponential distribution of rate μ_1 , while class-2 customers' service requirements follow a degenerate hyper-exponential distribution as defined in Equation (4.28) with parameters $m_2 = 2$, $\beta_{21} = w$, $\beta_{22} = 1 - w$, $\mathbb{E}[B_{21}] = 1/(\mu_2 w)$ and $\mathbb{E}[B_{22}] = 0$. We consider $p_1 = 2$, $p_2 = 5$, $\alpha_1 = 7/12$, $\alpha_2 = 5/12$, $\mathbb{E}[B_1] = 11/3$, $\mathbb{E}[B_2] = 44/3$. In Figure 6.2 we plot the relative error of the first and second moments of the total number of customers in the system with respect to w for different values of the load. Observe that as expected for $\rho \approx 0$ and $\rho \approx 1$ our interpolation approximation is exact. The absolute largest error occurs for intermediate values of the load, as w approaches 0.

6.3.2 RP versus DPS

In Figure 6.3 we plot the first and second moments of the number of customers in the system for RP and DPS. Recall that to measure the accuracy, for the RP model we use

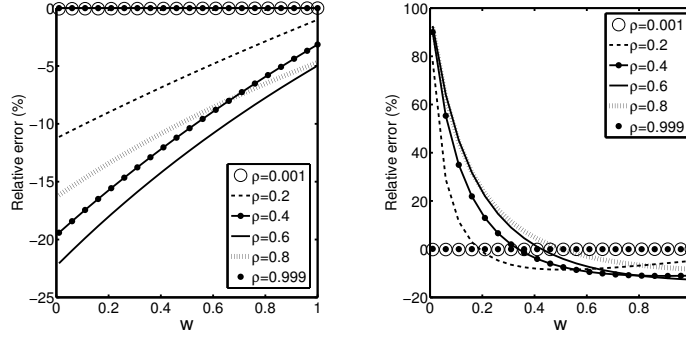


Figure 6.2: Scenario 2. Relative error in RP of the first (left) and second moments (right) of the number of customers in the system for different values of the load.

as reference the algorithm proposed by Kim et al. in [31, Section 3.2], that allows us to obtain the moments numerically for any service-time distribution, and we denote their results by $(*)^{KIM}$, where $*$ refers to the metric studied. For the DPS model we use the algorithm proposed by Rege et al. in [46, Section 1] which is only valid for exponential service times and we denote their results by $(*)^{REGE}$ where, again, $*$ refers to the metric under consideration.

Scenario 3. In Figure 6.3 we plot the first and second moments of the total number of customers in the system for RP and DPS. We plot both our interpolation approximation as well as the exact results obtained from the literature. We consider two classes, class 1 is exponentially distributed with $\mathbb{E}[B_1] = 5$, and class 2 is degenerate hyper-exponential with $\mathbb{E}[B_2] = 2$. We assume that an arriving customer is of class 1 (class 2) with probability $\alpha_1 = 8/12$ ($\alpha_2 = 4/12$). The weights of the DPS and RP are the same, namely, $g_1 = p_1 = 5$ and $g_2 = p_2 = 1$. We observe in Figure 6.3 that our approximation is rather accurate. In addition, as $w \rightarrow 0$, that is, as the coefficient of variation grows large, the performance of DPS is better than that of RP, both for our approximation as for the exact results. This is something we could expect, since as $w \rightarrow 0$, the second moment of class 2 tends to ∞ , and therefore the performance of RP (which is non-preemptive) is worse than DPS (which is time-sharing).

6.A Appendix

6.A.1 Proof of Lemma 6.1.2

We obtain the zeroth, first and second light-traffic derivatives for the probability generating function $\psi^{RP}(\lambda, \vec{z})$ of the stationary queue length vector at arbitrary time epochs.

From (5.16) it follows directly that the zeroth derivative in $\lambda = 0$ satisfies

$$\left(\psi^{RP}\right)^{(0)}(\lambda, \vec{z})\Big|_{\lambda=0} = 1. \quad (6.8)$$

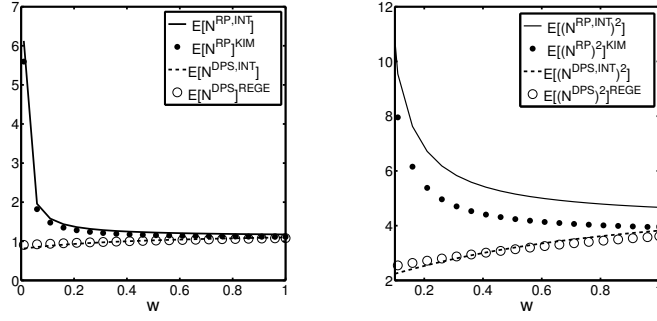


Figure 6.3: Scenario 3. First (left) and second moments (right) of the total number of customers in the system under DPS and RP for exponential service-time distribution.

Taking the derivative in (5.16) we obtain that the first derivative satisfies

$$\begin{aligned}
 (\psi^{RP})^{(1)}(0, \vec{z}) &= \frac{\partial \psi^{RP}(0, \vec{z})}{\partial \lambda} \\
 &= -\mathbb{E}[B] + \frac{1}{1 - \sum_{k=1}^K \alpha_k z_k} \sum_{i=1}^K \alpha_i z_i \left(\left(-\mathbb{E}[B] + \frac{p_i}{\alpha_i} \frac{\partial^2 r(\lambda, \vec{z})}{\partial \lambda \partial z_i} \right) \left(1 - B_i^* \left(\lambda - \lambda \sum_{k=1}^K \alpha_k z_k \right) \right) \right. \\
 &\quad \left. + \left(1 - \rho + \frac{p_i}{\alpha_i} \frac{\partial r(\lambda, \vec{z})}{\partial z_i} \right) \left(-B_i^* \left(\lambda - \lambda \sum_{k=1}^K \alpha_k z_k \right) \left(1 - \sum_{k=1}^K \alpha_k z_k \right) \right) \right) \Big|_{\lambda=0} \\
 &= -\mathbb{E}[B] + \sum_{i=1}^K \alpha_i \mathbb{E}[B_i] z_i \left(1 + \frac{p_i}{\alpha_i} \frac{\partial r(\lambda, \vec{z})}{\partial z_i} \Big|_{\lambda=0} \right) \\
 &= -\mathbb{E}[B] + \sum_{i=1}^K \alpha_i \mathbb{E}[B_i] z_i.
 \end{aligned} \tag{6.9}$$

In the last step we used that

$$\begin{aligned}
 \frac{\partial r(\lambda, \vec{z})}{\partial z_i} \Big|_{\lambda=0} &= \mathbb{E} \left[\frac{Q_i}{\sum_{k=1}^K Q_k p_k} \cdot \frac{z_1^{Q_1} \cdots z_K^{Q_K}}{z_i} \cdot \mathbf{1}_{(\sum_{k=1}^K Q_k > 0)} \right] \Big|_{\lambda=0} \\
 &= \mathbb{P}(\sum Q_k > 0) \Big|_{\lambda=0} \mathbb{E} \left[\frac{Q_i}{\sum_{k=1}^K Q_k p_k} \cdot \frac{z_1^{Q_1} \cdots z_K^{Q_K}}{z_i} \Big| \sum_{k=1}^K Q_k > 0 \right] \Big|_{\lambda=0} \\
 &= \rho \Big|_{\lambda=0} \mathbb{E} \left[\frac{Q_i}{\sum_{k=1}^K Q_k p_k} \cdot \frac{z_1^{Q_1} \cdots z_K^{Q_K}}{z_i} \Big| \sum_{k=1}^K Q_k > 0 \right] \Big|_{\lambda=0} \\
 &= 0,
 \end{aligned} \tag{6.10}$$

since the RP model is a work-conserving policy $\mathbb{P}(\sum Q_k > 0) = \rho$ is equal to the probability of the server being busy, which is independent of the scheduling policy. The

other term is finite since it satisfies

$$\begin{aligned}
& \mathbb{E} \left[\frac{Q_i}{\sum_{k=1}^K Q_k p_k} \cdot \frac{z_1^{Q_1} \cdots z_K^{Q_K}}{z_i} \middle| \sum_{k=1}^K Q_k > 0 \right] \bigg|_{\lambda=0} \\
&= \sum_{\substack{q_1=0, \dots, q_K=0 \\ \sum_{k=1}^K q_k > 0}}^{\infty} \frac{q_i}{\sum_{k=1}^K q_k p_k} \cdot \frac{z_1^{q_1} \cdots z_K^{q_K}}{z_i} \mathbb{P} \left(Q_1 = q_1, \dots, Q_K = q_K \middle| \sum_{k=1}^K Q_k > 0 \right) \bigg|_{\lambda=0} \\
&= \sum_{\substack{q_1=0, \dots, q_K=0 \\ \sum_{k=1}^K q_k = 1}}^{\infty} \frac{q_i}{\sum_{k=1}^K q_k p_k} \cdot \frac{z_1^{q_1} \cdots z_K^{q_K}}{z_i} \mathbb{P} \left(Q_1 = q_1, \dots, Q_K = q_K \middle| \sum_{k=1}^K Q_k > 0 \right) \bigg|_{\lambda=0} + o(\lambda^2) \\
&= \sum_{\substack{q_1=0, \dots, q_K=0 \\ \sum_{k=1}^K q_k = 1}}^{\infty} \frac{q_i}{q_i p_i} \cdot 1 \cdots 1 \cdot \mathbb{P} \left(\vec{Q} = e_i \middle| \sum_{k=1}^K Q_k > 0 \right) \bigg|_{\lambda=0} = \frac{\alpha_i}{p_i}, \tag{6.11}
\end{aligned}$$

due to

$$\begin{aligned}
& \mathbb{P} \left(\vec{Q} = e_i \middle| \sum_{k=1}^K Q_k > 0 \right) \bigg|_{\lambda=0} = \frac{\mathbb{P} \left(\vec{Q} = e_i \cap \sum_{k=1}^K Q_k > 0 \right)}{\mathbb{P} \left(\sum_{k=1}^K Q_k > 0 \right)} \bigg|_{\lambda=0} = \frac{\mathbb{P} \left(\vec{Q} = e_i \right)}{\mathbb{P} \left(\sum_{k=1}^K Q_k > 0 \right)} \bigg|_{\lambda=0} \\
&= \frac{\alpha_i \rho (1 - \rho)}{\rho} \bigg|_{\lambda=0} = \alpha_i
\end{aligned}$$

that follows from $A \cdot \mathbb{P}(\vec{Q} = \vec{0}) = B \cdot \mathbb{P}(\vec{Q} = e_i) + o(\lambda)$, where $A = \alpha_i \rho$ and $B = 1$ from [30, Equation 1] and again the fact that the RP model is a work-conserving policy.

The second derivative satisfies

$$\begin{aligned}
& (\psi^{RP})^{(2)}(\lambda, \vec{z}) \big|_{\lambda=0} = \frac{\partial^2 \psi^{RP}(\lambda, \vec{z})}{\partial \lambda^2} \bigg|_{\lambda=0} \\
&= \frac{1}{1 - \sum_{k=1}^K \alpha_k z_k} \sum_{i=1}^K \alpha_i z_i \left(\frac{p_i}{\alpha_i} \frac{\partial^3 r(\lambda, \vec{z})}{\partial \lambda^2 \partial z_i} \left(1 - B_i^* \left(\lambda - \lambda \sum_{k=1}^K \alpha_k z_k \right) \right) \right. \\
&\quad + 2 \left(-\mathbb{E}[B] + \frac{p_i}{\alpha_i} \frac{\partial^2 r(\lambda, \vec{z})}{\partial \lambda \partial z_i} \right) \left(-B_i^{*'} \left(\lambda - \lambda \sum_{k=1}^K \alpha_k z_k \right) \right) \left(1 - \sum_{k=1}^K \alpha_k z_k \right) \\
&\quad \left. + \left(1 - \rho + \frac{p_i}{\alpha_i} \frac{\partial r(\lambda, \vec{z})}{\partial z_i} \right) \left(-B_i^{*''} \left(\lambda - \lambda \sum_{k=1}^K \alpha_k z_k \right) \right) \left(1 - \sum_{k=1}^K \alpha_k z_k \right)^2 \right) \bigg|_{\lambda=0} \\
&= \sum_{i=1}^K \alpha_i z_i \left(2 \left(-\mathbb{E}[B] + \frac{p_i}{\alpha_i} \frac{\partial^2 r(\lambda, \vec{z})}{\partial \lambda \partial z_i} \right) \bigg|_{\lambda=0} \mathbb{E}[B_i] \right. \\
&\quad \left. - \left(1 + \alpha_i p_i \frac{\partial r(\lambda, \vec{z})}{\partial z_i} \right) \bigg|_{\lambda=0} \mathbb{E}[B_i^2] \left(1 - \sum_{k=1}^K \alpha_k z_k \right) \right)
\end{aligned}$$

$$= \sum_{i=1}^K \alpha_i z_i \mathbb{E}[B_i^2] \left(\sum_{k=1}^K \alpha_k z_k - 1 \right), \quad (6.12)$$

where in the last step we used

$$\begin{aligned} \frac{\partial^2 r(\lambda, \vec{z})}{\partial \lambda \partial z_i} \Big|_{\lambda=0} &= \frac{\partial \left(\rho \cdot \mathbb{E} \left[\frac{Q_i}{\sum_{k=1}^K Q_k p_k} \cdot \frac{z_1^{Q_1} \dots z_K^{Q_K}}{z_i} \mid \sum_{k=1}^K Q_k > 0 \right] \right)}{\partial \lambda} \Big|_{\lambda=0} \\ &= \mathbb{E}[B] \mathbb{E} \left[\frac{Q_i}{\sum_{k=1}^K Q_k p_k} \cdot \frac{z_1^{Q_1} \dots z_K^{Q_K}}{z_i} \mid \sum_{k=1}^K Q_k > 0 \right] \Big|_{\lambda=0} \\ &\quad + \rho \Big|_{\lambda=0} \cdot \frac{\partial \mathbb{E} \left[\frac{Q_i}{\sum_{k=1}^K Q_k p_k} \cdot \frac{z_1^{Q_1} \dots z_K^{Q_K}}{z_i} \mid \sum_{k=1}^K Q_k > 0 \right]}{\partial \lambda} \Big|_{\lambda=0} \\ &= \mathbb{E}[B] \frac{\alpha_i}{p_i}, \end{aligned}$$

which follows from Equation (6.11).

From Equations (6.8), (6.9) and (6.12) we obtain the result in Lemma 6.1.2 and conclude the proof.

6.A.2 Proof of Proposition 6.1.3

We obtain the interpolation approximation of the probability generating function of the queue length vector under the RP policy.

As explained in Section 2.3 we approximate $G(\lambda, \vec{z}^{(1-\rho)}) = \psi^{RP}(\lambda, \vec{z}^{(1-\rho)})$ by the polynomial

$$\hat{G}(\lambda, \vec{z}) = h_0(\vec{z}) + \lambda h_1(\vec{z}) + \lambda^2 h_2(\vec{z}) + \lambda^3 h_3(\vec{z}).$$

Undoing the normalisation, that is, for $f_\lambda^{-1} = \vec{z}^{(1-\rho)^{-1}}$, we have

$$\begin{aligned} \psi^{RP, INT}(\lambda, \vec{z}) &= \hat{G}(\lambda, \vec{z}^{(1-\rho)^{-1}}) \\ &= h_0(\vec{z}^{(1-\rho)^{-1}}) + \lambda h_1(\vec{z}^{(1-\rho)^{-1}}) + \lambda^2 h_2(\vec{z}^{(1-\rho)^{-1}}) + \lambda^3 h_3(\vec{z}^{(1-\rho)^{-1}}). \end{aligned}$$

Then, from the light-traffic conditions (2.12) we obtain $h_0(\vec{z}), h_1(\vec{z}), h_2(\vec{z})$. First we have, $\hat{G}(\lambda, \vec{z}^{(1-\rho)^{-1}}) \Big|_{\lambda=0} = \hat{G}(0, \vec{z}) = h_0(\vec{z})$. Together with (6.8) we obtain $h_0(\vec{z}) = 1$.

Second,

$$\begin{aligned}
& \left. \frac{d\hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{d\lambda} \right|_{\lambda=0} = \left. \frac{\partial \hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{\partial \lambda} \right|_{\lambda=0} + \sum_{i=1}^K \left. \frac{\partial \hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{\partial z_i} \right|_{\lambda=0} \cdot \left. \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \right|_{\lambda=0} \\
& = \left(h_1(\bar{z}^{(1-\rho)^{-1}}) + 2\lambda h_2(\bar{z}^{(1-\rho)^{-1}}) + 3\lambda^2 h_3(\bar{z}^{(1-\rho)^{-1}}) \right) \Big|_{\lambda=0} \\
& + \sum_{i=1}^K \left(\frac{dh_0(\bar{z}^{(1-\rho)^{-1}})}{dz_i} + \lambda \frac{dh_1(\bar{z}^{(1-\rho)^{-1}})}{dz_i} + \lambda^2 \frac{dh_2(\bar{z}^{(1-\rho)^{-1}})}{dz_i} + \lambda^3 \frac{dh_3(\bar{z}^{(1-\rho)^{-1}})}{dz_i} \right) \Big|_{\lambda=0} \\
& \quad \cdot \left. \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \right|_{\lambda=0} \\
& = h_1(\bar{z}) + \sum_{i=1}^K \left. \frac{dh_0(\bar{z}^{(1-\rho)^{-1}})}{dz_i} \right|_{\lambda=0} \left. \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \right|_{\lambda=0} \\
& = h_1(\bar{z}) + \sum_{i=1}^K \left. \frac{d(1)}{dz_i} \right|_{\lambda=0} \cdot \left. \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \right|_{\lambda=0} = h_1(\bar{z}).
\end{aligned}$$

Together with (6.9) we obtain $h_1(\bar{z}) = -\mathbb{E}[B] + \sum_{i=1}^K \alpha_i \mathbb{E}[B_i] z_i$.

Third,

$$\begin{aligned}
& \left. \frac{d^2 \hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{d\lambda^2} \right|_{\lambda=0} \\
& = \left. \frac{\partial^2 \hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{\partial \lambda^2} \right|_{\lambda=0} + \sum_{i=1}^K \left. \frac{\partial (\partial \hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}}) / \partial \lambda)}{\partial z_i} \right|_{\lambda=0} \cdot \left. \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \right|_{\lambda=0} \\
& + \sum_{i=1}^K \left(\left(\left. \frac{\partial (\partial \hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}}) / \partial z_i)}{\partial \lambda} \right|_{\lambda=0} + \left. \frac{\partial (\partial \hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}}) / \partial z_i)}{\partial z_i} \right|_{\lambda=0} \cdot \left. \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \right|_{\lambda=0} \right) \right. \\
& \quad \left. \cdot \left. \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \right|_{\lambda=0} + \left. \frac{\partial \hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{\partial z_i} \right|_{\lambda=0} \cdot \left. \frac{d^2(z_i^{(1-\rho)^{-1}})}{d\lambda^2} \right|_{\lambda=0} \right) \\
& = 2h_2(\bar{z}) + 2 \sum_{i=1}^K \frac{dh_1(\bar{z})}{dz_i} \cdot \left. \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \right|_{\lambda=0} = 2h_2(\bar{z}) + 2\mathbb{E}[B] \sum_{i=1}^K \alpha_i \mathbb{E}[B_i] z_i \ln(z_i).
\end{aligned}$$

Together with (6.12) we obtain

$$h_2(\bar{z}) = \frac{1}{2} \left(\sum_{i=1}^K \alpha_i z_i \mathbb{E}[B_i^2] \left(\sum_{k=1}^K \alpha_k z_k - 1 \right) - 2\mathbb{E}[B] \sum_{i=1}^K \alpha_i z_i \mathbb{E}[B_i] \ln(z_i) \right).$$

Finally, from Proposition 2.3.1 and noting that $G^{HT}(\bar{z})$ is equal to Equation (6.1), we conclude the proof.

6.A.3 Proof of Lemma 6.2.2

We obtain the zeroth and first light-traffic derivatives of the Laplace Transform of the waiting time of a class- l customer under DROS using the result presented in Theorem 5.4.1.

The zeroth derivative satisfies

$$\left(\widetilde{W}_k^{DROS}\right)^{(0)}(\lambda, u)\Big|_{\lambda=0} = [1 - \rho + T_l(u, \vec{1})]\Big|_{\lambda=0} = 1 + T_l(u, \vec{1})\Big|_{\lambda=0} = 1, \quad (6.13)$$

since from Equation (5.26) we obtain

$$T_l(u, \vec{1})\Big|_{\lambda=0} \left(1 + \sum_{i=1}^K \frac{p_i}{p_l} (1 - B_i^*(u))\right) = 0 \Rightarrow T_l(u, \vec{1})\Big|_{\lambda=0} = 0.$$

And the first derivative satisfies

$$\left(\widetilde{W}_k^{DROS}\right)^{(1)}(\lambda, u)\Big|_{\lambda=0} = -\mathbb{E}[B] + \frac{\partial T_l(u, \vec{1})}{\partial \lambda}\Big|_{\lambda=0} = -\mathbb{E}[B] + \left(\sum_{i=1}^K \alpha_i \frac{1 - B_i^*(u)}{u}\right), \quad (6.14)$$

since, again from Equation (5.26),

$$\begin{aligned} \frac{\partial T_l(u, \vec{1})}{\partial \lambda}\Big|_{\lambda=0} \left(1 + \sum_{i=1}^K \frac{p_i}{p_l} \mathbb{E}[Q_i^*]\Big|_{\lambda=0} (1 - B_i^*(u))\right) &= \sum_{i=1}^K \left(\alpha_i + p_i r(\lambda, \vec{1})\Big|_{\lambda=0}\right) \frac{1 - B_i^*(u)}{u} \\ \Rightarrow \frac{\partial T_l(u, \vec{1})}{\partial \lambda}\Big|_{\lambda=0} &= \left(\sum_{i=1}^K \alpha_i \frac{1 - B_i^*(u)}{u}\right), \end{aligned}$$

where $\mathbb{E}[Q_i^*]\Big|_{\lambda=0} = 0$ and $r(\lambda, \vec{1})\Big|_{\lambda=0} = 0$ from Equation (6.10).

6.A.4 Proof of Proposition 6.2.3

We obtain the interpolation approximatio for the waiting time under the RP policy. As explained in Section 2.3 we approximate $G(\lambda, (1 - \rho)u) = \widetilde{W}_k^{DROS}(\lambda, (1 - \rho)u)$ by the polynomial

$$\hat{G}(\lambda, u) = h_0(u) + \lambda h_1(u) + \lambda^2 h_2(u).$$

Undoing the normalisation, that is, for $f_\lambda^{-1} = (1 - \rho)^{-1}u$, we have

$$\begin{aligned} \widetilde{W}_k^{DROS, INT}(\lambda, u) &= \hat{G}(\lambda, (1 - \rho)^{-1}u) \\ &= h_0((1 - \rho)^{-1}u) + \lambda h_1((1 - \rho)^{-1}u) + \lambda^2 h_2((1 - \rho)^{-1}u). \end{aligned}$$

Then, from the light-traffic conditions (2.12) we obtain $h_0(u), h_1(u)$. First we have,

$$\hat{G}(\lambda, (1 - \rho)^{-1}u)\Big|_{\lambda=0} = \hat{G}(0, u) = h_0(u).$$

Together with (6.13) we obtain $h_0(u) = 1$.

Second,

$$\begin{aligned}
& \left. \frac{d\hat{G}(\lambda, (1-\rho)^{-1}u)}{d\lambda} \right|_{\lambda=0} \\
&= \left. \frac{d\hat{G}(\lambda, (1-\rho)^{-1}u)}{d\lambda} \right|_{\lambda=0} + \left. \frac{d\hat{G}(\lambda, (1-\rho)^{-1}u)}{du} \right|_{\lambda=0} \cdot \left. \frac{d((1-\rho)^{-1}u)}{d\lambda} \right|_{\lambda=0} \\
&= \left(h_1((1-\rho)^{-1}u) + 2\lambda h_2((1-\rho)^{-1}u) \right) \Big|_{\lambda=0} \\
&+ \left(\frac{dh_0((1-\rho)^{-1}u)}{du} + \lambda \frac{dh_1((1-\rho)^{-1}u)}{du} + \lambda^2 \frac{dh_2((1-\rho)^{-1}u)}{du} \right) \Big|_{\lambda=0} \cdot \left. \frac{d((1-\rho)^{-1}u)}{d\lambda} \right|_{\lambda=0} \\
&= h_1(u) + \left. \frac{dh_0((1-\rho)^{-1}u)}{du} \right|_{\lambda=0} \cdot \left. \frac{d((1-\rho)^{-1}u)}{d\lambda} \right|_{\lambda=0} \\
&= h_1(u) + \left. \frac{d(1)}{du} \right|_{\lambda=0} \cdot \left. \frac{d((1-\rho)^{-1}u)}{d\lambda} \right|_{\lambda=0} = h_1(u).
\end{aligned}$$

Together with (6.14) we obtain $h_1(u) = -\mathbb{E}[B] + \left(\sum_{i=1}^K \alpha_i \frac{1 - B_i^*(u)}{u} \right)$.

Finally, from Proposition 2.3.1 and noting that $G^{HT}(u)$ is equal to Equation (6.7), we conclude the proof.

Bibliography

- [1] E. Altman, K. Avrachenkov, and U. Ayesta. A survey on discriminatory processor sharing. *Queueing systems*, 53(1-2):53–63, 2006.
- [2] S. Asmussen. *Applied Probability and Queues*. Springer, 2003.
- [3] U. Ayesta, A. Izagirre, and I.M. Verloop. Heavy-traffic analysis of the discriminatory random-order-of-service discipline. *ACM SIGMETRICS Performance Evaluation Review*, 32(2):41–43, 2011.
- [4] M. Harchol Balter. Performance modeling and design of computer systems: Queueing theory in action. In *Cambridge University Press, Cambridge (UK)*, 2013.
- [5] P. Billingsley. *Convergence of Probability Measures*. Wiley, 1999.
- [6] S.C. Borst, R. Núñez-Queija, and A.P. Zwart. Sojourn time asymptotics in processor sharing queues. *Queueing Systems*, 53(1–2):31–51, 2006.
- [7] O.J. Boxma, S.G. Foss, J.-M. Lasgouttes, and R. Núñez-Queija. Waiting time asymptotics in the single server queue with service in random order. *Queueing Systems*, 46:35–73, 2004.
- [8] M. Bramson and J.G.Dai. Heavy traffic limits for some queueing networks. *Annals of Applied Probability*, 11:49–90, 2001.
- [9] D. Burman and D. Smith. An asymptotic analysis of a queueing system with markov-modulated arrivals. *Operations Research*, 34:105–119, 1986.
- [10] J.W. Cohen. The multiple phase service network with generalized processor sharing. *Acta Informatica*, 12:245–284, 1979.
- [11] J.L. Dorsman, R.D. van der Mei, and M. Vlasiou. Analysis of a two-layered network with correlated queues by means of the power-series algorithm. *Performance Evaluation*, 70:1072–1089, 2013.
- [12] R. Núñez-Queija E. Altman, K.E. Avrachenkov. Perturbation analysis for denumerable markov chains with application to queueing models. *Applied Probability Trust*, 36(3):839–853, 2004.

- [13] G. Fayolle, I. Mitrani, and R. Iasnogorodski. Sharing a processor among many job classes. *Journal of the ACM*, 27(3):519–532, 1980.
- [14] W. Feller. *An Introduction to Probability Theory and Its Applications, Vol. II*. Wiley, New York, 1971.
- [15] P.J. Fleming and B. Simon. Interpolation approximations of sojourn time distributions. *Operations Research*, 39(2):251–260, 1991.
- [16] S. Ben Fredj, T. Bonald, A. Proutiere, G. Regnie, and J. Roberts. Statistical bandwidth sharing: A study of congestion at flow level. In *SIGCOMM*, pages 111–122, 2001.
- [17] E. Gelenbe and I. Mitrani. *Analysis and Synthesis of Computer Systems*. Imperial College Press, 2010.
- [18] S. Grishechkin. On a relationship between processor sharing queues and Crump-Mode-Jagers branching processes. *Adv. Appl. Prob.*, 24(3):653–698, 1992.
- [19] V. Gupta, M. Harchol-Balter, K. Sigman, and W. Whitt. Analysis of join-the-shortest-queue routing for web server farms. In *Proceedings of Performance*, 2007.
- [20] M. Haviv and J. van der Wal. Equilibrium strategies for processor sharing and random queues with relative priorities. *Probability in the Engineering and Informational Sciences*, 11:403–412, 1997.
- [21] M. Haviv and J. van der Wal. Waiting times in queues with relative priorities. *Operations Research Letters*, 35:591–594, 2007.
- [22] A. Izagirre, U. Ayesta, and I.M. Verloop. Sojourn time approximations for a discriminatory-processor-sharing queue. *To appear in Transactions on Modeling and Performance Evaluation of Computing Systems*.
- [23] A. Izagirre, U. Ayesta, and I.M. Verloop. Sojourn time approximations in a multi-class time-sharing system. *IEEE INFOCOM*, 2014.
- [24] A. Izagirre, U. Ayesta, and I.M. Verloop. Heavy-traffic analysis of a multi-class queue with relative priorities. *Probability in Engineering and Informational Sciences*, 29(2):153–180, 2015.
- [25] A. Izagirre, U. Ayesta, and I.M. Verloop. Interpolation approximations for the steady-state distribution in multi-class resource-sharing systems. *Performance Evaluation*, <http://dx.doi.org/10.1016/j.peva.2015.06.005>, 2015.
- [26] A. Izagirre and A.M. Makowski. Light traffic performance under the power of two load balancing strategy: the case of server heterogeneity. *ACM SIGMETRICS Performance Evaluation Review*, 42:18–20, 2014.
- [27] F.P. Kelly. *Stochastic Networks and Reversibility*. Wiley, Chichester, 1979.

- [28] F.P. Kelly. Charging and rate control for elastic traffic. *European Transactions on Telecommunications*, 8:33–37, 1997.
- [29] D.G. Kendall. Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded markov chains. *Annals of Mathematical Statistics*, 24:338–354, 1953.
- [30] J. Kim. Queue length distribution in a queue with relative priorities. *Bull. Korean Math. Soc.*, 46:107–116, 2009.
- [31] J. Kim, J. Kim, and B. Kim. Analysis of the M/G/1 queue with discriminatory random order service policy. *Performance Evaluation*, 68(3):256–270, 2011.
- [32] J.F.C. Kingman. The single server queue in heavy traffic. *Proc. Cambridge Philos. Soc.*, 57:902–904, 1961.
- [33] J.F.C. Kingman. On queues in which customers are served in random order. *Proc. Cambridge Philos. Soc.*, 58:79–91, 1962.
- [34] J.F.C. Kingman. Queue disciplines in heavy traffic. *Mathematics of Operations Research*, 7(2):262–271, 1982.
- [35] L. Kleinrock. Analysis of a time-shared processor. *Naval Research Logistics Quarterly*, 11:59–73, 1964.
- [36] L. Kleinrock. Time-shared systems: A theoretical treatment. *Journal of the ACM*, 14(2):242–261, 1967.
- [37] L. Kleinrock. *Queueing Systems, vol. 2*. John Wiley and Sons, 1976.
- [38] X. Kang L. Ying, R. Srikant. The power of slightly more than one sample in randomized load balancing. *Infocom 2015*, 2015.
- [39] J.D.C. Little. A proof for the queueing formula: $L=\lambda w$. *Operations Research*, 9(3):383–387, 1961.
- [40] A.W. Marshall and I. Olkin. Inequalities: Theory of majorization and its applications. In *Academic Press, San Diego (CA)*, 1979.
- [41] M. Mitzenmacher. The power of two choices in randomized load balancing. *IEEE Transactions on Parallel and Distributed Systems*, 12(10):1094–1104, 2001.
- [42] A. Mukhopadhyay and R.R. Mazumdar. Analysis of load balancing in large heterogeneous processor sharing systems. *Submitted on 2013*.
- [43] R.L. Dobrushin N.D. Vvedenskaya and F.I. Karpelevich. Queueing system with selection of the shortest of two queues: An asymptotic approach. *Problems of Information Transmission*, 32:20–34, 1996.

- [44] R. Núñez-Queija. *Processor-Sharing Models for Integrated-Services Networks*. PhD thesis, Eindhoven University of Technology, 2000.
- [45] C. Palm. Waiting times with random served queue. *Tele1 (English edition; original 1938)*, 1–107, 1957.
- [46] K.M. Rege and B. Sengupta. Queue length distribution for the discriminatory processor sharing queue. *Operations Research*, 44(4):653–657, 1996.
- [47] M.I. Reiman and B. Simon. An interpolation approximation for queueing systems with Poisson input. *Operations Research*, 36:454–469, 1988.
- [48] M.I. Reiman and B. Simon. Light traffic limits of sojourn time distributions in Markovian queueing networks. *Stochastic Models*, 4:191–233, 1988.
- [49] M.I. Reiman and B. Simon. Open queueing systems in light traffic. *Oper. Res.*, 14:26–59, 1989.
- [50] J.W. Roberts. A survey on statistical bandwidth sharing. *Computer Networks*, 45:319–332, 2004.
- [51] A.L. Stolyar. Pull-based load distribution in large-scale heterogeneous service systems. *Queueing Systems*, 2015.
- [52] H.C. Tijms. *A First Course in Stochastic Models*. Wiley, 2003.
- [53] G. van Kessel, R. Núñez-Queija, and S.C. Borst. Differentiated bandwidth sharing with disparate flow sizes. In *Proceedings of IEEE INFOCOM*, 2005.
- [54] S. Varma and A.M. Makowski. Interpolation approximations for symmetric fork-join queues. *Performance Evaluation*, 20:245–265, 1994.
- [55] I.M. Verloop, U. Ayesta, and R. Núñez-Queija. Heavy-traffic analysis of a multiple-phase network with discriminatory processor sharing. *Operations Research*, 59(3):648–660, 2011.
- [56] J. Walrand. An introduction to queueing networks. In *Prentice-Hall, Englewood Cliffs, NJ*, 1988.
- [57] S.F. Yashkov. Processor sharing queues: Some progress in analysis. *Queueing Systems*, 2:1–17, 1987.
- [58] S.F. Yashkov. Mathematical problems in the theory of processor sharing queueing systems. *Journal of Soviet Mathematics*, 58:101–147, 1992.

Annex A

Interpolation des mesures de performance à l'état stationnaire

A.1 Introduction

Perdre du temps dans une file d'attente, les uns derrière les autres est aussi commun que désagréable. Nous attendons tous patiemment notre tour dans une file d'attente de supermarché, dans une salle d'urgence d'hôpital ou lorsque nous faisons appel à un service d'assistance téléphonique. À un niveau plus abstrait, ces files d'attente sont également rencontrés dans les systèmes de réseau de télécommunication. Par exemple, chaque e-mail que vous envoyez ou chaque fichier que vous téléchargez est divisé en différents paquets. Chaque paquet est ensuite envoyé à sa destination par le chemin le moins chargé afin d'éviter les files d'attente d'autres chemins déjà chargés par d'autres paquets.

La théorie des files d'attente est une approche mathématiques qui étudie les situations mentionnées ci-dessus (situation de création de file d'attente non souhaitées), et a deux objectifs principaux. D'une part, cela permet d'étudier les performances d'un système. D'autre part, cela permet de trouver la meilleure politique afin d'optimiser les performances de celui-ci. Le nombre de clients dans le système, aussi appelé, longueur de la file d'attente, et le temps de séjour (i.e. le temps total passé dans le système) font partis des métriques les plus fréquemment considérées dans la littérature de l'évaluation de performance.

Dans cette thèse, nous nous concentrons sur le premier objectif et nous étudions les performances d'une file d'attente multi-classes à serveur unique appliquant une politique de partage du processeur discriminatoire (DPS - *Discriminatory Processor Sharing*) et une politique de priorités relatives, puis le modèle de serveurs parallèles appliquant la politique *power-of-two*. Des détails complémentaires sur ces modèles sont exposés dans la section A.1.1.

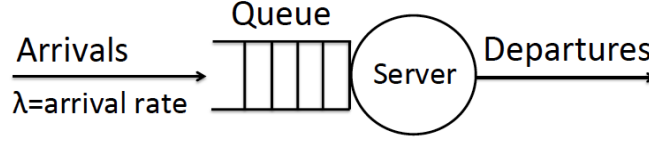


Figure A.1: Illustration de la file d'attente dans le cas d'un serveur unique.

L'idéal serait de pouvoir caractériser les métriques de performance d'une manière exacte. Cependant, en raison de la difficulté que cela représente, cela est souvent impossible à obtenir. Par conséquent, la théorie des files d'attente a montré un grand intérêt pour un rapprochement des indicateurs de performance qui analysent le système sous des régimes limitant, comme la décomposition par échelle de temps, les files asymptotique, limites à fort trafic, limites fluides, etc. Dans cette thèse, nous utilisons une technique d'approximation pour l'interpolation de faible et fort trafic dans le but d'en tirer une bonne approximation pour la distribution de la longueur de la file d'attente à l'état stationnaire, le temps d'attente et le temps de séjour. Plus précisément, premièrement une étude de la performance est proposée dans le cas d'un trafic faible, ce qui signifie que le taux d'arrivée tend vers 0. Deuxièmement, la performance est étudiée dans le cas d'un trafic fort, ce qui signifie que le système est proche de la saturation. Puis, les caractérisations des régimes de faible et fort trafics sont combinés afin de développer une approximation d'interpolation qui vise à estimer la performance pour n'importe quelle charge.

A.1.1 Modèles de files d'attente

Dans cette section nous introduisons les modèles de file d'attente étudiés au cours de ce Doctorat.

Système à serveur unique

La file d'attente d'un serveur unique classique est décrite par la Figure A.1. Il y a un processus d'arrivée de taux λ , de sorte que λ^{-1} est la moyenne de temps entre les instants d'arrivée des clients. À son arrivée, le client devra soit attendre dans la file d'attente, soit il sera directement traité. Cela dépendra de la politique d'ordonnancement appliquée. Nous supposons que la capacité du serveur est de 1. Les besoins de service sont désignés par la variable aléatoire B . Ceci représente le temps que le client va passer dans le système si le serveur lui attribue toute sa capacité.

Une hypothèse courante est que les instants d'inter-arrivées sont indépendants et identiquement distribués (i.i.d), les exigences de traitement sont i.i.d., les séquences des temps d'inter-arrivées et les exigences de service sont indépendantes. Ce modèle de file d'attente est connu sous le nom $G/G/1$, où G représente la distribution générale. Cette notation a été introduite par Kendall [13]. Dans cette thèse, nous nous concentrons sur

la file d'attente $M/G/1$, où M signifie Markov ou sans mémoire, ce qui signifie que les arrivées des clients suivent un processus de Poisson, ou en d'autres termes, que les temps inter-arrivées sont exponentiellement distribués.

La charge de travail totale au niveau du système est indépendante de la politique d'ordonnancement *work-conserving* qui est utilisée. Un système de type *work-conserving* fonctionne à sa vitesse maximum à chaque fois qu'il doit être utilisé. La stabilité est également indépendante de la politique d'ordonnancement *work-conserving* utilisée. La file d'attente sera stable aussi longtemps que $\lambda \mathbb{E}[B]$ est strictement inférieur à 1. En revanche, la longueur de la file d'attente dépend de la politique utilisée.

Les systèmes mono-serveur considérés dans cette thèse prennent en compte K différentes classes de clients. Les clients de classe- k arrivent selon un processus de Poisson de taux λ_k et donc $\alpha_k := \lambda_k / \lambda$ désigne la fraction des arrivées de la classe- k . $B_k, k = 1, \dots, K$ représente les besoins d'un service de classe- k . Naturellement, la charge des clients de classe- k est notée $\rho_k := \lambda_k \mathbb{E}[B_k]$, de sorte que $\rho := \sum_{k=1}^K \rho_k$ est la charge totale du trafic.

Nous allons maintenant présenter les politiques de *work-conserving* qui jouent un rôle très important dans cette thèse.

Processeur partagé

Dans le cas de la politique de partage du processeur (PS - *Processor Sharing* en anglais), la capacité du serveur est partagée entre les clients de manière égale. Plus précisément, si à l'instant t , $n(t)$ clients sont présents dans le système, alors chaque client est traité à un taux de $1/n(t)$. Nous nous référons aux articles [37] et [25] qui proposent un aperçu général de la littérature de ce domaine

Dans le cas d'un taux d'arrivée suivant une loi de Poisson, la distribution stationnaire du nombre de clients dans le système ne dépend que de la distribution des besoins de service en moyenne, et non par d'autres statistiques d'ordres supérieurs. Plus précisément, la répartition de la longueur de la file d'attente à l'état stationnaire a une distribution géométrique de paramètre ρ , i.e., la probabilité d'avoir n clients dans le système est égale à $(1-\rho)\rho^n, n = 0, 1, \dots$, [33]. Suivant la loi de Little, [22], l'insensibilité de la distribution de la longueur de la file d'attente traduit l'insensibilité du temps de séjour moyen.

Contrairement à la simple distribution géométrique de la longueur de la file d'attente, la distribution du temps de séjour n'a pas de caractérisation simple. Initié par l'analyse de Kleinrock d'une file d'attente de type $M/M/1$ PS [19, 20], de nombreuses études dans la littérature ont mis l'accent sur l'analyse des conditions (besoins de service) de temps de séjour. Pour les résultats concernant la distribution des temps de séjour dans une file d'attente $M/M/1$ PS, nous nous référons au résumé dans [2] et aux articles [38, 39]

pour les résultats concernant les files d'attente $M/G/1$ PS.

Pour une file d'attente multi-classes PS, une distribution géométrique de la longueur de la file d'attente est également applicable. Dans l'hypothèse d'un processus d'arrivée de Poisson, comme étudié dans [5, 12], la probabilité d'avoir n_k clients de classe- k dans le système, $k = 1, \dots, K$, est égal à

$$(1 - \rho) \cdot \frac{(n_1 + \dots + n_K)!}{n_1! \cdot \dots \cdot n_K!} \prod_{k=1}^K \rho_k^{n_k}.$$

Une file d'attente PS a désormais un rôle de premier plan dans le domaine de l'évaluation de performance, incluant une variété de mécanismes d'allocation de ressources [21, 12, 38]. Au cours des dernières années, ce domaine a de nouveau été reconnu comme un bon moyen d'abstraction pour la modélisation de la performance des flux réseau pour les protocoles de partage de bande passante dans les réseaux à commutation de paquets, comme TCP en particulier [8, 31].

Processeur partagé discriminatoire

Le modèle de file d'attente à processeur partagé discriminatoire (DPS - *Discriminatory Processor Sharing* en anglais) est un modèle de file d'attente polyvalent permettant de modéliser de manière pertinente les différenciations des types services. Il a été introduit par Kleinrock dans [20]. Il est une extension multi-classes de la politique de PS. Encore une fois, nous supposons qu'il y a K classes de clients, et ces différentes classes sont pondérées par des poids positifs, $g_1, \dots, g_K$. La capacité de service est partagée simultanément parmi tous les clients présents en respectant la proportionnalité des poids attribués à la classe à laquelle ils appartiennent. Plus précisément, étant donné K classes de clients, si au temps t , il y a $n_k(t)$ clients de classe- k présents dans le système, $k = 1, \dots, K$, alors en appliquant la politique DPS, chaque client de classe- k est servi à un taux égal à

$$\frac{g_k}{\sum_{j=1}^K g_j n_j(t)}.$$

Malgré la simplicité de la description du modèle et le fait que les propriétés de la file d'attente PS égalitaire sont assez bien comprises et que des solutions analytiques existent, l'analyse du DPS s'est avérée extrêmement difficile sans caractérisation de solution analytique. Nous nous référons à l'article [1] pour un aperçu approfondi de la littérature sur le modèle DPS. Ci-dessous, nous présentons les résultats liés au modèle DPS.

Dans [27] Rege et al. ont établi que la fonction de génération du vecteur de la longueur de la file d'attente satisfait une équation différentielle pour *des distributions de temps de service exponentielles*. À partir de cette équation, les auteurs montrent en outre que les instants peuvent être déterminés numériquement comme une solution d'un système d'équations.

Le régime à fort trafic analyse le système quand il est proche de la saturation. Dans [27], en supposant que les besoins de services suivent une distribution exponentielle,

Rege et al. ont mis en place une méthode d'effondrement de l'espace d'état pour la distribution de la longueur de la file d'attente dans un régime de fort trafic. Dans [35], Verloop et al. ont généralisé le résultat de *phase-type* distributions, qui est, dans la limite du vecteur de la longueur de la file d'attente à l'échelle est distribué comme le produit d'une variable aléatoire exponentielle distribué et un vecteur déterministe. Soit $N_k^{DPS}, k = 1, \dots, K$, représentent le nombre de clients de classe- k à l'état stationnaire, alors que $\rho \uparrow 1$

$$(1 - \lambda \mathbb{E}[B])(N_1^{DPS}, \dots, N_K^{DPS}) \xrightarrow{d} Y \cdot \left(\frac{\alpha_1 \mathbb{E}[B_1]}{g_1}, \frac{\alpha_2 \mathbb{E}[B_2]}{g_2}, \dots, \frac{\alpha_K \mathbb{E}[B_K]}{g_K} \right),$$

avec $\xrightarrow{d}$ la convergence de la distribution et Y une variable aléatoire exponentielle de moyenne $\mathbb{E}[B^2] / \mathbb{E}[B] \sum_{k=1}^K \alpha_k \mathbb{E}[B_k^2] / g_k$.

Dans [6], Fayolle et al. propose une étude de la moyenne conditionnelle (des besoins de service) et inconditionnelle du temps de séjour. Pour des distributions de temps de service générales, les auteurs ont obtenu le temps moyen de séjour conditionnelle comme une solution du système d'équations intégral-différentielles suivant :

$$\begin{aligned} \frac{\partial \overline{S}_k^{DPS}(\lambda, b)}{\partial b} &= 1 + \lambda \sum_{j=1}^K \int_0^\infty \alpha_j \frac{g_j}{g_k} \frac{\partial \overline{S}_j^{DPS}(\lambda, y)}{\partial y} \left[1 - F_j \left(y + \frac{g_j}{g_k} b \right) \right] dy \\ &\quad + \lambda \int_0^b \frac{\partial \overline{S}_k^{DPS}(\lambda, y)}{\partial y} \sum_{j=1}^K \alpha_j \frac{g_j}{g_k} \left[1 - F_j \left(\frac{g_j}{g_k} (b - y) \right) \right] dy, \end{aligned}$$

avec $\overline{S}_k^{DPS}(\lambda, b)$ la moyenne conditionnelle (pour un besoin de service b) du temps de séjour moyen d'un client arbitraire de classe- k avec la politique DPS. En outre, les auteurs fournissent une analyse approfondie pour besoins de service exponentiellement distribués. Toutefois, excepté pour le cas de deux classes, il n'existe pas d'expression analytique et une analyse numérique est nécessaire afin de calculer les temps de séjour moyens.

Priorités relatives

Les priorités relatives (RP - *Relative priorities* en anglais) est une file d'attente multi-classes. Cela fournit un environnement approprié pour modéliser la différenciation des services dans des systèmes non-préemptifs. Le service est non-préemptif et une fois le service terminé, la probabilité que le prochain client de classe- k soit servi est :

$$\frac{n_k p_k}{\sum_{j=1}^K n_j p_j}, \tag{A.1}$$

avec, $p_j > 0, j = 1, \dots, K$, les poids des différentes classes, et n_j le nombre de clients de classe- j à la date de décision. Une fois qu'une classe est choisie pour être

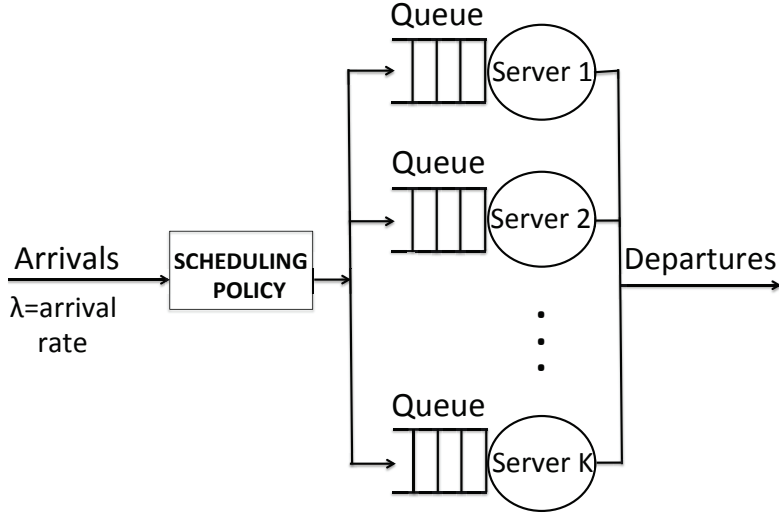


Figure A.2: A parallel-server model with K servers.

servie, une politique d'ordonnancement intra-classes détermine quel client appartenant à celle-ci sera servi.

Un cas particulier du modèle étudié est lorsque la politique d'ordonnancement intra-classes est uniformément aléatoire, alors, un client est sélectionné au hasard parmi les classes. Ce modèle a été proposé dans [9] et il est considéré DROS (*Discriminatory Random Order of Service* en anglais). Au cours des dernières années, plusieurs études intéressantes ont été publiées à propos de DROS, [10, 14, 15]. Des expressions pour le temps moyen d'attente d'un client suivant la classe à laquelle il appartient ont été obtenues par [10]. Dans [14] les auteurs dérivent des équations différentielles que la transformée des longueurs des files d'attente jointes doit satisfaire. Également, dans [15] les auteurs dérivent des équations différentielles que la transformée du temps d'attente à l'état stationnaire doit satisfaire. Cela permet aux auteurs de trouver les instants où la longueur de la file d'attente est une solution d'un système d'équations linéaires.

Dans le cas de classe unique, $K = 1$, DROS peut se réduire à la politique ROS (*Random Order of Service* en anglais) bien connue. Des articles incontournables sur la politique ROS sont par exemple [17, 18, 26]. La transformée de Laplace pour la distribution des temps d'attente a été obtenue dans [17]. Plus récemment, les auteurs de [3] ont obtenu la distribution des temps d'attente dans un cas de trafic fort pour certaines exigences de service ayant une variance infinie. En outre, des files d'attente asymptotiques de temps d'attente ont été obtenues.

A.1.2 Modèle de serveur parallèle

Le modèle de serveur parallèle se compose de K serveurs *hétérogènes* (i.e., chacun ayant une vitesse et une capacité différentes) pouvant travailler simultanément (Figure ??). Un client entrant dans le système est affecté à l'un des serveurs K selon la politique *power-of-two*, notée $SQ(d)$ avec $d = 2$: Le client entrant sélectionne au hasard 2 serveurs dans l'ensemble des K serveurs disponibles. La politique *Join-the-shortest-queue* (JSQ) est ensuite utilisée entre ces deux serveurs; les liens sont coupés au hasard. La politique *first-come-first-served* (FCFS) est ensuite appliquée au sein de chaque file d'attente.

Il est à noter que $SQ(1)$ est équivalent à une assignation aléatoire de serveur, tandis que $SQ(K)$ est équivalent à la politique JSQ entre tous les serveurs du système.

Le système *power-of-two* a été étudié depuis un certain temps; voir le bref historique dans [23, Section 1.1]. L'analyse du modèle parallèle serveur est rendu difficile en raison du couplage entre les files d'attente induits par les utilisateurs locaux de JSQ. Il en est ainsi même lorsque les clients arrivent selon un processus de Poisson, les serveurs sont *homogènes* (i.e., vitesse ou capacité identiques) et les exigences de service sont réparties de façon exponentielle. Dans ce cadre, Mitzenmacher [23] et Vvedenskaya et al. (avec $d = 2$) [24], ont étudié le système de limitation obtenu en laissant le nombre de serveurs tendre vers l'infini. Leurs résultats indiquent une amélioration substantielle de la performance du cas $d = 2$ sur $d = 1$ sans l'*overhead* de JSQ, tandis que $d = 3$ est seulement un facteur constant meilleur que $d = 2$. C'est de là que vient la terminologie "*power-of-two*". Plus récemment, Stolyar [32] a analysé un modèle serveurs parallèles hétérogènes utilisant une politique dénommée *pull-based*. En suivant cette politique chaque serveur envoie un "*pull-message*" au routeur quand il devient inactif; le routeur attribue un client à un serveur, si possible suivant un "*pull-message*" choisi au hasard, ou sinon le serveur est choisi au hasard. Stolyar a prouvé que cette politique *pull-based* fournit en outre des améliorations substantielles pour la politique *power-of-two*.

A.2 Motivation pour l'approximation de l'interpolation

Comme mentionné précédemment, la technique d'interpolation est une combinaison des caractérisations des trafics faible et fort. La technique impose le rapprochement pour être exact dans le trafic de lumière, à savoir, lorsque la charge ρ est proche de zéro et dans le cas d'un trafic fort, à savoir, lorsque la charge ρ du système est proche de la saturation. La technique d'interpolation fournit une approximation pour des valeurs intermédiaires de la charge. Un avantage important de l'approximation est qu'elle permet de fournir un aperçu de la dépendance de la performance sur les paramètres du système (poids, distributions du temps de service, etc.), et nous pensons que cela sera utile lors de leur mise en œuvre.

À titre d'exemple, dans la figure A.3 nous avons tracé la forme exacte et l'approximation de l'interpolation pour le nombre total moyen de clients dans un système mono-serveur utilisant la politique PS (à gauche) et DPS (à droite), notées respectivement N^{PS} et N^{DPS} . Pour l'expression exacte PS est disponible et donnée par $\mathbb{E}[N^{PS}] = \frac{\rho}{1 - \rho}$.

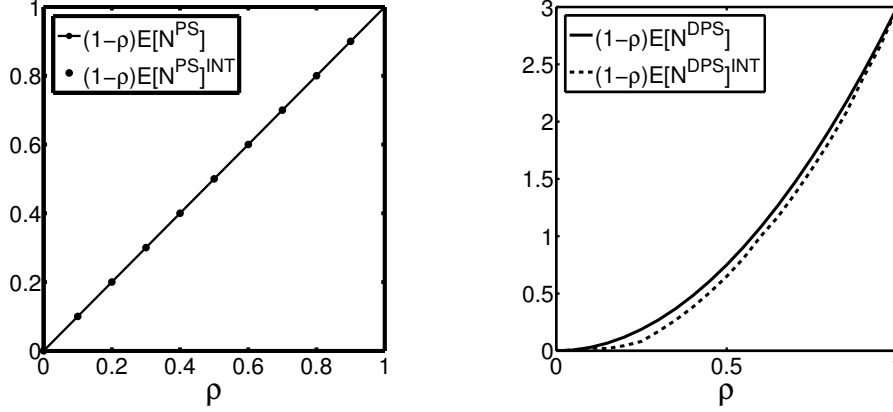


Figure A.3: Précision de l'approximation de interpolation pour un nombre total de clients pour le PS (à gauche) et DPS (à droite).

En raison de l'imposition de la technique d'interpolation sur l'exactitude des valeurs extrêmes, l'approximation de l'interpolation est donc exacte, voir la figure A.3 gauche. Pour DPS nous observons que l'approximation de l'interpolation est exacte sur les valeurs extrêmes de la charge, $\rho \downarrow 0$ et $\rho \uparrow 1$, mais pas pour des valeurs de charge intermédiaires, même si elle reste précise, voir Figure A.3 droite.

Nous allons maintenant décrire l'application de la technique d'interpolation aux trois modèles étudiés dans cette thèse.

Le modèle de serveurs parallèles utilisant la politique *power-of-two* se révèle difficile à analyser dû à des dépendances entre les files d'attente. Afin d'avoir des files d'attente asymptotiquement indépendantes, il est souvent considéré dans la littérature que le nombre de serveurs peut croître à l'infini. Sans l'hypothèse d'un nombre de serveurs infini, les résultats sous forme d'expressions explicites sont rares. Motivé par ce manque de résultats, nous avons analysé le temps de séjour moyen de la politique *power-of-two* pour un trafic faible. La méthode suivie pour construire les dérivés du trafic faible est celle proposée dans [30]. À de notre connaissance, il n'y a pas d'article dans la littérature qui étudie la politique *power-of-two* pour une régime de fort trafic. C'est pour cela que nous n'avons fait d'études d'approximation de l'interpolation de la politique *power-of-two*.

L'analyse du modèle DPS est connue pour être extrêmement difficile. Par exemple, comme indiqué dans la section A.1.1, les résultats sur une métrique de base importante comme le temps de séjour moyen dans le système ont seulement été dérivés d'une manière très implicite ou dans certains régimes limités (décomposition de l'échelle de temps, fort trafic, surcharge, etc.). Pour le vecteur des longueurs de file d'attente, des résultats analytiques sont obtenus uniquement pour des besoins de services exponentiels. Motivé par la difficulté de l'analyse du système sous forme exacte, nous dérivons une approximation du trafic faible pour la fonction génératrice jointe de la longueur de la file d'attente commune et pour la fonction de probabilité de distribution de la durée de séjour. La méthode

suivie pour générer les dérivés du trafic faible est celle proposés dans [30], comme ce fut également le cas pour le modèle de serveurs parallèles. Combinant nos résultats pour le trafic faible avec les résultats obtenus à fort trafic dans [27, 35], nous avons obtenu les approximations d'interpolation pour la distribution des métriques de performance.

Comme mentionné dans le chapitre précédent, pour des besoins de service généraux, Kim et al. [14, 15] dérivent respectivement pour les modèles RP et DROS, des équations différentielles que la transformation des longueurs de file d'attente communes et le temps d'attente à l'état stationnaire doivent satisfaire. Par conséquent, les approximations du trafic faible et les résultats à fort trafic pour les mesures mentionnées ci-dessus peuvent être obtenues directement à partir de ces équations différentielles.

Nous allons ensuite combiner les résultats ainsi obtenus afin de développer des approximations d'interpolation qui visent à obtenir la performance du système pour n'importe quelles valeurs de charge.

A.3 Approximation d'interpolation

Dans cette section, nous décrivons comment les métriques de performance peuvent être élaborées pour un régime de trafic faible, nous proposons une discussion à propos du régime à fort trafic et nous présentons comment l'interpolation des trafics faible et fort est effectuée. Cette approche a été lancée dans une série de documents par Reiman & Simon, voir par exemple [30], dans lequel l'objectif était le nombre moyen de clients ou le temps moyen de séjour, et étendu à la distribution du temps de séjour des files d'attente de Markov dans [7] et [29]. Cette technique a été appliquée avec succès à des modèles de processeurs partagés, exécutions parallèles, etc. [4, 34, 11, 7, 29]. Comme mentionné dans la section précédente, la technique d'interpolation impose que l'approximation soit exacte pour une trafic faible, correspondant à un taux d'arrivée λ proche de zéro, dans le cas d'un trafic fort, correspondant à un taux d'arrivée proche de la saturation, et vise à obtenir valeur de performance précise (si la valeur exacte ne peut être obtenue) pour n'importe quelle valeur de λ . Nous avons vu pour les modèles DPS et RP, que l'approximation se révèle être exacte pour tout λ avec des paramétrages particuliers du système.

Soit $G(\lambda, \vec{y})$ la métrique de performance à laquelle nous nous intéressons, qui dépend du taux d'arrivée λ et du vecteur $\vec{y}$. L'interprétation de la fonction G et du vecteur $\vec{y}$ vont changer en fonction de la métrique que nous voulons approximer. Dans cette thèse, la métrique G représentera l'un des choix suivants :

(i) la fonction génératrice du vecteur de la longueur de la file d'attente $(N_1, \dots, N_K)$, désignée par $\psi(\lambda, \vec{z}) := \mathbb{E} [z_1^{N_1} \dots z_K^{N_K}]$. Ainsi $\vec{y} = \vec{z}$,

(ii) la fonction de répartition complémentaire du temps d'attente conditionnel d'un client de classe- k notée $W(\lambda, b, x) := \mathbb{P}[W_k(b) > x]$, avec $W_k(b)$ le temps de séjour conditionnel d'un client de classe- k . Ainsi $\vec{y} = (b, x)$,

(iii) La transformé de Laplace-Stieltjes (*Laplace-Stieltjes Transform* - LST) du temps d'attente d'un client de classe- k notée $\widetilde{W}_k(\lambda, u) := \mathbb{E} \left[e^{-u(1-\lambda\mathbb{E}[B])W_k} \right]$, avec W_k le temps d'attente inconditionnel d'un client de classe- k . Ainsi $\vec{y} = u$.

Dans la section A.3.1 nous allons caractériser $G(\lambda, \vec{y})$ pour un régime trafic faible, $\lambda \downarrow 0$. Dans la section A.3.2 le régime de trafic fort est discuté. Enfin, la Section A.3.3 présente le paramétrage général pour l'approximation de l'interpolation.

A.3.1 Analyse du trafic faible

Le régime de faible trafic concerne la performance du système lorsque le taux d'arrivée λ tend vers zéro, ou en d'autres termes, lorsque la quantité de travail arrivant au système par unité de temps se rapproche de zéro.

L'analyse du trafic faible a été appliquée avec succès dans plusieurs articles, par exemple [34, 11, 29].

La métrique $G(\lambda, \vec{y})$ est approchée par une série de Taylor avec $\lambda = 0$. En supposant que les premières N dérivées de $G(\lambda, \vec{y})$ avec $\lambda = 0$ existent, nous avons l'approximation suivante pour $G(\lambda, \vec{y})$ quand λ est proche de zéro :

$$G^{LT}(\lambda, \vec{y}) := G^{(0)}(0, \vec{y}) + \lambda G^{(1)}(0, \vec{y}) + \cdots + \frac{\lambda^n}{n!} G^{(n)}(0, \vec{y}). \quad (\text{A.2})$$

Nous allons nous référer à cela pour l'approximation du trafic faible d'ordre n . Avec $G^{(0)}(0, \vec{y}) := G(0, \vec{y})$, que nous considérons comme la zéroième dérivée du trafic faible. En outre, $G^{(m)}(0, \vec{y})$, $m = 1, 2, \dots, n$, désigne la dérivée m -ième avec $\lambda = 0$, i.e., $G^{(m)}(0, \vec{y}) := \left. \frac{\partial^m G(\lambda, \vec{y})}{\partial \lambda^m} \right|_{\lambda=0}$. Le choix de la valeur de n dépendra du compromis entre la traçabilité et la précision visées.

Lorsqu'une caractérisation de $G(\lambda, \vec{y})$ existe, par exemple en termes d'équation différentielle qu'elle a besoin de satisfaire, celle-ci peut être exploitée pour mener à bien l'analyse des régime de faible et de fort trafics. Cependant, en général, une caractérisation de $G(\lambda, \vec{y})$ pourrait ne pas exister. Dans ce cas, nous utilisons les résultats de [30, Section 3.] et [36, Chapter 6.3.] où il est montré comment obtenir les dérivées du trafic faible d'ordre arbitraire m sous condition d'une recevabilité générale. Suite à la discussion dans [30, Appendix A] nous faisons l'hypothèse suivante sur les exigences de service B_k :

$$\mathbb{E}[e^{\eta B}] = \sum_{n=0}^{\infty} \frac{\eta^n}{n!} \mathbb{E}[B^n] < \infty, \quad (\text{A.3})$$

pour certains $\eta > 0$. Cette condition d'un moment exponentiel fini exige que tous les moments de l'exigence de service B soient également finis. L'Équation (A.3) implique l'admissibilité; cela est probablement plus fort que nécessaire, mais le but ici est de fournir un cadre idéal au sein duquel les calculs peuvent être justifiés.

Les expressions comme obtenues dans [30, 36] pour la zéroième, première et deuxième dérivées du trafic faible sont données dans la proposition ci-dessous. Nous voulons souligner que nous utilisons ces expressions pour obtenir les dérivées du trafic faible pour le modèle de serveurs parallèles et la file d'attente DPS. Dans le cas du modèle de RP, les dérivées du trafic faible sont obtenues directement à partir des équations différentielles de [15].

Proposition A.3.1. [30, Section 3], [36, Chapter 6.3]

Soit $A(s, t)$ représente le nombre d'arrivées dans l'intervalle $[s, t]$ en plus des clients étiquetés supposés arriver à l'instant 0, et $G(\lambda, \vec{y}|A)$ désigne la métrique de performance conditionnée sur A . Puis la zéroième, première et deuxième dérivées du trafic faible peuvent être écrites comme

$$G^{(0)}(0, \vec{y}) = G\left(0, \vec{y} \middle| A(-\infty, \infty) = 0\right), \quad (\text{A.4})$$

$$G^{(1)}(0, \vec{y}) = \int_{-\infty}^{\infty} \left(G\left(0, \vec{y} \middle| A(-\infty, \infty) = 1, \tau_1 = t\right) - G\left(0, \vec{y} \middle| A(-\infty, \infty) = 0\right) \right) dt \quad (\text{A.5})$$

et

$$\begin{aligned} G^{(2)}(0, \vec{y}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(G\left(0, \vec{y} \middle| A(-\infty, \infty) = 2, \tau_1 = t', \tau_2 = t''\right) - G\left(0, \vec{y} \middle| A(-\infty, \infty) = 1, \tau_1 = t'\right) \right. \\ &\quad \left. - G\left(0, \vec{y} \middle| A(-\infty, \infty) = 1, \tau_1 = t''\right) + G\left(0, \vec{y} \middle| A(-\infty, \infty) = 0\right) \right) dt' dt'' \quad (\text{A.6}) \end{aligned}$$

avec $\tau_i, i = 1, 2$, l'instant d'arrivée du i -ème client.

Nous fournissons maintenant un raisonnement permettant d'obtenir la zéroième et les premières dérivées du trafic faible. Ceci est basé sur l'analyse de J. Walrand dans [36, Chapter 6.3]. Les dérivées du deuxième ordre et des ordres supérieurs du trafic faible peuvent être obtenues d'une manière similaire.

Prenons un système qui commence à l'instant $-Z$ et que s'exécute jusqu'à l'instant T , avec $Z, T > 0$ donnés. Soit $G(\lambda, \vec{y}, -Z, T)$ désigne le terme qui nous intéresse d'approximer avec $\lim_{Z, T \rightarrow \infty} G(\lambda, \vec{y}, -Z, T) = G(\lambda, \vec{y})$. Soit $A(s, t)$ désigne le nombre d'arrivées dans l'intervalle $[s, t]$ en plus des client étiquetés qui sont supposés arriver à l'instant 0. Tout au long de cette section, nous supposons que les limites (par rapport à Z et T) et les attentes peuvent être inter-changées. Nous avons alors

$$G(\lambda, \vec{y}, -Z, T) = \sum_{a=0}^{\infty} G\left(\lambda, \vec{y}, -Z, T \middle| A(-Z, T) = a\right) \cdot \frac{(\lambda(T+Z))^a}{a!} e^{-\lambda(T+Z)}, \quad (\text{A.7})$$

avec $G(\lambda, \vec{y}, -Z, T | A(-Z, T) = a)$ conditionné par le fait qu'il y ait exactement a arrivées. L'évaluation de cela avec $\lambda = 0$ donne

$$G(\lambda, \vec{y}, -Z, T) \Big|_{\lambda=0} = G(\lambda, \vec{y}, -Z, T | A(-Z, T) = 0), \quad (\text{A.8})$$

et maintenant en prenant la limite $Z, T \rightarrow \infty$, on obtient la zéroième dérivée du trafic faible

$$G^{(0)}(0, \vec{y}) := \lim_{Z, T \rightarrow \infty} G(\lambda, \vec{y}, -Z, T) \Big|_{\lambda=0} = G(0, \vec{y}, -Z, T | A(-\infty, \infty) = 0)$$

avec la deuxième égalité qui découle de (A.8).

Ensuite, en considérant la dérivée par rapport à λ dans l'équation A.7 et en évaluant celle-ci avec $\lambda = 0$, cela donne

$$\begin{aligned} & \frac{\partial}{\partial \lambda} G(\lambda, \vec{y}, -Z, T) \Big|_{\lambda=0} \\ &= -G(\lambda, \vec{y}, -Z, T | A(-Z, T) = 0) \cdot (T + Z) + G(\lambda, \vec{y}, -Z, T | A(-Z, T) = 1) \cdot (T + Z) \\ &= \int_{-Z}^T \left(G(\lambda, \vec{y}, -Z, T | A(-Z, T) = 1, \tau_1 = t) - G(\lambda, \vec{y}, -Z, T | A(-Z, T) = 0) \right) dt, \end{aligned} \quad (\text{A.9})$$

avec τ_1 l'instant d'arrivée du premier client. La deuxième égalité est conservée parce que les arrivées suivent un processus de Poisson. Par conséquent, étant donné que le nombre d'arrivées dans $[-Z, T)$ est égal à 1 ($A(-Z, T) = 1$), nous avons τ qui est uniformément distribué sur l'intervalle $[-Z, T)$.

Maintenant, en prenant $Z, T \rightarrow \infty$, nous obtenons la dérivée première pour le trafic faible

$$\begin{aligned} G^{(1)}(0, \vec{y}) &:= \lim_{Z, T \rightarrow \infty} \frac{\partial}{\partial \lambda} G(\lambda, \vec{y}, -Z, T) \Big|_{\lambda=0} \\ &= \int_{-\infty}^{\infty} \left(G(0, \vec{y} | A(-\infty, \infty) = 1, \tau_1 = t) - G(0, \vec{y} | A(-\infty, \infty) = 0) \right) dt, \end{aligned}$$

avec la deuxième égalité qui découle de (A.8).

A.3.2 Régime à fort trafic

Le régime à fort trafic concerne la performance d'une file d'attente quand celle-ci est proche de la saturation. En d'autres termes, la quantité de travail qui arrive est proche

(mais toujours en dessous) de la capacité maximum du système. Dans ce régime, nous visons à dériver une expression pour notre indicateur de performance dans le régime à fort trafic après avoir appliqué une mise à l'échelle appropriée. On désigne cette mise à l'échelle par $f_\lambda(\vec{y})$ et nous voulons donc obtenir une expression pour $G(\lambda, f_\lambda(\vec{y}))$. Dans le reste de cette section, nous expliquons ce concept pour les indicateurs de performance d'un système à un seul serveur comme étudié dans cette thèse.

Un système à serveur unique est proche de la saturation lorsque la charge se rapproche de 1, $\rho \uparrow 1$. Par conséquent, ce régime peut être obtenu en laissant

$$\lambda \uparrow \hat{\lambda} := \frac{1}{\mathbb{E}[B]}.$$

L'analyse du trafic fort a été réalisée dans plusieurs articles comme [27, 16, 35].

Le principe de base est d'établir que les métriques de performance mises à l'échelle $(1 - \lambda\mathbb{E}[B])\vec{N}$, $(1 - \lambda\mathbb{E}[B])W_k(b)$ et $(1 - \lambda\mathbb{E}[B])W_k$, ont une limite tel que $\lambda \uparrow \frac{1}{\mathbb{E}[B]}$. Ainsi, dans le régime à fort trafic, nous avons des expressions pour les métriques de performance mises à l'échelle suivantes :

$$(i) \ \psi(\lambda, \vec{z}^{1-\lambda\mathbb{E}[B]}) = \mathbb{E}[\vec{z}^{(1-\lambda\mathbb{E}[B])\vec{N}}],$$

$$(ii) \ W_k^{DPS}(\lambda, b, x/(1 - \lambda\mathbb{E}[B])) = \mathbb{P}[(1 - \lambda\mathbb{E}[B])W_k^{DPS}(b) > x],$$

$$(iii) \ \widetilde{W}_k^{DROS}(\lambda, u(1 - \lambda\mathbb{E}[B])) = \mathbb{E}[e^{-u(1-\lambda\mathbb{E}[B])W_k^{DROS}}],$$

dans lesquelles nous avons utilisé la notation $\vec{z}^{\gamma\vec{N}} := (z_1^{\gamma N_1}, \dots, z_K^{\gamma N_K})$.

Dépendant des trois métriques décrites ci-dessus, la mise à l'échelle $f_\lambda(\vec{y})$ est donc

$$(i) \ f_\lambda(\vec{z}) = \vec{z}^{1-\lambda\mathbb{E}[B]},$$

$$(ii) \ f_\lambda(b, x) = (b, x/(1 - \lambda\mathbb{E}[B])),$$

$$(iii) \ f_\lambda(u) = u(1 - \lambda\mathbb{E}[B]).$$

A.3.3 Approximation de l'interpolation

Les expressions de métriques de performance sont connues pour les cas de trafics faible et fort. Une approximation arbitraire de λ peut être obtenue en suivant la méthode d'interpolation.

Cette technique a été popularisée par Reiman et Simon [28, 29, 30] et consiste à approcher la métrique de performance mise à l'échelle, $G(\lambda, f_\lambda(\vec{y}))$, par un polynôme $\hat{G}(\lambda, \vec{y})$ d'ordre $n + 1$:

$$\hat{G}(\lambda, \vec{y}) = h_0(\vec{y}) + h_1(\vec{y})\lambda + h_2(\vec{y})\lambda^2 + \dots + h_{n+1}(\vec{y})\lambda^{n+1}. \quad (\text{A.10})$$

On obtient alors une approximation d'interpolation pour la métrique performance non normalisée $G(\lambda, \vec{y})$, qui est :

$$G^{INT}(\lambda, \vec{y}) := \hat{G}(\lambda, f_\lambda^{-1}(\vec{y})) \quad (\text{A.11})$$

for $0 \leq \lambda < 1/\mathbb{E}[B]$.

Pour déterminer les coefficients $h_0(\vec{y}), \dots, h_n(\vec{y})$, nous prenons la m -ième dérivée, $m = 0, \dots, n$, dans (A.11) avec $\lambda = 0$ et fixons cette égalité à la m -ième dérivée de la métrique de performance approximée. Par conséquent, nous obtenons les conditions de faible trafic suivantes :

$$\left. \frac{\partial^m G^{INT}(\lambda, \vec{y})}{\partial \lambda^m} \right|_{\lambda=0} = G^{(m)}(0, \vec{y}), \text{ for } m = 0, \dots, n. \quad (\text{A.12})$$

Il est à noter que les expressions de $G^{(m)}(0, \vec{y})$ sont données dans la Proposition A.3.1.

Pour déterminer $h_{n+1}(\vec{y})$, nous utilise la condition du trafic fort :

$$\hat{G}(1/\mathbb{E}[B], \vec{y}) = G(1/\mathbb{E}[B], f_{1/\mathbb{E}[B]}(\vec{y})), \quad (\text{A.13})$$

avec $G(1/\mathbb{E}[B], \vec{y})$ le résultat du régime à fort trafic.

Dans ce qui suit, nous verrons comment déterminer les coefficients $h_0(\vec{y}), \dots, h_n(\vec{y})$ en pratique. Nous nous référons à l'approximation (A.11) comme approximation de l'interpolation d'ordre $n + 1$.

Ci-dessous, le résultat de l'approximation de l'interpolation utilisé tout au long de cette thèse.

Proposition A.3.2. *Pour un système à serveur unique l'approximation de l'interpolation d'ordre $n + 1$ de manière équivalente comme suit*

$$G^{INT}(\lambda, \vec{y}) = \sum_{i=0}^n \lambda^i \left(1 - (\lambda \mathbb{E}[B])^{n+1-i} \right) h_i \left(f_\lambda^{-1}(\vec{y}) \right) + (\lambda \mathbb{E}[B])^{n+1} G(1/\mathbb{E}[B], \vec{y}). \quad (\text{A.14})$$

Proof. Nous obtenons de l'équation de condition du régime à fort trafic

$$h_{n+1}(\vec{y}) = \mathbb{E}[B]^{n+1} \left(G(1/\mathbb{E}[B], f_{1/\mathbb{E}[B]}(\vec{y})) - \sum_{i=0}^n \frac{h_i(\vec{y})}{\mathbb{E}[B]^i} \right)$$

L'Équation (A.14) est obtenue en substituant cette expression de (A.10) puis en retirant la normalisation comme dans l'équation (A.11)

□

Il est à noter que dans ce cas, $G(\lambda, \vec{y})$ désigne la distribution des temps de séjour sans condition, $\tilde{S}(\lambda, x) := \mathbb{P}[(1 - \lambda\mathbb{E}[B])S_k > x]$, où S_k est le temps de séjour d'un client de classe- k , l'équation (A.14) se réduit à l'équation (1) de [7] après avoir retiré la normalisation :

$$\tilde{S}^{INT}(\lambda, x) = \sum_{i=0}^n \lambda^i \left(1 - (\lambda\mathbb{E}[B])^{n+1-i}\right) \tilde{S}^{(i)}(\lambda, (1 - \rho)x) + (\lambda\mathbb{E}[B])^{n+1} G(1/\mathbb{E}[B], x), \quad (\text{A.15})$$

avec $\tilde{S}^{(i)}(\lambda, (1 - \rho)x)$ le i -ème, $i = 0, \dots, n$, le coefficient obtenu à partir de l'interpolation et qui est équivalente à la fonction $h_i(f_\lambda^{-1}(\vec{y}))$, avec $\vec{y} = x$, dans l'Équation (A.14).

Dans un cas particulier dans lequel nous approximations la moyenne de V et que nous noterons $G(\lambda) := \mathbb{E}[V]$, et $G^{HT} := \lim_{\lambda \uparrow 1/\mathbb{E}[B]} f_\lambda \cdot G(\lambda)$ le trafic fort terme. Dans ce cas, l'équation (A.11) caractérise les dérivées du régime faible de trafic et l'équation du régime à fort trafic. De ce que nous connaissons, ceci est un nouveau résultat et il est applicable à tout modèle de serveur unique.

Corollary A.3.3. *L'approximation de l'interpolation d'ordre $n + 1$ de la moyenne peut être écrite comme*

$$\begin{aligned} G^{INT}(\lambda) &= \frac{1}{f_\lambda} \left(\sum_{i=0}^n \frac{\lambda^i}{i!} \left(1 - (\lambda\mathbb{E}[B])^{n+1-i}\right) \sum_{j=0}^i \binom{i}{j} \frac{d^{i-j} f_\lambda}{d\lambda^{i-j}} \Big|_{\lambda=0} G^{(j)}(0) + (\lambda\mathbb{E}[B])^{n+1} G^{HT} \right), \end{aligned} \quad (\text{A.16})$$

avec $G^{(i)}(\lambda)$ qui désigne la i -ième, $i = 0, \dots, n$ la dérivée du régime de faible trafic.

Proof. A partir de la condition du régime de faible trafic (A.12) nous obtenons :

$$h_i(f_\lambda^{-1}) = \frac{\sum_{j=0}^i \binom{i}{j} \frac{d^{i-j} f_\lambda}{d\lambda^{i-j}} \Big|_{\lambda=0} G^{(j)}(0)}{i! f_\lambda}, i = 0, 1, \dots, n, .$$

l'Équation (A.16) est obtenue après avoir substitué ces expressions de l'Équation (A.14).

□

A.4 Principaux résultats

Dans cette section nous présentons les principaux résultats obtenus durant cette thèse.

A.4.1 *Power of two*

Les dérivées du régime à faible trafic peuvent être exprimées de manière compacte à l'aide d'une variable aléatoire X qui est uniformément répartie sur l'ensemble des valeurs $\frac{1}{C_1}, \dots, \frac{1}{C_K}$, i.e.,

$$\mathbb{P}\left[X = \frac{1}{C_1}\right] = \dots = \mathbb{P}\left[X = \frac{1}{C_K}\right] = \frac{1}{K}.$$

It is plain that

$$\mathbb{E}[X^p] = \frac{1}{K} \sum_{k=1}^K \frac{1}{C_k^p}, \quad p = 1, 2, \dots$$

Proposition A.4.1. *Les dérivées du régime à fort trafic de la moyenne du temps de séjour en utilisant la politique power-of-two satisfont*

$$\begin{aligned} \overline{SQ2}^{(0)}(0) &= \mathbb{E}[X] \cdot \mathbb{E}[B], \\ \overline{SQ2}^{(1)}(0) &= -\frac{1}{K-1} \text{Var}[X] \cdot \mathbb{E}[B]^2 \end{aligned}$$

et

$$\overline{SQ2}^{(2)}(0) = \frac{2}{(K-1)^2} (\dots) \cdot \mathbb{E}[B]^3$$

avec

$$(\dots) = \mathbb{E}[X]^3 - 2\mathbb{E}[X] \mathbb{E}[X^2] + \mathbb{E}[X^3].$$

A.4.2 Processeur partagé discriminatoire

L'approximation de l'interpolation pour la distribution du temps d'attente en utilisant la politique DPS satisfait l'équation suivante :

Proposition A.4.2. *L'interpolation du régime à fort trafic (d'ordre 2) de la distribution complémentaire du temps d'attente sous condition d'un client de classe-k marqués avec une exigence de service donnée b est définie par*

$$\begin{aligned}
W_k^{DPS,INT}(\lambda, b, x) &= \lambda(1 - \rho) \sum_{j=1}^K \alpha_j \mathbb{E} \left[\left(1 + \frac{g_k}{g_j} \right) \left(-(1 - \rho)x + \min\{B_j, \frac{g_j}{g_k}b\} \right)^+ \right. \\
&\quad \left. + \mathbf{1} \left[\frac{g_j}{g_k}b > (1 - \rho)x \right] \left(B_j - \min\{B_j, \frac{g_j}{g_k}b\} \right) \right. \\
&\quad \left. + \mathbf{1} [B_j > (1 - \rho)x] \left(b - \frac{g_k}{g_j} \min\{B_j, \frac{g_j}{g_k}b\} \right) \right] + \rho^2 e^{-(1-\rho)x \frac{g_k}{b\mathbb{E}[V]}},
\end{aligned}$$

with

$$\mathbb{E}[V] = \frac{\mathbb{E}[B^2]}{\sum_{i=1}^K \alpha_i \mathbb{E}[B_i^2]/g_i}.$$

En corollaire de la proposition A.4.2, nous obtenons le résultat suivant pour le l'approximation de l'interpolation de la durée de séjour moyenne, sous condition d'un client étiqueté, celle est notée $\bar{S}_k^{DPS,INT}(\lambda, b)$

Corollary A.4.3. *L'approximation de l'interpolation (d'ordre 2) de la moyenne du temps de séjour sous condition pour un client étiqueté de classe-k et avec une exigence de service b est donnée par*

$$\begin{aligned}
\bar{S}_k^{DPS,INT}(\lambda, b) &:= \int_0^\infty S_k^{DPS,INT}(\lambda, b, x) dx \\
&= b(1 + \rho) \\
&\quad + \lambda \mathbb{E} \left[\frac{1}{2} \left(1 + \frac{g_k}{g_{U_t}} \right) \min\{B_{U_t}, b \frac{g_{U_t}}{g_k}\}^2 - \left(b \frac{g_{U_t}}{g_k} + \frac{g_k}{g_{U_t}} B_{U_t} \right) \min\{B_{U_t}, b \frac{g_{U_t}}{g_k}\} + \frac{g_{U_t}}{g_k} b B_{U_t} \right] \\
&\quad + \frac{(\lambda \mathbb{E}[B])^2}{(1 - \lambda \mathbb{E}[B])} \frac{b}{g_k} \frac{\mathbb{E}[B^2]}{\sum_{j=1}^K \alpha_j \mathbb{E}[B_j^2]/g_j}.
\end{aligned}$$

A.4.3 Priorités relatives

Nous présentons le résultat de l'effondrement de l'espace d'état pour la distribution de la longueur des files d'attente à l'état stable aux instants de départ. La proposition suivante montre le vecteur de la longueur de la file d'attente est le produit d'une variable aléatoire exponentiellement distribuée et d'un vecteur déterministe.

Proposition A.4.4. *Lorsque la mise à l'échelle est effectué par $1 - \rho$, le vecteur de la longueur de la file d'attente aux instants de départ a une distribution asymptotique exacte :*

$$(1 - \rho) (Q_1^{RP}, \dots, Q_K^{RP}) \xrightarrow{d} (\widehat{Q}_1^{RP}, \dots, \widehat{Q}_K^{RP}) \stackrel{d}{=} \left(\frac{\hat{\lambda}_1}{p_1}, \dots, \frac{\hat{\lambda}_K}{p_K} \right) \cdot Y,$$

avec $\xrightarrow{d}$ qui désigne la convergence de la distribution et avec Y exponentiellement distribué.

Nous présentons maintenant l'approximation de l'interpolation pour la longueur de la file d'attente.

Proposition A.4.5. *L'interpolation pour les régimes à faible et fort trafics (d'ordre 3) de la fonction génératrice de probabilité jointe de $(N_1^{RP}, \dots, N_K^{RP})$ est donnée par*

$$\begin{aligned}
& \psi^{RP,INT}(\lambda, \vec{z}) \\
&= (1 - \rho^3) + \lambda (1 - \rho^2) \left(-\mathbb{E}[B] + \sum_{i=1}^K \alpha_i \mathbb{E}[B_i] z_i^{(1-\rho)^{-1}} \right) \\
&+ \frac{\lambda^2 (1 - \rho)}{2} \left(-2\mathbb{E}[B] \sum_{i=1}^K \alpha_i z_i^{(1-\rho)^{-1}} \mathbb{E}[B_i] \frac{\ln(z_i)}{1 - \rho} \right. \\
&\quad \left. + \sum_{i=1}^K \alpha_i z_i^{(1-\rho)^{-1}} \mathbb{E}[B_i^2] \left(\sum_{k=1}^K \alpha_k z_k^{(1-\rho)^{-1}} - 1 \right) \right) \\
&+ \rho^3 \frac{\mathbb{E}[B] \nu(\vec{p})}{\mathbb{E}[B] \nu(\vec{p}) - \sum_{i=1}^K \frac{\alpha_i}{p_i} \ln(z_i^{(1-\rho)^{-1}})}.
\end{aligned}$$

Bibliography

- [1] E. Altman, K. Avrachenkov, and U. Ayesta. A survey on discriminatory processor sharing. *Queueing systems*, 53(1-2):53–63, 2006.
- [2] S.C. Borst, R. Núñez-Queija, and A.P. Zwart. Sojourn time asymptotics in processor sharing queues. *Queueing Systems*, 53(1-2):31–51, 2006.
- [3] O.J. Boxma, S.G. Foss, J.-M. Lasgouttes, and R. Núñez-Queija. Waiting time asymptotics in the single server queue with service in random order. *Queueing Systems*, 46:35–73, 2004.
- [4] D. Burman and D. Smith. An asymptotic analysis of a queueing system with markov-modulated arrivals. *Operations Research*, 34:105–119, 1986.
- [5] J.W. Cohen. The multiple phase service network with generalized processor sharing. *Acta Informatica*, 12:245–284, 1979.
- [6] G. Fayolle, I. Mitrani, and R. Iasnogorodski. Sharing a processor among many job classes. *Journal of the ACM*, 27(3):519–532, 1980.
- [7] P.J. Fleming and B. Simon. Interpolation approximations of sojourn time distributions. *Operations Research*, 39(2):251–260, 1991.
- [8] S. Ben Fredj, T. Bonald, A. Proutiere, G. Regnie, and J. Roberts. Statistical bandwidth sharing: A study of congestion at flow level. In *SIGCOMM*, pages 111–122, 2001.
- [9] M. Haviv and J. van der Wal. Equilibrium strategies for processor sharing and random queues with relative priorities. *Probability in the Engineering and Informational Sciences*, 11:403–412, 1997.
- [10] M. Haviv and J. van der Wal. Waiting times in queues with relative priorities. *Operations Research Letters*, 35:591–594, 2007.
- [11] A. Izagirre, U. Ayesta, and I.M. Verloop. Sojourn time approximations in a multi-class time-sharing system. *IEEE INFOCOM*, 2014.
- [12] F.P. Kelly. *Stochastic Networks and Reversibility*. Wiley, Chichester, 1979.

-
- [13] D.G. Kendall. Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded markov chains. *Annals of Mathematical Statistics*, 24:338–354, 1953.
- [14] J. Kim. Queue length distribution in a queue with relative priorities. *Bull. Korean Math. Soc.*, 46:107–116, 2009.
- [15] J. Kim, J. Kim, and B. Kim. Analysis of the M/G/1 queue with discriminatory random order service policy. *Performance Evaluation*, 68(3):256–270, 2011.
- [16] J.F.C. Kingman. The single server queue in heavy traffic. *Proc. Cambridge Philos. Soc.*, 57:902–904, 1961.
- [17] J.F.C. Kingman. On queues in which customers are served in random order. *Proc. Cambridge Philos. Soc.*, 58:79–91, 1962.
- [18] J.F.C. Kingman. Queue disciplines in heavy traffic. *Mathematics of Operations Research*, 7(2):262–271, 1982.
- [19] L. Kleinrock. Analysis of a time-shared processor. *Naval Research Logistics Quarterly*, 11:59–73, 1964.
- [20] L. Kleinrock. Time-shared systems: A theoretical treatment. *Journal of the ACM*, 14(2):242–261, 1967.
- [21] L. Kleinrock. *Queueing Systems, vol. 2*. John Wiley and Sons, 1976.
- [22] J.D.C. Little. A proof for the queueing formula: $L=\lambda w$. *Operations Research*, 9(3):383–387, 1961.
- [23] M. Mitzenmacher. The power of two choices in randomized load balancing. *IEEE Transactions on Parallel and Distributed Systems*, 12(10):1094–1104, 2001.
- [24] R.L. Dobrushin N.D. Vvedenskaya and F.I. Karpelevich. Queueing system with selection of the shortest of two queues: An asymptotic approach. *Problems of Information Transmission*, 32:20–34, 1996.
- [25] R. Núñez-Queija. *Processor-Sharing Models for Integrated-Services Networks*. PhD thesis, Eindhoven University of Technology, 2000.
- [26] C. Palm. Waiting times with random served queue. *Tele1 (English edition; original 1938)*, 1–107, 1957.
- [27] K.M. Rege and B. Sengupta. Queue length distribution for the discriminatory processor sharing queue. *Operations Research*, 44(4):653–657, 1996.
- [28] M.I. Reiman and B. Simon. An interpolation approximation for queueing systems with Poisson input. *Operations Research*, 36:454–469, 1988.

-
- [29] M.I. Reiman and B. Simon. Light traffic limits of sojourn time distributions in Markovian queueing networks. *Stochastic Models*, 4:191–233, 1988.
- [30] M.I. Reiman and B. Simon. Open queueing systems in light traffic. *Oper. Res.*, 14:26–59, 1989.
- [31] J.W. Roberts. A survey on statistical bandwidth sharing. *Computer Networks*, 45:319–332, 2004.
- [32] A.L. Stolyar. Pull-based load distribution in large-scale heterogeneous service systems. *Queueing Systems*, 2015.
- [33] H.C. Tijms. *A First Course in Stochastic Models*. Wiley, 2003.
- [34] S. Varma and A.M. Makowski. Interpolation approximations for symmetric fork-join queues. *Performance Evaluation*, 20:245–265, 1994.
- [35] I.M. Verloop, U. Ayesta, and R. Núñez-Queija. Heavy-traffic analysis of a multiple-phase network with discriminatory processor sharing. *Operations Research*, 59(3):648–660, 2011.
- [36] J. Walrand. An introduction to queueing networks. In *Prentice-Hall, Englewood Cliffs, NJ*, 1988.
- [37] S.F. Yashkov. Processor sharing queues: Some progress in analysis. *Queueing Systems*, 2:1–17, 1987.
- [38] S.F. Yashkov. Processor sharing queues: Some progress in analysis. *Queueing Systems*, 2:1–17, 1987.
- [39] S.F. Yashkov. Mathematical problems in the theory of processor sharing queueing systems. *Journal of Soviet Mathematics*, 58:101–147, 1992.

Annex B

Errendimendu metriken interpolazio bidezko hurbilketak oreka egoeran

Ane Izagirre

Zuzendariak: Dr. F.X. Albizuri , Dr. U. Ayesta eta Dr. I.M. Verloop

Institut National des Sciences Appliquées (INSA) eta Euskal Herriko
Unibertsitatea (EHU)

2015eko iraila

Hitzaurrea

I feel very lucky to have had the opportunity of carrying out the Ph.D thesis with Urtzi and Maaïke, my supervisors. All my gratitude for their continuous support, patience and motivation; without their help this would not have been possible. I would also like to thank EDSYS for the Ph.D fellowship.

Xabierri, nire tesi zuzendari denari, eskerrik beroenak eman nahi dizkiot aldi oro laguntzeko erakutsi duen prestutasunagatik. Berari esker EHU-ko egonaldiak errazagoak izan dira.

The three-month visit to College Park, which was financially supported by EDSYS and INP, has been a valuable experience abroad. I am grateful to Armand Makowski for offering me this opportunity and also for the hospitality extended to me while I was in DC.

I would also like to thank all the members of the jury for accepting the invitation to attend my PhD defence.

Moreover, I wish to thank all the members of the SARA group at LAAS-CNRS laboratory and in particular, to my officemates Maialen, Samir and Tom, who have always been happy to help.

Modu berean, EHU-ko informatika fakultateko KZAA departamentuari eta hirugarren pixuko ikasleei eskerrak eman nahi dizkiet; bereziki Jonathan, Leti eta Mendi mahai inguruko tertulia atseginengatik. Mikel bulegokideari ere goizak atseginagoak egiteagatik. Eta azkenik, Martini urte hauetan zehar igarotako momentuengatik.

Bukatzeko, gurasoi eskerrak eman nahi dizkiet beraien animo eta konfindantza guztiagatik. Baita kuadrilako lagunei ere, eta bereziki Olatz eta Maddaleni krisi une horietan entzun eta laguntzeko prest egoteagatik. Eta azkenik Mikeli bihotz-bihotzez eskerrak, bera izan baita nirekin batera tesi hau hurbilen bizi izan duena eta egoerak bultzaturiko distantzia medio hurrun bezain hurbil sentitu dut.

Gaien Aurkibidea

1	Sarrera	1
1.1	Ilara ereduak	2
1.1.1	Zerbitzari bakarreko sistema	2
1.1.2	Zerbitzari paraleloak	5
1.2	Interpolazio bidezko hurbilketaren motibazioa	6
1.3	Tesiaren laburpena	7
2	Interpolazio bidezko hurbilketa	9
2.1	Trafiko-arin analisia	10
2.2	Trafiko-geldo erregimena	12
2.3	Interpolazio bidezko hurbilketa	13
3	<i>Bi aukeren abantaila</i> politikaren trafiko arin analisia	15
3.1	Ereduen deskribapena	15
3.2	Sistemako egonaldiaren batezbesteko iraute denbora	16
3.3	Zerogarren trafiko-arin deribatua	18
3.4	Emaizta lagungarri bat	19
3.5	Lehenengo trafiko-arin deribatua	22
3.6	Bigarren trafiko-arin deribatua	24
3.7	Simulaturiko emaitzak	34
3.A	Eranskina	37
3.A.1	(3.18) Ekuazioaren frogapena	37
3.A.2	Bi kalkulu lagungarri	38
3.A.3	3.6.3 Lemaren frogapena	39
3.A.4	3.6.4 Lemaren frogapena	41
3.A.5	(3.84) Ekuazioaren frogapena	44
3.A.6	(3.86) Ekuazioaren frogapena	45
4	Interpolazio bidezko hurbilketak prozesadore-partekatze diskriminatzaile politikarentzat	47
4.1	Ereduen deskribapena	47
4.2	Emaizta ezagunak	48
4.2.1	Egonaldiaren batezbesteko iraute denbora baldintzatua	48
4.2.2	Trafiko geldo emaitzak	49
4.3	Bezero kopuruaren banaketa	50
4.4	Itxaron denboraren banaketa	54

4.5	Sistemako egonaldiaren batezbesteko iraute denbora	56
4.5.1	Sistemako egonaldiaren batezbesteko iraute denbora baldintzatua	56
4.5.2	Sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatua	60
4.5.3	Beharrezko zerbitzu esponenziala	60
4.6	Emaitza numerikoak	63
4.6.1	Bezero kopuruaren banaketa	65
4.6.2	Itxaron denboraren banaketa	66
4.6.3	Sistemako egonaldiaren iraute denbora	67
4.A	Eranskina	70
4.A.1	4.3.1 Lemaren frogapena	70
4.A.2	4.3.1 Proposizioaren frogapena	82
4.A.3	4.4.1 Lemaren frogapena	83
4.A.4	4.4.1 Proposizioaren frogapena	88
4.A.5	Trafiko arinpean deribatuak lortzeko beste modu bat	88
5	Lehentasun erlatibodun trafiko geldopeko analisia	93
5.1	Ereduen deskribapena	94
5.2	Ilararen luzera irteera unean	95
5.2.1	Trafiko-geldo eskalatua	96
5.2.2	5.2.1 Proposizioaren frogapena	98
5.3	Ilararen luzera une arbitrarioan	99
5.3.1	Trafiko-geldo eskalatua	100
5.3.2	5.3.1 Proposizioaren frogapena	101
5.4	Itxaron denbora	103
5.5	Pixu optimoak	108
5.5.1	<i>Mantentze-kostua</i>	108
5.5.2	Itxaron denboraren momentuak	109
5.6	Emaitza numerikoak	112
5.6.1	Egoera-espazioaren kolapsoa ilara luzerarentzat	112
5.6.2	Ilara luzera eta itxaron denborarean momentuak	113
5.6.3	Pixuen balio optimoak	114
5.A	Eranskina	114
5.A.1	5.2.1 Lemaren frogapena	114
5.A.2	5.2.4 Lemaren frogapena	116
5.A.3	5.4.2 Lemaren frogapena	118
5.A.4	(5.35) Ekuazio diferentzialaren ebazpena	118
6	Interpolazio bidezko hurbilketak lehentasun erlatibodun politikarentzat	121
6.1	Bezero kopuruaren banaketa	121
6.2	Itxaron denboraren banaketa	123
6.3	Alderaketa numerikoa	125
6.3.1	Interpolazio bidezko hurbilketaren zehaztasun maila aztertzen	125
6.3.2	RP versus DPS	126
6.A	Eranskina	127
6.A.1	6.1.2 Lemaren frogapena	127
6.A.2	6.1.1 Proposizioaren frogapena	129

6.A.3	6.2.2 Lemaren frogapena	131
6.A.4	6.2.1 Proposizioaren frogapena	131
Bibliografia		133

1. Kapitulu

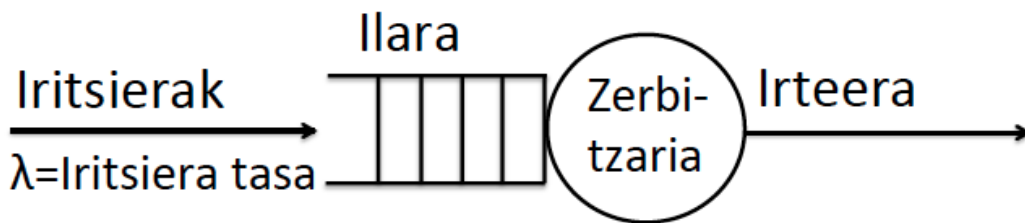
Sarrera

Ilaran itxarotea desatsegina bezain ohikoa da. Guztiok itxaron dugu supermerkatuko ilaran, ospitaleko larrialdi zerbitzuan edota dei-zentro batera hots egitean. Maila abstraktuago batean, ilara hauek ere telekomunikazio sisteman aurki daitezke. Esate baterako, bidaltzen dugun email bakoitza edo behera kargatzen dugun dokumentu bakoitza pakete desberdinetan zatitzen da. Ondoren, pakete bakoitza bere helmugara bidaltzen da erabilgarri dagoen bide onenetik beste paketeek sorturiko ilarak ekiditeko asmoz.

Ilaren teoria aipatu berri ditugun egoerak aztertzen dituen teoria matematikoa da eta bi helburu nagusi ditu. Alde batetik, sistemaren errendimendua aztertzea. Bestetik, errendimendua hobetzeko politika egokiena aurkitzea. Sistemako bezero kopurua eta itxaron denbora dira literatura kontsideratzen diren metrika ohikoenak.

Tesi honetan lehenengo kasuan jarriko dugu arreta eta klase anitzeko eta zerbitzari bakarreko ilaren errendimendua aztertuko dugu *Prozesadore-partekatze diskriminatzailea* eta *Lehentasun erlatibodun* politikenpean, eta zerbitzari paralelodun eredua *bi aukeren abantaila* deritzon politikapean. Ereduek hauen inguruko xehetasun gehiago eskainiko ditugu 1.1 Atalean.

Errendimendu metrikak zehaztasunez karakterizatzea gustatuko litzaiguke. Dena den, askotan, oso zaila izaten da. Horregatik, ilaren teoriak metrikak hurbiltzeko interes handia erakutsi du. Tesi honetan *trafiko arinaren eta geldoaren interpolazio* hurbilketa teknika erabiliko dugu bezero kopurua, itxaron denbora eta sistemako egonaldiaren iraute denbora hurbiltzeko. Zehazkiago, lehenik eta behin errendimendua aztertuko dugu *trafiko arin* erregimenpean, hau da, iritsiera tasak zeroruntz konbergitzen duenean. Bigarren, errendimendua *trafiko geldo* erregimenpean aztertuko dugu, hau da, sistema asetasun puntuan dagoenean. Ondoren, trafiko arinpeko eta trafiko geldopeko karakterizazioak konbinatuz interpolazio bidezko hurbilketa eraikiko dugu zeina kargaren edozein balioetarako baliogarria den. 1.2 Atalean motibatuko dugu teknika hau eta 2 Kapitulan azalduko dugu zehaztasunez.



1.1. Irudia: Zerbitzari bakarreko ilara.

1.1 Ilara ereduak

Atal honetan tesian zehar aztertuko ditugun ilara ereduak aurkeztuko ditugu.

1.1.1 Zerbitzari bakarreko sistema

Zerbitzari bakarreko ilara 1.1 Irudian deskribaturikoa da. λ tasako iritsiera prozesu bat dago eta, beraz, λ^{-1} bezeroen arteko batezbesteko denbora da. Bezeroa iristean ilaran itxarongo du edo zuzenean zerbitzatu izango da ezarritako zerbitzatzeko politikaren arabera. Zerbitzariaren kapazitatea bat dela onartzen dugu. Beharrezko zerbitzua B zorizko aldagaiak adieraziko dugu. Hau da, hain zuzen, zerbitzariak bere kapazitate guztia bezero honi eskeiniz gero sisteman igaroko duen denbora.

Ohiko onarpen bat iritsieren arteko denborak independenteak eta berdinki banatuak direla da, beharrezko zerbitzuak ere independenteak eta berdinki banatuak daude, eta beraien artean independenteak dira. Eredu hau $G/GI/1$ ilara bezala ezagutzen da, non G hizkiak banaketa orokorra adierazten duen eta GI hizkiek banaketa orokorra eta independentea. Notazio hau [25] Kendallek aurkeztu zuen. Tesi honetan $M/GI/1$ ilaran jarriko dugu arreta, non M hizkiak markobianoa ala memoriagabea adierazten duen, hau da, bezeroek Poisson iritsiera prozesu bat jarraitzen dutenean.

Sistemako lan totala erabilitako zerbitzatzeko politikaren independentea da. Oreka ere zerbitzatzeko politikaren independentea da. Ilara orekatua izango da $\lambda \mathbb{E}[B]$ bat baino txikiagoa den bitartean. Aldiz, sistemako bezero kopurua politikaren menpe dago.

Tesi honetan kontsideraturiko zerbitzari bakarreko sistemek K bezero klase dituzte. k klaseko bezeroak λ_k tasako Poisson prozesu baten arabera iristen dira eta horregatik, $\alpha_k := \lambda_k / \lambda$ aldagaiak k klaseko iritsieren zatikia adierazten du. $B_k, k = 1, \dots, K$ k klaseko bezeroen beharrezko zerbitzua da. k klaseko bezeroen karga $\rho_k := \lambda_k \mathbb{E}[B_k]$ da eta $\rho := \sum_{k=1}^K \rho_k$ karga totala.

Jarraian tesi honetan erabili eta aztertutako politikak aurkeztuko ditugu.

Prozesadore partekatzea

Prozesadore partekatze politikapean, literaturan PS bezala ezaguna, zerbitzariaren kapazitatea era berean partekatzen da sisteman dauden bezeroen artean. Zehazkiago, t unean $n(t)$ bezero badaude sisteman, PS politikapean bezero bakoitza $1/n(t)$ abiaduran da zerbitzatu. Ikusi [51] eta [38] literaturaren ideia orokor baterako.

Poisson iritsieren kasuan sistemako bezero kopuruaren banaketa beharrezko zerbitzuaren batezbestekoaren menpe dago soilik, eta ez maila altuagoko estatistiken menpe. Zehazkiago, bezero kopuruaren banaketak orekan, ρ parametroko banaketa geometrikoa du, hau da, sisteman n bezero egoteko probabilitatea da $(1 - \rho)\rho^n, n = 0, 1, \dots$, [45]. Little-en legearen bitartez, [33], bezero kopuruaren banaketaren sentsibiltate eza batezbesteko sistemako iraute denboran ere aplikatzen da.

Aldiz, sistemako iraute denboraren banaketak ez du karakterizazio arruntik. Kleinrock-ek hasita, M/M/1 PS ilararen azterketa asko egin dira, [31, 32, 52, 53, 6].

Klase anitzdun PS ilararentzat ere bezero kopuruaren banaketa geometrikoa da.

Poisson iritsieren onarpenpean, [9, 23] artikuluetan erakutsi bezala, k klaseko $n_k, k = 1, \dots, K$, bezero edukitzeko probabilitatea honakoa da

$$(1 - \rho) \cdot \frac{(n_1 + \dots + n_K)!}{n_1! \cdot \dots \cdot n_K!} \prod_{k=1}^K \rho_k^{n_k}.$$

Prozesadore-partekatze diskriminatzailea

Prozesadore-partekatze diskriminatzailea ilara, literaturan DPS bezala ezaguna, erabilera askoko eredua da. Kleinrock-ek aurkeztu zuen [32] artikuluan. PS politikaren orokortze bat da. Berriz ere, K bezero klase daudela onartzen dugu eta klase bakoitzari pixu bat ezartzen diogu, $g_1, \dots, g_K$. Zerbitzariaren kapazitatea sisteman dauden bezero guztien artean partekatzen da, ezarritako pixuen proportzioan. Zehazkiago, K bezero klase izanik, t unean k klaseko $n_k(t), k = 1, \dots, K$, bezero badaude sisteman, DPS politikapean k klaseko bezero bakoitza ondoko abiaduran da zerbitzatua

$$\frac{g_k}{\sum_{j=1}^K g_j n_j(t)}.$$

Ereduaren deskribapena arrunta izan arren eta PS politika ondo aztertu eta ulerturiko eredua izan arren, DPS ereduaren analisisa oso zaila da, karakterizazio itxirik gabekoa. Ikusi [1] artikulua literaturaren laburpen baterako. Jarraian, 4 Kapituluaren erabilitako emaitza batzuk aurkeztuko ditugu.

[40] artikuluan Rege et al.-ek bezero kopuruaren funtzio sortzaileak ekuazio diferentzial bat betetzen duela adierazi zuten. Ekuazio horietatik, autoreek momentuak numerikoki nola lortu erakusten dute.

Trafiko geldo erregimenak sistema asetasun puntuan aztertzen du. [40] artikuluan Rege et al.-ek, beharrezko zerbitzuaren banaketa esponentziala den kasurako, bezero kopuruarentzat *state-space-collapse* emaitza erakusten dute. [48] artikuluan Verloop et al.-ek, emaitza hori orokortu zuten banaketa orokorra duten beharrezko zerbitzuen kasurako. Izan bedi $N_k^{DPS}, k = 1, \dots, K, k$ klaseko bezero kopurua DPS politikapean, orduan $\rho \uparrow 1$ limitepean ondokoa lortzen dugu

$$(1 - \lambda \mathbb{E}[B])(N_1^{DPS}, \dots, N_K^{DPS}) \xrightarrow{d} Y \cdot \left(\frac{\alpha_1 \mathbb{E}[B_1]}{g_1}, \frac{\alpha_2 \mathbb{E}[B_2]}{g_2}, \dots, \frac{\alpha_K \mathbb{E}[B_K]}{g_K} \right),$$

non $\xrightarrow{d}$ ikurak konbergentzia banaketan adieratzen duen eta Y banaketa esponentziala jarraitzen duen zorizko aldagaia den, $\frac{\mathbb{E}[B^2]}{\mathbb{E}[B] \sum_{k=1}^K \alpha_k \mathbb{E}[B_k^2] / g_k}$ batezbestekoa duena.

[10] artikuluan Fayolle et al.-ek batezbesteko sistemako iraute denbora baldintzatua (beharrezko zerbitzuan) eta ez baldintzatua aztertzen dute. Beharrezko zerbitzuen banaketa orokorrerako, autoreek ondoko ekuazio integro-diferentziala lortzen dute batezbesteko sistemako iraute denbora baldintzaturako:

$$\begin{aligned} \frac{\partial \overline{S}_k^{DPS}(\lambda, b)}{\partial b} &= 1 + \lambda \sum_{j=1}^K \int_0^\infty \alpha_j \frac{g_j}{g_k} \frac{\partial \overline{S}_j^{DPS}(\lambda, y)}{\partial y} \left[1 - F_j \left(y + \frac{g_j}{g_k} b \right) \right] dy \\ &\quad + \lambda \int_0^b \frac{\partial \overline{S}_k^{DPS}(\lambda, y)}{\partial y} \sum_{j=1}^K \alpha_j \frac{g_j}{g_k} \left[1 - F_j \left(\frac{g_j}{g_k} (b - y) \right) \right] dy, \end{aligned}$$

non $\overline{S}_k^{DPS}(\lambda, b)$ k klaseko batezbesteko sistemako iraute denbora baldintzatua den DPS politika-pean. Hortaz gain, autoreek xehetasun handiko analisia egiten dute beharrezko zerbitzuek banaketa esponentziala duten kasurako. Dena den, sisteman bi klase dauden kasurako izan ezik, adierazpen itxirik ez dago eskuragarri eta analisi numerikoa ezinbestekoa da batezbesteko sistemako iraute denbora kalkulatzeko. 4.2.1. Kapitulan [10] artikuluko emaitzen xehetasun gehiago erakutsiko ditugu.

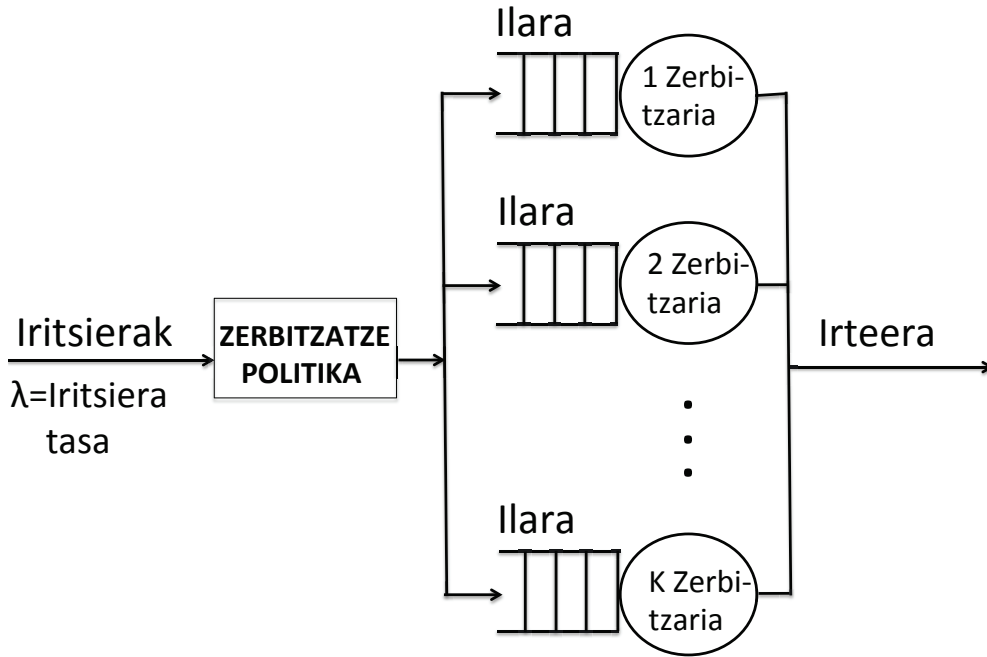
Lehentasun erlatiboak

Lehentasun erlatiboa, literaturan RP bezala ezaguna, klase anitzdun ilara da. Zerbitzatua izango den hurrengo bezeroa k klasekoa izateko probabilitatea honakoa da

$$\frac{n_k p_k}{\sum_{j=1}^K n_j p_j},$$

non $p_j > 0$, $j = 1, \dots, K$, pixuak diren eta n_j erabakia hartzeko unean sisteman dauden j klaseko bezero kopurua den. Behin zerbitzatua izango den klasea aukeratzean, klase barneko politikak erabakiko du zein bezero izango den zerbitzatua.

Deskribatzen ari garen ereduaren kasu zehatz bat klase barneko politika uniformeki zorizkoa denekoa da, hau da, klase barneko bezeroen artean bat aukeratzen da zoriz. Eredu hau [16] artikuluan aurkeztu zen eta *Zerbitzua auskazko orden diskriminatzailean* bezala ezagutzen da, literaturan DROS bezala ezaguna. Azken urteetan emaitza interesgarriak argitaratu dira eredu honen inguruan, [17, 26, 27]. Batezbesteko itxaron denbora [17] artikuluan lortu zen. [26, 27] artikuluan autoreek bezero kopuruak eta itxaron denborak bete beharreko ekuazio diferentzialak lortzen dituzte, hurrenez hurren. [27] artikuluan aurkezturiko emaitzen inguruko xehetasun gehiago erakutsiko ditugu 5. Kapitulan.

1.2. Irudia: K zerbitzari paralelodun eredua.

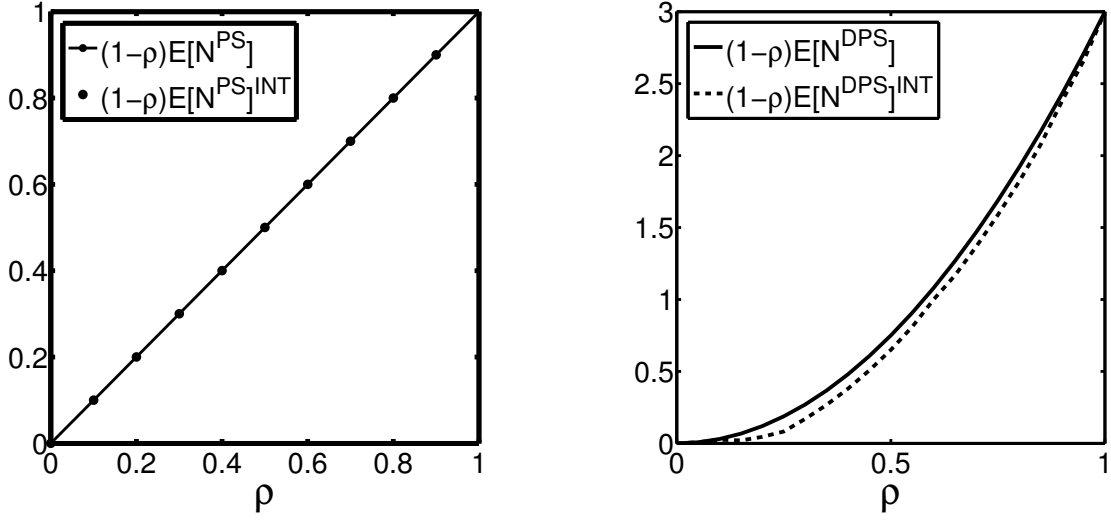
Klase bakarreko kasuan, $K = 1$, zerbitzua auskazko orden diskriminatzailean eredua aski ezaguna den *Zerbitzua auskazko ordenean* politikan bilakatzen da, literaturan ROS bezala ezaguna,. Zerbitzua auskazko ordenean politikaren inguruko artikulu klasikoak dira [29, 30, 39].

1.1.2 Zerbitzari paraleloak

Zerbitzari paralelodun ereduak aldi berean lan egiten duten K zerbitzari *heterogeneo* (i.e., denak kapazitate desberdinekoak) ditu, ikusi 1.2 Irudia. Bezero bat iristean, K zerbitzarietako bat hautatuko du *bi aukeren abantaila* politikaren arabera zeina $SQ(d)$, $d = 2$ bezala adierazten den: Bezeroak bi zerbitzari aukeratzen ditu zoriz. Jarraian, *ilara motzena batu*, literaturan JSQ bezala ezaguna, politika erabiltzen da aukeraturiko bi zerbitzari horien artean, hau da, ilaran bezero guxien dituen zerbitzarira igaroko da, berdinketak zoriz puskatuz. *Lehenengoa iristen, lehenengoa zerbitzua jasotzen*, literaturan FCFS bezala ezaguna, politika erabiltzen du zerbitzari bakoitzak.

Ohartu $SQ(1)$ zerbitzari bat zoriz aukeratzearen baliokidea dela. Aldiz, $SQ(K)$ JSQ politika K zerbitzarien artean aplikatzearen baliokidea da.

Denbora da *bi aukeren abantaila* sistema lehen aldiz aztertu zela, ikusi laburpena [35, 1.1. Atala] artikuluan. Zerbitzari paraleloen analisia zaila da JSQ politikak inplikatzan duen zerbitzarien arteko menpekotasuna dela eta. Arazoa bera da Poisson iritsiera prozesua eta zerbitzari *homogeneoak* (i.e., kapazitate berak) hautatu arren. Onarpen hauenpean Mitzenmacher [35] eta Vvedenskaya et al.-ek ($d = 2$) [37], independenteki aztertu dute limitean sistema bezero kopurua



1.3. Irudia: Bezero kopuruaren interpolazio bidezko hurbilketaren zehaztasuna PS (ezkerra) eta DPS (eskuina) ereduentzat.

infinitura doala onartuz. Beraien emaitzek zera erakutsi dute: Bi zerbitzari aukeratzeak ($d = 2$), bakar bat aukeratzea $d = 1$ baino emaitza hobek ematen dituela. Eta hiru aukeratzeak ($d = 3$), aldiz, errendimendua hobetzen du, baina oso gutxi $d = 2$ kasuarekin alderatuz. Hemendik dator “*bi aukeren abantaila*” terminologia.

Berriki, [44] artikuluan Stolyar-ek kapazitate heterogenoa duten zerbitzari paralelodun eredu bat aztertu du *pull* deritzon politikapean. Zerbitzari bakoitzak mezu bat bidaltzen dio router-ari hutsik geratzean eta router-ak iritsierak zerbitzarira bidaltzen ditu, mezua bidali dioten zerbitzarien artean bat zoriz aukeratuz, edo denak lanean ari badira, zerbitzari bat zoriz aukeratuz. Stolyar-ek frogatu du *pull* politikak *bi aukeren abantaila* politikak baino emaitza hobek ematen dituela.

1.2 Interpolazio bidezko hurbilketaren motibazioa

Lehenago aipatu bezala interpolazio teknika trafiko arinpeko eta trafiko geldopeko karakterizazioen konbinaketa bat da. Teknikak hurbilketa zehatza izatea ezartzen du trafiko arinpean, hau da, ρ karga zerotik hurbil dagoenean, eta trafiko geldopean, sistema asetatsun puntuan dagoenean. Interpolazio teknikak kargaren balio guztiak kontsideratzen dituen hurbilketa bat eraikitzen du. Hurbilketa teknikaren abantaila garrantzitsu bat errendimenduak sistemako parametroekiko duen menpekotasuna erakusten duela da. Interpolazio hurbilketa eraikitzeko erabilitako teknika [43] artikuluan proposaturikoa da, zeina 2. Kapituluan xehetasunez aurkezten den.

Adibide bezala, 1.3 Irudian bezero kopuru totalaren emaitza zehatza eta interpolazio bidezko hurbilketa irudikatu ditugu PS eta DPS polikenpean, N^{PS} eta N^{DPS} bezala adierazita, hurrenez hurren. PS-en kasurako emaitza zehatza eskuragarri dago eta $\mathbb{E}[N^{PS}] = \frac{\rho}{1-\rho}$ da. Hurbilketa eraikitzeko modua dela eta hurbilketa zehatza da $\rho = 0$ eta $\rho = 1$ puntuetan eta PS kasurako $(1-\rho)\mathbb{E}[N^{PS}] = \rho$ lerro zuzena denez, interpolazio bidezko hurbilketa zehatza da kasu honetan,

ikusi 1.3 Irudia, ezkerre. DPS kasurako interpolazio bidezko hurbilketa zehatza da kargaren murreko balioetarako, $\rho \downarrow 0$ eta $\rho \uparrow 1$, baina ez kargaren beste balioetarako, hurbilketa ona izan arren.

Jarraian hurbilketa teknikaren aplikagarritasuna azalduko dugu aztertutako hiru ereduetan.

Zerbitzari paralelodun eredua *bi aukeren abantaila* politikapean zaila dela aztertzen erakutsi du literaturak, ilaren artean duen menpekotasuna dela eta. Independentek diren ilarak edukitzeko, literaturan maiz zerbitzari kopurua limitean infinitoruntz doala onartu izan da, baina adierazpen itxiak oso murrizak dira. Emaizta gabeziak motibatuta batezbesteko sistemako iraute denbora aztertu dugu *trafiko arin* erregimenpean. Guk dakigula, *bi aukeren abantaila* politika hau ez da trafiko geldo erregimenpean aztertu. Hori dela eta ezin izan diogu interpolazio teknika aplikatu eredu honi.

DPS ereduaren analisia ere zaila dela ikusi da. Esate baterako, 1.1. Atalean aipatu bezala, batezbeste sisteman igarotako denbora bezalako metrika garrantzitsu baten inguruan oso emaitza gutxi ezagutzen dira eta denak era inplizituan ala limite batenpean. Bezero kopuruaren kasuan emaitzak beharrezko zerbitzuaren banaketa esponentziala den kasuan soilik ezagutzen dira. Zailtasun honek animatuta, ilara luzeraren probabilitate funtzio sortzailearen interpolazio bidezko hurbilketa eta itxaron denbora baldintzatuaren funtzio osagarriaren interpolazio bidezko hurbilketa lortzen ditugu.

Aurreko atalean aipatu bezala, Kim et al.-ek [26, 27] lehentasun erlatibodun eta DROS ereduaz, hurrenez hurren, bezero kopuruak eta itxaron denborak bete beharreko ekuazioak lortzen dituzte. Hori dela eta, trafiko arin eta trafiko geldopeko emaitzak zuzenean ekuazio horietatik lortu ditzakegu, ikusi 5 eta 6. Kapituluak. Ondoren, lortutako emaitzak konbinatuz interpolazio bidezko hurbilketa eraikiko dugu.

1.3 Tesiaren laburpena

Atal honetan tesiaren ideia orokor bat erakusten dugu. Kapitulu honetan aztertutako hiru ilara ereduak deskribatu eta erabiliriko hurbilketa teknika motibatu dugu.

2. Kapituluak trafiko-arin hurbilketa nola lortu, trafiko geldo erregimena aipatu eta interpolazio bidezko hurbilketa nola eraiki erakusten dugu.

3. Kapituluak zerbitzari paraleloen ereduak fokatzeko da, *bi aukeren abantaila* politikapean. Batez beste sisteman igarotako denboraren lehen hiru trafiko arinpeko deribatuak aurkezten ditugu. Ikusten dugu batez beste sisteman igarotako denbora beherakorra dela λ parametroarekiko. Hau, espero ez genuen aurkikuntza bat da, izan ere, normalean sisteman igarotako denbora monotonoki hazten da trafiko intentsitatea handitzen den heinean. Kapitulu honetako emaitzak [22] artikuluan argitaratuak izan dira.

4. Kapituluak DPS ereduak aztertzen dugu. Ilara luzeraren probabilitate funtzio sortzailea eta itxaron denbora baldintzatuaren funtzio osagarriaren trafiko-arin emaitzak lortzen ditugu. Eta

hauek [40] eta [48] artikuluetan lorturiko trafiko geldopeko emaitzekin konbinatuz interpolazio bidezko hurbilketak eraikitzen ditugu. Egonaldiaren batezbesteko iraute denboraren interpolazio bidezko hurbilketa sakonki aurkezten dugu. Analitikoki eta numerikoki hurbilketak nahikoa zehatzak direla ikusten dugu. Kapitulu honetako emaitzak [19, 18, 21] artikuluetan argitaratuak izan dira.

5. eta 6. Kapituluetan lehentasun erlatibodun ilara aztertzen dugu. 5. Kapitulan lehentasun erlatibodun ilararen analisisia aurkezten dugu. Limitean bezero kopuru eskalatua esponentzialki banaturiko zorizko aldagai baten eta bektore determinista baten biderkadura bezala banaturik dago. k klaseko itxaron denbora eskalatua ere esponentzialki banaturiko bi zorizko aldagaien biderkadura bezala banaturik dagoela erakusten dugu. 6. Kapitulan ilara luzeraren probabilitate funtzio sortzailearen eta itxaron denbora baldintzatuaren funtzio osagarriaren trafiko-arin emaitzak lortzen ditugu lehentasun erlatibodun eredurako. Hauek 5. Kapitulan lorturiko trafiko geldopeko emaitzekin konbinatuz interpolazio bidezko hurbilketak lortzen ditugu. Hurbilketen zehaztasuna lehen eta bigarren momentuen bitartez neurtzen dugu. Kapitulu honetako emaitzak [3, 20, 21] artikuluetan argitaratuak izan dira.

Tesi honetan erabilitako metodoa eraikitzailea da eta beste ilara sistema batzuei aplikagarria. Hori dela eta, teknika bera tesiaren ekarpen garrantzitsu bat dela uste dugu.

2. Kapituluia

Interpolazio bidezko hurbilketa

Kapitulu honetan *trafiko arin* erregimenpean errendimendua nola hurbildu deskribatzen dugu, *trafiko geldo* erregimena eztabaidatzen dugu eta *trafiko arinaren eta geldoaren interpolazioa* nola eraiki aurkezten dugu. Teknika hau Reimanek eta Simonek aurkeztu zuten, [43], hauen helburua batezbesteko bezero kopurua eta sistemako iraupen denbora hurbiltzea izanik. Gerora, teknika sistemako iraupen denboraren banaketa kalkulatzeko orokortua izan zen. Teknika hau hainbat sistemetara aplikatua izan da arrakastaz: *prozesadorea partekatzea*, etab.; ikusi literaturako adibideak [8, 47, 19, 12, 42] artikuluetan.

Aurreko kapituluan aipatu bezala interpolazio teknikak hurbilketa zehatza izatea ezartzen du trafiko-arin erregimenpean, λ 0-tik hurbil dagoenean, eta trafiko geldo erregimenpean, iritsiera tasa asetatsun puntutik hurbil dagoenean. 4. eta 6. Kapituluetan ikusiko dugu orokorrean hurbilketa nahikoa zehatza dela eta hainbat kasutan guztiz zehatza.

Izan bedi $G(\lambda, \vec{y})$ hurbildu nahi dugun metrika, λ iritsiera tasa eta $\vec{y}$ bektorearen menpekoa. G funtzioaren eta $\vec{y}$ bektorearen esanahia aldatu egingo da hurbildu nahi dugun metrikaren arabera. Tesi honetan hurrengo kasuak aztertuko ditugu:

(i) $(N_1, \dots, N_K)$ ilara luzeraren probabilitate funtzio sortailea, $\psi(\lambda, \vec{z}) := \mathbb{E} [z_1^{N_1} \dots z_K^{N_K}]$ bezala adierazia. Beraz, $\vec{y} = \vec{z}$,

(ii) k klaseko bezero baten itxaron denbora baldintzatuaren funtzio osagarria $W(\lambda, b, x) := \mathbb{P} [W_k(b) > x]$ bezala adierazia, non $W_k(b)$ k klaseko bezero baten itxaron denbora baldintzatua den. Beraz, $\vec{y} = (b, x)$,

(iii) k klaseko bezero baten itxaron denboraren Laplace-Stieltjes Transformatua (LST) $\widetilde{W}_k(\lambda, u) := \mathbb{E} [e^{-u(1-\lambda\mathbb{E}[B])W_k}]$ bezala adierazia, non W_k k klaseko bezero baten itxaron denbora ez-baldintzatua den. Beraz $\vec{y} = u$.

2.1. Atalean $G(\lambda, \vec{y})$ trafiko-arin erregimenpean karakterizatuko dugu, $\lambda \downarrow 0$. 2.2. Atalean trafiko geldo erregimena aurkezten dugu eta 2.3. Atalean interpolazio teknika.

2.1 Trafiko-arin analisia

Trafiko-arin erregimena sistemaren errendimendua λ iritsiera tasa oso txikia denean aztertzean datza, hau da, sistema ia hutsik dagoenean. $G(\lambda, \vec{y})$ Taylor serien bitartez hurbilduko dugu. λ aldagaiarekiko $G(\lambda, \vec{y})$ funtzioaren lehenengo n deribatuak $\lambda = 0$ puntuan existitzen direla onartuz, k klaseko sistemako egonaldiaren batezbesteko iraute denboraren honako hurbilketa dugu λ zerotik hurbil dagoen kasuan:

$$G^{LT}(\lambda, \vec{y}) := G^{(0)}(0, \vec{y}) + \lambda G^{(1)}(0, \vec{y}) + \cdots + \frac{\lambda^n}{n!} G^{(n)}(0, \vec{y}). \quad (2.1)$$

Honi n mailako trafiko-arin hurbilketa deituko diogu. $G^{(0)}(0, \vec{y}) := G(0, \vec{y})$ eta zerogarren trafiko arinpeko deribatua deituko diogu. Gainera, $G^{(m)}(0, \vec{y})$, $m = 1, 2, \dots, n$ λ -rekiko m -garren deribatua da $\lambda = 0$ puntuan, hau da, $G^{(m)}(0, \vec{y}) := \left. \frac{\partial^m G(\lambda, \vec{y})}{\partial \lambda^m} \right|_{\lambda=0}$.

Gure analisia Reiman eta Simon-en [43] artikuluan oinarritzen da, non deribatuak nola lortu erakusten den, zenbait baldintzapean. [43, A Eranskina] artikuluko analisia jarraituz honako hipotesia egiten dugu B beharrezko zerbitzuetarako

$$\mathbb{E}[e^{\eta B}] = \sum_{n=0}^{\infty} \frac{\eta^n}{n!} \mathbb{E}[B^n] < \infty, \quad (2.2)$$

$\eta > 0$ baterako. Momentu esponenziala finitua izateak beharrezko baldintza bat betetzea ziurtatzen du; behar baino gogorragoa da, baina gure analisia justifikatzen du.

Zero, bat eta bi trafiko-arin deribatuen adierazpenak jarraian ikus daitekeen proposizioan erakusten dira, [43, 50] artikuluetan lortu bezala. Azpimarratu nahi dugu adierazpen hauek zerbitzari paraleloko ereduarentzat (ikus 3.2.1 Proposizioa) eta DPS ilararentzat (ikus 4.3.1 eta 4.4.1 Lemak) erabili ditugula. Lehentasun erlatibodun ereduaren kasuan trafiko arinpeko deribatuak zuzenean [27] artikuluko ekuazio diferentzialetatik lortzen ditugu (ikus 6.1.2 eta 6.2.2 Lemak).

2.1.1 Proposizioa [43, 3. Atala], [50, 6.3. Kapitulu] *Izan bedi $A(s, t)$ $[s, t]$ tarteko iritsiera kopurua etiketatutako bezeroaz gain, zeina 0 unean iristen den. Izan bedi $G(\lambda, \vec{y}|A)$ errendimendu metrika A -n baldintzatua. Orduan, zero, bat eta bi trafiko arin deribatuak honela kalkula daitezke*

$$G^{(0)}(0, \vec{y}) = G\left(0, \vec{y} \middle| A(-\infty, \infty) = 0\right), \quad (2.3)$$

$$G^{(1)}(0, \vec{y}) = \int_{-\infty}^{\infty} \left(G\left(0, \vec{y} \middle| A(-\infty, \infty) = 1, \tau_1 = t\right) - G\left(0, \vec{y} \middle| A(-\infty, \infty) = 0\right) \right) dt \quad (2.4)$$

eta

$$\begin{aligned} & G^{(2)}(0, \vec{y}) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(G\left(0, \vec{y} \middle| A(-\infty, \infty) = 2, \tau_1 = t', \tau_2 = t''\right) - G\left(0, \vec{y} \middle| A(-\infty, \infty) = 1, \tau_1 = t'\right) \right. \\ & \quad \left. - G\left(0, \vec{y} \middle| A(-\infty, \infty) = 1, \tau_1 = t''\right) + G\left(0, \vec{y} \middle| A(-\infty, \infty) = 0\right) \right) dt' dt'', \end{aligned} \quad (2.5)$$

non τ_i , $i = 1, 2$, i -garren bezeroaren iritsiera unea den.

Jarraian zerogarren eta lehenengo trafiko arinpeko deribatuak nola lortu erakusten dugu. Garapen hau J. Walrand [50, 6.3. Kapitulu]n oinarriturik dago. Maila altuagoko trafiko-arin deribatuak era berean lor daitezke.

Kontsidera dezagun $-Z$ unean hasi eta T uneraino irauten duen sistema bat, $Z, T > 0$ ematik. Izan bedi $G(\lambda, \vec{y}, -Z, T)$ hurbildu nahi dugun metrika. Ohartu $\lim_{Z, T \rightarrow \infty} G(\lambda, \vec{y}, -Z, T) = G(\lambda, \vec{y})$. Izan bedi $A(s, t)$ $[s, t]$ iritsiera kopurua etiketatutako bezeroaz gain. Atal honetan Z eta T -rekiko limiteak eta itxaropenak trukakorrak direla onartzen dugu. Orduan,

$$G(\lambda, \vec{y}, -Z, T) = \sum_{a=0}^{\infty} G\left(\lambda, \vec{y}, -Z, T \middle| A(-Z, T) = a\right) \cdot \frac{(\lambda(T+Z))^a}{a!} e^{-\lambda(T+Z)}, \quad (2.6)$$

non $G\left(\lambda, \vec{y}, -Z, T \middle| A(-Z, T) = a\right)$ zeina a iritsiera izatean baldintzaturik dagoen. $\lambda = 0$ puntuan ebaluatuz

$$G(\lambda, \vec{y}, -Z, T) \Big|_{\lambda=0} = G\left(0, \vec{y}, -Z, T \middle| A(-Z, T) = 0\right), \quad (2.7)$$

eta orain $Z, T \rightarrow \infty$ limitea hartuz zerogarren trafiko arinpeko deribatua lortzen dugu

$$G^{(0)}(0, \vec{y}) := \lim_{Z, T \rightarrow \infty} G(\lambda, \vec{y}, -Z, T) \Big|_{\lambda=0} = G\left(0, \vec{y} \middle| A(-\infty, \infty) = 0\right)$$

non bigarren berdintza (2.7) ekuaziotik betetzen den.

Jarraian, (2.6) Ekuazioan λ -rekiko deribatua hartuz eta $\lambda = 0$ puntuan ebaluatuz, lortzen duguna da

$$\begin{aligned} & \frac{\partial}{\partial \lambda} G(\lambda, \vec{y}, -Z, T) \Big|_{\lambda=0} \\ &= -G\left(0, \vec{y}, -Z, T \middle| A(-Z, T) = 0\right) \cdot (T+Z) + G\left(0, \vec{y}, -Z, T \middle| A(-Z, T) = 1\right) \cdot (T+Z) \\ &= \int_{-Z}^T \left(G\left(0, \vec{y}, -Z, T \middle| A(-Z, T) = 1, \tau_1 = t\right) - G\left(0, \vec{y}, -Z, T \middle| A(-Z, T) = 0\right) \right) dt, \end{aligned} \quad (2.8)$$

non τ_1 lehenengo bezeroaren iritsiera unea den. Bigarren berdintza iritsierek Poisson prozesu bat jarraitzen dutelako betetzen da. Beraz, $[-Z, T)$ tartean iritsiera bakarra izan denez ($A(-Z, T) = 1$), τ_1 $[-Z, T)$ tartean uniformeki banatua dagoela ondoriozta dezakegu.

Orain $Z, T \rightarrow \infty$ limiteak hartuz lehenengo trafiko arinpeko deribatua lortzen dugu

$$\begin{aligned}
G^{(1)}(0, \vec{y}) &:= \lim_{Z, T \rightarrow \infty} \frac{\partial}{\partial \lambda} G(\lambda, \vec{y}, -Z, T) \Big|_{\lambda=0} \\
&= \int_{-\infty}^{\infty} \left(G\left(0, \vec{y} \mid A(-\infty, \infty) = 1, \tau_1 = t\right) - G\left(0, \vec{y} \mid A(-\infty, \infty) = 0\right) \right) dt,
\end{aligned}$$

non bigarren berdintza (2.8) Ekuaziotik betetzen den.

2.2 Trafiko-geldo erregimena

Trafiko-geldo erregimenak sistema asetahun puntutik gertu aztertzen du. Beste modu batean esanda, sistemara iristen den lan kopurua sistemaren kapazitatetik hurbil dagoenean. Helburua aztertu nahi dugun metrika eskalatua lortzea da trafiko geldo erregimenpean. Eskala hori $f_\lambda(\vec{y})$ bezala adieraziko dugu eta, beraz, $G(\lambda, f_\lambda(\vec{y}))$ adierazpena lortzea izango da gure helburua.

Zerbitzari bakarreko sistema asetahun puntutik gertu egongo da karga batetik gertu dagoenean, $\rho \uparrow 1$. Beraz, erregimen hau honela lor daiteke

$$\lambda \uparrow \hat{\lambda} := \frac{1}{\mathbb{E}[B]}.$$

Trafiko geldopeko analisia hainbat artikuluetan gauzatu da, ikusi [40, 28, 48] artikulua.

4. Kapituluhan DPS ereduarentzat lortu diren trafiko geldopeko emaitzen laburpen bat egiten dugu eta 5 . Kapituluhan lehen tasun erlatibodun ereduarentzat trafiko-geldo emaitzak lortzen ditugu.

Trafiko-geldo erregimenpean metrika eskalatuen ondoko adierazpenak ditugu:

- (i) $\psi(\lambda, \vec{z}^{1-\lambda\mathbb{E}[B]}) = \mathbb{E}[\vec{z}^{(1-\lambda\mathbb{E}[B])\vec{N}}],$
- (ii) $W_k(\lambda, b, x/(1 - \lambda\mathbb{E}[B])) = \mathbb{P}[(1 - \lambda\mathbb{E}[B])W_k(b) > x],$
- (iii) $\widetilde{W}_k(\lambda, u(1 - \lambda\mathbb{E}[B])) = \mathbb{E}[e^{-u(1-\lambda\mathbb{E}[B])W_k}],$

non $\vec{z}^{\gamma\vec{N}} := (z_1^{\gamma N_1}, \dots, z_K^{\gamma N_K})$ notazioa erabiltzen dugun. Metrikaren arabera, $f_\lambda(\vec{y})$ eskala desberdina izan daiteke

- (i) $f_\lambda(\vec{z}) = \vec{z}^{1-\lambda\mathbb{E}[B]},$
- (ii) $f_\lambda(b, x) = (b, x/(1 - \lambda\mathbb{E}[B])),$
- (iii) $f_\lambda(u) = u(1 - \lambda\mathbb{E}[B]).$

Izan bedi $G^{HT}(\vec{y})$ jarraian definituriko trafiko-geldo terminoa

$$G^{HT}(\vec{y}) := \lim_{\lambda \uparrow 1/\mathbb{E}[B]} G(\lambda, f_\lambda(\vec{y})). \quad (2.9)$$

2.3 Interpolazio bidezko hurbilketa

Metrikaren trafiko arin eta trafiko-geldoa adierazpenak jakinak diren kasuan, interpolazio teknikaren bitartez edozein λ -rako balio dezakeen hurbilketa bat eraiki daiteke. Teknika hau Reiman eta Simon-ek aurkeztu zuten, [41, 42, 43], eta metrika eskalatua, $G(\lambda, f_\lambda(\vec{y}))$, $n+1$ mailako polinomio batez hurbiltzean datza, $\hat{G}(\lambda, \vec{y})$:

$$\hat{G}(\lambda, \vec{y}) = h_0(\vec{y}) + h_1(\vec{y})\lambda + h_2(\vec{y})\lambda^2 + \dots + h_{n+1}(\vec{y})\lambda^{n+1}. \quad (2.10)$$

Normalizazioa deseginez $G(\lambda, \vec{y})$ metrikaren interpolazio bidezko hurbilketa lortzen dugu:

$$G^{INT}(\lambda, \vec{y}) := \hat{G}(\lambda, f_\lambda^{-1}(\vec{y})) \quad (2.11)$$

non $0 \leq \lambda < 1/\mathbb{E}[B]$.

$h_0(\vec{y}), \dots, h_n(\vec{y})$ koefizienteak lortzeko (2.11) Ekuazioan m -garren, $m = 0, \dots, n$, deribatua hartu eta hurbildu nahi dugun metrikaren m -garren deribatuarekin berdinduko dugu. Eta ondoko *trafiko arin* baldintzak lortzen ditugu:

$$\left. \frac{\partial^m G^{INT}(\lambda, \vec{y})}{\partial \lambda^m} \right|_{\lambda=0} = G^{(m)}(0, \vec{y}), \text{ for } m = 0, \dots, n. \quad (2.12)$$

Ohartu $G^{(m)}(0, \vec{y})$ funtzioaren adierazpenak 2.1.1 Proposizioan aurkeztu ditugula. $h_{n+1}(\vec{y})$ zehazteko trafiko-geldoa baldintza erabiliko dugu:

$$\hat{G}(1/\mathbb{E}[B], \vec{y}) = G^{HT}(\vec{y}), \quad (2.13)$$

non $G^{HT}(\vec{y})$ (2.9) Ekuazioan definituriko trafiko-geldoa emaitza den. Ondoren ikusiko dugu $h_0(\vec{y}), \dots, h_{n+1}(\vec{y})$ koefizienteak nola zehaztu praktikan. (2.11) hurbilketa $n+1$ mailako interpolazio bidezko hurbilketa moduan deituko dugu.

Jarraian, tesian zehar erabiliko dugun interpolazio bidezko hurbilketaren inguruko emaitza bat aurkeztuko dugu.

2.3.1 Proposizioa $n+1$ mailako interpolazio bidezko hurbilketa honela idatz daiteke

$$G^{INT}(\lambda, \vec{y}) = \sum_{i=0}^n \lambda^i (1 - (\lambda \mathbb{E}[B])^{n+1-i}) h_i(f_\lambda^{-1}(\vec{y})) + (\lambda \mathbb{E}[B])^{n+1} G(1/\mathbb{E}[B], \vec{y}). \quad (2.14)$$

Frogapena: (2.13) trafiko-geldoa baldintzatik ondokoa lortzen dugu

$$h_{n+1}(\vec{y}) = \mathbb{E}[B]^{n+1} \left(G(1/\mathbb{E}[B], f_{1/\mathbb{E}[B]}(\vec{y})) - \sum_{i=0}^n \frac{h_i(\vec{y})}{\mathbb{E}[B]^i} \right)$$

(2.14) Ekuazioa adierazpen hau (2.10) Ekuazioan ordezkatu eta (2.11) ekuazioan bezala normalizazioa deseginez lortzen da. ■

$G(\lambda, \vec{y})$ funtzioak sistemako egonaldiaren denbora ez-baldintzatuen banaketa adierazten duenean, $\bar{S}(\lambda, x) := \mathbb{P}[(1 - \lambda \mathbb{E}[B])S_k > x]$, non S_k k klaseko sistemako egonaldiaren denbora

den, (2.14) Ekuazioak [12] artikuluko lehenengo ekuazioarekin bat egiten du normalizazioa desegin ostean.

Behaketa garrantzitsu bat da metriken probabilitate funtzio sortzaileek eta LST-en interpolazio hurbilketek ez dutela beti zorizko aldagai batekin bat egiten, hau da, ez dira beti monotonoak izango, ikusi [11, XIII.4. Atala]. Dena den, momentuen hurbilketa egokiak izan daitezkeela ikusiko dugu.

V aldagaiaren batazbestekoa hurbiltzen ari garen kasuan (2.11) Ekuazioa trafiko-arin eta trafiko-geldo termininoen menpe idatz daiteke. Kasu honetan $G(\lambda) := \mathbb{E}[V]$ definitzen dugu eta trafiko-geldo terminoa $G^{HT} := \lim_{\lambda \uparrow 1/\mathbb{E}[B]} f_\lambda \cdot G(\lambda)$ da, non f_λ den trafiko geldo eskala egokia. Beraz, $\vec{y}$ aldagaiarekiko menpekotasuna galtzen da. Gure jakintzaren baitan, emaitza hau berria da eta zerbitzari bakarreko edozein ereduri aplikagarria.

2.3.1 Korolaria $n+1$ mailako batazbestekoaren interpolazio bidezko hurbilketa honela idatz daiteke

$$\begin{aligned} G^{INT}(\lambda) &= \frac{1}{f_\lambda} \left(\sum_{i=0}^n \frac{\lambda^i}{i!} (1 - (\lambda \mathbb{E}[B])^{n+1-i}) \sum_{j=0}^i \binom{i}{j} \frac{d^{i-j} f_\lambda}{d\lambda^{i-j}} \Big|_{\lambda=0} G^{(j)}(0) + (\lambda \mathbb{E}[B])^{n+1} G^{HT} \right), \end{aligned} \quad (2.15)$$

non $G^{(i)}(\lambda)$ i -garren, $i = 0, \dots, n$ trafiko arinpeko deribatua den.

Frogapena: (2.12) trafiko-arin baldintzatik eta normalizazioa desegin ostean ondokoa lortzen dugu

$$h_i(f_\lambda^{-1}) = \frac{\sum_{j=0}^i \binom{i}{j} \frac{d^{i-j} f_\lambda}{d\lambda^{i-j}} \Big|_{\lambda=0} G^{(j)}(0)}{i! f_\lambda}, i = 0, 1, \dots, n.$$

(2.15) Ekuazioa adierazpen hauek (2.14) Ekuazioan ordezkatzuz lortzen da. ■

3. Kapitulu

Power-of-two politikaren trafikoko arin analisi¹

In this chapter we focus on the heterogeneous parallel-server model under the power-of-two policy, described in Section 1.1.2. Due to intractability of the system in general, we apply the light-traffic approximation technique described in Section 2.1 and we point to interesting structural features associated with server heterogeneity in light traffic.

This chapter is organised as follow. In Section 3.1 we describe the model analysed and provide the needed mathematical framework to calculate the light-traffic derivatives. In Section 3.2 we gather the main results of this chapter. Sections 3.3, 3.5 and 3.6 contain the proofs of the zeroth, first and second light-traffic derivatives, respectively. In Section 3.7 we present the corresponding numerical results.

3.1 Ereduaren deskribapena

The system comprises K parallel servers labelled $k = 1, \dots, K$. Server k , which has capacity C_k , is attended by an infinite capacity buffer and operates in a FCFS manner. Customers arrive according to a Poisson process of rate $\lambda > 0$, each customer bringing a random amount of work distributed according to the probability distribution F . Upon arrival, an incoming customer is assigned to one of the K servers according to the power-of-two load balancing policy: *Two* servers are selected at random from the available pool of K servers. The JSQ policy is then used in isolation with these two servers; ties are broken randomly (but other choices are possible). As usual, the Poisson process of arrivals, the sequence of customer requirement random variables and the sequence of server selection random variables are mutually independent. Moreover, each of the two sequences is a sequence of i.i.d. random variables (across customers).

Let the random variable B be the service requirement of an arbitrary arriving customer. Let S^{SQ2} denote the stationary sojourn time of a customer at arbitrary epoch. Set the mean sojourn time

$$\overline{S^{SQ2}}(\lambda) = \mathbb{E} [S^{SQ2}] .$$

We expect $\overline{S^{SQ2}}(\lambda) < \infty$ over some non-degenerate interval $[0, \lambda^*)$ for some finite $\lambda^* > 0$; see also

¹Kapitulu hau ez da euskarara itzulita izan, baina ez du eragozpenik eragiten tesiaren gainerakoa irakurri eta ulertzeko.

[35, Section 2.1, Lemma 1] and [36, Lemma 6]. In what follows we shall not be concerned with this issue any further since we are mainly interested in the situation where $\lambda \simeq 0$.

We find it convenient to introduce the following terminology and notation: With t in $\mathbb{R}$, a customer arriving at time t , hereafter referred to as a t -customer, has two random variables B_t and Σ_t associated with it – The $\mathbb{R}_+$ -valued random variable B_t stipulates the amount of work requested by the t -customer from the system, while the random variable Σ_t is an (unordered) pair of servers from amongst the K available servers. The t -customer is assigned to a server ν_t selected in Σ_t according to the power-of-two policy (with a random tie-breaker). We shall refer to the random variables (B_t, Σ_t) as the characteristic pair of the t -customer.

As expected, we sometimes refer to the 0-customer with characteristics (B_0, Σ_0) as the *tagged* customer. The Reiman-Simon approach to light traffic focuses on the performance of this tagged customer under scenarios of increasing complexity. To define them, fix $n = 0, 1, \dots$. Interpret every n -uple $(t_1, \dots, t_n)$ in $\mathbb{R}^n$ as the arrival epochs of n customers into the system. For each $i = 1, \dots, n$, we lighten the notation by denoting the characteristic pair (B_{t_i}, Σ_{t_i}) of the t_i -customer arriving at time t_i simply by (B_i, Σ_i) . Throughout the following conditions are assumed to be enforced:

1. The random variables $\{B_0, B_1, \dots, B_n\}$ are i.i.d. $\mathbb{R}_+$ -valued random variables, each distributed according to the probability distribution F , namely

$$\mathbb{P}[B_i \leq x] = F(x), \quad \begin{array}{l} x \geq 0, \\ i = 0, 1, \dots, n. \end{array}$$

2. The random variables $\{\Sigma_0, \Sigma_1, \dots, \Sigma_n\}$ are $\mathcal{P}_2(K)$ -valued uniformly distributed random variables with

$$\mathbb{P}[\Sigma_i = T] = \frac{1}{\binom{K}{2}}, \quad T \in \mathcal{P}_2(K), \quad i = 0, 1, \dots, n.$$

3. The collections of random variables $\{B_0, B_1, \dots, B_n\}$ and $\{\Sigma_0, \Sigma_1, \dots, \Sigma_n\}$ are mutually independent.

We shall also have use for the random variables $\nu_0^*, \nu_1^*, \dots, \nu_n^*$ associated with the random pairs $\Sigma_0, \Sigma_1, \dots, \Sigma_n$, and defined in the following manner: For each $i = 0, 1, \dots, n$, the random variable ν_i^* is an Σ_i -valued random variable with

$$[\nu_i^* | \Sigma_i] \sim \mathcal{U}(\Sigma_i).$$

It is always understood that the random variables $\nu_0^*, \nu_1^*, \dots, \nu_n^*$ are conditionally mutually independent given the $2(n+1)$ random variables $B_0, B_1, \dots, B_n, \Sigma_0, \Sigma_1, \dots, \Sigma_n$. Under the enforced assumptions, we readily conclude that the random variables $\nu_0^*, \nu_1^*, \dots, \nu_n^*$ are i.i.d. random variables, each of which is uniformly distributed on $\{1, \dots, K\}$.

3.2 Sistemako egonaldiaren batezbesteko iraute denbora

When computing the light-traffic derivatives, for each $(t_1, \dots, t_n)$ in $\mathbb{R}^n$, let the random variable $S_n(t_1, \dots, t_n)$ denote the sojourn time of the tagged customer under the scenario that *in addition* to the tagged customer, only n customers are allowed to enter the system over $\mathbb{R}$,

say at times $t_1, \dots, t_n$, with characteristic pairs $(B_1, \Sigma_1), \dots, (B_n, \Sigma_n)$ as defined earlier. Note that $S_n(t_1, \dots, t_n)$ depends on the random variables $\{B_0, B_1, \dots, B_n\}$, $\{\Sigma_0, \Sigma_1, \dots, \Sigma_n\}$ and $\{\nu_0^*, \nu_1^*, \dots, \nu_n^*\}$ in a complicated manner through the scheduling policy used. We shall write

$$\widehat{S}_n(t_1, \dots, t_n) := \mathbb{E}[S_n(t_1, \dots, t_n)]. \quad (3.1)$$

For the power-of-two policy there is no characterisation available for the mean sojourn time with heterogeneous servers. Thus, in the next proposition we state the second order light-traffic approximation using the result given in Proposition 2.1.1. The proof method is constructive and it could be applied to other queueing systems for which no analytical results are available.

The results can be expressed compactly with the help of a random variable X which is uniformly distributed over the set of values $\frac{1}{C_1}, \dots, \frac{1}{C_K}$, i.e.,

$$\mathbb{P}\left[X = \frac{1}{C_1}\right] = \dots = \mathbb{P}\left[X = \frac{1}{C_K}\right] = \frac{1}{K}.$$

It is plain that

$$\mathbb{E}[X^p] = \frac{1}{K} \sum_{k=1}^K \frac{1}{C_k^p}, \quad p = 1, 2, \dots$$

3.2.1 Propozizioa *The light-traffic approximation of the mean sojourn time under the power-of-two policy satisfies*

$$\overline{SQ2,LT}(\lambda) = \left(\overline{SQ2}\right)^{(0)}(0) + \lambda \left(\overline{SQ2}\right)^{(1)}(0) + \frac{\lambda^2}{2} \left(\overline{SQ2}\right)^{(2)}(0) \quad (3.2)$$

with

$$\left(\overline{SQ2}\right)^{(0)}(0) = \mathbb{E}[X] \cdot \mathbb{E}[B], \quad (3.3)$$

$$\left(\overline{SQ2}\right)^{(1)}(0) = -\frac{1}{K-1} \text{Var}[X] \cdot \mathbb{E}[B]^2 \quad (3.4)$$

and

$$\left(\overline{SQ2}\right)^{(2)}(0) = \frac{2}{(K-1)^2} \left(\mathbb{E}[X]^3 - 2\mathbb{E}[X] \mathbb{E}[X^2] + \mathbb{E}[X^3] \right) \cdot \mathbb{E}[B]^3. \quad (3.5)$$

Frogapena: The proof follows from Proposition 2.3.1, together with the expressions obtained for the zeroth, first and second light-traffic derivatives, derived in Sections 3.3, 3.5.2 and 3.6.3, respectively. ■

Several observations readily flow from Proposition 3.2.1.

Equal capacities: From (3.4) it is plain that $\left(\overline{SQ2}\right)^{(1)}(0) \leq 0$, with $\left(\overline{SQ2}\right)^{(1)}(0) = 0$ if and only if $\text{Var}[X] = 0$, or equivalently, $C_1 = \dots = C_K \equiv C$ – In that case we also have $\left(\overline{SQ2}\right)^{(2)}(0) = 0$ with

$$\overline{SQ2,LT}(\lambda) = \frac{\mathbb{E}[B]}{C} + o(\lambda^2).$$

Unequal capacities: When the capacities are different, then $\left(\overline{S^{SQ2}}\right)^{(1)}(0) < 0$ and $\overline{S^{SQ2,LT}}(\lambda)$ is *decreasing* for small values of λ . This is a somewhat unexpected finding because most queueing systems are “monotone” in the sense that increasing the traffic intensity λ results in an increase in a performance metric such as the mean sojourn time.

This fact can be explained as follows: On the average, a customer entering an empty system experiences a sojourn time given by $\left(\overline{S^{SQ2}}\right)^{(0)}(0)$ since in such a circumstance the scheduling policy will assign it to *any* of the K servers with probability $\frac{1}{K}$. However, in the presence of different server capacities, the assigned server may not have been the fastest and it is possible for subsequent customers to be served by faster servers by the luck of the draw. This will result in a decrease in the mean sojourn time if the traffic intensity increases slightly but still allows for some faster server to be starved with some non-negligible probability.

Intuitively we expect this decrease to be more pronounced if the values of $C_1, \dots, C_k$ are more “spread out”. Indeed, for a *given* average value

$$\frac{1}{K} \sum_{k=1}^K \frac{1}{C_k},$$

the decrease will be steeper the more “unbalanced” the values of $\frac{1}{C_1}, \dots, \frac{1}{C_k}$ are in the sense of majorization; this follows from (3.4) by majorization arguments and Schur-convexity (using the convexity of $t \rightarrow t^2$) [34, Prop. C.1, p. 64].

Only $\mathbb{E}[B]$ matters – The two first derivatives depend only on the first moment of B , and could be read as a form of insensitivity in light traffic. This is in sharp contrast with other systems where the first light-traffic derivative depends on $\mathbb{E}[B^2]$, e.g., $M/G/1$ -like queues [41] and the discriminatory processor sharing model [19]. This is rather unexpected because the variance of B is known to be a key factor in shaping JSQ performance with homogeneous servers under FCFS scheduling [4, Chapter 24]. See next item for a possible explanation.

FCFS vs. PS – Proposition 3.2.1 was established under the assumption that the servers operate under the FCFS discipline. However, it is easy to see that both (3.3) and (3.4) (but not (3.5)) are still valid if the servers all use the PS discipline. That the variability of B seems to play little role in light traffic is therefore consistent with the aforementioned fact that performance under the PS discipline is nearly insensitive to service variability [15].

3.3 Zerogarren trafiko-arin deribatua

The zeroth light-traffic derivative corresponds to the scenario when besides the tagged customer, no other customer enters over the entire horizon $(-\infty, \infty)$. Let S_0 denote the sojourn time of the tagged customer under these circumstances. Obviously, under the power-of-two scheduling strategy, we have

$$S_0 = \frac{B_0}{C_{\nu_0}} \quad \text{with } \nu_0 = \nu_0^*, \quad (3.6)$$

because in the absence of any other customer in the system, the tagged customer is necessarily assigned to server ν_0^* . In analogy with earlier notation we write

$$\widehat{S}_0 = \mathbb{E}[S_0(t)].$$

3.3.1 Proposizioa *Under the enforced assumptions, the random variable ν_0^* is uniformly distributed over $\{1, \dots, K\}$ with*

$$\mathbb{P}[\nu_0^* = k] = \frac{1}{K}, \quad k = 1, \dots, K \quad (3.7)$$

and the relation

$$\widehat{S}_0 = \left(\frac{1}{K} \sum_{k=1}^K \frac{1}{C_k} \right) \cdot \mathbb{E}[B] \quad (3.8)$$

holds.

It will often be convenient to write (3.8) more compactly as

$$\widehat{S}_0 = \frac{\Gamma}{K} \cdot \mathbb{E}[B], \quad (3.9)$$

with

$$\Gamma = \sum_{k=1}^K \frac{1}{C_k} \quad (3.10)$$

Proof of Proposition 3.3.1 For each $k = 1, \dots, K$, the definition of ν_0^* gives

$$\begin{aligned} \mathbb{P}[\nu_0^* = k] &= \sum_{\ell=1, \ell \neq k}^K \mathbb{P}[\Gamma_0 = \{k, \ell\}, \nu_0^* = k] \\ &= \sum_{\ell=1, \ell \neq k}^K \mathbb{P}[\nu_0^* = k | \Gamma_0 = \{k, \ell\}] \mathbb{P}[\Gamma_0 = \{k, \ell\}] \\ &= (K-1) \cdot \frac{1}{2} \cdot \frac{2}{K(K-1)} = \frac{1}{K}. \end{aligned} \quad (3.11)$$

As pointed earlier, we necessarily have $\nu_0 = \nu_0^*$. The random variables ν_0^* and B_0 being independent, we then obtain from (3.6) that

$$\widehat{S}_0 = \mathbb{E} \left[\frac{B_0}{C_{\nu_0^*}} \right] = \mathbb{E}[B_0] \cdot \mathbb{E} \left[\frac{1}{C_{\nu_0^*}} \right],$$

and the conclusion (3.8) readily follows from (3.7). ■

We have $\left(\overline{S^{SQ2}} \right)^{(0)}(0) = \widehat{S}_0$ according to Equation (2.3), hence Equation (3.3) is established with the help of (3.8).

3.4 Emaizta lagungarri bat

Obtaining the light-traffic derivatives of order 1 and 2 is computationally more involved. The technical result discussed next will simplify the presentation by isolating an evaluation which is repeatedly carried out during the analysis. This auxiliary result is given in a setting that mimics power-of-two scheduling with only two customers present:

Fix $y < 0$. In addition to the tagged customer arriving at time $t = 0$ with characteristic pair (B_0, Γ_0) , assume that another customer arrives at time y with (random) service requirement B . The y -customer is then assigned to the server γ , with γ being some $\{1, \dots, K\}$ -valued random variable, while the tagged customer is assigned to the server γ_0 (in Γ_0) in accordance with the power-of-two scheduling policy. Thus, if $y + \frac{B}{C_\gamma} \leq 0$, then $\gamma_0 = \nu_0^*$. On the other hand, if $y + \frac{B}{C_\gamma} > 0$, then the operational rules of the power-of-two scheduling policy preclude the tagged customer to be assigned to server γ : If γ is not in Γ_0 , then $\gamma_0 = \nu_0^*$ again, while if γ is an element of Γ_0 , then γ_0 is necessarily the other server in the pair Γ_0 , i.e., the one different from γ . In this scenario γ is given *a priori*, and should be thought as a place holder for a server assignment random variable determined via power-of-two scheduling under various circumstances. On the other hand, γ_0 depends on y , B , γ and Γ_0 (as well as ν_0^*). The explicit dependence on these quantities will be dropped from the notation.

For reasons that will become apparent in subsequent developments, we also introduce an event E (to be specified later).

3.4.1 Lema *Given are the random variables $\mathbf{1}[E]$, B , γ , B_0 , Γ_0 and ν_0^* . We assume that (i) the random variable ν_0^* is uniformly distributed on Γ_0 conditionally on all the other random variables $\mathbf{1}[E]$, B , γ , B_0 and Γ_0 ; (ii) the collections of random variables $\{\mathbf{1}[E], B, \gamma\}$ and $\{\nu_0^*, \Gamma_0, B_0\}$ are independent; and (iii) the random variables Γ_0 and B_0 are independent. Then, for each $y < 0$ and each $k = 1, \dots, K$, we have*

$$\begin{aligned} \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \frac{B_0}{C_{\gamma_0}} \right] &= \mathbb{P} \left[E, \gamma = k, y + \frac{B}{C_k} \leq 0 \right] \cdot \widehat{S}_0 \\ &\quad + \frac{1}{K-1} \mathbb{P} \left[E, \gamma = k, y + \frac{B}{C_k} > 0 \right] \left(\Gamma - \frac{1}{C_k} \right) \cdot \mathbb{E}[B_0], \end{aligned} \quad (3.12)$$

with γ_0 defined earlier.

Note that under the enforced assumptions, the random variable B_0 is independent of the collection of random variables $\{\nu_0^*, \Gamma_0\}$.

Proof of Lemma 3.4.1: Fix $k = 1, \dots$. We start with the natural decomposition

$$\begin{aligned} &\mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \frac{B_0}{C_{\gamma_0}} \right] \\ &= \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_\gamma} \leq 0 \right] \frac{B_0}{C_{\gamma_0}} \right] + \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_\gamma} > 0 \right] \frac{B_0}{C_{\gamma_0}} \right]. \end{aligned} \quad (3.13)$$

For the first term, the definition of γ_0 given above leads to

$$\begin{aligned} \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_\gamma} \leq 0 \right] \frac{B_0}{C_{\gamma_0}} \right] &= \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} \leq 0 \right] \frac{B_0}{C_{\nu_0^*}} \right] \\ &= \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} \leq 0 \right] \right] \mathbb{E} \left[\frac{B_0}{C_{\nu_0^*}} \right] \\ &= \mathbb{P} \left[E, \gamma = k, y + \frac{B}{C_k} \leq 0 \right] \cdot \widehat{S}_0, \end{aligned} \quad (3.14)$$

since the collections $\{\mathbf{1}[E], \gamma, B\}$ and $\{\nu_0^*, B_0\}$ are independent under the enforced assumptions.

We further decompose the second term to obtain

$$\begin{aligned}
& \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_\gamma} > 0 \right] \frac{B_0}{C_{\gamma_0}} \right] \\
&= \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} > 0 \right] \mathbf{1}[k \notin \Gamma_0] \frac{B_0}{C_{\nu_0^*}} \right] \\
&\quad + \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} > 0 \right] \mathbf{1}[k \in \Gamma_0] \frac{B_0}{C_{\gamma_0}} \right]. \tag{3.15}
\end{aligned}$$

It is plain that

$$\begin{aligned}
& \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} > 0 \right] \mathbf{1}[k \in \Gamma_0] \frac{B_0}{C_{\gamma_0}} \right] \\
&= \sum_{\ell=1, \ell \neq k}^K \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} > 0 \right] \mathbf{1}[\Gamma_0 = \{k, \ell\}] \frac{B_0}{C_{\gamma_0}} \right] \\
&= \sum_{\ell=1, \ell \neq k}^K \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} > 0 \right] \mathbf{1}[\Gamma_0 = \{k, \ell\}] \frac{B_0}{C_\ell} \right] \\
&= \frac{2}{K(K-1)} \sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[E, \gamma = k, y + \frac{B}{C_k} > 0 \right] \frac{\mathbb{E}[B_0]}{C_\ell} \\
&= \frac{2}{K(K-1)} \left(\sum_{\ell=1, \ell \neq k}^K \frac{\mathbb{E}[B_0]}{C_\ell} \right) \mathbb{P} \left[E, \gamma = k, y + \frac{B}{C_k} > 0 \right] \\
&= \frac{2}{K(K-1)} \left(\Gamma - \frac{1}{C_k} \right) \mathbb{P} \left[E, \gamma = k, y + \frac{B}{C_k} > 0 \right] \cdot \mathbb{E}[B_0], \tag{3.16}
\end{aligned}$$

since $\gamma_0 = \ell$ if $\Gamma_0 = \{k, \ell\}$ when $\gamma = k$ and $y + \frac{B}{C_k} > 0$.

On the other hand, the definition of ν_0^* implies

$$\begin{aligned}
& \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} > 0 \right] \mathbf{1}[k \notin \Gamma_0] \frac{B_0}{C_{\nu_0^*}} \right] \\
&= \sum_{T \in \mathcal{P}_2(K)} \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} > 0 \right] \mathbf{1}[k \notin \Gamma_0] \mathbf{1}[\Gamma_0 = T] \frac{B_0}{C_{\nu_0^*}} \right] \\
&= \sum_{a=1, a \neq k}^K \left(\sum_{b=1, b \neq k}^{a-1} \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} > 0 \right] \mathbf{1}[\Gamma_0 = \{a, b\}] \frac{B_0}{C_{\nu_0^*}} \right] \right) \\
&= \sum_{a=1, a \neq k}^K \left(\sum_{b=1, b \neq k}^{a-1} \mathbb{P} \left[E, \gamma = k, y + \frac{B}{C_k} > 0 \right] \mathbb{E} \left[\mathbf{1}[\Gamma_0 = \{a, b\}] \frac{B_0}{C_{\nu_0^*}} \right] \right) \\
&= \sum_{a=1, a \neq k}^K \mathbb{P} \left[E, \gamma = k, y + \frac{B}{C_k} > 0 \right] \left(\sum_{b=1, b \neq k}^{a-1} \mathbb{P}[\Gamma_0 = \{a, b\}] \cdot \frac{1}{2} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \right) \cdot \mathbb{E}[B_0] \\
&= \frac{1}{K(K-1)} \sum_{a=1, a \neq k}^K \mathbb{P} \left[E, \gamma = k, y + \frac{B}{C_k} > 0 \right] \left(\sum_{b=1, b \neq k}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \right) \cdot \mathbb{E}[B_0]. \quad (3.17)
\end{aligned}$$

In Appendix 3.A.1 we show that

$$\sum_{a=1, a \neq k}^K \left(\sum_{b=1, b \neq k}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \right) = (K-2) \left(\Gamma - \frac{1}{C_k} \right), \quad (3.18)$$

and (3.17) can be written more compactly as

$$\begin{aligned}
& \mathbb{E} \left[\mathbf{1}[E] \mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_k} > 0 \right] \mathbf{1}[k \notin \Gamma_0] \frac{B_0}{C_{\nu_0^*}} \right] \\
&= \frac{K-2}{K(K-1)} \left(\Gamma - \frac{1}{C_k} \right) \mathbb{P} \left[E, \gamma = k, y + \frac{B}{C_k} > 0 \right] \cdot \mathbb{E}[B_0]. \quad (3.19)
\end{aligned}$$

To conclude the proof, substitute (3.16) and (3.19) into (3.15), yielding

$$\mathbb{E} \left[\mathbf{1}[\gamma = k] \mathbf{1} \left[y + \frac{B}{C_\gamma} > 0 \right] \frac{B_0}{C_{\gamma_0}} \right] = \frac{1}{K-1} \left(\Gamma - \frac{1}{C_k} \right) \mathbb{P} \left[\gamma = k, y + \frac{B}{C_k} > 0 \right] \cdot \mathbb{E}[B_0].$$

Combining this last expression with (3.14) we get the desired result (3.12) with the help of (3.13). $\blacksquare$

3.5 Lehenengo trafiko-arin deribatua

The analysis of the first derivative is associated with the following scenario: The tagged customer arrives at time $t = 0$ with characteristic pair (B_0, Γ_0) . With t in $\mathbb{R}$, in addition to the tagged customer, only a single customer arrives during the entire horizon $(-\infty, \infty)$, say at time t with characteristic pair (B_t, Γ_t) . The tagged customer and this t -customer are assigned to the servers ν_0 (in Γ_0) and ν_t (in Γ_t), respectively, in accordance with the power-of-two scheduling policy.

3.5.1 Evaluating $\widehat{S}_1(t)$

For each t in $\mathbb{R}$, we have

$$\widehat{S}_1(t) = \mathbb{E}[S_1(t)] \quad \text{with } S_1(t) = \frac{B_0}{C_{\nu_0}}. \quad (3.20)$$

However, with the presence of the t -customer, ν_0 does not always coincide with ν_0^* , as the determination of ν_0 may be affected by whether the t -customer completed service at the time the tagged customer arrives.

First some notation: With t arbitrary in $\mathbb{R}$, set

$$H_k(t) = \mathbb{P}[C_k t + B_t \leq 0] \cdot \widehat{S}_0 + \frac{1}{K-1} \left(\Gamma - \frac{1}{C_k} \right) \mathbb{P}[C_k t + B_t > 0] \cdot \mathbb{E}[B_0] \quad (3.21)$$

for each $k = 1, \dots, K$.

3.5.1 Proposizioa *Under the enforced independence assumptions, it holds that*

$$\widehat{S}_1(t) = \widehat{S}_0, \quad t > 0 \quad (3.22)$$

and

$$\widehat{S}_1(t) = \frac{1}{K} \sum_{k=1}^K H_k(t), \quad t < 0. \quad (3.23)$$

Frogapena: Fix t in $\mathbb{R}$. As we seek to evaluate $\widehat{S}_1(t)$ as given by (3.20), two cases need to be examined: If $0 < t$, then $\nu_0 = \nu_0^*$, whence $S_1(t) = S_0$, and the conclusion (3.22) follows.

If $t < 0$, then $\nu_t = \nu_t^*$ and we are in the setting of Lemma 3.4.1 with $y = t$, $E = \Omega$, $\tau = B_t$ and $\gamma = \nu_t^*$ (so that $\gamma_0 = \nu_0$): For each $k = 1, \dots, K$, the expression (3.12) becomes

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1}[\nu_t^* = k] \frac{B_0}{C_{\nu_0}} \right] \\ &= \mathbb{P} \left[\nu_t^* = k, t + \frac{B_t}{C_k} \leq 0 \right] \cdot \widehat{S}_0 + \frac{1}{K-1} \mathbb{P} \left[\nu_t^* = k, t + \frac{B_t}{C_k} > 0 \right] \left(\Gamma - \frac{1}{C_k} \right) \cdot \mathbb{E}[B_0] \\ &= \frac{H_k(t)}{K}, \end{aligned} \quad (3.24)$$

since the random variable ν_t^* is independent of B_t and uniformly distributed on $\{1, \dots, K\}$ (as pointed out in Proposition 3.3.1). The desired result (3.23) now follows from (3.20) upon noting the decomposition

$$\widehat{S}_1(t) = \sum_{k=1}^K \mathbb{E} \left[\mathbf{1}[\nu_t^* = k] \frac{B_0}{C_{\nu_0}} \right].$$

■

3.5.2 Proof of Equation (3.4)

We can now complete the proof of Equation (3.4). Our starting point is the expression

$$\left(\overline{S^{SQ2}}\right)^{(1)}(0) = \int_{\mathbb{R}} \left(\widehat{S}_1(t) - \widehat{S}_0\right) dt = \int_{-\infty}^0 \left(\widehat{S}_1(t) - \widehat{S}_0\right) dt \quad (3.25)$$

as we use (3.22). Next, for $t < 0$, with the help of (3.23), we can rewrite the integrand as

$$\begin{aligned} \widehat{S}_1(t) - \widehat{S}_0 &= \frac{1}{K} \sum_{k=1}^K H_k(t) - \widehat{S}_0 \\ &= -\frac{1}{K} \sum_{k=1}^K \left(\widehat{S}_0 - \frac{1}{K-1} \left(\Gamma - \frac{1}{C_k} \right) \cdot \mathbb{E}[B] \right) \mathbb{P}[C_k t + B > 0]. \end{aligned} \quad (3.26)$$

Inserting this expression back into (3.25) we get

$$\begin{aligned} \left(\overline{S^{SQ2}}\right)^{(1)}(0) &= \int_{-\infty}^0 \left(\widehat{S}_1(t) - \widehat{S}_0\right) dt \\ &= -\frac{1}{K} \left(\sum_{k=1}^K \frac{\mathbb{E}[B]}{C_k} \right) \widehat{S}_0 + \frac{1}{K(K-1)} \sum_{k=1}^K \frac{\mathbb{E}[B]}{C_k} \left(\Gamma - \frac{1}{C_k} \right) \cdot \mathbb{E}[B] \end{aligned} \quad (3.27)$$

upon noting that

$$\int_{-\infty}^0 \mathbb{P}[C_k t + B > 0] dt = \frac{1}{C_k} \int_0^{\infty} \mathbb{P}[B > x] dx = \frac{\mathbb{E}[B]}{C_k}, \quad k = 1, \dots, K \quad (3.28)$$

by a simple change of variable. Uninteresting algebra readily yield (3.4) with the help of (3.9), and this completes the proof of Equation (3.4). $\blacksquare$

3.6 Bigarren trafiko-arin deribatua

The computation of the second derivative is given under the following scenario: The tagged customer arrives at time $t = 0$ with characteristic pair (B_0, Γ_0) . With s and t in $\mathbb{R}$, in addition to the tagged customer, exactly two customers arrive over the entire horizon $(-\infty, \infty)$, say at times s and t with characteristic pairs (B_s, Γ_s) and (B_t, Γ_t) , respectively. The tagged customer, the s -customer and the t -customer are assigned to their respective servers ν_0 (in Γ_0), ν_s (in Γ_s) and ν_t (in Γ_t) in accordance with the power-of-two load balancing scheduling policy.

3.6.1 Evaluating $\widehat{S}_2(s, t)$

For each s and t in $\mathbb{R}$, we have

$$\widehat{S}_2(s, t) = \mathbb{E}[S_2(s, t)] \quad \text{with } S_2(s, t) = \frac{B_0}{C_{\nu_0}}. \quad (3.29)$$

However, the server assignment random variables ν_0 , ν_s and ν_t do not always coincide with ν_0^* , ν_s^* and ν_t^* , respectively. This is due to the fact that these random variables may be affected by whether earlier customers have completed service by the time server selection needs to be determined.

3.6.1 Proposizioa *Under the enforced independence assumptions, we have*

$$\widehat{S}_2(s, t) = \widehat{S}_0, \quad 0 < s < t \quad (3.30)$$

and

$$\widehat{S}_2(s, t) = \widehat{S}_1(s) \quad s < 0 < t. \quad (3.31)$$

For $s < t < 0$, it holds that

$$\begin{aligned} \widehat{S}_2(s, t) &= \left(\frac{1}{K} \sum_{k=1}^K \mathbb{P} \left[s + \frac{B_s}{C_k} \leq t \right] \right) \cdot \widehat{S}_1(t) \\ &\quad + \frac{1}{K(K-1)} \sum_{k=1}^K \left(\sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \right) \cdot H_k(t) \\ &\quad + \frac{1}{K(K-1)^2} \sum_{k=1}^K \sum_{\ell=1, \ell \neq k}^K \left(\Gamma - \frac{1}{C_\ell} \right) \mathbb{P}[C_\ell s + B_s > 0] \mathbb{P}[C_k t + B_t \leq 0] \cdot \mathbb{E}[B_0] \\ &\quad + \frac{1}{K^2(K-1)^2} \sum_{k=1}^K \sum_{\ell=1, \ell \neq k}^K \Gamma_{k\ell} \mathbb{P}[C_k s + B_s > 0] \mathbb{P}[C_\ell t + B_t > 0] \cdot \mathbb{E}[B_0], \end{aligned} \quad (3.32)$$

with

$$\Gamma_{k\ell} = (K+1)\Gamma - K \left(\frac{1}{C_k} + \frac{1}{C_\ell} \right), \quad k, \ell = 1, \dots, K. \quad (3.33)$$

Frogapena: The proof of Proposition 3.6.1 starts in Section 3.6.2. ■

We now obtain a more compact expression for (3.32).

As we focus on the last two terms in (3.32) we readily check that their sum is given by

$$\begin{aligned} &\frac{1}{K(K-1)^2} \sum_{k=1}^K \sum_{\ell=1, \ell \neq k}^K \left(\Gamma - \frac{1}{C_\ell} \right) \mathbb{P}[C_\ell s + B_s > 0] \mathbb{P}[C_k t + B_t \leq 0] \cdot \mathbb{E}[B_0] \\ &\quad + \frac{1}{K^2(K-1)^2} \sum_{k=1}^K \sum_{\ell=1, \ell \neq k}^K \Gamma_{k\ell} \cdot \mathbb{P}[C_k s + B_s > 0] \mathbb{P}[C_\ell t + B_t > 0] \cdot \mathbb{E}[B_0] \\ &= \frac{1}{K^2(K-1)^2} \sum_{k=1}^K \sum_{\ell=1, \ell \neq k}^K G_{k\ell}(s, t) \cdot \mathbb{E}[B_0], \end{aligned} \quad (3.34)$$

with

$$\begin{aligned} G_{k\ell}(s, t) &= K \left(\Gamma - \frac{1}{C_k} \right) \mathbb{P}[C_k s + B_s > 0] \mathbb{P}[C_\ell t + B_t \leq 0] \\ &\quad + \left((K+1)\Gamma - K \left(\frac{1}{C_k} + \frac{1}{C_\ell} \right) \right) \cdot \mathbb{P}[C_k s + B_s > 0] \mathbb{P}[C_\ell t + B_t > 0] \\ &= K \left(\Gamma - \frac{1}{C_k} \right) \mathbb{P}[C_k s + B_s > 0] \\ &\quad + K \left(\frac{\Gamma}{K} - \frac{1}{C_\ell} \right) \mathbb{P}[C_k s + B_s > 0] \mathbb{P}[C_\ell t + B_t > 0], \end{aligned} \quad (3.35)$$

for every $k, \ell = 1, \dots, K$. Upon substitution into (3.32), we then conclude that

$$\begin{aligned}
\widehat{S}_2(s, t) &= \left(\frac{1}{K} \sum_{k=1}^K \mathbb{P} \left[s + \frac{B_s}{C_k} \leq t \right] \right) \cdot \widehat{S}_1(t) \\
&\quad + \frac{1}{K(K-1)} \sum_{k=1}^K \left(\sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \right) \cdot H_k(t) \\
&\quad + \frac{1}{K(K-1)^2} \sum_{k=1}^K \sum_{\ell=1, \ell \neq k}^K \left(\frac{\Gamma}{K} - \frac{1}{C_\ell} \right) \mathbb{P}[C_k s + B_s > 0] \mathbb{P}[C_\ell t + B_t > 0] \cdot \mathbb{E}[B_0] \\
&\quad + \frac{1}{K(K-1)} \sum_{k=1}^K \left(\Gamma - \frac{1}{C_k} \right) \mathbb{P}[C_k s + B_s > 0] \cdot \mathbb{E}[B_0]. \tag{3.36}
\end{aligned}$$

Next, for each $k = 1, \dots, K$, we note that

$$\begin{aligned}
H_k(t) &= \widehat{S}_0 \cdot (1 - \mathbb{P}[C_k t + B_t > 0]) + \frac{1}{K-1} \left(\Gamma - \frac{1}{C_k} \right) \mathbb{E}[B_0] \cdot \mathbb{P}[C_k t + B_t > 0] \\
&= \widehat{S}_0 + \left(\frac{1}{K-1} \left(\Gamma - \frac{1}{C_k} \right) - \frac{\Gamma}{K} \right) \mathbb{E}[B_0] \cdot \mathbb{P}[C_k t + B_t > 0] \\
&= \widehat{S}_0 + \frac{1}{K-1} \left(\frac{\Gamma}{K} - \frac{1}{C_k} \right) \mathbb{P}[C_k t + B_t > 0] \cdot \mathbb{E}[B_0], \tag{3.37}
\end{aligned}$$

as we make use of the expression (3.9). Therefore,

$$\begin{aligned}
&\sum_{k=1}^K \left(\sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \right) \cdot H_k(t) \\
&= \sum_{\ell=1}^K \left(\sum_{k=1, k \neq \ell}^K \mathbb{P} \left[t < s + \frac{B_s}{C_k} \leq 0 \right] \right) \cdot H_\ell(t) \\
&= \sum_{\ell=1}^K \left(\sum_{k=1, k \neq \ell}^K \mathbb{P} \left[t < s + \frac{B_s}{C_k} \leq 0 \right] \right) \left(\widehat{S}_0 + \frac{1}{K-1} \left(\frac{\Gamma}{K} - \frac{1}{C_\ell} \right) \mathbb{P}[C_\ell t + B_t > 0] \cdot \mathbb{E}[B_0] \right),
\end{aligned}$$

and the second term in (3.36) becomes

$$\begin{aligned}
&\frac{1}{K(K-1)} \sum_{k=1}^K \left(\sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \right) \cdot H_k(t) \\
&= \frac{1}{K(K-1)} \sum_{\ell=1}^K \left(\sum_{k=1, k \neq \ell}^K \mathbb{P} \left[t < s + \frac{B_s}{C_k} \leq 0 \right] \right) \widehat{S}_0 \\
&\quad + \frac{1}{K(K-1)^2} \sum_{\ell=1}^K \sum_{k=1, k \neq \ell}^K \left(\frac{\Gamma}{K} - \frac{1}{C_\ell} \right) \mathbb{P} \left[t < s + \frac{B_s}{C_k} \leq 0 \right] \mathbb{P}[C_\ell t + B_t > 0] \cdot \mathbb{E}[B_0].
\end{aligned}$$

Substituting this last expression into (3.36) we readily get the following more compact expression for the second half of Proposition 3.6.1.

3.6.2 Proposizioa *Under the enforced independence assumptions, for $s < t < 0$, it holds that*

$$\begin{aligned} \widehat{S}_2(s, t) &= \left(\frac{1}{K} \sum_{\ell=1}^K \mathbb{P} \left[s + \frac{B_s}{C_\ell} \leq t \right] \right) \cdot \widehat{S}_1(t) + \left(\frac{1}{K} \sum_{\ell=1}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \right) \cdot \widehat{S}_0 \\ &\quad + H(s, t) + \frac{1}{K(K-1)} \sum_{k=1}^K \left(\Gamma - \frac{1}{C_k} \right) \mathbb{P}[C_k s + B_s > 0] \cdot \mathbb{E}[B_0]. \end{aligned} \quad (3.38)$$

where we have set

$$\begin{aligned} H(s, t) &= \frac{1}{K(K-1)^2} \sum_{\ell=1}^K \sum_{k=1, k \neq \ell}^K \left(\frac{\Gamma}{K} - \frac{1}{C_\ell} \right) \mathbb{P}[C_k t < C_k s + B_s] \mathbb{P}[C_\ell t + B_t > 0] \cdot \mathbb{E}[B_0]. \end{aligned} \quad (3.39)$$

3.6.2 Proof of Proposition 3.6.1

The cases $0 < s < t$ and $s < 0 < t$ are straightforward by virtue of the operational assumptions of the power-of-two load balancing policy. Indeed, when $0 < s < t$, $\nu_0 = \nu_0^*$, hence $S_2(s, t) = S_0$ and (3.30) holds. On the other hand, when $s < 0 < t$, the future t -customer does not affect the selection of ν_0 , hence has no impact on the performance of the tagged customer. As only the s -customer can possibly affect the choice of ν_0 , we get $S_2(s, t) = S_1(s)$ and (3.31) is established.

From now on we assume $s < t < 0$, in which case we have $\nu_s = \nu_s^*$. The selection of ν_t can in principle be affected by whether the s -customer has completed its service by time t , while that of ν_0 will be determined by whether the s -customer and t -customer have completed service by the time the tagged customer enters the system. Therefore, as the s -customer completes at time $s + \frac{B_s}{C_{\nu_s^*}}$, several possibilities arise; they are captured in the decomposition

$$\begin{aligned} \mathbb{E}[S_2(s, t)] &= \mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} \leq t \right] S_2(s, t) \right] \\ &\quad + \mathbb{E} \left[\mathbf{1} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0 \right] S_2(s, t) \right] \\ &\quad + \mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} > 0 \right] \mathbf{1} \left[t + \frac{B_t}{C_{\nu_t}} \leq 0 \right] S_2(s, t) \right] \\ &\quad + \mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} > 0 \right] \mathbf{1} \left[t + \frac{B_t}{C_{\nu_t}} > 0 \right] S_2(s, t) \right]. \end{aligned} \quad (3.40)$$

These four terms are evaluated separately in the next four lemmas.

3.6.1 Lema *With $s < t < 0$, we have*

$$\mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} \leq t \right] S_2(s, t) \right] = \left(\frac{1}{K} \sum_{k=1}^K \mathbb{P} \left[s + \frac{B_s}{C_k} \leq t \right] \right) \cdot \widehat{S}_1(t). \quad (3.41)$$

Frogapena: When $s + \frac{B_s}{C_{\nu_s^*}} \leq t$, the s -customer will have completed service by the time the t -customer arrives. Therefore, conditionally on $s + \frac{B_s}{C_{\nu_s^*}} \leq t$, it holds that $S_2(s, t) =_{st} S_1(t)$, whence

$$\mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} \leq t \right] S_2(s, t) \right] = \mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} \leq t \right] S_1(t) \right] = \mathbb{P} \left[s + \frac{B_s}{C_{\nu_s^*}} \leq t \right] \cdot \hat{S}_1(t) \quad (3.42)$$

with

$$\mathbb{P} \left[s + \frac{B_s}{C_{\nu_s^*}} \leq t \right] = \frac{1}{K} \sum_{k=1}^K \mathbb{P} \left[s + \frac{B}{C_k} \leq t \right],$$

by the usual arguments. This completes the proof of (3.41). ■

3.6.2 Lema *With $s < t < 0$, we have*

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0 \right] S_2(s, t) \right] \\ &= \frac{1}{K(K-1)} \sum_{k=1}^K \left(\sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \right) \cdot H_k(t), \end{aligned} \quad (3.43)$$

with $H_k(t)$ given by (3.21) for all $k = 1, \dots, K$.

Frogapena: When $t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0$, the s -customer has not completed its service by time t , but will have completed it by the time the tagged customer arrives. Thus, only the t -customer can affect the definition of ν_0 (through B_t and ν_t).

With this in mind, consider the decomposition

$$\mathbb{E} \left[\mathbf{1} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0 \right] S_2(s, t) \right] = \sum_{k=1}^K \mathbb{E} \left[\mathbf{1} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0 \right] \mathbf{1}[\nu_t = k] \frac{B_0}{C_{\nu_0}} \right]. \quad (3.44)$$

Fix $k = 1, \dots, K$. We are in the setting of Lemma 3.4.1 with $y = t$, $E = [t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0]$, $\tau = B_t$ and $\gamma = \nu_t$ so that $\gamma_0 = \nu_0$: The expression (3.12) becomes

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0 \right] \mathbf{1}[\nu_t = k] \frac{B_0}{C_{\nu_0}} \right] \\ &= \mathbb{P} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0, \nu_t = k, t + \frac{B_t}{C_k} \leq 0 \right] \cdot \hat{S}_0 \\ & \quad + \frac{1}{K-1} \mathbb{P} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0, \nu_t = k, t + \frac{B_t}{C_k} > 0 \right] \left(\Gamma - \frac{1}{C_k} \right) \cdot \mathbb{E}[B_0] \\ &= \mathbb{P} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0, \nu_t = k \right] H_k(t), \end{aligned} \quad (3.45)$$

with $H_k(t)$ defined at (3.21). In the last step we used the fact that under the enforced independence assumptions, the random variable B_t is independent of the random variables $\{B_s, \nu_s^*, \nu_t\}$ when ν_t is generated by the power-of-two load balancing policy.

In Lemma 3.A.1 of Appendix 3.A.2 we show that

$$\mathbb{P} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0, \nu_t = k \right] = \frac{1}{K(K-1)} \sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right]. \quad (3.46)$$

Inserting (3.46) back into (3.45) yields

$$\mathbb{E} \left[\mathbf{1} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0 \right] \mathbf{1} [\nu_t = k] \frac{B_0}{c_{\nu_0}} \right] = \left(\frac{1}{K(K-1)} \sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \right) \cdot H_k(t),$$

and the desired result is now obtained by making use of (3.44). $\blacksquare$

The last two terms in the decomposition (3.40) are more cumbersome to evaluate. Their expressions are given in the next two lemmas whose proofs can be found in Appendix 3.A.3 and 3.A.4, respectively.

3.6.3 Lema *With $s < t < 0$, we have*

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} > 0 \right] \mathbf{1} \left[t + \frac{B_t}{C_{\nu_t}} \leq 0 \right] S_2(s, t) \right] \\ &= \frac{1}{K(K-1)^2} \sum_{k=1}^K \sum_{\ell=1, \ell \neq k}^K \left(\Gamma - \frac{1}{C_\ell} \right) \mathbb{P}[C_\ell s + B_s > 0] \mathbb{P}[C_k t + B_t \leq 0] \cdot \mathbb{E}[B_0]. \end{aligned} \quad (3.47)$$

3.6.4 Lema *With $s < t < 0$, we have*

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} > 0 \right] \mathbf{1} \left[t + \frac{B_t}{C_{\nu_t}} > 0 \right] S_2(s, t) \right] \\ &= \frac{1}{K^2(K-1)^2} \sum_{k=1}^K \sum_{\ell=1, \ell \neq k}^K \Gamma_{k\ell} \cdot \mathbb{P}[C_k s + B_s > 0] \mathbb{P}[C_\ell t + B_t > 0] \cdot \mathbb{E}[B_0], \end{aligned} \quad (3.48)$$

with the constants $\Gamma_{k\ell}$, $k, \ell = 1, \dots$, given by (3.33).

3.6.3 Proof of Equation (3.5)

For notational simplicity we shall write

$$R^*(s, t) = \widehat{S}_2(s, t) - \widehat{S}_1(s) - \widehat{S}_1(t) + \widehat{S}_0, \quad s, t \in \mathbb{R}.$$

We start with

$$\begin{aligned} \left(\overline{S^{SQ^2}} \right)^{(2)}(0) &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} R^*(s, t) dt \right) ds \\ &= \int_{\mathbb{R}} \left(\int_{-\infty}^s R^*(s, t) dt \right) ds + \int_{\mathbb{R}} \left(\int_s^{\infty} R^*(s, t) dt \right) ds. \end{aligned} \quad (3.49)$$

Focusing on the second term in this last expression, we write

$$\int_{\mathbb{R}} \left(\int_s^\infty R^*(s, t) dt \right) ds = \int_{-\infty}^0 \left(\int_s^\infty R^*(s, t) dt \right) ds + \int_0^\infty \left(\int_s^\infty R^*(s, t) dt \right) ds$$

Now, by Propositions 3.5.1 and 3.6.1 we have

$$R^*(s, t) = \widehat{S}_0 - \widehat{S}_0 - \widehat{S}_0 + \widehat{S}_0 = 0, \quad 0 < s < t$$

and the conclusion

$$\int_0^\infty \left(\int_s^\infty R^*(s, t) ds \right) dt = 0$$

follows.

Next, consider the decomposition

$$\int_{-\infty}^0 \left(\int_s^\infty R^*(s, t) dt \right) ds = \int_{-\infty}^0 \left(\int_s^0 R^*(s, t) dt \right) ds + \int_{-\infty}^0 \left(\int_0^\infty R^*(s, t) dt \right) ds.$$

On the range $s < 0 < t$, Propositions 3.5.1 and 3.6.1 yield $\widehat{S}_2(s, t) = \widehat{S}_1(s)$ and $\widehat{S}_1(t) = \widehat{S}_0$, whence

$$R^*(s, t) = \widehat{S}_2(s, t) - \widehat{S}_1(t) - \widehat{S}_1(s) + \widehat{S}_0 = \widehat{S}_1(s) - \widehat{S}_0 - \widehat{S}_1(s) + \widehat{S}_0 = 0,$$

so that

$$\int_{-\infty}^0 \left(\int_0^\infty R^*(s, t) dt \right) ds = 0,$$

and we conclude to

$$\int_{\mathbb{R}} \left(\int_s^\infty R^*(s, t) dt \right) ds = \int_{-\infty}^0 \left(\int_s^0 R^*(s, t) dt \right) ds.$$

On the way to evaluating this last integral we consider $R^*(s, t)$ for $s < t < 0$. On that range,

upon using first (3.38) and then (3.26) (with t replaced by s), we readily get

$$\begin{aligned}
R^*(s, t) &= \widehat{S}_2(s, t) - \widehat{S}_1(s) - \widehat{S}_1(t) + \widehat{S}_0 \\
&= \widehat{S}_0 - \widehat{S}_1(s) - \left(\widehat{S}_1(t) - \frac{1}{K} \sum_{\ell=1}^K \mathbb{P} \left[s + \frac{B_s}{C_\ell} \leq t \right] \cdot \widehat{S}_1(t) \right) \\
&\quad + \left(\frac{1}{K} \sum_{\ell=1}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \right) \cdot \widehat{S}_0 + H(s, t) \\
&\quad + \frac{1}{K(K-1)} \sum_{k=1}^K \left(\Gamma - \frac{1}{C_k} \right) \mathbb{P} [C_k s + B_s > 0] \cdot \mathbb{E} [B_0] \\
&= \left(\frac{1}{K} \sum_{k=1}^K \mathbb{P} [C_k s + B_s > 0] \right) \widehat{S}_0 - \left(\frac{1}{K} \sum_{\ell=1}^K \mathbb{P} \left[s + \frac{B_s}{C_\ell} > t \right] \right) \cdot \widehat{S}_1(t) \\
&\quad + \left(\frac{1}{K} \sum_{k=1}^K \mathbb{P} [C_k t < C_k s + B_s \leq 0] \right) \cdot \widehat{S}_0 + H(s, t) \\
&= \left(\frac{1}{K} \sum_{k=1}^K \mathbb{P} [C_k t < C_k s + B_s] \right) \cdot \widehat{S}_0 - \left(\frac{1}{K} \sum_{k=1}^K \mathbb{P} \left[s + \frac{B_s}{C_k} > t \right] \right) \cdot \widehat{S}_1(t) + H(s, t) \\
&= \left(\frac{1}{K} \sum_{k=1}^K \mathbb{P} [C_k t < C_k s + B_s] \right) \cdot \left(\widehat{S}_0 - \widehat{S}_1(t) \right) + H(s, t).
\end{aligned}$$

Next, after a change of order of integration and a change of variable we note that

$$\begin{aligned}
&\int_{-\infty}^0 \left(\int_s^0 \left(\frac{1}{K} \sum_{k=1}^K \mathbb{P} [C_k t < C_k s + B] \right) \cdot \left(\widehat{S}_0 - \widehat{S}_1(t) \right) dt \right) ds. \\
&= \frac{1}{K} \sum_{k=1}^K \int_{-\infty}^0 \left(\int_s^0 \mathbb{P} [C_k t < C_k s + B] \cdot \left(\widehat{S}_0 - \widehat{S}_1(t) \right) dt \right) ds \\
&= \frac{1}{K} \sum_{k=1}^K \int_{-\infty}^0 \left(\int_{-\infty}^t \mathbb{P} [C_k t < C_k s + B] \cdot \left(\widehat{S}_0 - \widehat{S}_1(t) \right) ds \right) dt \\
&= \frac{1}{K} \sum_{k=1}^K \int_{-\infty}^0 \left(\int_{-\infty}^t \mathbb{P} [C_k t < C_k s + B] ds \right) \cdot \left(\widehat{S}_0 - \widehat{S}_1(t) \right) dt
\end{aligned} \tag{3.50}$$

$$\begin{aligned}
&= \frac{1}{K} \sum_{k=1}^K \int_{-\infty}^0 \left(\int_0^{\infty} \mathbb{P}[C_k x < B] dx \right) \cdot (\hat{S}_0 - \hat{S}_1(t)) dt \\
&= \frac{1}{K} \sum_{k=1}^K \int_{-\infty}^0 \frac{\mathbb{E}[B]}{C_k} \cdot (\hat{S}_0 - \hat{S}_1(t)) dt \\
&= \left(\int_{-\infty}^0 (\hat{S}_0 - \hat{S}_1(t)) dt \right) \cdot \frac{\Gamma}{K} \mathbb{E}[B] \\
&= \left(\frac{1}{K} \left(\sum_{k=1}^K \frac{\mathbb{E}[B]}{C_k} \right) \hat{S}_0 - \frac{1}{K(K-1)} \sum_{k=1}^K \frac{\mathbb{E}[B]}{C_k} \left(\Gamma - \frac{1}{C_k} \right) \cdot \mathbb{E}[B] \right) \cdot \frac{\Gamma}{K} \mathbb{E}[B] \quad (3.51)
\end{aligned}$$

where the last step made used of the expression (3.27). Uninteresting calculations lead to

$$\begin{aligned}
&\left(\frac{1}{K} \left(\sum_{k=1}^K \frac{\mathbb{E}[B]}{C_k} \right) \hat{S}_0 - \frac{1}{K(K-1)} \sum_{k=1}^K \frac{\mathbb{E}[B]}{C_k} \left(\Gamma - \frac{1}{C_k} \right) \cdot \mathbb{E}[B] \right) \cdot \frac{\Gamma}{K} \mathbb{E}[B] \\
&= \left(\left(\frac{\Gamma}{K} \right)^2 - \frac{1}{K(K-1)} \left(\Gamma^2 - \sum_{k=1}^K \frac{1}{C_k^2} \right) \right) \cdot \frac{\Gamma}{K} (\mathbb{E}[B])^3 \\
&= \left(-\frac{\Gamma^2}{K^2(K-1)} + \frac{1}{K(K-1)} \sum_{k=1}^K \frac{1}{C_k^2} \right) \cdot \frac{\Gamma}{K} (\mathbb{E}[B])^3 \\
&= \left(-\left(\frac{\Gamma}{K} \right)^2 + \frac{1}{K} \sum_{k=1}^K \frac{1}{C_k^2} \right) \cdot \frac{\Gamma}{K(K-1)} (\mathbb{E}[B])^3. \quad (3.52)
\end{aligned}$$

In a similar vein, we find that

$$\int_{-\infty}^0 \left(\int_s^0 H(s, t) dt \right) ds = \frac{1}{K(K-1)^2} \sum_{\ell=1}^K \sum_{k=1, k \neq \ell}^K \left(\frac{\Gamma}{K} - \frac{1}{C_\ell} \right) I_{k\ell} \cdot \mathbb{E}[B] \quad (3.53)$$

with

$$\begin{aligned}
I_{k\ell} &= \int_{-\infty}^0 \left(\int_s^0 \mathbb{P}[C_k t < C_k s + B] \mathbb{P}[C_\ell t + B > 0] dt \right) ds \\
&= \int_{-\infty}^0 \left(\int_{-\infty}^t \mathbb{P}[C_k t < C_k s + B] ds \right) \mathbb{P}[C_\ell t + B > 0] dt \\
&= \int_{-\infty}^0 \left(\int_0^{\infty} \mathbb{P}[C_k x < B] dx \right) \mathbb{P}[C_\ell t + B > 0] dt \\
&= \left(\int_0^{\infty} \mathbb{P}[C_k x < B] dx \right) \left(\int_{-\infty}^0 \mathbb{P}[C_\ell t + B > 0] dt \right) \\
&= \frac{\mathbb{E}[B]}{C_k} \cdot \frac{\mathbb{E}[B]}{C_\ell}. \quad (3.54)
\end{aligned}$$

Therefore,

$$\int_{-\infty}^0 \left(\int_s^0 H(s, t) dt \right) ds = \frac{1}{K(K-1)^2} \sum_{\ell=1}^K \sum_{k=1, k \neq \ell}^K \left(\frac{\Gamma}{K} - \frac{1}{C_\ell} \right) \left(\frac{\mathbb{E}[B]}{C_k} \cdot \frac{\mathbb{E}[B]}{C_\ell} \right) \cdot \mathbb{E}[B]$$

with

$$\begin{aligned}
\sum_{\ell=1}^K \sum_{k=1, k \neq \ell}^K \frac{\mathbb{E}[B]}{C_k} \cdot \frac{\mathbb{E}[B]}{C_\ell} &= \sum_{\ell=1}^K \frac{\mathbb{E}[B]}{C_\ell} \left(\sum_{k=1, k \neq \ell}^K \frac{\mathbb{E}[B]}{C_k} \right) \\
&= \sum_{\ell=1}^K \frac{1}{C_\ell} \left(\Gamma - \frac{1}{C_\ell} \right) \cdot (\mathbb{E}[B])^2 \\
&= \left(\Gamma^2 - \sum_{\ell=1}^K \frac{1}{C_\ell^2} \right) \cdot (\mathbb{E}[B])^2
\end{aligned} \tag{3.55}$$

and

$$\begin{aligned}
\sum_{\ell=1}^K \sum_{k=1, k \neq \ell}^K \frac{1}{C_\ell} \left(\frac{\mathbb{E}[B]}{C_k} \cdot \frac{\mathbb{E}[B]}{C_\ell} \right) &= \sum_{\ell=1}^K \frac{\mathbb{E}[B]}{C_\ell^2} \left(\sum_{k=1, k \neq \ell}^K \frac{\mathbb{E}[B]}{C_k} \right) \\
&= \sum_{\ell=1}^K \frac{1}{C_\ell^2} \left(\Gamma - \frac{1}{C_\ell} \right) \cdot (\mathbb{E}[B])^2 \\
&= \left(\Gamma \sum_{\ell=1}^K \frac{1}{C_\ell^2} - \sum_{\ell=1}^K \frac{1}{C_\ell^3} \right) \cdot (\mathbb{E}[B])^2.
\end{aligned} \tag{3.56}$$

Collecting terms we conclude that

$$\begin{aligned}
&\int_{-\infty}^0 \left(\int_s^0 H(s, t) dt \right) ds \\
&= \frac{1}{K(K-1)^2} \left(\frac{\Gamma}{K} \left(\Gamma^2 - \sum_{\ell=1}^K \frac{1}{C_\ell^2} \right) - \left(\Gamma \sum_{\ell=1}^K \frac{1}{C_\ell^2} - \sum_{\ell=1}^K \frac{1}{C_\ell^3} \right) \right) \cdot (\mathbb{E}[B])^3 \\
&= \frac{1}{K(K-1)^2} \left(\frac{\Gamma^3}{K} - \frac{K+1}{K} \Gamma \sum_{\ell=1}^K \frac{1}{C_\ell^2} + \sum_{\ell=1}^K \frac{1}{C_\ell^3} \right) \cdot (\mathbb{E}[B])^3.
\end{aligned} \tag{3.57}$$

Finally,

$$\begin{aligned}
& \int_{-\infty}^0 \left(\int_s^0 R_2^*(s, t) dt \right) ds \\
&= \left(- \left(\frac{\Gamma}{K} \right)^2 + \frac{1}{K} \sum_{k=1}^K \frac{1}{C_k^2} \right) \cdot \frac{\Gamma}{K(K-1)} (\mathbb{E}[B])^3 \\
&\quad + \frac{1}{K(K-1)^2} \left(\frac{\Gamma^3}{K} - \frac{K+1}{K} \Gamma \sum_{\ell=1}^K \frac{1}{C_\ell^2} + \sum_{\ell=1}^K \frac{1}{C_\ell^3} \right) (\mathbb{E}[B])^3 \\
&= \left(- \frac{1}{K^3(K-1)} + \frac{1}{K^2(K-1)^2} \right) \Gamma^3 (\mathbb{E}[B])^3 \\
&\quad + \left(\frac{1}{K^2(K-1)} - \frac{K+1}{K^2(K-1)^2} \right) \left(\sum_{k=1}^K \frac{1}{C_k^2} \right) \Gamma (\mathbb{E}[B])^3 \\
&\quad + \frac{1}{K(K-1)^2} \left(\sum_{\ell=1}^K \frac{1}{C_\ell^3} \right) (\mathbb{E}[B])^3 \\
&= \frac{1}{(K-1)^2} (\dots) (\mathbb{E}[B])^3 \tag{3.58}
\end{aligned}$$

where

$$(\dots) = \left(\frac{\Gamma}{K} \right)^3 - 2 \left(\frac{1}{K} \sum_{k=1}^K \frac{1}{C_k^2} \right) \left(\frac{\Gamma}{K} \right) + \frac{1}{K} \sum_{\ell=1}^K \frac{1}{C_\ell^3} \tag{3.59}$$

In other words,

$$\begin{aligned}
& \int_{-\infty}^0 \left(\int_s^0 R_2^*(s, t) dt \right) ds \\
&= \frac{1}{(K-1)^2} \left(\left(\frac{\Gamma}{K} \right)^3 - 2 \left(\frac{1}{K} \sum_{k=1}^K \frac{1}{C_k^2} \right) \left(\frac{\Gamma}{K} \right) + \frac{1}{K} \sum_{\ell=1}^K \frac{1}{C_\ell^3} \right) (\mathbb{E}[B])^3. \tag{3.60}
\end{aligned}$$

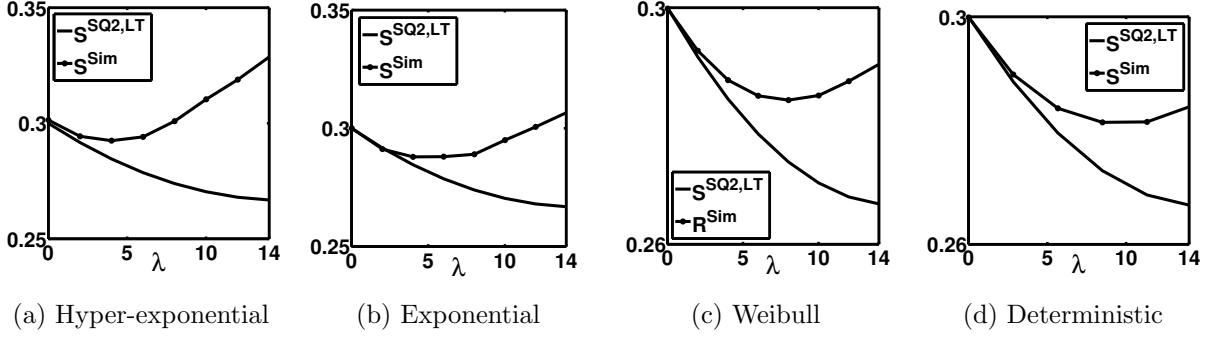
3.7 Simulaturiko emaitzak

In this section we analyse numerically the behaviour of the light-traffic approximation obtained in Proposition 3.2.1. As already pointed out by Reiman and Simon [41, 43], without additional information (e.g., heavy traffic information), we should not expect $S^{SQ2,LT}(\lambda)$ to act as an accurate proxy for mean sojourn time in medium to heavy traffic. This lack of accuracy is certainly apparent in the simulation results reported below. We have carried out simulations for different distributions of B , all with unit mean, namely hyperexponential (obtained by mixing the exponential random variables $\text{Exp}(1/2)$ and $\text{Exp}(2)$ with probability $1/3$ and $2/3$, respectively), exponential $\text{Exp}(1)$ (of parameter 1), Weibull (with shape parameter 2 and scale parameter $\Gamma(3/2)^{-1}$) and deterministic.

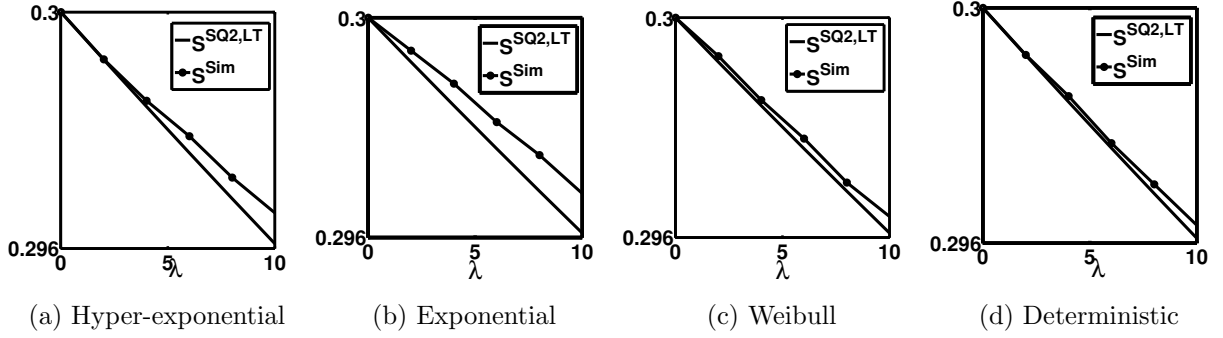
We next recall the distributions used:

- We say that B has an exponential distribution if

$$\mathbb{P}(B \leq b) = 1 - e^{-b/\mathbb{E}[B]}.$$



3.1. Irudia: Scenario 1



3.2. Irudia: Scenario 2

- We say that B has a hyper-exponential distribution with m phases if

$$\mathbb{P}(B \leq b) = 1 - \sum_{k=1}^m \beta_k e^{(-b/\mathbb{E}[B_k])},$$

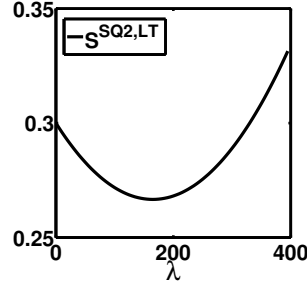
with $\sum_{k=1}^m \beta_k = 1$, where β_k is the probability that a customer is exponentially distributed with mean $\mathbb{E}[B_k]$.

- We say that B has a Weibull distribution if

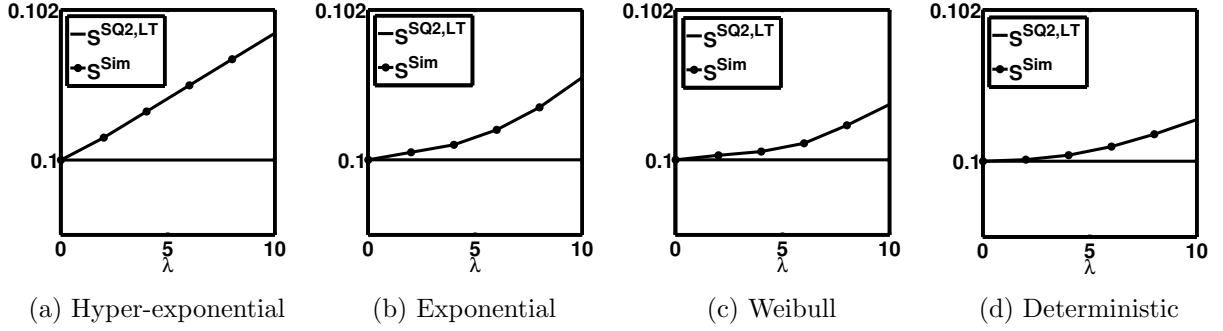
$$\mathbb{P}(B \leq b) = 1 - e^{(-b/\alpha)^k},$$

where $k > 0$ is the shape parameter and $\alpha > 0$ the scale parameter of the distribution.

The simulation results are based on averaging 10 runs with each run comprising 10^5 busy periods. A busy period is defined as the interval of time between two consecutive time epochs when the system becomes empty, such points being regenerative points for the stochastic process of interest. We have verified that the simulation results obtained for a system with $K = 100$ homogeneous servers and exponential service requirements agree with those given by Mitzenmacher [35, Table 1]. We are interested in the behaviour of the mean sojourn time in lightly loaded situations, and stability is therefore not a concern here as mentioned earlier. Three different scenarios were explored. In Scenarios 1 and 2 there are two types of servers, namely slow servers with capacity C_{slow} bytes/sec and fast servers with capacity C_{fast} bytes/sec. In Scenario 3 all the servers have the same capacity.



3.3. Irudia: Scenario 2.



3.4. Irudia: Scenario 3

- Scenario 1: $K = 10$ servers. 5 slow servers with capacity $C_{\text{slow}} = 2$ bytes/sec and 5 fast servers with capacity $C_{\text{fast}} = 10$ bytes/sec. See Figure 3.1.
- Scenario 2: $K = 100$ servers. 50 slow servers with capacity $C_{\text{slow}} = 2$ bytes/sec and 50 fast servers with capacity $C_{\text{fast}} = 10$ bytes/sec. See Figure 3.2.
- Scenario 3: $K = 10$ servers with $C_1 = \dots = C_{10} = 10$ bytes/sec. See Figure 3.4.

In the figures we use S^{Sim} to denote the mean sojourn time obtained by simulation. Also the subscript p^* in the quantities $\left(\overline{S_{p^*}^{SQ2}}\right)^{(0)}$ and $\left(\overline{S_{p^*}^{SQ2}}\right)^{(1)}$ refers to Scenario p under $*$, where $*$ correspond with the hyper-exponential (H), exponential (E), Weibull (W) or deterministic (D) distribution. Let CV_{p^*} denote the coefficient of variation corresponding to Scenario p under $*$, with $*$ again equal to H, E, W or D distribution. Then $CV_{iH} = 1.4$, $CV_{iE} = 1$, $CV_{iW} = 0.52$ and $CV_{iD} = 0$, $\forall i = 1, 2, 3$. The simulations do confirm the aforementioned structural insights gleaned from the light traffic derivatives for non-homogeneous servers:

- For all distributions, the mean sojourn time decreases as λ increases over a small neighbourhood of $\lambda = 0$,
- Over that small interval, performance seems nearly insensitive to the variability of B (as measured by its coefficient of variation).

Moreover, we observe that although in Scenario 1 and Scenario 2 there is an equal proportion of slow and fast servers, with $\left(\overline{S_{p^*}^{SQ2}}\right)^{(0)}(0) = 0.3000, \forall p = 1, 2, \forall * = H, E, W, D$, the impact of the

variability in server speeds is seen to diminish with increasing K since $\left(\overline{S_{1*}^{SQ2}}\right)^{(1)}(0) = -0.0044$ and $\left(\overline{S_{2*}^{SQ2}}\right)^{(1)}(0) = -4.0404 \cdot 10^{-4}, \forall * = H, E, W, D$.

Note that Figure 3.2 is a zoom of Figure 3.3. Although the sojourn time seems to be a straight line in a small interval of λ , after a while it increases. In Figure 3.4 we observe the aforementioned property for homogeneous servers; $\overline{S^{SQ2,LT}}(\lambda)$ becomes a constant line while the simulation results show that the mean sojourn time of a customer is increasing.

3.A Eranskina

3.A.1 (3.18) Ekuazioaren frogapena

Fix $k = 1, 2, \dots, K$. Elementary calculations give

$$\begin{aligned}
& \sum_{a=1, a \neq k}^K \left(\sum_{b=1, b \neq k}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \right) \\
&= \sum_{a=1}^{k-1} \left(\sum_{b=1, b \neq k}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \right) + \sum_{a=k+1}^K \left(\sum_{b=1, b \neq k}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \right) \\
&= \sum_{a=1}^{k-1} \left(\sum_{b=1}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \right) + \sum_{a=k+1}^K \left(-\frac{1}{C_a} - \frac{1}{C_k} + \sum_{b=1}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \right) \\
&= \sum_{a=1}^{k-1} \frac{a-1}{C_a} + \sum_{a=k+1}^K \frac{a-1}{C_a} + \sum_{a=1}^{k-1} \left(\sum_{b=1}^{a-1} \frac{1}{C_b} \right) + \sum_{a=k+1}^K \left(-\frac{1}{C_a} - \frac{1}{C_k} + \sum_{b=1}^{a-1} \frac{1}{C_b} \right) \\
&= \sum_{a=1}^K \frac{a-1}{C_a} - \frac{k-1}{C_k} + \sum_{a=1}^K \left(\sum_{b=1}^{a-1} \frac{1}{C_b} \right) - \sum_{b=1}^{k-1} \frac{1}{C_b} - \sum_{a=k+1}^K \left(\frac{1}{C_a} + \frac{1}{C_k} \right) \\
&= \sum_{a=1}^K \frac{a-1}{C_a} - \frac{k-1}{C_k} + \sum_{a=1}^K \left(\sum_{b=1}^{a-1} \frac{1}{C_b} \right) - \sum_{a=1}^K \frac{1}{C_a} - \frac{K-k-1}{C_k} \\
&= \sum_{a=1}^K \frac{a-1}{C_a} - \frac{K-2}{C_k} + \sum_{a=1}^K \left(\sum_{b=1}^{a-1} \frac{1}{C_b} \right) - \sum_{a=1}^K \frac{1}{C_a} \\
&= \sum_{a=1}^K \frac{a-1}{C_a} - \frac{K-2}{C_k} + \sum_{b=1}^{K-1} \left(\sum_{a=b+1}^K \frac{1}{C_b} \right) - \sum_{a=1}^K \frac{1}{C_a} \\
&= \sum_{a=1}^K \frac{a-1}{C_a} - \frac{K-2}{C_k} + \sum_{b=1}^{K-1} \frac{K-b}{C_b} - \sum_{a=1}^K \frac{1}{C_a} \\
&= (K-1)\Gamma - \frac{K-2}{C_k} - \Gamma
\end{aligned} \tag{3.61}$$

and the proof of (3.18) is complete. ■

3.A.2 Bi kalkulu lagungarri

We present two results that will simplify some of the calculations.

3.A.1 Lema *With $s < t < 0$, we have*

$$\mathbb{P} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0, \nu_t = k \right] = \frac{1}{K(K-1)} \sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \quad (3.62)$$

for each $k = 1, \dots, K$.

Frogapena: Fix $k = 1, \dots, K$. Our point of departure is the obvious decomposition

$$\mathbb{P} \left[t < s + \frac{B_s}{C_{\nu_s^*}} \leq 0, \nu_t = k \right] = \sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \nu_t = k \right]. \quad (3.63)$$

Pick $\ell = 1, \dots, K$ distinct from k , and note that

$$\begin{aligned} \mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \nu_t = k \right] &= \mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \ell \in \Gamma_t, \nu_t = k \right] \\ &\quad + \mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \ell \notin \Gamma_t, \nu_t = k \right]. \end{aligned} \quad (3.64)$$

We examine each term in turn: First, when ℓ belongs to Γ_t with $\nu_s^* = \ell$, then $\nu_t = k$ happens only if $\nu_s^* = \ell$ and $\Gamma_t = \{k, \ell\}$, whence

$$\begin{aligned} \mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \ell \in \Gamma_t, \nu_t = k \right] &= \mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \Gamma_t = \{k, \ell\}, \nu_t = k \right] \\ &= \mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \Gamma_t = \{k, \ell\} \right] \\ &= \frac{2}{K^2(K-1)} \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right]. \end{aligned} \quad (3.65)$$

Next, $\nu_t = \nu_t^*$ when ν_s^* is not in Γ_t , so that

$$\begin{aligned}
& \mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \ell \notin \Gamma_t, \nu_t = k \right] \\
&= \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \ell \notin \Gamma_t, \Gamma_t = \{k, a\}, \nu_t^* = k \right] \\
&= \frac{1}{2} \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \Gamma_t = \{k, a\} \right] \\
&= \frac{1}{2} \sum_{a=1, a \neq k, a \neq \ell}^K \frac{1}{K} \frac{2}{K(K-1)} \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \\
&= \frac{1}{K^2(K-1)} \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \\
&= \frac{K-2}{K^2(K-1)} \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right]. \tag{3.66}
\end{aligned}$$

Inserting (3.65) and (3.66) back into (3.64) we get

$$\mathbb{P} \left[\nu_s^* = \ell, t < s + \frac{B_s}{C_\ell} \leq 0, \nu_t = k \right] = \frac{1}{K(K-1)} \mathbb{P} \left[t < s + \frac{B_s}{C_\ell} \leq 0 \right] \tag{3.67}$$

and the desired conclusion (3.62) follows with the help of (3.63). ■

The exact same arguments lead to the following expression.

3.A.2 Lema *With $s < t < 0$, we have*

$$\mathbb{P} \left[s + \frac{B_s}{C_{\nu_s^*}} > 0, \nu_t = k \right] = \frac{1}{K(K-1)} \sum_{\ell=1, \ell \neq k}^K \mathbb{P} \left[s + \frac{B_s}{C_\ell} > 0 \right] \tag{3.68}$$

for each $k = 1, \dots, K$.

3.A.3 3.6.3 Lemaren frogapena

If $s + \frac{B_s}{C_{\nu_s^*}} > 0$ (hence $s + \frac{B_s}{C_{\nu_s^*}} > t$), then the s -customer completes its service only after the tagged arrives, so that both the s -customer and t -customer can possibly affect the definition of ν_0 . If in addition we have $t + \frac{B_t}{C_{\nu_t}} \leq 0$, then only the s -customer can affect the selection ν_0 .

In the usual manner we have the decomposition

$$\begin{aligned}
& \mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} > 0 \right] \mathbf{1} \left[t + \frac{B_t}{C_{\nu_t}} \leq 0 \right] S_2(s, t) \right] \\
&= \sum_{k=1}^K \mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} > 0 \right] \mathbf{1} \left[t + \frac{B_t}{C_k} \leq 0 \right] \mathbf{1} [\nu_t = k] \frac{B_0}{C_{\nu_0}} \right] \\
&= \sum_{k=1}^K \sum_{\ell=1, \ell \neq k}^K \mathbb{E} \left[\mathbf{1} [\nu_s^* = \ell] \mathbf{1} \left[s + \frac{B_s}{C_\ell} > 0 \right] \mathbf{1} \left[t + \frac{B_t}{C_k} \leq 0 \right] \mathbf{1} [\nu_t = k] \frac{B_0}{C_{\nu_0}} \right]. \tag{3.69}
\end{aligned}$$

Pick distinct $k, \ell = 1, \dots, K$. This time we apply Lemma 3.4.1 with $y = s$, $E = [s + \frac{B_s}{C_\ell} > 0, \nu_t = k, t + \frac{B_t}{C_k} \leq 0]$, $\tau = B_s$ and $\gamma = \nu_s^*$ (so that $\gamma_0 = \nu_0$). This leads to

$$\begin{aligned}
& \mathbb{E} \left[\left(\mathbf{1} \left[s + \frac{B_s}{C_\ell} > 0 \right] \mathbf{1} \left[t + \frac{B_t}{C_k} \leq 0 \right] \mathbf{1} [\nu_t = k] \right) \mathbf{1} [\nu_s^* = \ell] \frac{B_0}{C_{\nu_0}} \right] \\
&= \mathbb{P} \left[s + \frac{B_s}{C_\ell} > 0, t + \frac{B_t}{C_k} \leq 0, \nu_t = k, \nu_s^* = \ell, s + \frac{B_s}{C_\ell} \leq 0 \right] \cdot \widehat{S}_0 \\
&\quad + \frac{1}{K-1} \mathbb{P} \left[s + \frac{B_s}{C_\ell} > 0, t + \frac{B_t}{C_k} \leq 0, \nu_t = k, \nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0 \right] \left(\Gamma - \frac{1}{C_\ell} \right) \cdot \mathbb{E}[B_0] \\
&= \frac{1}{K-1} \mathbb{P} \left[s + \frac{B_s}{C_\ell} > 0, t + \frac{B_t}{C_k} \leq 0, \nu_t = k, \nu_s^* = \ell \right] \left(\Gamma - \frac{1}{C_\ell} \right) \cdot \mathbb{E}[B_0] \\
&= \frac{1}{K-1} \mathbb{P} \left[t + \frac{B_t}{C_k} \leq 0 \right] \mathbb{P} \left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \nu_t = k \right] \left(\Gamma - \frac{1}{C_\ell} \right) \cdot \mathbb{E}[B_0] \tag{3.70}
\end{aligned}$$

since the random variables B_t is independent of the collection $\{\nu_s^*, B_s, \nu_t\}$ under the enforced independence assumptions. Next, we write

$$\begin{aligned}
& \mathbb{P} \left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \nu_t = k \right] \\
&= \mathbb{P} \left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \ell \in \Gamma_t, \nu_t = k \right] + \mathbb{P} \left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \ell \notin \Gamma_t, \nu_t = k \right] \tag{3.71}
\end{aligned}$$

Taking terms in turn we first get

$$\begin{aligned}
\mathbb{P} \left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \ell \in \Gamma_t, \nu_t = k \right] &= \mathbb{P} \left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \Gamma_t = \{\ell, k\}, \nu_t = k \right] \\
&= \mathbb{P} \left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \Gamma_t = \{\ell, k\} \right] \\
&= \frac{2}{K^2(K-1)} \mathbb{P} \left[s + \frac{B_s}{C_\ell} > 0 \right] \tag{3.72}
\end{aligned}$$

since under the constraint $s + \frac{B_s}{C_\ell} > 0$, the fact that ν_s^* is an element of Γ_t forces ν_t to be the other element in Γ_t . In a similar way, under the constraint $s + \frac{B_s}{C_\ell} > 0$, ν_s^* not being in Γ_t implies $\nu_t = \nu_t^*$, and this leads to

$$\begin{aligned}
& \mathbb{P} \left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \ell \notin \Gamma_t, \nu_t = k \right] \\
&= \mathbb{P} \left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \ell \notin \Gamma_t, \nu_t^* = k \right] \\
&= \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{P} \left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \Gamma_t = \{a, k\}, \nu_t^* = k \right] \\
&= \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{P} [\nu_s^* = \ell] \mathbb{P} \left[s + \frac{B_s}{C_\ell} > 0 \right] \frac{1}{2} \cdot \frac{2}{K(K-1)} \\
&= \frac{K-2}{K^2(K-1)} \mathbb{P} \left[s + \frac{B_s}{C_\ell} > 0 \right]. \tag{3.73}
\end{aligned}$$

Collecting (3.72) and (3.73) gives

$$\mathbb{P} \left[\nu_s^* = \ell, s + \frac{B_s}{C_\ell} > 0, \nu_t = k \right] = \frac{1}{K(K-1)} \mathbb{P} \left[s + \frac{B_s}{C_\ell} > 0 \right], \quad (3.74)$$

and with the help of (3.70) we conclude that

$$\begin{aligned} & \mathbb{E} \left[\left(\mathbf{1} \left[s + \frac{B_s}{C_\ell} > 0 \right] \mathbf{1} \left[t + \frac{B_t}{C_k} \leq 0 \right] \mathbf{1} [\nu_t = k] \right) \mathbf{1} [\nu_s^* = \ell] \frac{B_0}{C_{\nu_0}} \right] \\ &= \frac{1}{K(K-1)^2} \mathbb{P} \left[t + \frac{B_t}{C_k} \leq 0 \right] \mathbb{P} \left[s + \frac{B_s}{C_\ell} > 0 \right] \left(\Gamma - \frac{1}{C_\ell} \right) \cdot \mathbb{E}[B_0]. \end{aligned} \quad (3.75)$$

Inserting this last expression into (3.69) we obtain (3.47) as desired. $\blacksquare$

3.A.4 3.6.4 Lemaren fropapena

If $s + \frac{B_s}{C_{\nu_s^*}} > 0$ and $t + \frac{B_t}{C_{\nu_t}} > 0$, then ν_t is determined by the s -customer. At time $t = 0$, when the tagged customer arrives, both $\nu_s (= \nu_s^*)$ and ν_t would have already been selected with both s -customer and t -customer still in service when ν_0 needs to be selected. Furthermore, $\nu_s^* \neq \nu_t$. In order to establish (3.48), we begin with the observation that

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} > 0 \right] \mathbf{1} \left[t + \frac{B_t}{C_{\nu_t}} > 0 \right] S_2(s; t) \right] \\ &= \mathbb{E} \left[\mathbf{1} \left[s + \frac{B_s}{C_{\nu_s^*}} > 0 \right] \mathbf{1} \left[t + \frac{B_t}{C_{\nu_t}} > 0 \right] \frac{B_0}{C_{\nu_0}} \right] \\ &= \sum_{k=1}^K \sum_{\ell=1, \ell \neq k}^K \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \frac{B_0}{C_{\nu_0}} \right]. \end{aligned} \quad (3.76)$$

To take advantage of this decomposition, pick distinct $k, \ell = 1, \dots, K$. As we keep in mind whether ν_s^* and ν_t are in Γ_0 , we shall have to consider four possible cases: First, if both ν_s^* and ν_t are in Γ_0 , then $\nu_0 = \nu_s^*$ and we have

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [k \in \Gamma_0, \ell \in \Gamma_0] \frac{B_0}{C_{\nu_0}} \right] \\ &= \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [\Gamma_0 = \{k, \ell\}] \frac{B_0}{C_{\nu_0^*}} \right] \\ &= \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \mathbb{P} [\Gamma_0 = \{k, \ell\}] \frac{1}{2} \cdot \left(\frac{1}{C_k} + \frac{1}{C_\ell} \right) \mathbb{E}[B_0] \\ &= \frac{1}{K(K-1)} \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \left(\frac{1}{C_k} + \frac{1}{C_\ell} \right) \cdot \mathbb{E}[B_0]. \end{aligned} \quad (3.77)$$

Next, if ν_s^* is not in Γ_0 but ν_t is in Γ_0 , then ν_0 is the other element in Γ_0 , and we get

$$\begin{aligned}
& \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [k \notin \Gamma_0, \ell \in \Gamma_0] \frac{B_0}{C_{\nu_0}} \right] \\
&= \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [\Gamma_0 = \{\ell, a\}] \frac{B_0}{C_{\nu_0}} \right] \\
&= \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [\Gamma_0 = \{\ell, a\}] \frac{B_0}{C_a} \right] \\
&= \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \mathbb{P} [\Gamma_0 = \{\ell, a\}] \frac{\mathbb{E}[B_0]}{C_a} \\
&= \frac{2}{K(K-1)} \left(\sum_{a=1, a \neq k, a \neq \ell}^K \frac{1}{C_a} \right) \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \mathbb{E}[B_0] \\
&= \frac{2}{K(K-1)} \left(\Gamma - \frac{1}{C_k} - \frac{1}{C_\ell} \right) \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \cdot \mathbb{E}[B_0]. \quad (3.78)
\end{aligned}$$

In a similar way, if ν_s^* is in Γ_0 but ν_t is not in Γ_0 , then ν_0 is the other element in Γ_0 , and we get

$$\begin{aligned}
& \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [k \in \Gamma_0, \ell \notin \Gamma_0] \frac{B_0}{C_{\nu_0}} \right] \\
&= \sum_{b=1, b \neq k, b \neq \ell}^K \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [\Gamma_0 = \{k, b\}] \frac{B_0}{C_{\nu_0}} \right] \\
&= \sum_{b=1, b \neq k, b \neq \ell}^K \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [\Gamma_0 = \{k, b\}] \frac{B_0}{C_b} \right] \\
&= \sum_{b=1, b \neq k, b \neq \ell}^K \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \mathbb{P} [\Gamma_0 = \{k, b\}] \frac{\mathbb{E}[B_0]}{C_b} \\
&= \frac{2}{K(K-1)} \left(\sum_{b=1, b \neq k, b \neq \ell}^K \frac{1}{C_b} \right) \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \mathbb{E}[B_0] \\
&= \frac{2}{K(K-1)} \left(\Gamma - \frac{1}{C_k} - \frac{1}{C_\ell} \right) \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \cdot \mathbb{E}[B_0]. \quad (3.79)
\end{aligned}$$

Finally, when neither ν_s^* and ν_t are in Γ_0 , then $\nu_0 = \nu_0^*$, whence

$$\begin{aligned}
& \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [k \notin \Gamma_0, \ell \notin \Gamma_0] \frac{B_0}{C_{\nu_0}} \right] \\
&= \mathbb{E} \left[\mathbf{1} [\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1} [\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1} [k \notin \Gamma_0, \ell \notin \Gamma_0] \frac{B_0}{C_{\nu_0^*}} \right] \\
&= \sum_{a=1, a \neq k, a \neq \ell}^K \sum_{b=1, b \neq a, b \neq k, b \neq \ell}^{a-1} (\dots)_{k\ell} \quad (3.80)
\end{aligned}$$

with

$$\begin{aligned}
(\dots)_{k\ell} &= \mathbb{E} \left[\mathbf{1}[\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1}[\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1}[\Gamma_0 = \{a, b\}] \frac{B_0}{C_{\nu_0^*}} \right] \\
&= \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \mathbb{P}[\Gamma_0 = \{a, b\}] \frac{1}{2} \cdot \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \mathbb{E}[B_0] \\
&= \frac{1}{K(K-1)} \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \cdot \mathbb{E}[B_0]. \quad (3.81)
\end{aligned}$$

It then follows that

$$\begin{aligned}
&\mathbb{E} \left[\mathbf{1}[\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1}[\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \mathbf{1}[k \notin \Gamma_0, \ell \notin \Gamma_0] \frac{B_0}{C_{\nu_0}} \right] \\
&= \frac{H_{k\ell}}{K(K-1)} \cdot \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \cdot \mathbb{E}[B_0] \quad (3.82)
\end{aligned}$$

where we have set

$$H_{k\ell} = \sum_{a=1, a \neq k, a \neq \ell}^K \left(\sum_{b=1, b \neq a, b \neq k, b \neq \ell}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \right).$$

Collecting terms (3.78)-(3.82), we conclude from (3.77) that

$$\begin{aligned}
&\mathbb{E} \left[\mathbf{1}[\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1}[\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \frac{B_0}{C_{\nu_0}} \right] \\
&= \frac{H_{k\ell}^*}{K(K-1)} \cdot \mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] \cdot \mathbb{E}[B_0] \quad (3.83)
\end{aligned}$$

with

$$H_{k\ell}^* = H_{k\ell} + 4\Gamma - 3 \left(\frac{1}{C_k} + \frac{1}{C_\ell} \right).$$

In Appendix 3.A.5 we show that

$$\mathbb{P} \left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0 \right] = \frac{1}{K(K-1)} \mathbb{P} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbb{P} \left[t + \frac{B_t}{C_\ell} > 0 \right] \quad (3.84)$$

and the conclusion

$$\begin{aligned}
&\mathbb{E} \left[\mathbf{1}[\nu_s^* = k] \mathbf{1} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbf{1}[\nu_t = \ell] \mathbf{1} \left[t + \frac{B_t}{C_\ell} > 0 \right] \frac{B_0}{C_{\nu_0}} \right] \\
&= \frac{H_{k\ell}^*}{K^2(K-1)^2} \cdot \mathbb{P} \left[s + \frac{B_s}{C_k} > 0 \right] \mathbb{P} \left[t + \frac{B_t}{C_\ell} > 0 \right] \cdot \mathbb{E}[B_0] \quad (3.85)
\end{aligned}$$

follows. In Appendix 3.A.6 we also show that

$$H_{k\ell} = (K-3) \left(\Gamma - \frac{1}{C_k} - \frac{1}{C_\ell} \right) \quad (3.86)$$

so that

$$\begin{aligned}
H_{k\ell}^* &= H_{k\ell} + 4\Gamma - 3\left(\frac{1}{C_k} + \frac{1}{C_\ell}\right) \\
&= (K-3)\left(\Gamma - \frac{1}{C_k} - \frac{1}{C_\ell}\right) + 4\Gamma - 3\left(\frac{1}{C_k} + \frac{1}{C_\ell}\right) \\
&= (K+1)\Gamma - K\left(\frac{1}{C_k} + \frac{1}{C_\ell}\right).
\end{aligned} \tag{3.87}$$

Inserting this last expression into (3.85) yields the desired conclusion (3.48). ■

3.A.5 (3.84) Ekuazioaren frogapena

Fix distinct $k, \ell = 1, \dots, K$. We need to show that

$$\mathbb{P}\left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0\right] = \frac{1}{K(K-1)} \mathbb{P}\left[s + \frac{B_s}{C_k} > 0\right] \mathbb{P}\left[t + \frac{B_t}{C_\ell} > 0\right] \tag{3.88}$$

By arguments used earlier we get

$$\begin{aligned}
&\mathbb{P}\left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, k \in \Gamma_t, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0\right] \\
&= \mathbb{P}\left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \Gamma_t = \{k, \ell\}, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0\right] \\
&= \mathbb{P}\left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \Gamma_t = \{k, \ell\}, t + \frac{B_t}{C_\ell} > 0\right] \\
&= \frac{2}{K^2(K-1)} \mathbb{P}\left[s + \frac{B_s}{C_k} > 0\right] \mathbb{P}\left[t + \frac{B_t}{C_\ell} > 0\right]
\end{aligned} \tag{3.89}$$

under the enforced independence assumptions.

In a similar way, we find

$$\begin{aligned}
&\mathbb{P}\left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, k \notin \Gamma_t, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0\right] \\
&= \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{P}\left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \Gamma_t = \{a, \ell\}, \nu_t = \ell, t + \frac{B_t}{C_\ell} > 0\right] \\
&= \frac{1}{2} \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{P}\left[\nu_s^* = k, s + \frac{B_s}{C_k} > 0, \Gamma_t = \{a, \ell\}, t + \frac{B_t}{C_\ell} > 0\right] \\
&= \frac{1}{K^2(K-1)} \sum_{a=1, a \neq k, a \neq \ell}^K \mathbb{P}\left[s + \frac{B_s}{C_k} > 0\right] \mathbb{P}\left[t + \frac{B_t}{C_\ell} > 0\right] \\
&= \frac{K-2}{K^2(K-1)} \mathbb{P}\left[s + \frac{B_s}{C_k} > 0\right] \mathbb{P}\left[t + \frac{B_t}{C_\ell} > 0\right]
\end{aligned} \tag{3.90}$$

under the enforced independence assumptions. Collecting (3.89) and (3.90) we conclude to the validity of (3.88). ■

3.A.6 (3.86) Ekuazioaren frogapena

To show (3.86) it suffices to establish this fact for $k = 1$ and $\ell = 2$. Thus,

$$\begin{aligned}
& \sum_{a=3}^K \sum_{b=3}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) \\
&= \sum_{a=3}^K \left(\sum_{b=1}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) - \left(\frac{1}{C_a} + \frac{1}{C_1} + \frac{1}{C_a} + \frac{1}{C_2} \right) \right) \\
&= \sum_{a=3}^K \left(\frac{a-3}{C_a} + \sum_{b=1}^{a-1} \frac{1}{C_b} - \frac{1}{C_1} - \frac{1}{C_2} \right) \\
&= \sum_{a=3}^K \frac{a-3}{C_a} + \sum_{a=3}^K \left(\sum_{b=1}^{a-1} \frac{1}{C_b} \right) - (K-2) \left(\frac{1}{C_1} + \frac{1}{C_2} \right) \\
&= \sum_{a=3}^K \frac{a-3}{C_a} + \sum_{a=1}^K \left(\sum_{b=1}^{a-1} \frac{1}{C_b} \right) - (K-2) \left(\frac{1}{C_1} + \frac{1}{C_2} \right) - \frac{1}{C_1} \\
&= \sum_{a=1}^K \frac{a-3}{C_a} + \sum_{a=1}^K \left(\sum_{b=1}^{a-1} \frac{1}{C_b} \right) - (K-2) \left(\frac{1}{C_1} + \frac{1}{C_2} \right) - \frac{1}{C_1} - \left(-\frac{2}{C_1} - \frac{1}{C_2} \right) \\
&= \sum_{a=1}^K \frac{a-3}{C_a} + \sum_{a=1}^K \left(\sum_{b=1}^{a-1} \frac{1}{C_b} \right) - \frac{K-3}{C_1} - \frac{K-3}{C_2} \\
&= \sum_{a=1}^K \frac{a-3}{C_a} + \sum_{b=1}^{K-1} \left(\sum_{a=b+1}^K \frac{1}{C_b} \right) - \frac{K-3}{C_1} - \frac{K-3}{C_2} \\
&= \sum_{a=1}^K \frac{a-3}{C_a} + \sum_{b=1}^{K-1} \frac{K-b}{C_b} - \frac{K-3}{C_1} - \frac{K-3}{C_2} \\
&= \sum_{a=1}^{K-1} \frac{K-3}{C_a} + \frac{K-3}{C_K} - \frac{K-3}{C_1} - \frac{K-3}{C_2}, \tag{3.91}
\end{aligned}$$

whence

$$\sum_{a=3}^K \sum_{b=3}^{a-1} \left(\frac{1}{C_a} + \frac{1}{C_b} \right) = (K-3) \left(\Gamma - \frac{1}{C_1} - \frac{1}{C_2} \right)$$

as desired.

4. Kapitulu

Interpolazio bidezko hurbilketak prozesadore-partekatze diskriminatzaile politikarentzat

Kapitulu honetan 1.1.1 Atalean deskribaturiko prozesadore-partekatze diskriminatzaile politika aztertzen dugu eta ilara luzeraren eta itxaron denbora baldintzatuaren karakterizazioak lortzea da helburu. Sistemaren maneigarritasun ezak bultzatuta, 2. Kapituluaren deskribaturiko interpolazio bidezko hurbilketa teknika aplikatzen dugu errendimendu metrika hauen banaketen hurbilketak lortzeko.

Lehenik eta behin, ilara luzeraren probabilitate funtzio sortzailearen eta itxaron denbora baldintzatuaren funtzio osagarriaren trafiko-arin hurbilketa ezartzen dugu. Lorturiko trafiko-arin hurbilketak eta trafiko geldo karakterizazioak konbinatuz interpolazio bidezko hurbilketa eraikitzen dugu, zeina beharrezko zerbitzuaren banaketa orokor batentzat betetzen den.

Kapitulu hau honela antolatuta dago. 4.1. Atalean ereduaren deskribapena ematen dugu. 4.2. Atalean kapitulu honetan erabiliko ditugun [10], [48] eta [14] artikuluetako emaitzak biltzen ditugu. 4.3. Atalean ilara luzeraren probabilitate funtzio sortzailearen interpolazio bidezko hurbilketa lortzen dugu. 4.4. Atalean itxaron denbora baldintzatuaren funtzio osagarriaren interpolazio bidezko hurbilketa aurkezten dugu eta 4.5. Atalean egonaldiaren batezbesteko iraute denboraren interpolazio bidezko hurbilketa aurkezten dugu. 4.6. Atalean lorturiko hurbilketen zehaztasuna aztertzen dugu numerikoki.

4.1 Ereduaren deskribapena

Kapitulu honetan klaseanizdun eta zerbitzari bakarra duen ilara aztertzen dugu. K bezero klase daude. k klaseko bezeroak, $k = 1, \dots, K$, $\lambda_k \geq 0$ tasako Poisson prozesu independenteen arabera iristen dira. Izan bedi iritsiera tasa totala $\lambda = \sum_{k=1}^K \lambda_k$. Onar dezagun k klaseko bezeroek banaketa orokorra duten beharrezko zerbitzuak dituztela, B_k , eta $\mathbb{E}[B_k^2] < \infty$, $k = 1, \dots, K$. k klaseko bezeroen karga da $\rho_k := \lambda_k \mathbb{E}[B_k]$ eta karga totala

$$\rho := \sum_{k=1}^K \rho_k = \sum_{k=1}^K \lambda_k \mathbb{E}[B_k] = \lambda \sum_{k=1}^K \alpha_k \mathbb{E}[B_k] = \lambda \mathbb{E}[B],$$

non $\alpha_k = \lambda_k/\lambda$ iritsiera bat k klasekoa izateko probabilitatea den eta B zorizko aldagaia arbitrarioki iristen den bezero baten beharrezko zerbitzua.

K klaseko bezeroek bat kapazitatea duen zerbitzari bat partekatzen dute. Klase bakoitzari hertsiki positiboa den pixu bat ezartzen diogu $g_1, \dots, g_K$. Sistemari k klaseko n_k , $k = 1, \dots, K$, bezero egonik, k klaseko bezero bakoitza ondoko abiaduran da zerbitzatua

$$\frac{g_k}{\sum_{j=1}^K n_j g_j}.$$

Izan bedi N_k^{DPS} k klaseko bezero kopurua sisteman momentu arbitrario batean. $\psi^{DPS}(\lambda, \vec{z})$, non $\vec{z} = (z_1, \dots, z_K)$, $(N_1^{DPS}, \dots, N_K^{DPS})$ bektorearen probabilitate funtzio sortzailea da. Itxaron denbora, sistemako egonaldia ken beharrezko zerbitzua bezala definiturik dago. $W_k(b)$ eta W_k aldagaiek k klaseko bezero arbitrario baten itxaron denbora baldintzatua eta ez-baldintzatua adierazten dute, hurrenez hurren. $W_k(\lambda, b, x) := \mathbb{P}[W_k(b) > x]$ itxaron denbora baldintzatua-ren funtzio osagarria da. $S_k^{DPS}(b)$ eta S_k^{DPS} aldagaiek sistemako egonaldiaren denbora baldintzatua eta ez-baldintzatua adierazten dute, hurrenez hurren. $\bar{S}_k^{DPS}(\lambda, b) := \mathbb{E}[S_k^{DPS}(b)]$ eta $\bar{S}_k^{DPS}(\lambda) := \mathbb{E}[S_k^{DPS}]$ definitzen ditugu λ parametroarekiko duten menpekotasuna adierazteko.

4.2 Emaizak ezagunak

Lehenago aipatu bezala, DPS ereduaren analisi zehatz bat beharrezko zerbitzuaren banaketa orokor batentzat oso zaila da. Atal honetan ezagunak diren eta ondoren kapituluan zehar erabiliko ditugun emaitzak biltzen ditugu.

4.2.1 Egondaldiaren batezbesteko iraute denbora baldintzatua

Fayolle *et al.*-ek [10] artikuluan lortzen dute sistemako egondaldiaren batezbesteko iraute denboraren deribatuek honako ekuazio sistema integro-diferentziala betetzen dutela:

$$\begin{aligned} & \frac{\partial \bar{S}_k(\lambda, b)}{\partial b} \\ &= 1 + \lambda \sum_{j=1}^K \int_0^\infty \alpha_j \frac{g_j}{g_k} \frac{\partial \bar{S}_j(\lambda, y)}{\partial y} [1 - F_j(y + \frac{g_j}{g_k} b)] dy \\ &+ \lambda \int_0^b \frac{\partial \bar{S}_k(\lambda, y)}{\partial y} \sum_{j=1}^K \alpha_j \frac{g_j}{g_k} [1 - F_j(\frac{g_j}{g_k} (b - y))] dy, \end{aligned} \quad (4.1)$$

$k = 1, \dots, K$. Beharrezko baldintza da $\bar{S}_k(\lambda, 0) = 0$, $k = 1, \dots, K$.

Sistema honentzat lortu den emaitza analitiko bakarra, beharrezko zerbitzuen banaketa esponentziala dela onartuz izan da. Kasu honetan, izan bedi $\mu_j := 1/\mathbb{E}[B_j]$, $\forall j$. [10] artikuluan frogatu da

$$\bar{S}_k(\lambda, b) = \frac{b}{1 - \rho} + \sum_{j=1}^m \frac{g_k h_j \beta_j + d_j}{\beta_j^2} \left(1 - e^{-\beta_j b / g_k}\right), \quad (4.2)$$

non $-\beta_j$, $j = 1, 2, \dots, m$, honako ekuazioaren m erro negatibo desberdinak diren

$$\sum_{j=1}^K \frac{\lambda_j g_j}{\mu_j g_j + s} = 1, \quad (4.3)$$

eta non h_j eta d_j , $j = 1, \dots, m$, β_j , $j = 1, \dots, m$ eta sarrera parametroen funtzioak diren.

Honetaz gain, sistemako egonaldiaren batezbesteko iraute denbora beharrezko zerbitzuaren banaketa esponentziala denean, [10] artikuluan erakutsi da $\bar{S}_k(\lambda)$, $k = 1, \dots, K$, hurrengo ekuazio sistemaren emaitza bakarra dela:

$$\bar{S}_k(\lambda) \left(1 - \sum_{j=1}^K \frac{\lambda_j g_j}{\mu_j g_j + \mu_k g_k} \right) - \sum_{j=1}^K \frac{\lambda_j g_j \bar{S}_j(\lambda)}{\mu_j g_j + \mu_k g_k} = \frac{1}{\mu_k}. \quad (4.4)$$

(4.4) Ekuazio sistemarentzat emaitza itxia $K = 2$ kasurako soilik dago eskuragarri eta honako da

$$\bar{S}_1(\lambda) = \frac{1}{\mu_1(1-\rho)} \left(1 + \frac{\mu_1 \rho_2 (g_2 - g_1)}{D} \right) \quad (4.5)$$

eta

$$\bar{S}_2(\lambda) = \frac{1}{\mu_2(1-\rho)} \left(1 + \frac{\mu_2 \rho_1 (g_1 - g_2)}{D} \right), \quad (4.6)$$

non $D = \mu_1 g_1 (1 - \rho_1) + \mu_2 g_2 (1 - \rho_2)$.

4.2.2 Trafiko geldo emaitzak

Atal honetan ilara luzeraren eta sistemako egonaldiaren batezbesteko iraute denbora baldintzuaren trafiko geldo emaitzak aurkezten ditugu.

Jarraian erakusten dugun bezala, *state-space collapse* fenomenoak agertzen da ilara luzeraren bektore eskalatuarentzat. Hau da, limitean, ilara luzera eskalatuaren bektorea banaketa esponentziala duen zorizko aldagai baten eta bektore determinista baten biderketa moduan banaturik dago.

4.2.1 Proposizioa [48, 2.1. Proposizioa] *Ilara luzeraren bektorearen probabilitate funtzio sortzaile eskalatuak, $\psi^{DPS}(\lambda, \bar{z}^{(1-\lambda\mathbb{E}[B])})$, trafiko geldo erregimenpean honako berdintza betetzen du*

$$\begin{aligned} \lim_{\lambda \uparrow 1/\mathbb{E}[B]} \psi^{DPS}(\lambda, \bar{z}^{(1-\lambda\mathbb{E}[B])}) &= \lim_{\lambda \uparrow 1/\mathbb{E}[B]} \mathbb{E} \left[z_1^{(1-\lambda\mathbb{E}[B])N_1^{DPS}} \dots z_K^{(1-\lambda\mathbb{E}[B])N_K^{DPS}} \right] \\ &= \frac{\mathbb{E}[B]/\mathbb{E}[Y]}{\mathbb{E}[B]/\mathbb{E}[Y] - \sum_{i=1}^K \frac{\alpha_i \mathbb{E}[B_i]}{g_i} \ln(z_i)}, \end{aligned} \quad (4.7)$$

non

$$\mathbb{E}[Y] = \frac{\mathbb{E}[B^2]}{\mathbb{E}[B] \sum_{k=1}^K \alpha_k \mathbb{E}[B_k^2] / g_k}. \quad (4.8)$$

Beste hitz batzuetan, $\lambda \uparrow 1/\mathbb{E}[B]$,

$$(1 - \lambda\mathbb{E}[B])(N_1^{DPS}, \dots, N_K^{DPS}) \xrightarrow{d} Y \cdot \left(\frac{\alpha_1 \mathbb{E}[B_1]}{g_1}, \frac{\alpha_2 \mathbb{E}[B_2]}{g_2}, \dots, \frac{\alpha_K \mathbb{E}[B_K]}{g_K} \right),$$

$\text{non} \xrightarrow{d}$ ikurrak konbergentzia banaketan adierazten duen eta Y banaketa esponentziala duen zorizko aldagaia den, $\mathbb{E}[Y]$ batezbestekoa (4.8) Ekuazioan emanda dagoelarik.

[14] artikuluan erakutsi zen trafiko geldo erregimenpean k klaseko bezeroen itxaron denbora baldintzatua, $W_k^{DPS}(b)$, banaketa esponentzialdun zorizko aldagai baten eta faktore determinista baten biderketa moduan banaturik dagoela.

4.2.2 Proposizioa [14, 4.2 Teorema] k klaseko bezeroaren itxaron denbora eskalatua trafiko geldo erregimenpean honakoa betetzen du

$$\lim_{\lambda \uparrow 1/\mathbb{E}[B]} \mathbb{P}[(1 - \lambda\mathbb{E}[B])W_k^{DPS}(\lambda, b) > x] = e^{-x \frac{g_k}{b\mathbb{E}[V]}}, \quad (4.9)$$

non

$$\mathbb{E}[V] = \frac{\mathbb{E}[B^2]}{\sum_{i=1}^K \alpha_i \mathbb{E}[B_i^2]/g_i}. \quad (4.10)$$

Beste hitz batzuetan, $\lambda \uparrow 1/\mathbb{E}[B]$, $(1 - \lambda\mathbb{E}[B])W_k^{DPS}(\lambda, b) \xrightarrow{d} \frac{b}{g_k}V$, $\text{non} \xrightarrow{d}$ ikurrak konbergentzia banaketan adierazten duen eta V banaketa esponentziala duen zorizko aldagaia den, $\mathbb{E}[V]$ batezbestekoa (4.10) Ekuazioan emanda dagoelarik.

4.3 Bezero kopuruaren banaketa

Atal honetan ilara luzeraren probabilitate funtzio sortzailearen interpolazio bidezko hurbilketa lortzea da gure helburua. Horreagatik, 2. Kapituluaren aurkezturiko notazioarekin bat eginez $\vec{y} = \vec{z}$ finkatzen dugu eta izan bedi $G(\lambda, \vec{z}) = \psi^{DPS}(\lambda, \vec{z})$, $(N_1^{DPS}, \dots, N_K^{DPS})$ bektorearen probabilitate funtzio sortzailea DPS politikapean.

Karakterizaziorik ez dagoenez hurrengo leman trafiko-arin hurbilketa lortzen dugu 2.1.1 Proposizioan aurkezturiko emaitzez baliatuz. Frogapenaren metodoa eraikitzailea da eta beste sistema batzuei aplikagarria.

4.3.1 Lema $(N_1^{DPS}, \dots, N_K^{DPS})$ bektorearen probabilitate funtzio sortzailearen trafiko-arin hurbil-

keta honakoa da

$$\begin{aligned}
\psi^{DPS,LT}(\lambda, \vec{z}) &= (\psi^{DPS})^{(0)}(\lambda, \vec{z})\Big|_{\lambda=0} + \lambda (\psi^{DPS})^{(1)}(\lambda, \vec{z})\Big|_{\lambda=0} + \frac{\lambda^2}{2} (\psi^{DPS})^{(2)}(\lambda, \vec{z})\Big|_{\lambda=0} \\
&= 1 - \rho + \lambda \sum_{i=1}^K \alpha_i z_i \mathbb{E}[B_i] \\
&\quad + \frac{\lambda^2}{2} \cdot 2 \left(\sum_{i,j=1}^K \alpha_i \alpha_j (z_i - 1) \mathbb{E} \left[\left(B_i - B_j \frac{g_i}{g_j} \right) \left(B_i - \min\{B_i, B_j \frac{g_i}{g_j}\} \right) \right. \right. \\
&\quad \quad \quad \left. \left. - \frac{1}{2} \left(B_i - \min\{B_i, B_j \frac{g_i}{g_j}\} \right)^2 - \frac{B_i^2}{2} \right] \right. \\
&\quad \quad + \sum_{i,j=1}^K \alpha_i \alpha_j (z_i \cdot z_j - 1) \mathbb{E} \left[B_j \left(1 + \frac{g_i}{g_j} \right) \left(B_i - \min\{B_i, \frac{g_i}{g_j} B_j\} \right) \right. \\
&\quad \quad \quad \left. \left. + \frac{1}{2} \left(1 + \frac{g_j}{g_i} \right) \min\{B_i, \frac{g_i}{g_j} B_j\}^2 \right] \right. \\
&\quad \quad \left. + \sum_{i,j=1}^K \alpha_i \alpha_j (z_j - 1) \mathbb{E} \left[\frac{g_i}{2g_j} \min\{\frac{g_j}{g_i} B_i, B_j\}^2 - B_i \min\{\frac{g_j}{g_i} B_i, B_j\} \right] \right). \tag{4.11}
\end{aligned}$$

Frogapena: Probabilitate funtzio sortzailearen zero, bat eta bi trafiko arinpeko deribatua kalkulatzeko etiketatutako bezeroa iristean (0 unean) sisteman zenbat bezero dauden neurtzen dugu, gehienez 0, 1 edo 2 bezero iristen direla onartuz, hurrenez hurren. Esate baterako, zero deribatua kalkulatzean etiketatutako bezeroak sistema hutsa ikusiko du. Horregatik, probabilitate funtzio sortzaileak $\mathbb{E}[z_1^0 \cdots z_K^0] = 1$ beteko du. Lehenengo deribatua kalkulatzeko sistemara bezero bat iristen dela onartu behar dugu, ikusi 2.1.1 Proposizioa ((2.3) Ekuazioa): t unean iristen den bezeroa etiketatutako bezeroa baino lehen iritsi daiteke eta sistema bere iritsiera baino lehenago edo beranduago utz dezake. Kasu honetan, probabilitate funtzio sortzailea $\mathbb{E}[z_{U_t}^0] = 1$ edo $\mathbb{E}[z_{U_t}^1]$ izango da, hurrenez hurren. Gerta liteke etiketatutako bezeroaren iritsiera baina beranduago iristea, kasu honetan $\mathbb{E}[z_{U_t}^0] = 1$. Bigarren trafiko arinpeko deribatua lortzeko ikusi 2.1.1 Proposizioa ((2.4) Ekuazioa). Kasu honetan 6 kasu desberdin kontsideratu behar dira. Ikusi 4.A.1 Eranskina frogapenaren zehaztasun gehiagorako. ■

Jarraian, ilara luzeraren bektorearen probabilitate funtzio sortzailearen interpolazio bidezko hurbilketa lortzen dugu.

4.3.1 Proposizioa ($N_1^{DPS}, \dots, N_K^{DPS}$) bektorearen probabilitate funtzio sortzailearen interpolazio

bidezko hurbilketa honakoa da

$$\begin{aligned}
& \psi^{DPS,INT}(\lambda, \vec{z}) \\
&= (1 - \rho^3) + \lambda (1 - \rho^2) \left(\sum_{i=1}^K \alpha_i z_i^{(1-\rho)^{-1}} \mathbb{E}[B_i] - \mathbb{E}[B_{U_t}] \right) + \lambda^2 (1 - \rho) \sum_{i,j=1}^K \alpha_i \alpha_j \left(\right. \\
& \quad \mathbb{E} \left[\left(z_i^{(1-\rho)^{-1}} - 1 \right) \left(B_j \frac{g_i}{g_j} \left(\min\{B_i, B_j \frac{g_i}{g_j}\} - B_i \right) - \frac{1}{2} \min\{B_i, B_j \frac{g_i}{g_j}\}^2 \right) \right] \\
& \quad + \mathbb{E} \left[\left(z_i^{(1-\rho)^{-1}} \cdot z_j^{(1-\rho)^{-1}} - 1 \right) \left(B_j \left(1 + \frac{g_i}{g_j} \right) \left(B_i - \min\{B_i, \frac{g_i}{g_j} B_j\} \right) \right. \right. \\
& \quad \quad \left. \left. + \frac{1}{2} \left(1 + \frac{g_j}{g_i} \right) \min\{B_i, \frac{g_i}{g_j} B_j\}^2 \right) \right] \\
& \quad \left. + \mathbb{E} \left[\left(z_j^{(1-\rho)^{-1}} - 1 \right) \left(\frac{g_i}{2g_j} \min\{\frac{g_j}{g_i} B_i, B_j\}^2 - B_i \min\{\frac{g_j}{g_i} B_i, B_j\} \right) \right] \right. \\
& \quad \left. - \mathbb{E}[B] \sum_{i=1}^K \alpha_i \mathbb{E}[B_i] z_i^{(1-\rho)^{-1}} \ln \left(z_i^{(1-\rho)^{-1}} \right) \right) + \rho^3 \frac{\mathbb{E}[B]/\mathbb{E}[Y]}{\mathbb{E}[B]/\mathbb{E}[Y] - \sum_{i=1}^K \frac{\alpha_i \mathbb{E}[B_i]}{g_i} \ln \left(z_i^{(1-\rho)^{-1}} \right)},
\end{aligned}$$

non $\mathbb{E}[Y]$ batezbestekoa (4.8) Ekuazioan emandakoa den.

Frogapena: Frogapena 2.3.1 Proposizioa erabiliz lortzen da zuzenean, (4.7) Ekuazioan lorturiko trafiko geldo emaitza $G^{HT}(\vec{z})$ eta trafiko-arin emaitza (4.11) konbinatuz. Ikusi 4.A.2 Eranskina xehetasun gehiagorako. ■

Azpinarratu nahi dugu lorturiko interpolazio bidezko hurbilketa ez dagokiola zorizko aldagai bati. Kontsideratu, esate baterako, hurbilketa bezero kopuru totalarentzat, hau da, $z_i = z, \forall i$. Kasu honetan ikus daiteke $\left. \frac{\partial^n \psi^{DPS,INT}(\lambda, \vec{z})}{\partial z^n} \right|_{z=0} = \infty$ betetzen dela, beraz n bezero edukitzearen probabilitateak esplotatu egiten du. Dena den, 4.3.1 Proposizioak momentuen hurbilketa zehatzak eskaintzen ditu, ikusi 4.6 Atala emaitza numerikoetarako. Momentuen zehaztasuna honela azal daiteke: $z = 0$ puntuan, probabilitate bakoitiek $+\infty$ -ra jotzen duten bitartean, bikoitiek $-\infty$ -ra jotzen dute.

Jarraian lehenengo eta bigarren momentuak lortzen ditugu. Lehenengo momentuaren hurbilketa honakoa da:

$$\begin{aligned}
\mathbb{E} [N^{DPS,INT}] &= \mathbb{E} [N_1^{DPS,INT} + \dots + N_K^{DPS,INT}] = \frac{\partial \left(\psi^{DPS,INT}(\lambda, \vec{z}) \Big|_{z_i=z_j=z} \right)}{\partial z} \Big|_{z=1} \\
&= \rho + \lambda^2 \sum_{i,j=1}^K \alpha_i \alpha_j \mathbb{E} \left[\left(2 + \frac{g_i}{g_j} \right) B_i B_j - \left(2 + \frac{g_i}{g_j} \right) B_j \min\{B_i, B_j \frac{g_i}{g_j}\} \right. \\
&\quad \left. + \left(\frac{1}{2} + \frac{3g_j}{2g_i} \right) \min\{B_i, B_j \frac{g_i}{g_j}\}^2 - \frac{g_j}{g_i} B_i \min\{B_i, \frac{g_i}{g_j} B_j\} \right] + \frac{\rho^3}{1-\rho} \frac{\mathbb{E}[Y]}{\mathbb{E}[B]} \sum_{i=1}^K \frac{\alpha_i \mathbb{E}[B_i]}{g_i}. \quad (4.12)
\end{aligned}$$

$\psi^{DPS,INT}(\lambda, \vec{z})$ adierazpenaren bigarren deribatua z aldagaiarekiko $z = 1$ puntuan hau da

$$\begin{aligned}
\frac{\partial^2 \left(\psi^{DPS,INT}(\lambda, \vec{z}) \Big|_{z_i=z_j=z} \right)}{\partial z^2} \Big|_{z=1} &= \mathbb{E} [(N^{DPS,INT})^2] - \mathbb{E} [N^{DPS,INT}] \\
&= \lambda^2 \sum_{i,j=1}^K \alpha_i \alpha_j \left(\frac{\rho}{(1-\rho)} \mathbb{E} \left[B_j \frac{g_i}{g_j} \left(-B_i + \min\{B_i, B_j \frac{g_i}{g_j}\} \right) - \frac{1}{2} \min\{B_i, B_j \frac{g_i}{g_j}\}^2 \right] \right. \\
&\quad \left. + \frac{2(1+\rho)}{(1-\rho)} \mathbb{E} \left[\left(B_j \left(1 + \frac{g_i}{g_j} \right) \left(B_i - \min\{B_i, \frac{g_i}{g_j} B_j\} \right) + \frac{1}{2} \left(1 + \frac{g_j}{g_i} \right) \min\{B_i, \frac{g_i}{g_j} B_j\}^2 \right) \right] \right. \\
&\quad \left. + \frac{\rho}{(1-\rho)} \mathbb{E} \left[\left(\frac{g_j}{2g_i} \min\{B_i, \frac{g_i}{g_j} B_j\}^2 - \frac{g_j}{g_i} B_i \min\{B_i, \frac{g_i}{g_j} B_j\} \right) \right] \right) \\
&\quad + \rho^3 \frac{\mathbb{E}[Y]}{\mathbb{E}[B]} \left(\frac{2\mathbb{E}[Y]}{\mathbb{E}[B]} \left(\sum_{i=1}^K \frac{\alpha_i \mathbb{E}[B_i]}{g_i} (1-\rho)^{-1} \right)^2 - \left(\sum_{i=1}^K \frac{\alpha_i \mathbb{E}[B_i]}{g_i} (1-\rho)^{-1} \right) \right). \quad (4.13)
\end{aligned}$$

Hori dela eta, bigarren momentua (4.12) eta (4.13) Ekuazioetatik lortzen dugu

$$\mathbb{E} [(N^{DPS,INT})^2] = \frac{\partial^2 \left(\psi^{DPS,INT}(\lambda, \vec{z}) \Big|_{z_i=z_j=z} \right)}{\partial z^2} \Big|_{z=1} + \mathbb{E} [N^{DPS,INT}]. \quad (4.14)$$

Azpmarratu nahi dugu, $1/\mu$ batezbestekoko banaketa esponentziala duten beharrezko zerbitzuen kasuan, lehenengo eta bigarren momentuak $M/M/1$ ilara baten baliokideak direla eta hurbilketak honako emaitza ematen digu $\mathbb{E} [N^{DPS,INT}] = \frac{\rho}{1-\rho}$ eta $\mathbb{E} [(N^{DPS,INT})^2] = \frac{2\rho^2}{(1-\rho)^2} + \frac{\rho}{1-\rho}$, zeinak zehatzak diren.

4.6 Atalean lehenengo eta bigarren momentuak erabiltzen ditugu numerikoki interpolazio bidezko hurbilketaren zehaztasuna neurtzeko.

4.4 Itxaron denboraren banaketa

Atal honetan itxaron denbora baldintzatuaren hurbilketa lortzea da gure helburua. 2. Kapitulu-
luko notazioarekin bat eginez, $\vec{y} = (b, x)$ finkatu eta izan bedi $G(\lambda, b, x) = W_k^{DPS}(\lambda, b, x) =$
 $\mathbb{P}[W_k^{DPS}(b) > x]$ itxaron denbora baldintzatuaren funtzio osagarria.

Ilara luzeraren kasuan bezala DPS politikarentzat ez dago itxaron denboraren banaketaren
karakterizaziorik banaketa orokorreko beharrezko zerbitzuen kasuan. Horregatik, hurrengo leman
lehenengo mailako trafiko-arin hurbilketa lortzen dugu 2.1.1 Proposizioan aurkezturiko emaitza
erabiliz lortu duguna. Ilara luzeraren kasuan bezala frogapena konstruktiboa da eta beste sistema
batzuetara aplikagarria.

Beraz, $\mathbb{P}(W_k^{DPS}(b) > x)$ -rako hurbilketa lortu nahi dugu. Ohartu, itxaron denborak $1 -$
 $\mathbb{P}(W_k^{DPS}(b) > 0)$ tamainako atomo bat duela $x = 0$ puntuan.

Izan bedi $W_k^{DPS}(b | A(-\infty, \infty) = 1, \tau_1 = t)$ k klaseko etiketatutako bezeroaren itxaron den-
bora iritsiera bat dagoenean $t \in \mathbb{R}$ unean. $W_k^{DPS}(b | A(-\infty, \infty) = 1, \tau_1 = t)$ honela idatziko dugu:

$$W_k^{DPS}(b | A(-\infty, \infty) = 1, \tau_1 = t) = \begin{cases} t + b_{u_t} & \text{if } t \leq 0 \leq t + b_{u_t} \text{ eta } \frac{b}{g_k} > \frac{t + b_{u_t}}{g_{u_t}} \\ \frac{g_{u_t} b}{g_k} & \text{if } t \leq 0 \leq t + b_{u_t} \text{ eta } \frac{b}{g_k} \leq \frac{t + b_{u_t}}{g_{u_t}} \\ 0 & \text{if } t + b_{u_t} < 0 \\ b_{u_t} & \text{if } 0 < t < b \text{ eta } \frac{b-t}{g_k} > \frac{b_{u_t}}{g_{u_t}} \\ (b-t) \frac{g_{u_t}}{g_k} & \text{if } 0 < t < b \text{ eta } \frac{b-t}{g_k} \leq \frac{b_{u_t}}{g_{u_t}} \\ 0 & \text{if } 0 < b < t, \end{cases} \quad (4.15)$$

non u_t aldagaiak t unean iritsitako bezeroaren klasea deskribatzen duen eta b_{u_t} aldagaiak t unean
iritsiriko bezeroaren beharrezko zerbitzua adierazten duen.

Lehenengo adierazpenak beste bezeroa etiketatutako bezeroa baina lehenago iritsi eta etiketa-
tutako bezeroa iritsi ondoren joaten dela deskribatzen du. Beraz, lan kontserbazioaren propietatea
dela eta, etiketatutako bezeroa sisteman egongo da 0 uneko lan guztia amaitu arte (etiketatutako
bezeroaren beharrezko zerbitzua izan ezik), hau da, $b_{u_t} - (-t)$. Gogoratu lan kontserbazioaren
propietateak zera diola, sistema hutsik ez dagoen bitartean, zerbitzariak lanean jarduten duela.
Bigarren adierazpenak beste bezeroa etiketatutako bezeroa baino lehenago iritsi eta etiketatutako
bezeroa joatean bertan jarraitzen duen kasua deskribatzen du. Beraz, k klaseko etiketatutako beze-
roa $\frac{g_k}{g_k + g_{u_t}}$ tasan zerbitzatu izango da eta sistemako egonaldiaren iraute denbora $b \left(\frac{g_k}{g_k + g_{u_t}} \right)^{-1} - b$
izango da. Laugarren adierazpenak beste bezeroa etiketatutako bezeroaren ondoren iritsi eta eti-
ketatutako bezeroa baino lehen alde egiten duen egoera deskribatzen du. Beraz, sistemako lan
kontserbazioaren propietateari esker, sistemako egonaldiaren iraute denbora b_{u_t} izango da. Bos-
garren adierazpenak beste bezeroa etiketatutako bezeroaren ondoren iritsi eta alde egiten duela
deskribatzen du. Beraz, sistemako egonaldiaren iraute denbora $t + (b - t) \left(\frac{g_k}{g_k + g_{u_t}} \right)^{-1} - b$ da. Hi-
rugarren eta seigarren kasuek etiketatutako bezeroak beste bezeroarekin bat egiten ez duen kasua
deskribatzen dute. Beraz, etiketatutako bezeroaren itxaron denbora 0 da.

Jarraian bat mailako trafiko-arin hurbilketa lortzen dugu.

4.4.1 Lema *Itxaron denbora baldintzatuaren funtzio osagarriaren k klaseko etiketatutako bezeroaren trafiko-arin hurbilketa honakoa da*

$$\begin{aligned}
 & W_k^{DPS,LT}(\lambda, b, x) \\
 &= \lambda \sum_{j=1}^K \alpha_j \mathbb{E} \left[\left(1 + \frac{g_k}{g_j} \right) \left(-x + \min\{B_j, \frac{g_j}{g_k} b\} \right)^+ \right. \\
 & \quad \left. + \mathbf{1} \left[\frac{g_j}{g_k} b > x \right] \left(B_j - \min\{B_j, \frac{g_j}{g_k} b\} \right) + \mathbf{1} [B_j > x] \left(b - \frac{g_k}{g_j} \min\{B_j, \frac{g_j}{g_k} b\} \right) \right].
 \end{aligned} \tag{4.16}$$

Frogapena: Trafiko arinpeko bat deribatua kalkulatzeko sisteman iritsiera bat dagoen kasua aztertzen dugu. Sei kasu desberdin gerta daitezke. Ikusi 4.A.3 Eranskina xehetasun gehiagorako. ■

Jarraian itxaron denbora baldintzatuaren funtzio osagarriaren interpolazio bidezko hurbilketa aurkezten dugu.

4.4.1 Proposizioa *k klaseko etiketatutako bezeroaren itxaron denbora baldintzatuaren funtzio osagarriaren interpolazio bidezko hurbilketa honakoa da*

$$\begin{aligned}
 & W_k^{DPS,INT}(\lambda, b, x) \\
 &= \lambda(1-\rho) \sum_{j=1}^K \alpha_j \mathbb{E} \left[\left(1 + \frac{g_k}{g_j} \right) \left(-(1-\rho)x + \min\{B_j, \frac{g_j}{g_k} b\} \right)^+ \right. \\
 & \quad \left. + \mathbf{1} \left[\frac{g_j}{g_k} b > (1-\rho)x \right] \left(B_j - \min\{B_j, \frac{g_j}{g_k} b\} \right) \right. \\
 & \quad \left. + \mathbf{1} [B_j > (1-\rho)x] \left(b - \frac{g_k}{g_j} \min\{B_j, \frac{g_j}{g_k} b\} \right) \right] + \rho^2 e^{-(1-\rho)x \frac{g_k}{b\mathbb{E}[V]}},
 \end{aligned} \tag{4.17}$$

non $\mathbb{E}[V]$ batezbestekoa (4.10) Ekuazioan emandakoa den.

Frogapena: Frogapena 2.3.1 Proposizioa erabiliz lortzen da zuzenean, (4.9) Ekuazioan lorturiko trafiko geldoko emaitza $G^{HT}(b, x)$ eta trafiko arinpeko emaitza (4.16) konbinatuz. Ikusi 4.A.4 Eranskina xehetasun gehiagorako. ■

Azpinarratu nahi dugu itxaron denbora baldintzatuaren funtzio osagarriaren interpolazio bidezko hurbilketa ez dela banaketa bat. Esate baterako, $x = 0$ puntuan, $W_k^{DPS,INT}(\lambda, b, x) = \lambda(1-\rho) (\mathbb{E}[B] + b) + \rho^2$ betetzen da, zeina bat baino handiagoa izan daitekeen. Dena den, 4.4.1 Proposizioa emaitzak lehenengo momentuaren emaitza zehatzak eskaintzen ditu, ikusi emaitza numerikoak 4.6 Atalean.

4.4.1 Proposiziotik honako korolariora lortzen dugu.

4.4.1 Korolaria *k klaseko etiketatutako bezeroaren itxaron denboraren funtzio osagarriaren interpolazio bidezko hurbilketa honakoa da*

$$\begin{aligned}
 W_k^{DPS,INT}(\lambda, x) &:= \int_0^\infty W_k^{DPS,INT}(\lambda, b, x) dF_k(b) \\
 &= \lambda(1 - \rho) \sum_{j=1}^K \alpha_j \left(\left(1 + \frac{g_k}{g_j} \right) \int_{(1-\rho)x \frac{g_k}{g_j}}^\infty \left(\int_{\frac{g_j}{g_k} b}^{\frac{g_j}{g_k} b} (1 - F_j(b_j)) db_j \right) dF_k(b) \right. \\
 &\quad + \left(\int_{(1-\rho)x \frac{g_k}{g_j}}^\infty \left(\int_{\frac{g_j}{g_k} b}^\infty (1 - F_j(b_j)) db_j \right) dF_k(b) \right) \\
 &\quad + \int_{(1-\rho)x g_k/g_j}^\infty \left(\left(b - \frac{g_k}{g_j} (1 - \rho)x \right) (1 - F_j((1 - \rho)x)) \right. \\
 &\quad \left. \left. - \frac{g_k}{g_j} \int_{(1-\rho)x}^{\frac{g_j}{g_k} b} (1 - F_j(b_j)) db_j \right) dF_k(b) \right) + \rho^2 e^{-(1-\rho)x \frac{g_k}{\mathbb{E}[B_k] \mathbb{E}[V]}}.
 \end{aligned} \tag{4.18}$$

4.5 Sistemako egonaldiaren batezbesteko iraute denbora

Atal honetan sistemako egonaldiaren batezbesteko iraute denboraren interpolazio bidezko hurbilketa aurkezten dugu eta zehaztasunez aztertu. Atal honetako emaitzak 4.4.1 Proposizioaren ondorio zuzenak dira, izan ere, sistemako egonaldiaren batezbesteko iraute denbora da batezbesteko itxaron denboraren eta bezeroaren beharrezko zerbitzuaren, b , batura.

4.5.1 Sistemako egonaldiaren batezbesteko iraute denbora baldintzatua

4.4.1 Lemaren eta 4.4.1 Proposizioaren korolario bezala etiketatutako bezeroaren sistemako egonaldiaren batezbesteko iraute denbora baldintzatuaren trafiko-arin hurbilketa eta interpolazio bidezko hurbilketa lortzen ditugu, $\overline{S}_k^{DPS,LT}(\lambda, b)$ eta $\overline{S}_k^{DPS,INT}(\lambda, b)$ bidez adieraziak, hurrenez hurren.

4.5.1 Korolaria *k klaseko etiketatutako bezeroaren sistemako egonaldiaren batezbesteko iraute denbora baldintzatuaren trafiko-arin hurbilketa (1 mailakoa) eta interpolazio bidezko hurbilketa (2 mailakoa), b beharrezko zerbitzua izanik, honakoak dira*

$$\begin{aligned}
 \overline{S}_k^{DPS,LT}(\lambda, b) &= b(1 + \rho) \\
 &\quad + \lambda \mathbb{E} \left[\frac{1}{2} \left(1 + \frac{g_k}{g_{U_t}} \right) \min \left\{ B_{U_t}, b \frac{g_{U_t}}{g_k} \right\}^2 - \left(b \frac{g_{U_t}}{g_k} + \frac{g_k}{g_{U_t}} B_{U_t} \right) \min \left\{ B_{U_t}, b \frac{g_{U_t}}{g_k} \right\} + \frac{g_{U_t}}{g_k} b B_{U_t} \right].
 \end{aligned} \tag{4.19}$$

eta

$$\begin{aligned}
& \overline{S_k^{DPS,INT}}(\lambda, b) \\
&= b(1 + \rho) \\
&+ \lambda \mathbb{E} \left[\frac{1}{2} \left(1 + \frac{g_k}{g_{U_t}} \right) \min\{B_{U_t}, b \frac{g_{U_t}}{g_k}\}^2 - \left(b \frac{g_{U_t}}{g_k} + \frac{g_k}{g_{U_t}} B_{U_t} \right) \min\{B_{U_t}, b \frac{g_{U_t}}{g_k}\} + \frac{g_{U_t}}{g_k} b B_{U_t} \right] \\
&+ \frac{(\lambda \mathbb{E}[B])^2}{(1 - \lambda \mathbb{E}[B])} \frac{b}{g_k} \frac{\mathbb{E}[B^2]}{\sum_{j=1}^K \alpha_j \mathbb{E}[B_j^2]/g_j}. \tag{4.20}
\end{aligned}$$

Frogapena: (4.19) Ekuazioa 4.4.1 Lematik nola ondorioztatu daitekeen erakusten dugu. Honako ekuazioa dugu

$$\begin{aligned}
& \overline{W_k^{DPS,LT}}(\lambda, b) := \int_0^\infty W_k^{DPS,LT}(\lambda, b, x) dx \\
&= \lambda \sum_{j=1}^K \alpha_j \mathbb{E} \left[\frac{1}{2} \left(1 + \frac{g_k}{g_j} \right) \min\{B_j, b \frac{g_j}{g_k}\}^2 - \left(b \frac{g_j}{g_k} + \frac{g_k}{g_j} B_j \right) \min\{B_j, b \frac{g_j}{g_k}\} + \left(1 + \frac{g_j}{g_k} \right) b B_j \right], \tag{4.21}
\end{aligned}$$

eta sistemako egonaldiaren definizioz, $\overline{S_k^{DPS,LT}}(\lambda, b) = \overline{W_k^{DPS,LT}}(\lambda, b) + b$, bilatzen ari ginen trafiko-arin hurbilketa adierazpena lortzen dugu (4.20). Itxaron denborarentzat aurkezturiko trafiko geldo emaitza, (4.9), sistemako egonaldiarentzat ere baliogarria da. Horregatik, trafiko-arin eta geldo emaitzak konbinatuz, (4.20) Ekuazioko interpolazio bidezko hurbilketa lortzen dugu. ■

4.A.5 Eranskinean trafiko arinpeko deribatuak kalkulatzeko beste modu bat erakusten dugu. Metodo honetan (4.1) Ekuazioa erabiltzen dugu, eta beraz DPS ereduarentzat soilik da erabilgarria. Numerikoki ikus dezakegu, ikusi 4.18-4.21 Irudiak, bigarren mailako trafiko-arin hurbilketa zehatzagoa dela, baina, aldiz, hirugarren mailako trafiko arinaren eta geldoaren interpolazioa ez dela beti hobea bihurtzen kargaren zenbait balioetarako.

Jarraian, ondorio interesgarriak lortzen ditugu sistemako egonaldiaren batezbesteko iraute denbora baldintzatuarentzat lorturiko hurbilketa inguruan. 4.6 Atalean numerikoki ebaluatuko dugu 4.20 Ekuazioan lorturiko hurbilketa zehaztasuna.

Zerbitzaria partekatzea

Zerbitzaria partekatzea ilararentzat sistemako egonaldiaren batezbesteko iraute denbora baldintza-tua ezaguna da, $b/(1 - \rho)$, [24]. Baldin eta (i) klase bakarra kontsideratzen badugu edo (ii) pixu guztiak berdinak badira, gure DPS eredu PS ereduaren baliokidea da. Jarraian erakusten dugu gure hurbilketa zehatza dela bi kasu hauetan, hau da, $b/(1 - \rho)$ dela.

Lehenik eta behin, klase bakarra kontsideratzen dugu, hau da, $\alpha_i = 0, \forall i \neq k$ eta $\alpha_k = 1$. Orduan, (4.20) Ekuazioak honako adierazpena betetzen du

$$\begin{aligned}
& b(1 + \rho) + \lambda_k \mathbb{E} \left[\min\{B_{U_t}, b\}^2 - (b + B_{U_t}) \min\{B_{U_t}, b\} + b B_{U_t} \right] + b \frac{\rho^2}{(1 - \rho)} \\
&= b(1 + \rho + \frac{\rho^2}{(1 - \rho)}) = \frac{b}{1 - \rho},
\end{aligned}$$

$\min\{B_{U_t}, b\}^2 - (b + B_{U_t}) \min\{B_{U_t}, b\} + bB_{U_t} = 0$ erabili dugun.

Orain, pixu guztiak berdinak direla onartzen dugu, hau da, $g_i = g_k, \forall i, k = 1, \dots, K$. Orduan, (4.20) Ekuazioak honako adierazpena betetzen du

$$\begin{aligned} & b(1 + \rho) + \lambda \mathbb{E} \left[\min\{B_{U_t}, b\}^2 - (b + B_{U_t}) \min\{B_{U_t}, b\} + bB_{U_t} \right] + \frac{b\rho^2}{(1 - \rho)} \frac{\mathbb{E}[B^2]}{\sum_{j=1}^K \alpha_j \mathbb{E}[B_j^2]} \\ & = b(1 + \rho + \frac{\rho^2}{(1 - \rho)}) = \frac{b}{1 - \rho}. \end{aligned}$$

Beraz, bi kasuek PS ereduaren emaitzarekin bat egiten dute.

Lehentasunezko ilara

Orain, etiketatutako bezeroaren pixua handitzen den kasua aztertzen dugu, hau da, $g_k \rightarrow \infty$. Beraz, k klaseak, limitean, lehentasuna du. Orduan, hurbilketa honako adierazpenera sinplifikatzen da

$$\begin{aligned} & \lim_{g_k \rightarrow \infty} \left(b(1 + \rho) \right. \\ & + \sum_{u_t=1}^K \lambda_{u_t} \mathbb{E} \left[\frac{1}{2} \left(\min\{B_{u_t}, b \frac{g_{u_t}}{g_k}\}^2 + \frac{g_k}{g_{u_t}} \min\{B_{u_t}, b \frac{g_{u_t}}{g_k}\}^2 \right) \right. \\ & - \left(b \frac{g_{u_t}}{g_k} \min\{B_{u_t}, b \frac{g_{u_t}}{g_k}\} + B_{u_t} \min\{B_{u_t} \frac{g_k}{g_{u_t}}, b\} \right) \\ & \left. \left. + \frac{g_{u_t}}{g_k} bB_{u_t} \right] + \frac{b(\lambda \mathbb{E}[B])^2}{(1 - \lambda \mathbb{E}[B])} \frac{\mathbb{E}[B^2]}{g_k \sum_{j=1, j \neq k}^K \alpha_j \mathbb{E}[B_j^2]/g_j + \alpha_k \mathbb{E}[B_k^2]} \right) \\ & = b(1 + \rho) + \mathbb{E} \left[\sum_{\substack{u_t=1 \\ u_t \neq k}}^K \lambda_{u_t} \left(\frac{1}{2} (0 + 0) - (0 + bB_{u_t}) + 0 \right) \right] \\ & = b(1 + \rho_k). \end{aligned}$$

Ohartu, $g_k \rightarrow \infty$ limitepean sistemako egonaldiaren batezbesteko iraute denbora baldintzatua ezagutzen dela eta $b/(1 - \rho_k)$ adierazpena betetzen duela. $1/(1 - \rho_k) = \sum_{i=0}^{\infty} \rho_k^i$ berdintzatik, zuzenean ikus dezakegu gure hurbilketa, adierazpen zuzenaren lehen mailako hurbilketa dela. Errore erlatiboa $100\% (b/(1 - \rho_k) - b(1 + \rho_k)) / b/(1 - \rho_k) = \rho_k^2 100\%$ da, eta beraz, ikus dezakegu errore erlatiboa handitu egiten dela k klaseko karga handitzen denean.

Pixuen monotonizitatea

Ikus daiteke, k klaseko etiketatutako bezeroaren sistemako egonaldiaren batezbesteko iraute denbora baldintzatua, $\overline{S}_k^{DPS,INT}(\lambda, b)$, txikitu egiten dela g_k -n eta handitu g_i -n, $i \neq k$.

Hau honela ikus daiteke, U_t aldagaian baldintzatuz idatz dezakegu

$$\begin{aligned} \overline{S}_k^{DPS,INT}(\lambda, b) &= b(1 + \rho) + \sum_{i=1, i \neq k}^K \lambda_i \mathbb{E} \left[\frac{1}{2} \left(1 + \frac{g_k}{g_i} \right) \min\{B_i, b \frac{g_i}{g_k}\}^2 \right. \\ &\quad \left. - \left(b \frac{g_i}{g_k} + \frac{g_k}{g_i} B_i \right) \min\{B_i, b \frac{g_i}{g_k}\} + b \frac{g_i}{g_k} B_i \right] \\ &\quad + \frac{(\lambda \mathbb{E}[B])^2}{(1 - \lambda \mathbb{E}[B])} \frac{b}{g_k} \frac{\mathbb{E}[B^2]}{\sum_{j=1}^K \alpha_j \mathbb{E}[B_j^2]/g_j}, \end{aligned}$$

non $U_t = k$ -rentzat $\min\{B_k, b\}^2 - (b + B_k) \min\{B_k, b\} + bB_k = 0$ erabiltzen dugun.

Orain, baldin $B_i \leq \frac{g_i}{g_k} b$,

$$\frac{1}{2} \left(1 + \frac{g_k}{g_i} \right) \min\{B_i, b \frac{g_i}{g_k}\}^2 - \left(b \frac{g_i}{g_k} + \frac{g_k}{g_i} B_i \right) \min\{B_i, b \frac{g_i}{g_k}\} + b \frac{g_i}{g_k} B_i = \frac{1}{2} B_i^2 \left(1 - \frac{g_k}{g_i} \right),$$

zeina beherakorra den g_k -n eta gorakorra g_i -n.

Baldin $B_i > \frac{g_i}{g_k} b$, orduan

$$\begin{aligned} &\frac{1}{2} \left(1 + \frac{g_k}{g_i} \right) \min\{B_i, b \frac{g_i}{g_k}\}^2 - \left(b \frac{g_i}{g_k} + \frac{g_k}{g_i} B_i \right) \min\{B_i, b \frac{g_i}{g_k}\} + b \frac{g_i}{g_k} B_i \\ &= \frac{1}{2} b^2 \frac{g_i}{g_k} \left(1 - \frac{g_i}{g_k} \right) + b B_i \left(\frac{g_i}{g_k} - 1 \right), \end{aligned}$$

zeina beherakorra den g_k aldagaiarekiko eta gorakorra g_i aldagaiarekiko. Orain, $\overline{S}_k^{INT}(\lambda, b)$ adierazpenaren monotonizitatea g_k eta g_i pixuekiko zuzenean ikus daiteke.

Uniformeki bornatua bigarren momentuan

PS ereduaren propietate garrantzitsu bat da sistemako egonaldiaren batezbesteko iraute denbora soilik beharrezko zerbitzuaren itxaropenaren menpe dagoela, [23]. Hau argumentu indartsua izan da bezeroen artean zerbitzua partekatzen duten ereduaren garrantzia erakusteko FCFS beza-lako politika klasikoaren aurrean. Noski, Pollaczek-Khinchine-en formula klasikoak erakusten du batezbesteko itxaron denbora infinitoruntz handitzen dela, beharrezko zerbitzuaren bigarren momentua handitzen denean. DPS ereduarentzat, (4.1) Ekuazioak ez digu ahalbidetzen sistemako egonaldiaren batezbesteko iraute denborak beharrezko zerbitzuaren momentuekiko duen menpekotasunarekiko ondoriorik ateratzen.

Beraz, oso interesgarria da (4.20) hurbilketa beharrezko zerbitzuen bigarren momentuan uniformeki bornatua egotea. Hau erakusteko, lehendabizi ohartu $\min\{B_{U_t}, b \frac{g_{U_t}}{g_k}\}^2 \leq B_{U_t} b \frac{g_{U_t}}{g_k}$, zeinak zuzenean inplikutzen duen (4.20) Ekuazioko lehen hiru terminoak uniformeki bornatuak direla beharrezko zerbitzuaren lehenengo momentuaren menpe dagoen termino batengatik. Trafiko geldo terminoa dugu aztertzeke, $\frac{\mathbb{E}[B^2]}{\sum_{j=1}^K \alpha_j \mathbb{E}[B_j^2]/g_j}$. Izan bedi j^* non $\mathbb{E}[B_{j^*}^2] \geq \mathbb{E}[B_j^2]$, $\forall j$. Orduan

$$\frac{\mathbb{E}[B^2]}{\sum_{j=1}^K \alpha_j \mathbb{E}[B_j^2]/g_j} = \frac{\sum_j \alpha_j \mathbb{E}[B_j^2]}{\sum_{j=1}^K \alpha_j \mathbb{E}[B_j^2]/g_j} \leq \frac{\mathbb{E}[B_{j^*}^2]}{\alpha_{j^*} \mathbb{E}[B_{j^*}^2]/g_{j^*}} = \frac{g_{j^*}}{\alpha_{j^*}}.$$

Amaitzeko, ondorioztatzen dugu (4.20) goitik bornatu daitekeela beharrezko zerbitzuaren itxaropenarekiko menpekotasuna soilik duen termino batengatik. Honek erakusten du, DPS ereduak sistemaren errendimendu egoki bat eskaintzen duela, beharrezko zerbitzuen banaketak oso aldakorak izan arren.

4.5.2 Sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatua

(4.20) Ekuazioaren korolariorik bezala k klaseko etiketatutako bezeroaren sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatua lortzen dugu.

4.5.2 Korolariorik k klaseko etiketatutako bezeroaren sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatuaren bigarren mailako trafikoko arinaren eta geldoaren interpolazioa honakoa da

$$\begin{aligned} \overline{S}_k^{DPS,INT}(\lambda) &:= \int_0^\infty \overline{S}_k^{INT}(\lambda, b) dF_k(b) \\ &= \mathbb{E}[B_k](1 + \rho) + \lambda \mathbb{E} \left[\frac{1}{2} \left(1 + \frac{g_k}{g_{U_t}} \right) \min \left\{ B_{U_t}, B_k \frac{g_{U_t}}{g_k} \right\}^2 \right. \\ &\quad \left. - \left(B_k \frac{g_{U_t}}{g_k} + \frac{g_k}{g_{U_t}} B_{U_t} \right) \min \left\{ B_{U_t}, B_k \frac{g_{U_t}}{g_k} \right\} + B_k \frac{g_{U_t}}{g_k} B_{U_t} \right] \\ &\quad + \frac{(\lambda \mathbb{E}[B])^2}{(1 - \lambda \mathbb{E}[B])} \frac{\mathbb{E}[B_k]}{g_k} \frac{\mathbb{E}[B^2]}{\sum_{j=1}^K \alpha_j \mathbb{E}[B_j^2]/g_j}. \end{aligned} \quad (4.22)$$

4.5.3 Beharrezko zerbitzu esponontziala

Atal honetan bezeroen beharrezko zerbitzuak banaketa esponontziala duen kasuan jarriko dugu arreta. Gogoratu, B_i zorizko aldagai batek banaketa esponontziala duela baldin eta $\mathbb{P}(B_i \leq b_i) = 1 - e^{-b_i/\mathbb{E}[B_i]}$. 4.5.3. Atalean trafikoko arinaren eta geldoaren interpolazioaren bitartez lorturiko sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatuaren hurbilketa sinplifikatzen dugu eta sisteman bi bezero klase dauden kasuan (4.5) eta (4.6) Ekuazio zehatzekin alderatzen dugu. 4.5.3 Atalean, errore erlatiboa (beharrezko zerbitzu desberdinentzat) kalkulatu dugu eta sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatuaren hurbilketa bezero arbitrario batentzat zehatza dela egiaztatzen dugu.

Sistemako egonaldiaren batezbesteko iraute denbora baldintzatua eta ez-baldintzatua

Beharrezko zerbitzuen banaketa esponontziala den kasuan, sistemako egonaldiaren batezbesteko iraute denbora baldintzatuaren eta ez-baldintzatuaren hurbilketak asko sinplifikatzen dira. Honako hau, (4.20) Ekuazioa eta 4.5.2 Korolarioraren ondorio zuzena da.

4.5.3 Korolariorik *Onar dezagun, k klaseko bezeroen beharrezko zerbitzuak $1/\mu_k, k = 1, \dots, K$, batezbestekoa duen banaketa esponontziala duela. k klaseko etiketatutako bezeroaren sistemako egonaldiaren batezbesteko iraute denbora baldintzatuaren bigarren mailako trafikoko arinaren eta gel-*

doaren interpolazioa, b beharrezko zerbitzua duena, honakoa da.

$$\begin{aligned} \overline{S}_k^{DPS,INT}(\lambda, b) &= b + \lambda \mathbb{E}[B]b + \lambda \sum_{j=1}^K \frac{\alpha_j}{\mu_j^2} \left(1 - \frac{g_k}{g_j}\right) \left(1 - e^{-b \frac{g_j}{g_k} \mu_j}\right) \\ &\quad + \frac{(\lambda \mathbb{E}[B])^2}{(1 - \lambda \mathbb{E}[B])} \frac{b}{g_k} \frac{\sum_{j=1}^K \alpha_j / \mu_j^2}{\sum_{j=1}^K \alpha_j / (\mu_j^2 g_j)}, \end{aligned}$$

eta sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatua honakoa

$$\begin{aligned} \overline{S}_k^{DPS,INT}(\lambda) &:= \int_0^\infty \overline{S}_k^{INT}(\lambda, b) dF_k(b) \\ &= \frac{1}{\mu_k} + \frac{1}{\mu_k} \lambda \mathbb{E}[B] + \lambda \sum_{j=1}^K \alpha_j \frac{1}{\mu_j} \frac{(g_j - g_k)}{g_j \mu_j + g_k \mu_k} \\ &\quad + \frac{(\lambda \mathbb{E}[B])^2}{(1 - \lambda \mathbb{E}[B])} \frac{1}{g_k \mu_k} \frac{\sum_{j=1}^K \alpha_j / \mu_j^2}{\sum_{j=1}^K \alpha_j / (\mu_j^2 g_j)}, \end{aligned} \quad (4.23)$$

$$\text{non } \mathbb{E}[B] = \sum_{j=1}^K \alpha_j / \mu_j.$$

Frogapena: Eraitza hauek (4.20) Ekuaziotik eta 4.5.2 Korolariotik ondoriozta daitezke kalkulu batzuk egin ondoren. ■

Bi bezero klaseren kasuan, Fayolle *et al* [10] artikuluan sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatuaren adierazpen itxiak erakusten dira, ikusi (4.5) eta (4.6) Ekuazioak. (4.23) $K = 2$ kasurako berridatziz, (4.5) eta (4.6) Ekuazioekiko antzekotasuna argia da:

$$\begin{aligned} \overline{S}_k^{DPS,INT}(\lambda) &= \frac{1}{\mu_k(1 - \rho)} \left(1 + \rho^2 \left(-1 + \frac{\sum_{j=1}^K \alpha_j / \mu_j^2}{g_k \sum_{j=1}^K \alpha_j / (\mu_j^2 g_j)} \right) \right. \\ &\quad \left. + \frac{\mu_k \rho_{-k} (g_{-k} - g_k)}{D} \frac{(1 - \rho) D}{\mu_1 g_1 + \mu_2 g_2} \right), \end{aligned} \quad (4.24)$$

non $k = 1, 2$, $-k = \text{mod}(k, 2) + 1$ eta $D = \mu_1 g_1 (1 - \rho_1) + \mu_2 g_2 (1 - \rho_2)$.

Zuzenean ikus daiteke eraitza zuzenekin alderatuz, $\rho^2 \left(-1 + \frac{\sum_{j=1}^K \alpha_j / \mu_j^2}{g_k \sum_{j=1}^K \alpha_j / (\mu_j^2 g_j)} \right)$ eta $\frac{(1 - \rho) D}{\mu_1 g_1 + \mu_2 g_2}$ terminoetan dagoela desberdintasuna.

Erraz egiaztatu daiteke gure hurbilketa zuzena dela kargaren balioen bi muturretan, $\rho = 0$ eta $\rho = 1$. Hau da, $\overline{S}_k^{DPS,INT}(0) = 1/\mu_k$ eta $\lim_{\lambda \rightarrow 1/\mathbb{E}[B]} (1 - \rho) \overline{S}_k^{DPS,INT}(\lambda) = \frac{1}{\mu_k} \left(1 + \frac{\mu_k \rho_{-k} (g_{-k} - g_k)}{D} \right)$ adierazpenak betetzen dira.

Izan bedi Rel.Error_k k klaseko bezero baten errore erlatiboa, hau da,

$$\text{Rel.Error}_k = \frac{\overline{S}_k^{DPS}(\lambda) - \overline{S}_k^{DPS,INT}(\lambda)}{\overline{S}_k^{DPS}(\lambda)},$$

$k = 1, 2$. Kontsidera dezagun $g_1 + g_2 = 1$. Orduan, lortzen dugu $\lim_{g_1 \uparrow 1} \text{Rel.Error}_1 = \rho_1^2 \cdot 100\%$ eta

$$\lim_{g_1 \uparrow 1} \text{Rel.Error}_2 = \frac{\mu_2 \rho_1 - (1 - \rho_1) \left(\rho^2 \frac{\rho_1 \mu_2}{\rho_2} + \mu_2 \rho_1 (1 - \rho) \right)}{\mu_1 (1 - \rho_1) + \mu_2 \rho_1} \cdot 100\%.$$

Beraz, 1 klaseko bezeroen errore erlatiboa ($g_1 \uparrow 1$ limitepean) handitu egiten da 1 klaseko karga handitzen den heinean, baina ez da 2 klaseko parametroen menpekkoa. Emaiza berbera lortu da 4.5.1 Atalean sistemako egonaldiaren batezbesteko iraute denbora baldintzatuarentzat, beharrezko zerbitzu banaketa orokorra duen bezero arbitrario batentzat. Gainera, 2 klaseko errore erlatibo absolutua ($g_1 \uparrow 1$ limitepean) handitu egiten da 2 klaseko karga txikitzean.

4.1 Irudian $K = 2$ kasurako sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatua irudikatzen dugu g_1 -rekiko. Kontsideraturiko parametroak dira, $\rho_1 = 0.2, \rho_2 = 0.4, \mu_1 = 1, \mu_2 = 1, g_2 = 1 - g_1$, eta lehenago aurkezturiko formuletatik lortzen dugu $\lim_{g_1 \uparrow 1} \text{Rel. Error}_1 = 4\%$, $\lim_{g_1 \uparrow 1} \text{Rel. Error}_2 = -0.8\%$, $\lim_{g_1 \downarrow 0} \text{Rel. Error}_1 = -12.8\%$, $\lim_{g_1 \downarrow 0} \text{Rel. Error}_2 = 16\%$, zeinak irudiko muturreko puntuekin bat egiten duten.

Bezero arbitrario baten sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatua

Atal honetan sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatuaren errore erlatiboa kalkulatzen dugu bezero arbitrario batentzat, $K = 2$ kasurako eta μ_1 edo μ_2 aldagaietako batek muturreko balioak hartzen dituenean. Eta egiaztatzen dugu gure hurbilketa zehatza dela bezero arbitrario baten sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatuarentzat, bezero guztien beharrezko zerbitzuak berdinak diren kasuan.

Izan bedi Rel.Error bezero arbitrario baten errore erlatiboa, hau da,

$$\text{Rel.Error} = \left(1 - \frac{\sum_{k=1}^2 \alpha_k \overline{S_k^{DPS,INT}}(\lambda)}{\sum_{k=1}^2 \alpha_k \overline{S_k^{DPS}}} \right) \cdot 100\%,$$

non $\overline{S_1^{DPS}}(\lambda)$ eta $\overline{S_2^{DPS}}(\lambda)$, (4.5) eta (4.6) Ekuazioetan ematen diren, hurrenez hurren.

Orduan, ρ_1 eta ρ_2 konstante mantentzen baditugu, $\lambda_1 = \rho_1 \mu_1$ eta $\lambda_2 = \rho_2 \mu_2$ aldatuz, lortzen dugu

$$\lim_{\mu_2 \downarrow 0} \text{Rel.Error} = \frac{\rho_1 \left(\frac{\rho_2 (g_2 - g_1)}{g_1} \left(\frac{1}{1 - \rho_1} - 1 + \rho \right) - \rho^2 \left(-1 + \frac{g_2}{g_1} \right) \right)}{\rho_1 \left(1 + \frac{\rho_2 (g_2 - g_1)}{g_1 (1 - \rho_1)} \right) + \rho_2} \cdot 100\% \quad (4.25)$$

eta $\lim_{\mu_1 \uparrow \infty} \text{Rel.Error} = \lim_{\mu_2 \downarrow 0} \text{Rel.Error}$. Ekuazio honen atzean uler daitekeena horrela azal dezakegu: $\mu_1 \rightarrow \infty$ eta $\lambda_1 \rightarrow \infty$, hau da, tamaina txikiko 1 klaseko bezero asko edukitzea eta $\mu_2 \rightarrow 0$ eta $\lambda_2 \rightarrow 0$, hau da, tamaina handiko 2 klaseko bezero gutxi izatea baliokideak direla.

4.2 Irudian bezero arbitrario baten sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatuaren errore erlatiboa irudikatzen dugu. Finkatzen ditugu ρ_1, ρ_2 eta μ_1 , eta aldatzen dugu $\lambda_2 = \rho_2 \mu_2$. Aukeraturiko parametroak dira $\rho_1 = 0.2, \rho_2 = 0.6, \mu_1 = 1, g_1 = 1, g_2 = 2$.

(4.25) Ekuaziotik lorturiko emaitzek, $\lim_{\mu_2 \downarrow 0} \text{Rel.Error} = -0.2105\%$ eta $\lim_{\mu_2 \uparrow \infty} \text{Rel.Error} = 8.307\%$ irudiko muturreko puntuekin bat egiten dutela ikus dezakegu.

Ikus dezagun orain, bezero arbitrario baten sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatua zehatza dela bezero guztien batezbesteko beharrezko zerbitzua berdina denean, hau da, $\mathbb{E}[B_j] = 1/\mu, \forall j = 1, \dots, K$.

Lehenago aipatu bezala, bezero arbitrario baten sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatua honela definiturik dago $\overline{S^{DPS}}(\lambda) := \sum_{k=1}^K \alpha_k \overline{S_k^{DPS}}(\lambda)$. Banaketa esponentziala duten beharrezko zerbitzuak eta $\mathbb{E}[B_k] = 1/\mu, \forall k = 1, \dots, K$ onartzen ditugunez, sistemako bezero kopuru totala PS ilararen banaketa bera du $\lambda = \sum_{k=1}^K \lambda_k$ iritsiera tasa eta μ zerbitzu tasarekin. Beraz, Little-en legetik ondoriozta dezakegu, sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatu totala $M/M/1$ ilarakoaren bera dela, hau da, $\frac{1/\mu}{1-\rho}$.

trafiko arinaren eta geldoaren interpolaziotik dugu

$$\begin{aligned} \overline{S^{DPS,INT}}(\lambda) &= \sum_{k=1}^K \alpha_k \overline{S_k^{DPS,INT}}(\lambda) \\ &= \frac{1+\rho}{\mu} + \lambda \sum_{k=1}^K \alpha_k \sum_{j=1}^K \frac{\alpha_j}{\mu} \frac{(g_j - g_k)}{g_j \mu + g_k \mu} \\ &\quad + \frac{(\lambda/\mu)^2}{(1-\lambda/\mu)} \sum_{k=1}^K \alpha_k \frac{1}{\mu g_k} \frac{\sum_{j=1}^K \alpha_j / \mu^2}{\sum_{j=1}^K \alpha_j / (\mu^2 g_j)} \\ &= \frac{1+\rho}{\mu} + \frac{1}{\mu} \frac{\rho^2}{1-\rho} \sum_{k=1}^K \frac{\alpha_k}{g_k} \frac{1}{\sum_{j=1}^K \alpha_j / g_j} = \frac{1/\mu}{1-\rho}, \end{aligned} \quad (4.26)$$

non erabili dugun

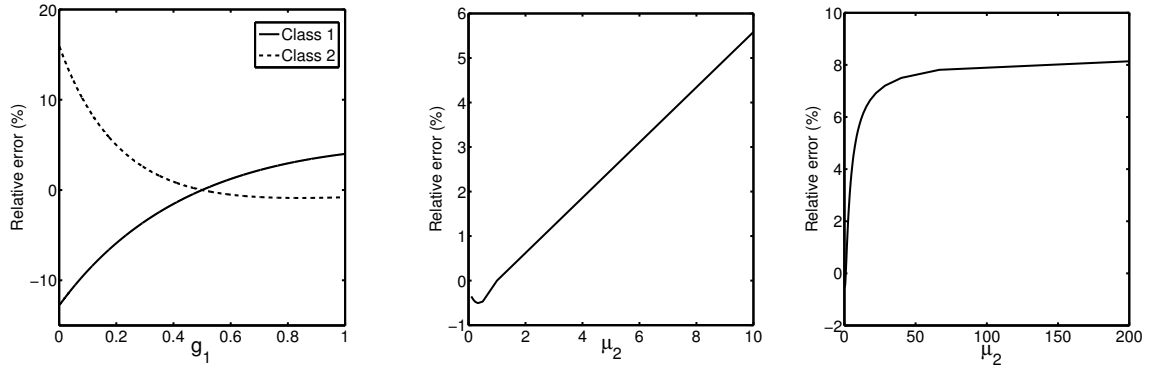
$$\begin{aligned} \sum_{k=1}^K \alpha_k \sum_{j=1}^K \alpha_j \frac{(g_j - g_k)}{g_j \mu + g_k \mu} &= \frac{1}{\mu} \sum_{k=1}^K \sum_{j=1}^K \alpha_k \alpha_j \frac{g_j - g_k}{g_j + g_k} \\ &= \frac{1}{\mu} \sum_{k=1}^K \sum_{j=1}^{K-1} \alpha_k \alpha_j \left(\frac{g_j - g_k}{g_j + g_k} + \frac{g_k - g_j}{g_j + g_k} \right) = 0. \end{aligned}$$

Honegatik, trafiko arinaren eta geldoaren interpolazio bidezko hurbilketa zehatza da.

4.2 Irudian (*ezkerra*) ikus daiteke $\mu_2 = 1$ denean, hau da, bi klaseetako beharrezko zerbitzua berbera denean, errore erlatiboa zero dela, (4.26) Ekuazioan frogatu bezala.

4.6 Emaitza numerikoak

Atal honetan numerikoki aztertzen dugu kapitulu honetan lorturiko hurbilketaren zehaztasuna. 4.6.3 Atalean sistemako egonaldiaren batezbesteko iraute denbora baldintzatua kontsideratzen dugu eta 4.6.3 Atalean sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatua, zeinen



4.1. Irudia: $K = 2$ de-
nean, sistemako egonaldiaren
batezbesteko iraute denbora
ez-baldintzatuaren errore erla-
tiboia g_1 -rekiko.

4.2. Irudia: Bezero arbitrario baten sistemako egonaldiaren
batezbesteko iraute denbora ez-baldintzatu.

hurbilketak (4.20) Ekuazioan eta 4.5.2 Korolarioan aipatzen ditugun, hurrenez hurren.

4.1 Atalean aipatu bezala, Fayolle et al.-ek [10] artikuluan sistemako egonaldiaren batezbes-
teko iraute denbora baldintzatuaren eta ez-baldintzatuaren adierazpen analitikoak soilik beharrez-
ko zerbitzuaren banaketa esponontziala den kasuan lortzen dituzte. Batezbesteko esponontziala
duten beharrezko zerbitzuen kasuan, hurbilketaren zehaztasuna [10] artikuluan lorturiko emaitza
zehatzekin alderatuz ebaluatuko dugu.

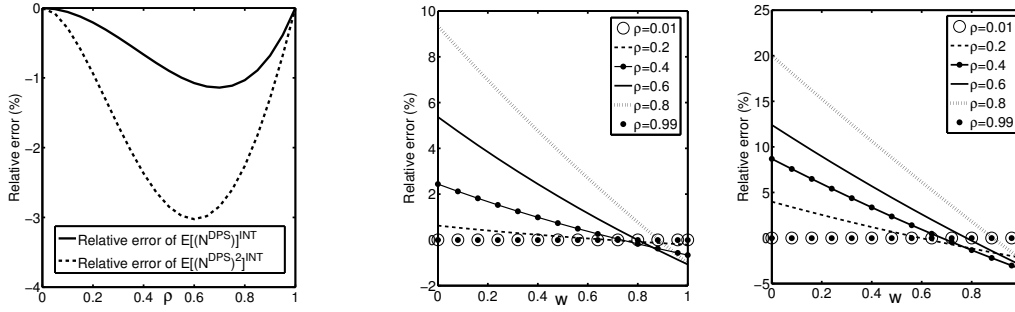
Hurbilketaren zehaztasunaren ulermen zabalagoa izateko, banaketa hiperesponontziala eta
Pareto ere kontsideratuko ditugu.

Hiperesponontzial eta Pareto banaketek *hazard-rate* beherakorra dute, eta beraien bigarren
momentuak oso handiak bihurtu daitezke. Honegatik interneteko beharrezko zerbitzuen banaketa
modelizatzeko egokiak dira.

B_i aldagaiak m_i faseko banaketa hiperesponontziala duela esaten dugu baldin $\mathbb{P}(B_i \leq b_i) =$
 $1 - \sum_{k=1}^{m_i} p_{ik} e^{(-b_i/\mathbb{E}[B_{ik}])}$, non p_{ik} i klaseko bezero batek $\mathbb{E}[B_{ik}]$ banaketa esponontziala izateko
probabilitatea den. Sistemako egonaldiaren batezbesteko iraute denboraren adierazpen zehatzak
lortzeko beharrezko zerbitzuak hiperesponontzialki banatua daudenean, honako behaketa egiten
dugu. $k = 1, \dots, m_i$ klaseek banaketa esponontziala badute eta DPS pixu berdinak badituzte, $g_1 =$
 $\dots = g_{m_i}$, orduan, i klase bakar batekoak balira bezala uler daiteke $p_{ik} = \lambda_k / \sum_{l=1}^{m_i} \lambda_l$ eta $\mathbb{E}[B_{ik}] =$
 $\mathbb{E}[B_k]$ parametroekin, $k = 1, \dots, m_i$ fase bakoitzarentzat. Honek sistemaren errendimendua (4.2)
eta (4.4) Ekuazioak erabiliz kalkulatzeko ahalbidetzen digu.

B_i aldagaiak d_i eta γ_i parametroetako Pareto banaketa duela esaten dugu baldin $\mathbb{P}(B_i \leq b_i) =$
 $1 - \left(\frac{1}{1 + d_i b_i} \right)^{\gamma_i}$. Sistemako egonaldiaren batezbesteko iraute denboraren adierazpen zehatzak
lortzeko (4.1) Ekuazioa numerikoki ebatzi genuen. Dena den, lorturiko emaitza ez zen orekatua
eta, horregatik, azkenean, DPS erdua simulatu genuen (MATLAB erabiliz).

Ohartu, banaketa hiperesponontzialak (2.2) Ekuazioko baldintza betetzen duela. Aldiz, Pa-
reto banaketak ez. Pareto banaketako γ_i parametroaren edozein balioetarako, γ_i baina handiagoak
diren momentuak ez dira bornatuak. Hala ere, ondoren ikusiko dugun bezala, (2.2) Ekuazioa bete



4.3. Irudia: 1 Eszenatokia. Bezero kopuru totalaren lehen eta bigarren momentuen errore erlatiboak, banaketa esponentziala duten beharrezko zerbitzuetarako.

4.4. Irudia: 2 Eszenatokia. Bezero kopuru totalaren lehen (ezkerre) eta bigarren (eskuia) momentuen errore erlatiboa irudikatzen dugu w parametroarekiko kargaren balio desberdinetarako

ez arren, trafiko arinaren eta geldoaren interpolazioa nahikoa zehatza da.

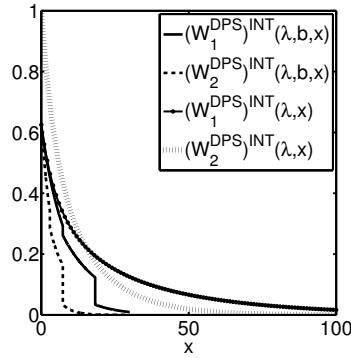
Atal honetan zehar, errendimendua neurtzeko erabiliko dugun irizpidea errore erlatiboa izango da. Esate baterako, bezero kopuruaren lehen eta bigarren momentuak $100\% \times \frac{\mathbb{E}[N^{DPS}] - \mathbb{E}[N^{DPS,INT}]}{\mathbb{E}[N^{DPS}]}$ eta $100\% \times \frac{\mathbb{E}[(N^{DPS})^2] - \mathbb{E}[(N^{DPS,INT})^2]}{\mathbb{E}[(N^{DPS})^2]}$ formulen bitartez kalkulatu ditugu, hurrenez hurren. Eta batezbesteko sistemako egonaldiaren iraupen baldintzaturik eta ez-baldintzaturik honela $100\% \times \frac{S_k^{DPS}(\lambda, b) - S_k^{DPS,INT}(\lambda, b)}{S_k^{DPS}(\lambda, b)}$ eta $100\% \times \frac{S_k^{DPS}(\lambda) - S_k^{DPS,INT}(\lambda)}{S_k^{DPS}(\lambda)}$, hurrenez hurren.

Lorturiko emaitza numerikoak erakutsi baino lehen, ondorio nagusiak laburbilduko ditugu:

- Hurbilketa nahikoa zehatza da parametroen balio desberdin askotarako.
- Parametro zehatz batzuentzat, sistemako egonaldiaren batezbesteko iraute denbora baldintzatuaren errore erlatiboa handitu egiten da etiketatutako bezeroaren beharrezko zerbitzua handitzen den heinean.
- Errorrea handitu egiten da pixuen arteko desberdintasuna handitzen den heinean.
- Edozein kasutan, errore erlatibo handiena kargaren 0 eta 1 balioen artean jazotzen da.
- Sistemako egonaldiaren batezbesteko iraute denbora baldintzatuaren errore erlatibo handiena b beharrezko zerbitzuaren balio ez-ohikoentzat gertatzen da. Honek azaltzen du, sistemako egonaldiaren batezbesteko iraute denbora baldintzatuaren hurbilketaren zehaztasuna.

4.6.1 Bezero kopuruaren banaketa

4.3.1 Proposizioan lorturiko hurbilketaren zehaztasuna neurtzen dugu banaketa esponentziala duten beharrezko zerbitzuetarako, (4.12) eta (4.14) ekuazioetan lorturiko lehen eta bigarren momentuak erabiliz, hurrenez hurren.



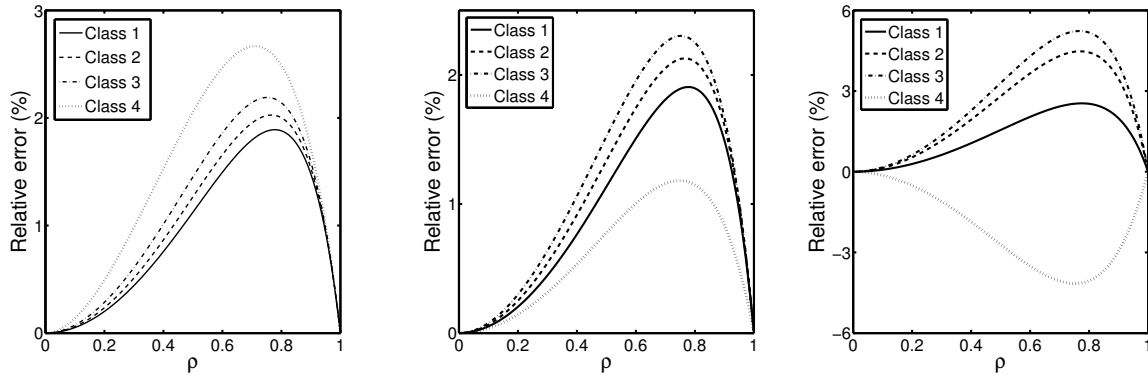
4.5. Irudia: 1 Eszenatokia. k klaseko bezeroen itxaron denbora baldintzatuaren eta ez-baldintzatuaren funtzio osagarria.

1 *Eszenatokia.* 4.3 Irudian bezero kopuru totalaren lehen eta bigarren momentuen errore erlatiboak irudikatzen ditugu kargarekiko. Bi klase kontsideratzen ditugu $\mathbb{E}[B_1] = 11/3$, $\mathbb{E}[B_2] = 44/3$ batezbesteko zerbitzuekin. Iritsiera 1 klasekoa (2 klasekoa) dela onartzen dugu $\alpha_1 = 8/12$ ($\alpha_2 = 4/12$) probabilitatearekin. Pixuak finkatzen ditugu $g_1 = 2$, $g_2 = 5$. Ikus daiteke lehen eta bigarren momentuak nahikoa zehatzak direla gehienez 1.2% eta 3% errore erlatibo absolutuarekin, hurrenez hurren.

2 *Eszenatokia.* 4.4 Irudian 2 klase kontsideratzen ditugu. 1 klaseko bezeroen beharrezko zerbitzuak μ_1 tasako banaketa esponentziala du eta 2 klasekoenak banaketa hiperesponentzial degeneratu bat du, (3.7) Ekuazioan definiturikoa, honako parametroekin: $m_2 = 2$, $\beta_{21} = w$, $\beta_{22} = 1 - w$, $\mathbb{E}[B_{21}] = 1/(\mu_2 w)$ eta $\mathbb{E}[B_{22}] = 0$. Finka ditzagun $g_1 = 2$, $g_2 = 5$, $\alpha_1 = 7/12$, $\alpha_2 = 5/12$, $\mathbb{E}[B_1] = 11/3$, $\mathbb{E}[B_2] = 44/3$. 6.2 Irudian bezero kopuru totalaren lehen eta bigarren momentuen errore erlatiboa irudikatzen dugu w parametroarekiko kargaren balio desberdinetarako. Ikusi $\rho \approx 0$ eta $\rho \approx 1$ kasuetan gure hurbilketa zehatza dela, espero genuen bezala. Errore handiena kargaren erdibideko balioetarako eta $w \downarrow 0$ kasurako gertatzen da.

4.6.2 Itxaron denboraren banaketa

4.5 Irudian 4.4.1 Proposizioan erakutsiriko emaitza erabiltzen dugu itxaron denboraren banaketaren hurbilketa irudikatzeko. Itxaron denbora baldintzatuarentzat $b = 11/3$ finkatzen dugu. 2 klaseari pixu handiagoa dagokio ($g_1 = 2$, $g_2 = 5$), eta ondorioz, 4.5 Irudian ikus daitekeen bezala, 2 klaseko itxaron denbora baldintzatua estokastikoki txikiagoa da 1 klasekoena baino. Dena den, 1 klaseko bezeroen beharrezko zerbitzua 2 klasekoena baino txikiagoa da eta ondorioz, 1 klaseko bezeroen itxaron denbora ez-baldintzatuaren probabilitatea x baino handiagoa izatea 2 klasekoena baino handiagoa da, x txikiatarako.



4.6. Irudia: 3 Eszenatokia: i klaseko etiketatutako bezeroaren sistemako egonaldiaren batezbesteko iraute denbora baldintzatuaren hurbilketaren errore erlatiboa b_i beharrezko zerbitzuak $\mathbb{P}(B_i \leq b_i) = 0.01$ (ezkerra), $\mathbb{P}(B_i \leq b_i) = 0.50$ (erdia), $\mathbb{P}(B_i \leq b_i) = 0.99$ (eskuia) betetzen duelarik.

4.6.3 Sistemako egonaldiaren iraute denbora

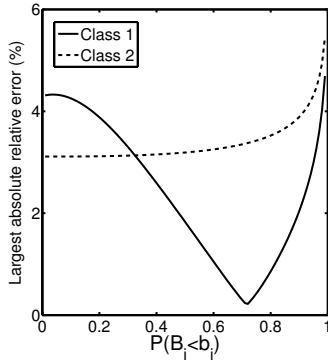
Atal honetan (4.20) Ekuazioan eta 4.5.2 Korolarioan lorturiko sistemako egonaldiaren batezbesteko iraute denbora baldintzatuaren eta ez-baldintzatuaren hurbilketaren zehaztasuna neurtzen dugu, hurrenez hurren.

Sistemako egonaldiaren batezbesteko iraute denbora baldintzatua

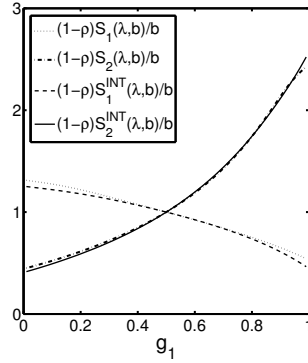
Atal honetan 4.20 Ekuazioan lorturiko sistemako egonaldiaren batezbesteko iraute denbora baldintzatuaren hurbilketaren zehaztasuna neurtzen dugu.

3 Eszenatokia. 4.6 Irudian $K = 4$ klase kontsideratzen ditugu beharrezko zerbitzuek banaketa esponentziala dutelarik. Klaseei dagozkien parametroak finkatuak daude eta iritsiera tasa totala aldatzen dugu kargaren balio guztiak kontuan hartzeko. Honako parametroak kontsideratzen ditugu $\mathbb{E}[B_1] = 2$, $\mathbb{E}[B_2] = 5$, $\mathbb{E}[B_3] = 7$, $\mathbb{E}[B_4] = 10$, $g_1 = 30$, $g_2 = 25$, $g_3 = 20$, $g_4 = 10$, eta $\alpha_1 = 10/36$, $\alpha_2 = 5/36$, $\alpha_3 = 8/36$, $\alpha_4 = 13/36$ non $\lambda_i = \alpha_i * \lambda$, $i = 1, \dots, 4$, λ iritsiera tasa totala izanik. 4.6 Irudian i klaseko etiketatutako bezeroaren sistemako egonaldiaren batezbesteko iraute denbora baldintzatuaren hurbilketaren errore erlatiboa irudikatzen dugu, non b_i i klaseko etiketatutako bezeroaren tamaina den, eta $\mathbb{P}(B_i \leq b_i) = 0.01$, $\mathbb{P}(B_i \leq b_i) = 0.50$ eta $\mathbb{P}(B_i \leq b_i) = 0.99$ probabilitateak jazotzeko aukeraturik dagoen, hurrenez hurren. Irudian ikus daitekeen bezala, sistemako egonaldiaren batezbesteko iraute denbora baldintzatuaren errore erlatiboa txikia mantentzen da, %6-aren azpitik.

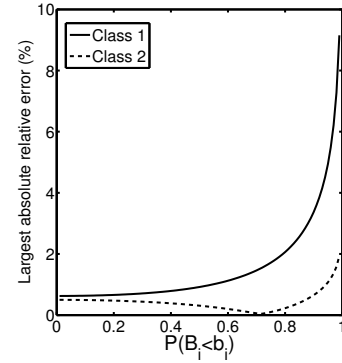
4 Eszenatokia. 4.7 Irudian beharrezko zerbitzua esponentzialki banaturik duten bi klase $K = 2$ kontsideratzen ditugu. Honako parametroak finkatzen ditugu: $\mathbb{E}[B_1] = 2$, $\mathbb{E}[B_2] = 1$, $g_1 = 1$, $g_2 = 3$, $\alpha_1 = 0.415$, $\alpha_2 = 0.585$ non $\lambda_i = \alpha_i * \lambda$, i klaseko etiketatutako bezeroaren beharrezko zerbitzua 0 eta $b_{i,max}$ artean mugi dadin ahalbidetzen dugu $\mathbb{P}(B_i \leq b_{i,max}) = 0.99$ izanik eta b bakoitzarentzat errore erlatibo absolutu handiena irudikatzen dugu $\rho \in [0, 1)$ balioetarako. Errore erlatibo handiena %6-aren azpitik mantentzen da.



4.7. Irudia: 4 Eszenatokia. Sistemako egonaldiaren batezbesteko iraute denbora baldintzatuaren errore erlatibo absolutu handiena $\mathbb{P}(B_i \leq b_i)$ funtzio gisa.



4.8. Irudia: 5 Eszenatokia. Sistemako egonaldiaren batezbesteko iraute denbora baldintzaturik g_1 aldagaiarekiko.

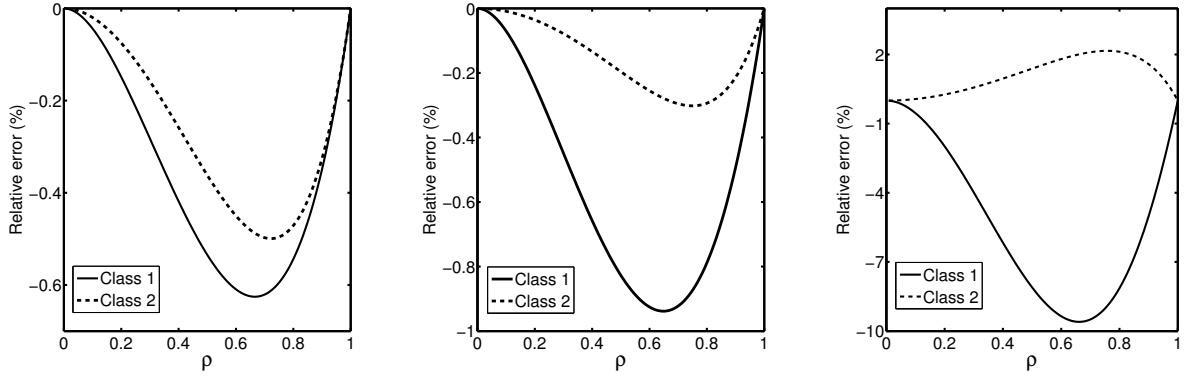


4.9. Irudia: 6 Eszenatokia. Sistemako egonaldiaren batezbesteko iraute denbora baldintzatuaren errore erlatibo absolutu handiena $\mathbb{P}(B_i \leq b_i)$ funtzio gisa.

5 Eszenatokia. 4.8 Irudian berriz ere beharrezko zerbitzua esponentzialki banaturik duten bi klase $K = 2$ kontsideratzen ditugu. Honako parametroak finkatzen ditugu: $\mathbb{E}[B_1] = 2$, $\mathbb{E}[B_2] = 1$, $\lambda_1 = 0.2$, $\lambda_2 = 1.5\lambda_1$ eta $b = 1$. $g_2 = 1 - g_1$ aukeratu eta g_1 aldatzen dugu ardatz horizontalean. Irudian, sistemako egonaldiaren batezbesteko iraute denbora baldintzaturik eta hurbilketa irudikatzen ditugu. 4.5.1 Atalean aipaturiko propietatea betetzen dela ikusten dugu, bereziki g_1 handitzean $\bar{S}_1^{INT}(\lambda, b)$ txikiagotu egiten da eta $\bar{S}_2^{INT}(\lambda, b)$ handitu. Gainera, irudian ikus daiteke hurbilketak zehaztasuna galtzen duela klase bati lehentasuna ematen diogunean, hau da, $g_1 \rightarrow 0$ edo $g_1 \rightarrow 1$.

6 Eszenatokia. 4.14 Irudian beharrezko zerbitzua hiperesponentzialki banaturik duten bi klase $K = 2$ kontsideratzen ditugu $\mathbb{E}[B_1] = 11/3$, $\mathbb{E}[B_2] = 44/3$ izanik. Banaketa hiperesponentzial bakoitzak 4 fase ditu. Parametroak honakoak dira: 1 klasearentzat $\mathbb{E}[B_{11}] = 3.5$, $\mathbb{E}[B_{12}] = 2$, $\mathbb{E}[B_{13}] = 5$, $p_{11} = 10/21$, $p_{12} = 5/21$, $p_{13} = 6/21$, eta 2 klasearentzat $\mathbb{E}[B_{21}] = 10$, $\mathbb{E}[B_{22}] = 15$, $\mathbb{E}[B_{23}] = 20$, $p_{21} = 4/15$, $p_{22} = 8/15$, $p_{23} = 3/15$. Pixuak $g_1 = 2$ eta $g_2 = 5$. Onartzen dugu, sistemara iristen den bezero bat 1 klasekoa (2 klasekoa) dela $\alpha_1 = 7/12$ ($\alpha_2 = 5/12$) probabilitatearekin. 1 Eszenatokian bezala, etiketatutako bezeroaren beharrezko zerbitzuak aukeratzen ditugu $\mathbb{P}(B_i \leq b_i) = 0.01, 0.5$ eta 0.99 bete dadin. Etiketatutako bezeroaren tamaina handitzen den heinean errorea handitu egiten dela ikus daiteke. Dena den, aipagarria da zein zehatza den hurbilketa.

4.9 Irudian 6 Eszenatokia kontsideratzen dugu. i klaseko etiketatutako bezeroaren beharrezko zerbitzua aldatzen dugu 0 eta $b_{i,max}$ tartean, $\mathbb{P}(B_i \leq b_{i,max}) = 0.99$ izanik. b bakoitzarentzat $\rho \in [0, 1]$ tartean aurki daitekeen errore erlatibo absolutu handiena irudikatzen dugu. Ikusi dezakegu errorea handitu egiten dela etiketatutako bezeroaren tamaina handitzen den heinean. Errore erlatibo absolutu handiena %22 da pixu txikiena duen klasearentzat eta %3 pixu handiena duen klasearentzat.



4.10. Irudia: 6 Eszenatokia : i klaseko etiketatutako bezzeroaren sistemako egonaldiaren batezbesteko iraute denbora baldintzatuaren errore erlatiboa b_i beharrezko zerbitzuak $\mathbb{P}(B_i \leq b_i) = 0.01$ (ezkerra), $\mathbb{P}(B_i \leq b_i) = 0.50$ (erdia), $\mathbb{P}(B_i \leq b_i) = 0.99$ (eskubia) betetzen duelarik.

Sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatua

Atal honetan 4.5.2 Korolarioan lorturiko sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatuaren zehaztasuna neurtzen dugu.

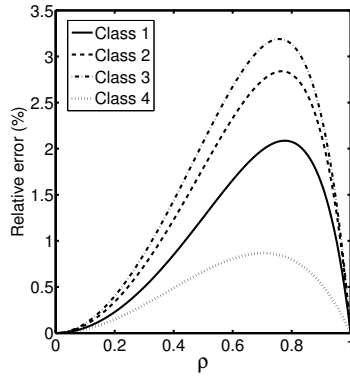
4.11 Irudian 3 Eszenatokiko parametro berdinak kontsideratzen ditugu eta ikusten dugu sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatuaren errore erlatiboa %3.5 baino txikiagoa dela.

4.12 Irudian beharrezko zerbitzua hiperesponentzialki banaturik duten bi klase $K = 2$ kontsideratzen ditugu. 6 Eszenatokiko parametro berdinak kontsideratzen ditugu. Ondorioztatzen dugu, sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatuaren errore erlatiboa %3 ingurukoa dela.

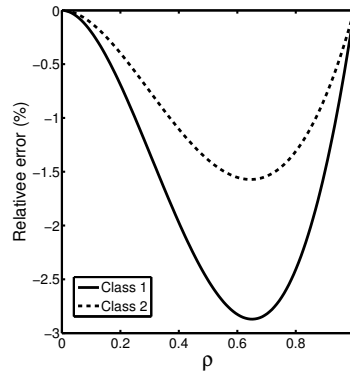
Atalaren hasieran aipatu bezala, ohartzen gara sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatuarentzat lortzen ari garen errore erlatiboak sistemako egonaldiaren batezbesteko iraute denbora baldintzatuarentzat lorturikoak baino txikiagoak direla. Honek izan dezakeen azalpen bat da sistemako egonaldiaren batezbesteko iraute denbora baldintzatuaren errore handiena, jazotzeko oso probabilitate txikia duten beharrezko zerbitzuen kasuan gertatzen dela.

4.13 Irudian beharrezko zerbitzua hiperesponentzialki banaturik duten bi klase $K = 2$ kontsideratzen ditugu. 6 Eszenatokiko parametro berdinak kontsideratzen ditugu. $g_2 = 1 - g_1$ aukeratu eta g_1 aldatzen dugu ardatz horizontalean. g_1 bakoitzarentzat sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatuarentzat errore erlatibo absolutu handiena kalkulatu dugu $\rho \in [0, 1)$ tartean. Ikus dezakegu sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatuaren errore erlatiboa gehienez %30 dela, eta hau 2 klaseak lehentasun osoa jasotzean gertatzen da.

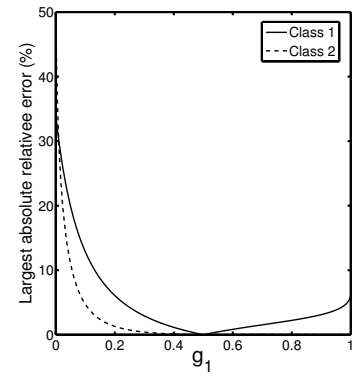
7 Eszenatokia. 4.16 Irudian 3 Eszenatokia kontsideratzen dugu baina Pareto banaketa duten beharrezko zerbitzuekin. Lau klase kontsideratzen ditugu honako parametroekin: $d_1 = 1/4, d_2 = 1/10, d_3 = 1/14, d_4 = 1/20$ eta $\gamma_1 = 3, \gamma_2 = 3, \gamma_3 = 3, \gamma_4 = 3$, eta ondorioz, $\mathbb{E}[B_1] = 2, \mathbb{E}[B_2] = 5$,



4.11. Irudia: Sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatuaren errore erlatiboa 3 Eszenatokian.



4.12. Irudia: Sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatuaren errore erlatiboa 6 Eszenatokian.



4.13. Irudia: Sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatuarentzat errore erlatibo absolutu handiena g_1 pixuarekiko 6 Eszenatokian.

$\mathbb{E}[B_3] = 7$, $\mathbb{E}[B_4] = 10$, eta pixuak $g_1 = 30$, $g_2 = 25$, $g_3 = 20$, $g_4 = 10$ eta $\alpha_1 = 10/36$, $\alpha_2 = 5/36$, $\alpha_3 = 8/36$, $\alpha_4 = 13/36$ non $\lambda_i = \alpha_i * \lambda$, $i = 1, \dots, 4$, λ iritsiera tasa totala izanik. Ohartu, $\mathbb{E}[B_i]$, g_i , α_i , $i = 1, \dots, 4$, 1 Eszenatokiko berberak direla. Azpimarratu nahi dugu $\rho = 1$ puntua ez dugula simulazioen bitartez lortu, trafiko geldo baldintzatik ordea, eta horregatik da zehatza. Ondoriozta dezakegu sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatuaren errore erlatibo absolutu handiena %5 ingurukoa dela.

8 Eszenatokia. 4.17 Irudian 6 Eszenatokia kontsideratzen dugu baina Pareto banaketa duten beharrezko zerbitzuekin. Bi klase kontsideratzen ditugu honako parametroekin: $d_1 = 3/22$, $d_2 = 3/88$ eta $\gamma_1 = 3$, $\gamma_2 = 3$, eta ondorioz, $\mathbb{E}[B_1] = 11/3$, $\mathbb{E}[B_2] = 44/3$, eta pixuak $g_1 = 2$ eta $g_2 = 5$. Onartzen dugu, sistemara iristen den bezero bat 1 klasekoa (2 klasekoa) dela $\alpha_1 = 7/12$ ($\alpha_2 = 5/12$) probabilitatearekin. Ohartu, $\mathbb{E}[B_i]$, g_i , α_i , $i = 1, \dots, 4$, 6 Eszenatokiko berberak direla. Azpimarratu nahi dugu $\rho = 1$ puntua ez dugula simulazioen bitartez lortu, trafiko geldo baldintzatik ordea, eta horregatik da zehatza. Ondoriozta dezakegu sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatuaren errore erlatibo absolutu handiena %5 baino txikiagoa dela.

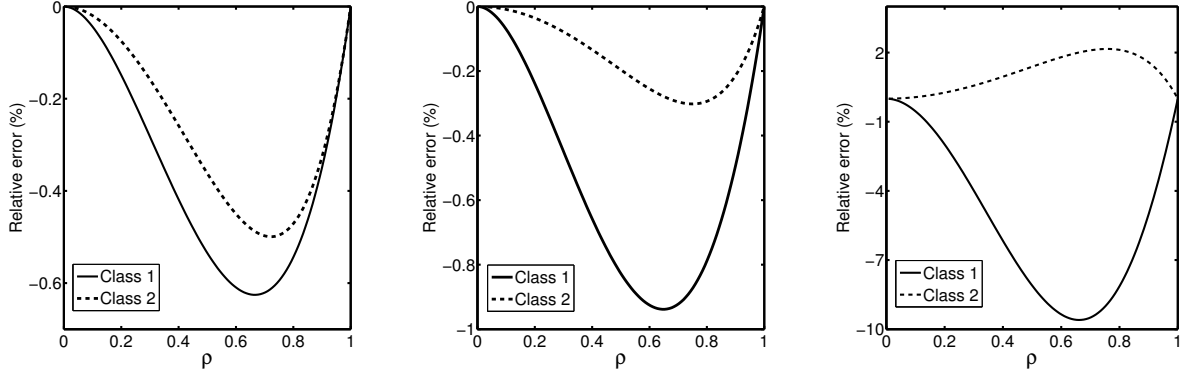
4.A Eranskina

4.A.1 4.3.1 Lemaren frogapena

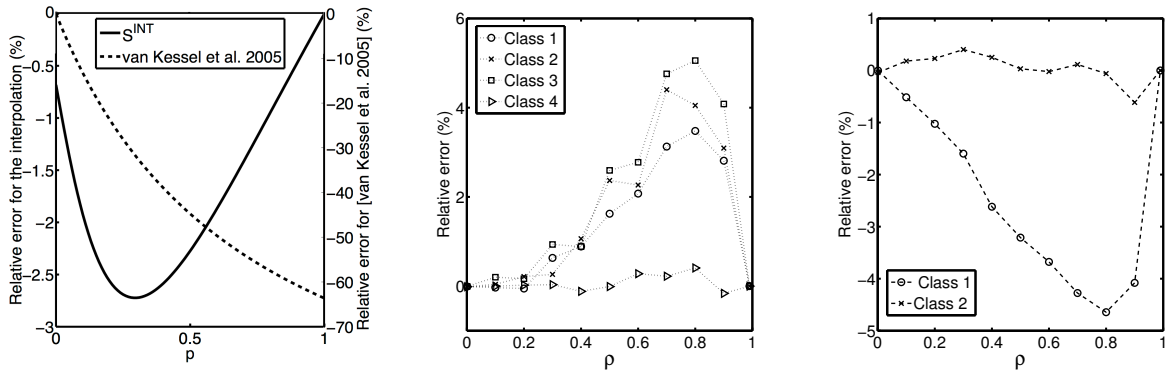
Bezero kopuru totalaren probabilitate funtzio sortzailearen zero, bat eta bi trafiko arinpeko deribatuak lortuko ditugu jarraian DPS ereduarentzat. Zero deribatuak honako berdintza betetzen du

$$(\psi^{DPS})^{(0)}(\lambda, \vec{z}) \Big|_{\lambda=0} = G\left(0, \vec{z} \Big| A(-\infty, \infty) = 0\right) = z_1^0 \cdots z_K^0 = 1. \quad (4.27)$$

Izan bedi t bezero bat sistemara iristen den unea, eta U_t aldagaiak bere klasea adierazten du. Bat deribatuarentzat ondoko bi kasuak bereiz ditzakegu:



4.14. Irudia: 6 Eszenatokia: i klaseko etiketatutako bezeroaren sistemako egonaldiaren batezbesteko iraute denbora baldintzatuaren errore erlatiboa b_i , beharrezko zerbitzua, honako berdintzak bete daitezten aukeratua izanik $\mathbb{P}(B_i \leq b_i) = 0.01$ (ezkerra), $\mathbb{P}(B_i \leq b_i) = 0.50$ (erdia), $\mathbb{P}(B_i \leq b_i) = 0.99$ (eskuia).



4.15. Irudia: 5 Eszenatokia: Batezbesteko sistemako egonaldiaren iraupenaren errore erlatiboa edozein bezerorentzat gure hurbilketarako eta [46] artikuluan lorturiko hurbilketarako.

4.16. Irudia: 6 Eszenatokia: kiko sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatuaren errore erlatiboa.

4.17. Irudia: 7 Eszenatokia: kiko sistemako egonaldiaren batezbesteko iraute denbora ez-baldintzatuaren errore erlatiboa.

Baldin $t > 0$, $z_{U_t}^0 = 1$ betetzen da. Ondorioz,

$$G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t\right) - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 0\right) = 0.$$

Baldin $t < 0$, honakoa dugu

$$\begin{cases} z_{U_t}^0 & \text{if } t + B_{U_t} < 0 \\ z_{U_t} & \text{if } t + B_{U_t} > 0, \end{cases}$$

hau da,

$$G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t\right) = \mathbb{E}[\mathbf{1}[t + B_{U_t} < 0] + \mathbf{1}[t + B_{U_t} > 0] z_{U_t}].$$

Ondorioz,

$$\begin{aligned} & (\psi^{DPS})^{(1)}(\lambda, \vec{z}) \Big|_{\lambda=0} \\ &= \int_{-\infty}^0 \left(G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t\right) - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 0\right) \right) dt \\ &= \mathbb{E} \left[\int_{-\infty}^0 (\mathbf{1}[t + B_{U_t} < 0] + \mathbf{1}[t + B_{U_t} > 0] z_{U_t} - 1) dt \right] = \mathbb{E}[(z_{U_t} - 1)B_{U_t}]. \end{aligned} \tag{4.28}$$

Bigarren deribatua kalkulatzeko onar dezagun $t' < t''$ betetezn dela, non t' eta t'' aldagaiek bezeroak iristen diren unea adierazten duten. Bukaeran, simetria dela eta, azken emaitza 2-gatik biderkatuko dugu. Hiru kasu bereizten ditugu:

Baldin $0 < t' < t''$, sistema hutsa egongo da etiketatutako bezeroaren iritsieran. Beraz, ondo-koia lortzen dugu

$$\begin{aligned} & G\left(0, \vec{z} \middle| A(-\infty, \infty) = 2, \tau_1 = t', \tau_2 = t''\right) - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t'\right) \\ & - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t''\right) + G\left(0, \vec{z} \middle| A(-\infty, \infty) = 0\right) = 0. \end{aligned}$$

Baldin $t' < 0$ & $0 < t''$ bi kasu bereiz ditzakegu:

$$\begin{cases} z_{U_{t'}}^0 \cdot z_{U_{t''}}^0 & \text{if } t' + B_{U_{t'}} < 0 \\ z_{U_{t'}} & \text{if } t' + B_{U_{t'}} > 0, \end{cases}$$

hau da,

$$\begin{aligned} & G\left(0, \vec{z} \middle| A(-\infty, \infty) = 2, \tau_1 = t', \tau_2 = t''\right) = \mathbb{E}[\mathbf{1}[t' + B_{U_{t'}} < 0] + \mathbf{1}[t' + B_{U_{t'}} > 0] z_{U_{t'}}] \\ & = G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t'\right). \end{aligned}$$

Ondorioz,

$$\begin{aligned} & G\left(0, \vec{z} \middle| A(-\infty, \infty) = 2, \tau_1 = t', \tau_2 = t''\right) - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t'\right) \\ & - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t''\right) + G\left(0, \vec{z} \middle| A(-\infty, \infty) = 0\right) = 0. \end{aligned}$$

Baldin $t' < t'' < 0$ kasu desberdinak jazo daitezke, jarraian erakusten dugun moduan:
Lehenik eta behin, baldin $t' + B_{U_{t'}} < t''$ honakoa dugu

$$\begin{cases} z_{U_{t'}}^0 \cdot z_{U_{t''}}^0 & \text{if } t'' + B_{U_{t''}} < 0 \\ z_{U_{t''}} & \text{if } t'' + B_{U_{t''}} > 0. \end{cases}$$

Ondorioz,

$$\begin{aligned} & G\left(0, \vec{z} \middle| A(-\infty, \infty) = 2, \tau_1 = t', \tau_2 = t''\right) - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t'\right) \\ & - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t''\right) + G\left(0, \vec{z} \middle| A(-\infty, \infty) = 0\right) = 0. \end{aligned}$$

Bigarrenez, baldin $t' + B_{U_{t'}} > t''$, ondoko sei kasuak bereiz daitezke

$$\left\{ \begin{array}{ll} z_{U_{t'}} & \text{if } (B_{U_{t'}} - t'' + t') \frac{g_{U_{t''}}}{g_{U_{t'}}} > B_{U_{t''}} \quad \& \quad t'' + B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} < 0 \\ & \& \quad t' + B_{U_{t'}} + B_{U_{t''}} > 0 \\ 1 & \text{if } (B_{U_{t'}} - t'' + t') \frac{g_{U_{t''}}}{g_{U_{t'}}} > B_{U_{t''}} \quad \& \quad t'' + B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} < 0 \\ & \& \quad t' + B_{U_{t'}} + B_{U_{t''}} < 0 \\ z_{U_{t'}} \cdot z_{U_{t''}} & \text{if } (B_{U_{t'}} - t'' + t') \frac{g_{U_{t''}}}{g_{U_{t'}}} > B_{U_{t''}} \quad \& \quad t'' + B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} > 0 \\ z_{U_{t''}} & \text{if } (B_{U_{t'}} - t'' + t') \frac{g_{U_{t''}}}{g_{U_{t'}}} < B_{U_{t''}} \quad \& \quad t'' + (B_{U_{t'}} - t'' + t') \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t'}}} < 0 \\ & \& \quad t' + B_{U_{t'}} + B_{U_{t''}} > 0 \\ 1 & \text{if } (B_{U_{t'}} - t'' + t') \frac{g_{U_{t''}}}{g_{U_{t'}}} < B_{U_{t''}} \quad \& \quad t'' + (B_{U_{t'}} - t'' + t') \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t'}}} < 0 \\ & \& \quad t' + B_{U_{t'}} + B_{U_{t''}} < 0 \\ z_{U_{t'}} \cdot z_{U_{t''}} & \text{if } (B_{U_{t'}} - t'' + t') \frac{g_{U_{t''}}}{g_{U_{t'}}} < B_{U_{t''}} \quad \& \quad t'' + (B_{U_{t'}} - t'' + t') \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t'}}} > 0. \end{array} \right.$$

Orduan,

$$\begin{aligned} & \int_{-\infty}^{\infty} \left(\int_{t'}^{\infty} \left(G\left(0, \vec{z} \middle| A(-\infty, \infty) = 2, \tau_1 = t', \tau_2 = t''\right) - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t'\right) \right. \right. \\ & \quad \left. \left. - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t''\right) + G\left(0, \vec{z} \middle| A(-\infty, \infty) = 0\right) \right) dt'' \right) dt' \\ & = \int_{-\infty}^0 \left(\int_{t'}^0 \left(G\left(0, \vec{z} \middle| A(-\infty, \infty) = 2, \tau_1 = t', \tau_2 = t''\right) - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t'\right) \right. \right. \\ & \quad \left. \left. - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = t''\right) + G\left(0, \vec{z} \middle| A(-\infty, \infty) = 0\right) \right) dt'' \right) dt' \\ & = \int_0^{\infty} \left(\int_0^r \left(G\left(0, \vec{z} \middle| A(-\infty, \infty) = 2, \tau_1 = -r, \tau_2 = -s\right) - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = -r\right) \right. \right. \\ & \quad \left. \left. - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = -s\right) + G\left(0, \vec{z} \middle| A(-\infty, \infty) = 0\right) \right) ds \right) dr \end{aligned}$$

non

$$\begin{aligned}
& G\left(0, \vec{z} \middle| A(-\infty, \infty) = 2, \tau_1 = -r, \tau_2 = -s\right) - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = -r\right) \\
& - G\left(0, \vec{z} \middle| A(-\infty, \infty) = 1, \tau_1 = -s\right) + G\left(0, \vec{z} \middle| A(-\infty, \infty) = 0\right) \\
& = \mathbb{E} \left[\mathbf{1}[-r + B_{U_{t'}} > -s] \left(\right. \right. \\
& \quad \mathbf{1} \left[(B_{U_{t'}} + s - r) \frac{g_{U_{t''}}}{g_{U_{t'}}} > B_{U_{t''}}, -s + B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} < 0 \right] \\
& \quad \cdot \mathbf{1}[-r + B_{U_{t'}} + B_{U_{t''}} > 0] (z_{U_{t'}} - 1) \\
& \quad + \mathbf{1} \left[(B_{U_{t'}} + s - r) \frac{g_{U_{t''}}}{g_{U_{t'}}} > B_{U_{t''}}, -s + B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} > 0 \right] (z_{U_{t'}} \cdot z_{U_{t''}} - 1) \\
& \quad + \mathbf{1} \left[(B_{U_{t'}} + s - r) \frac{g_{U_{t''}}}{g_{U_{t'}}} < B_{U_{t''}}, -s + (B_{U_{t'}} + s - r) \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t'}}} < 0 \right] \\
& \quad \cdot \mathbf{1}[-r + B_{U_{t'}} + B_{U_{t''}} > 0] (z_{U_{t''}} - 1) \\
& \quad + \mathbf{1} \left[(B_{U_{t'}} + s - r) \frac{g_{U_{t''}}}{g_{U_{t'}}} < B_{U_{t''}}, -s + (B_{U_{t'}} + s - r) \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t'}}} > 0 \right] (z_{U_{t'}} \cdot z_{U_{t''}} - 1) \\
& \quad + \mathbb{E} [\mathbf{1}[-r + B_{U_{t'}} > 0] (1 - z_{U_{t'}})] \\
& \quad \left. \left. + \mathbb{E} [\mathbf{1}[-s + B_{U_{t''}} > 0] (1 - z_{U_{t''}})] \right) \right].
\end{aligned}$$

Kalkula ditzagun sei integralak. Lehenengo,

$$\begin{aligned}
& \mathbb{E} \left[(z_{U_{t'}} - 1) \int_0^\infty \left(\int_s^\infty \mathbf{1}[-r + B_{U_{t'}} > -s] \right. \right. \\
& \quad \mathbf{1} \left[(B_{U_{t'}} + s - r) \frac{g_{U_{t''}}}{g_{U_{t'}}} > B_{U_{t''}}, -s + B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} < 0 \right] \\
& \quad \left. \left. \mathbf{1}[-r + B_{U_{t'}} + B_{U_{t''}} > 0] dr \right) ds \right] \\
& = \mathbb{E} \left[(z_{U_{t'}} - 1) \int_0^\infty \left(\int_s^\infty \mathbf{1}[s + B_{U_{t'}} > r] \right. \right. \\
& \quad \mathbf{1} \left[B_{U_{t'}} + s - \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}} > r, B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} < s \right] \mathbf{1}[B_{U_{t'}} + B_{U_{t''}} > r] dr \Big) ds \Big] \\
& = \mathbb{E} \left[(z_{U_{t'}} - 1) \int_0^\infty \mathbf{1} \left[B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} < s \right] \right. \\
& \quad \left. \left(\int_s^\infty \mathbf{1} \left[B_{U_{t'}} + s - \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}} > r, B_{U_{t'}} + B_{U_{t''}} > r \right] dr \right) ds \right]
\end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left[(z_{U_{t'}} - 1) \int_0^\infty \mathbf{1} \left[B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} < s \right] \right. \\
&\quad \left. \left(\int_s^{\max\{s, \min\{B_{U_{t'}} + s - \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, B_{U_{t'}} + B_{U_{t''}}\}} dr \right) ds \right] \\
&= \mathbb{E} \left[(z_{U_{t'}} - 1) \int_0^\infty \mathbf{1} \left[B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} < s \right] \right. \\
&\quad \left. \left(\max\{s, \min\{B_{U_{t'}} + s - \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, B_{U_{t'}} + B_{U_{t''}}\} - s \right) ds \right] \\
&= \mathbb{E} \left[(z_{U_{t'}} - 1) \int_0^\infty \mathbf{1} \left[B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} < s \right] \right. \\
&\quad \left. \max\{0, B_{U_{t'}} + \min\{-\frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, B_{U_{t''}} - s\}\} ds \right] \\
&= \mathbb{E} \left[(z_{U_{t'}} - 1) \int_0^\infty \mathbf{1} \left[B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} < s \right] \max\{0, B_{U_{t'}} + B_{U_{t''}} - s\} ds \right] \\
&= \mathbb{E} \left[(z_{U_{t'}} - 1) \int_{B_{U_{t''}}(g_{U_{t'}} + g_{U_{t''}})/g_{U_{t''}}}^{\max\{B_{U_{t'}} + B_{U_{t''}}, B_{U_{t''}}(g_{U_{t'}} + g_{U_{t''}})/g_{U_{t''}}\}} (B_{U_{t'}} + B_{U_{t''}} - s) ds \right] \\
&= \mathbb{E} \left[(z_{U_{t'}} - 1) \left((B_{U_{t'}} + B_{U_{t''}}) \max\{B_{U_{t'}} + B_{U_{t''}} - B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}}, 0\} \right. \right. \\
&\quad \left. \left. - \frac{s^2}{2} \Big|_{B_{U_{t''}}(g_{U_{t'}} + g_{U_{t''}})/g_{U_{t''}}}^{\max\{B_{U_{t'}} + B_{U_{t''}}, B_{U_{t''}}(g_{U_{t'}} + g_{U_{t''}})/g_{U_{t''}}\}} \right) \right] \\
&= \mathbb{E} \left[(z_{U_{t'}} - 1) \left((B_{U_{t'}} + B_{U_{t''}}) \max\{B_{U_{t'}} - B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}}, 0\} \right. \right. \\
&\quad \left. \left. - \frac{1}{2} \left(\left(B_{U_{t''}} + \max\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}}\} \right)^2 - \left(B_{U_{t''}} \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \right)^2 \right) \right) \right] \\
&= \mathbb{E} \left[(z_{U_{t'}} - 1) \left((B_{U_{t'}} + B_{U_{t''}}) \max\{B_{U_{t'}} - B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}}, 0\} \right. \right. \\
&\quad \left. \left. - \frac{1}{2} \left(\left(B_{U_{t''}} \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) + B_{U_{t'}} - \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\} \right)^2 \right. \right. \right. \\
&\quad \left. \left. \left. - \left(B_{U_{t''}} \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \right)^2 \right) \right) \right]
\end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left[(z_{U_{t'}} - 1) \left((B_{U_{t'}} + B_{U_{t''}}) \left(B_{U_{t'}} - \min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \} \right) \right. \right. \\
&\quad \left. \left. - \frac{1}{2} \left(\left(B_{U_{t'}} - \min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \} \right)^2 \right. \right. \right. \\
&\quad \left. \left. + 2B_{U_{t''}} \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \left(B_{U_{t'}} - \min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \} \right) \right) \right) \right] \\
&= \mathbb{E} \left[(z_{U_{t'}} - 1) \left(\left(B_{U_{t'}} - B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \left(B_{U_{t'}} - \min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \} \right) \right. \right. \\
&\quad \left. \left. - \frac{1}{2} \left(B_{U_{t'}} - \min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \} \right)^2 \right) \right]. \tag{4.29}
\end{aligned}$$

Bigarrenenez,

$$\begin{aligned}
&\mathbb{E} \left[\int_0^\infty \left(\int_s^\infty \mathbf{1}[-r + B_{U_{t'}} > -s] \mathbf{1} \left[(B_{U_{t'}} + s - r) \frac{g_{U_{t''}}}{g_{U_{t'}}} > B_{U_{t''}}, -s + B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} > 0 \right] \right. \right. \\
&\quad \left. \left. \cdot (z_{U_{t'}} \cdot z_{U_{t''}} - 1) dr \right) ds \right] \\
&= \mathbb{E} \left[\int_0^\infty \mathbf{1} \left[B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} > s \right] \right. \\
&\quad \left. \cdot \left(\int_s^\infty \mathbf{1} \left[s + B_{U_{t'}} > r, B_{U_{t'}} + s - \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}} > r \right] (z_{U_{t'}} \cdot z_{U_{t''}} - 1) dr \right) ds \right] \\
&= \mathbb{E} \left[(z_{U_{t'}} \cdot z_{U_{t''}} - 1) \int_0^\infty \mathbf{1} \left[B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} > s \right] \left(\int_s^{\max\{s, B_{U_{t'}} + s - \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\}} dr \right) ds \right] \\
&= \mathbb{E} \left[(z_{U_{t'}} \cdot z_{U_{t''}} - 1) B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} \max\{0, B_{U_{t'}} - \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\} \right] \\
&= \mathbb{E} \left[(z_{U_{t'}} \cdot z_{U_{t''}} - 1) B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} \left(B_{U_{t'}} - \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\} \right) \right]. \tag{4.30}
\end{aligned}$$

Hirugarrenenez,

$$\begin{aligned}
&\mathbb{E} \left[(z_{U_{t''}} - 1) \int_0^\infty \left(\int_0^r \mathbf{1}[-r + B_{U_{t'}} > -s] \right. \right. \\
&\quad \cdot \mathbf{1} \left[(B_{U_{t'}} + s - r) \frac{g_{U_{t''}}}{g_{U_{t'}}} < B_{U_{t''}}, -s + (B_{U_{t'}} + s - r) \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t'}}} < 0 \right] \\
&\quad \left. \cdot \mathbf{1}[-r + B_{U_{t'}} + B_{U_{t''}} > 0] \right) ds dr \right]
\end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left[(z_{U_{t''}} - 1) \int_0^\infty \left(\mathbf{1} [B_{U_{t'}} + B_{U_{t''}} > r] \right. \right. \\
&\quad \left. \left. \int_0^r \mathbf{1} [s > r - B_{U_{t'}}] \mathbf{1} \left[s < r - B_{U_{t'}} + \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, s < (r - B_{U_{t'}}) \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \right] \right) ds dr \right] \\
&= \mathbb{E} \left[(z_{U_{t''}} - 1) \int_0^\infty \mathbf{1} [B_{U_{t'}} + B_{U_{t''}} > r] \left(\int_{(r - B_{U_{t'}})^+}^{\max\{(r - B_{U_{t'}})^+, \min\{r - B_{U_{t'}} + \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, (r - B_{U_{t'}}) \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right)\}\}} \right) ds dr \right] \\
&= \mathbb{E} \left[(z_{U_{t''}} - 1) \int_0^\infty \mathbf{1} [B_{U_{t'}} + B_{U_{t''}} > r] \left(\max\{(r - B_{U_{t'}})^+, \min\{r - B_{U_{t'}} + \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, (r - B_{U_{t'}}) \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right)\}\} - (r - B_{U_{t'}})^+ \right) dr \right] \\
&= \mathbb{E} \left[(z_{U_{t''}} - 1) \int_0^{B_{U_{t'}} + B_{U_{t''}}} \left(\max\{0, r + \min\{0, -B_{U_{t'}} + \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, r \frac{g_{U_{t'}}}{g_{U_{t''}}} - B_{U_{t'}} \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right)\} - (r - \min\{r, B_{U_{t'}}\}) \right) dr \right] \\
&= \mathbb{E} \left[(z_{U_{t''}} - 1) \left(\int_0^{B_{U_{t'}}} \max\{0, \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, (r - B_{U_{t'}}) \frac{g_{U_{t'}}}{g_{U_{t''}}}\} - B_{U_{t'}} + \min\{r, B_{U_{t'}}\}\} dr \right. \right. \\
&\quad \left. \left. + \int_{B_{U_{t'}}}^{B_{U_{t'}} + B_{U_{t''}}} \max\{0, \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, (r - B_{U_{t'}}) \frac{g_{U_{t'}}}{g_{U_{t''}}}\} - B_{U_{t'}} + \min\{r, B_{U_{t'}}\}\} dr \right) \right] \\
&= \mathbb{E} \left[(z_{U_{t''}} - 1) \left(\int_0^{B_{U_{t'}}} \max\{0, (r - B_{U_{t'}}) \frac{g_{U_{t'}}}{g_{U_{t''}}} - B_{U_{t'}} + r\} dr \right. \right. \\
&\quad \left. \left. + \int_{B_{U_{t'}}}^{B_{U_{t'}} + B_{U_{t''}}} \left(\min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, (r - B_{U_{t'}}) \frac{g_{U_{t'}}}{g_{U_{t''}}}\} - B_{U_{t'}} + B_{U_{t'}} \right) dr \right) \right] \\
&= \mathbb{E} \left[(z_{U_{t''}} - 1) \left(\int_0^{B_{U_{t'}}} 0 dr + \int_{B_{U_{t'}}}^{B_{U_{t'}} + B_{U_{t''}}} \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, (r - B_{U_{t'}}) \frac{g_{U_{t'}}}{g_{U_{t''}}}\} dr \right) \right] \\
&= \mathbb{E} \left[(z_{U_{t''}} - 1) \left(\int_{B_{U_{t'}}}^{B_{U_{t'}} + \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\}} (r - B_{U_{t'}}) \frac{g_{U_{t'}}}{g_{U_{t''}}} dr \right. \right. \\
&\quad \left. \left. + \int_{B_{U_{t'}} + \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\}}^{B_{U_{t'}} + B_{U_{t''}}} \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\} dr \right) \right]
\end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left[(z_{U_{t''}} - 1) \left(\int_{B_{U_{t'}}}^{B_{U_{t'}} + \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\}} (r - B_{U_{t'}}) \frac{g_{U_{t'}}}{g_{U_{t''}}} dr \right. \right. \\
&\quad \left. \left. + \int_{B_{U_{t'}} + \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\}}^{B_{U_{t'}} + B_{U_{t''}}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\} dr \right) \right] \\
&= \mathbb{E} \left[(z_{U_{t''}} - 1) \left(\int_{B_{U_{t'}}}^{B_{U_{t'}} + \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\}} (r - B_{U_{t'}}) \frac{g_{U_{t'}}}{g_{U_{t''}}} dr \right. \right. \\
&\quad \left. \left. + \int_{B_{U_{t'}} + \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\}}^{B_{U_{t'}} + B_{U_{t''}}} B_{U_{t'}} dr \right) \right] \\
&= \mathbb{E} \left[(z_{U_{t''}} - 1) \left(\frac{g_{U_{t'}}}{2g_{U_{t''}}} \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\}^2 - B_{U_{t'}} \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\} + B_{U_{t'}} B_{U_{t''}} \right) \right]. \tag{4.31}
\end{aligned}$$

Laugarrenez,

$$\begin{aligned}
&\mathbb{E} \left[(z_{U_{t'}} \cdot z_{U_{t''}} - 1) \int_0^\infty \left(\int_0^r \mathbf{1}[-r + B_{U_{t'}} > -s] \right. \right. \\
&\quad \left. \left. \cdot \mathbf{1} \left[(B_{U_{t'}} + s - r) \frac{g_{U_{t''}}}{g_{U_{t'}}} < B_{U_{t''}}, -s + (B_{U_{t'}} + s - r) \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t'}}} > 0 \right] ds \right) dr \right] \\
&= \mathbb{E} \left[(z_{U_{t'}} \cdot z_{U_{t''}} - 1) \int_0^\infty \left(\int_0^r \mathbf{1}[s > r - B_{U_{t'}}] \right. \right. \\
&\quad \left. \left. \cdot \mathbf{1} \left[s < r - B_{U_{t'}} + \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, s > (r - B_{U_{t'}}) \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \right] ds \right) dr \right] \\
&= \mathbb{E} \left[(z_{U_{t'}} \cdot z_{U_{t''}} - 1) \int_0^\infty \left(\int_0^r \mathbf{1} \left[s < r - B_{U_{t'}} + \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}, s > (r - B_{U_{t'}}) \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \right] ds \right) dr \right] \\
&= \mathbb{E} \left[(z_{U_{t'}} \cdot z_{U_{t''}} - 1) \int_0^\infty \left(\int_{\max\{0, r - B_{U_{t'}}\}}^{\max\{0, (r - B_{U_{t'}}) \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right), \min\{r, r - B_{U_{t'}} + \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\}\}} ds \right) dr \right] \\
&= \mathbb{E} \left[(z_{U_{t'}} \cdot z_{U_{t''}} - 1) \right. \\
&\quad \left. \int_0^\infty \left(\max\{0, (r - B_{U_{t'}}) \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right), \min\{r, r - B_{U_{t'}} + \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\}\} \right. \right. \\
&\quad \left. \left. - \max\{0, r - B_{U_{t'}}\} \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \right) dr \right]
\end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left[(z_{U_{t'}} \cdot z_{U_{t''}} - 1) \left(\int_0^{B_{U_{t'}}} \max\{0, r + \min\{0, -B_{U_{t'}} + \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\}\} dr \right. \right. \\
&\quad \left. \left. + \int_{B_{U_{t'}}}^{\infty} \left(\max\{(r - B_{U_{t'}}) \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}}\right), r + \min\{0, -B_{U_{t'}} + \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\}\} \right. \right. \right. \\
&\quad \left. \left. \left. - (r - B_{U_{t'}}) \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}}\right) \right) dr \right) \right] \\
&= \mathbb{E} \left[(z_{U_{t'}} \cdot z_{U_{t''}} - 1) \right. \\
&\quad \left(\int_0^{B_{U_{t'}} - \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\}} 0 dr \right. \\
&\quad \left. + \int_{B_{U_{t'}} - \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\}}^{B_{U_{t'}}} (r + \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\} - B_{U_{t'}}) dr \right. \\
&\quad \left. + \int_{B_{U_{t'}}}^{B_{U_{t'}} + \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\}} \left(\min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\} + (B_{U_{t'}} - r) \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) dr \right. \\
&\quad \left. \left. + \int_{B_{U_{t'}} + \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\}}^{\infty} 0 dr \right) \right] \\
&= \mathbb{E} \left[(z_{U_{t'}} \cdot z_{U_{t''}} - 1) \frac{1}{2} \left(1 + \frac{g_{U_{t''}}}{g_{U_{t'}}}\right) \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\}^2 \right]. \tag{4.32}
\end{aligned}$$

Bosgarrenez,

$$\begin{aligned}
&\mathbb{E} \left[(1 - z_{U_{t'}}) \int_0^{\infty} \left(\int_0^r \mathbf{1}[-r + B_{U_{t'}} > -s] \mathbf{1}[-r + B_{U_{t'}} > 0] ds \right) dr \right] \\
&= \mathbb{E} \left[(1 - z_{U_{t'}}) \int_0^{\infty} \mathbf{1}[B_{U_{t'}} > r] \left(\int_0^r \mathbf{1}[s > r - B_{U_{t'}}] ds \right) dr \right] \\
&= \mathbb{E} \left[(1 - z_{U_{t'}}) \int_0^{\infty} \mathbf{1}[B_{U_{t'}} > r] \left(\int_{(r - B_{U_{t'}})^+}^r ds \right) dr \right] \\
&= \mathbb{E} \left[(1 - z_{U_{t'}}) \int_0^{\infty} \mathbf{1}[B_{U_{t'}} > r] (r - (r - \min\{r, B_{U_{t'}}\})) dr \right] \\
&= \mathbb{E} \left[(1 - z_{U_{t'}}) \int_0^{B_{U_{t'}}} \min\{r, B_{U_{t'}}\} dr \right] \\
&= \mathbb{E} \left[(1 - z_{U_{t'}}) \int_0^{B_{U_{t'}}} r dr \right] \\
&= \mathbb{E} \left[(1 - z_{U_{t'}}) \frac{B_{U_{t'}}^2}{2} \right]. \tag{4.33}
\end{aligned}$$

Seigarrenez,

$$\begin{aligned}
 & \mathbb{E} \left[(1 - z_{U_{t''}}) \int_0^\infty \left(\int_s^\infty \mathbf{1}[-r + B_{U_{t'}} > -s] \mathbf{1}[-s + B_{U_{t''}} > 0] dr \right) ds \right] \\
 &= \mathbb{E} \left[(1 - z_{U_{t''}}) \int_0^\infty \mathbf{1}[B_{U_{t'}} > s] \left(\int_s^\infty \mathbf{1}[s + B_{U_{t''}} > r] dr \right) ds \right] \\
 &= \mathbb{E} \left[(1 - z_{U_{t''}}) \int_0^{B_{U_{t''}}} \left(\int_s^{s+B_{U_{t'}}} dr \right) ds \right] \\
 &= \mathbb{E} [(1 - z_{U_{t''}}) B_{U_{t''}} B_{U_{t'}}]. \tag{4.34}
 \end{aligned}$$

(4.29)-(4.34) Ekuazioak batuz ondokoa lortzen dugu

$$\begin{aligned}
 & \mathbb{E} \left[(z_{U_{t'}} - 1) \left(\left(B_{U_{t'}} - B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \left(B_{U_{t'}} - \min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}}\} \right) \right. \right. \\
 & \quad \left. \left. - \frac{1}{2} \left(B_{U_{t'}} - \min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}}\} \right)^2 \right) \right] \\
 &+ \mathbb{E} \left[(z_{U_{t'}} \cdot z_{U_{t''}} - 1) B_{U_{t''}} \frac{g_{U_{t'}} + g_{U_{t''}}}{g_{U_{t''}}} \left(B_{U_{t'}} - \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\} \right) \right] \\
 &+ \mathbb{E} \left[(z_{U_{t''}} - 1) \left(\frac{g_{U_{t'}}}{2g_{U_{t''}}} \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\}^2 - B_{U_{t'}} \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\} + B_{U_{t'}} B_{U_{t''}} \right) \right] \\
 &+ \mathbb{E} \left[(z_{U_{t'}} \cdot z_{U_{t''}} - 1) \frac{1}{2} \left(1 + \frac{g_{U_{t''}}}{g_{U_{t'}}} \right) \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\}^2 \right] + \mathbb{E} \left[(1 - z_{U_{t'}}) \frac{B_{U_{t'}}^2}{2} \right] \\
 &+ \mathbb{E} [(1 - z_{U_{t''}}) B_{U_{t''}} B_{U_{t'}}] \\
 &= \mathbb{E} \left[(z_{U_{t'}} - 1) \left(B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \left(\min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}}\} - B_{U_{t'}} \right) - \frac{1}{2} \min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}}\}^2 \right) \right] \\
 &+ \mathbb{E} \left[(z_{U_{t'}} \cdot z_{U_{t''}} - 1) \left(B_{U_{t''}} \left(1 + \frac{g_{U_{t''}}}{g_{U_{t'}}} \right) \left(B_{U_{t'}} - \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\} \right) \right. \right. \\
 & \quad \left. \left. + \frac{1}{2} \left(1 + \frac{g_{U_{t''}}}{g_{U_{t'}}} \right) \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\}^2 \right) \right] \\
 &+ \mathbb{E} \left[(z_{U_{t''}} - 1) \left(\frac{g_{U_{t'}}}{2g_{U_{t''}}} \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\}^2 - B_{U_{t'}} \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\} \right) \right].
 \end{aligned}$$

Ohartu $t'' < t'$ kasua orain aztertu berri dugun kasuaren simetrikoa dela. Beraz,

$$\begin{aligned}
& (\psi^{DPS})^{(2)}(\lambda, \vec{z}) \Big|_{\lambda=0} \\
&= 2 \left(\mathbb{E} \left[(z_{U_{t'}} - 1) \left(B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \left(\min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}}\} - B_{U_{t'}} \right) - \frac{1}{2} \min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}}\}^2 \right) \right] \right. \\
&+ \mathbb{E} \left[(z_{U_{t'}} \cdot z_{U_{t''}} - 1) \left(B_{U_{t''}} \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \left(B_{U_{t'}} - \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\} \right) \right. \right. \\
&\quad \left. \left. + \frac{1}{2} \left(1 + \frac{g_{U_{t''}}}{g_{U_{t'}}} \right) \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\}^2 \right) \right] \\
&\left. + \mathbb{E} \left[(z_{U_{t''}} - 1) \left(\frac{g_{U_{t'}}}{2g_{U_{t''}}} \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\}^2 - B_{U_{t'}} \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\} \right) \right] \right). \tag{4.35}
\end{aligned}$$

4.A.2 4.3.1 Proposizioaren frogapena

Bezera kopuru totalaren probabilitate funtzio sortzailearen interpolazio bidezko hurbilketa lortuko dugu.

2.3 Atalean azaldu bezala $G(\lambda, \vec{z}^{(1-\rho)}) = \psi^{DPS}(\lambda, \vec{z}^{(1-\rho)})$ hurbiltzen dugu ondoko polinoioarekin

$$\hat{G}(\lambda, \vec{z}) = h_0(\vec{z}) + \lambda h_1(\vec{z}) + \lambda^2 h_2(\vec{z}) + \lambda^3 h_3(\vec{z}).$$

Normalizazioa deseginez, hau da, $f_\lambda^{-1} = \vec{z}^{(1-\rho)^{-1}}$ kontsideratuz, ondokoa dugu

$$\begin{aligned}
\psi^{DPS,INT}(\lambda, \vec{z}) &= \hat{G}(\lambda, \vec{z}^{(1-\rho)^{-1}}) \\
&= h_0(\vec{z}^{(1-\rho)^{-1}}) + \lambda h_1(\vec{z}^{(1-\rho)^{-1}}) + \lambda^2 h_2(\vec{z}^{(1-\rho)^{-1}}) + \lambda^3 h_3(\vec{z}^{(1-\rho)^{-1}}).
\end{aligned}$$

Orduan, (2.12) trafiko-arin baldintzatik $h_0(\vec{z}), h_1(\vec{z}), h_2(\vec{z})$ lortzen ditugu. Lehenik eta behin honakoa dugu, $\hat{G}(\lambda, \vec{z}^{(1-\rho)^{-1}}) \Big|_{\lambda=0} = \hat{G}(0, \vec{z}) = h_0(\vec{z})$. (4.27) Ekuazioarekin batera $h_0(\vec{z}) = 1$ lortzen dugu.

Bigarrenenez,

$$\begin{aligned}
& \left. \frac{d\hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{d\lambda} \right|_{\lambda=0} = \left. \frac{d\hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{d\lambda} \right|_{\lambda=0} + \sum_{i=1}^K \left. \frac{d\hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{dz_i} \right|_{\lambda=0} \cdot \left. \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \right|_{\lambda=0} \\
& = \left(h_1(\bar{z}^{(1-\rho)^{-1}}) + 2\lambda h_2(\bar{z}^{(1-\rho)^{-1}}) + 3\lambda^2 h_3(\bar{z}^{(1-\rho)^{-1}}) \right) \Big|_{\lambda=0} \\
& \quad + \sum_{i=1}^K \left(\frac{dh_0(\bar{z}^{(1-\rho)^{-1}})}{dz_i} + \lambda \frac{dh_1(\bar{z}^{(1-\rho)^{-1}})}{dz_i} + \lambda^2 \frac{dh_2(\bar{z}^{(1-\rho)^{-1}})}{dz_i} + \lambda^3 \frac{dh_3(\bar{z}^{(1-\rho)^{-1}})}{dz_i} \right) \Big|_{\lambda=0} \\
& \quad \cdot \left. \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \right|_{\lambda=0} \\
& = h_1(\bar{z}) + \sum_{i=1}^K \left. \frac{dh_0(\bar{z}^{(1-\rho)^{-1}})}{dz_i} \right|_{\lambda=0} \cdot \left. \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \right|_{\lambda=0} \\
& = h_1(\bar{z}) + \sum_{i=1}^K \left. \frac{d(1)}{dz_i} \right|_{\lambda=0} \cdot \left. \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \right|_{\lambda=0} = h_1(\bar{z}).
\end{aligned}$$

(4.28) Ekuazioarekin batera $h_1(\bar{z}) = -\mathbb{E}[B] + \sum_{i=1}^K \alpha_i \mathbb{E}[B_i] z_i$ lortzen dugu.

Hirugarrenenez,

$$\begin{aligned}
& \left. \frac{d^2 \hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{d\lambda^2} \right|_{\lambda=0} \\
& = \left. \frac{d^2 \hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{d\lambda^2} \right|_{\lambda=0} + \sum_{i=1}^K \left. \frac{d(d\hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})/d\lambda)}{dz_i} \right|_{\lambda=0} \cdot \left. \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \right|_{\lambda=0} \\
& \quad + \sum_{i=1}^K \left(\left(\left. \frac{d(d\hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})/dz_i)}{d\lambda} \right|_{\lambda=0} + \left. \frac{d(d\hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})/dz_i)}{dz_i} \right|_{\lambda=0} \cdot \left. \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \right|_{\lambda=0} \right) \right. \\
& \quad \cdot \left. \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \right|_{\lambda=0} + \left. \frac{d\hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{dz_i} \right|_{\lambda=0} \cdot \left. \frac{d^2(z_i^{(1-\rho)^{-1}})}{d\lambda^2} \right|_{\lambda=0} \Big) \\
& = 2h_2(\bar{z}) + 2 \sum_{i=1}^K \frac{dh_1(\bar{z})}{dz_i} \cdot \left. \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \right|_{\lambda=0} = 2h_2(\bar{z}) + 2\mathbb{E}[B] \sum_{i=1}^K \alpha_i \mathbb{E}[B_i] z_i \ln(z_i).
\end{aligned}$$

(4.35) Ekuazioarekin batera ondokoa lortzen dugu

$$\begin{aligned}
& h_2(\vec{z}) \\
&= \mathbb{E} \left[\left(z_{U_{t'}}^{(1-\rho)^{-1}} - 1 \right) \left(B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}} \left(\min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}}\} - B_{U_{t'}} \right) - \frac{1}{2} \min\{B_{U_{t'}}, B_{U_{t''}} \frac{g_{U_{t'}}}{g_{U_{t''}}}\}^2 \right) \right] \\
&+ \mathbb{E} \left[\left(z_{U_{t'}}^{(1-\rho)^{-1}} \cdot z_{U_{t''}}^{(1-\rho)^{-1}} - 1 \right) \left(B_{U_{t''}} \left(1 + \frac{g_{U_{t'}}}{g_{U_{t''}}} \right) \left(B_{U_{t'}} - \min\{B_{U_{t'}}, \frac{g_{U_{t'}}}{g_{U_{t''}}} B_{U_{t''}}\} \right) \right. \right. \\
&\quad \left. \left. + \frac{1}{2} \left(1 + \frac{g_{U_{t''}}}{g_{U_{t'}}} \right) \min\{B_{U_{t'}}, \frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t''}}\}^2 \right) \right] \\
&+ \mathbb{E} \left[\left(z_{U_{t''}}^{(1-\rho)^{-1}} - 1 \right) \left(\frac{g_{U_{t''}}}{2g_{U_{t'}}} \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\}^2 - B_{U_{t'}} \min\{\frac{g_{U_{t''}}}{g_{U_{t'}}} B_{U_{t'}}, B_{U_{t''}}\} \right) \right] \\
&- \mathbb{E}[B] \mathbb{E} \left[B_{U_{t'}} z_{U_{t'}}^{(1-\rho)^{-1}} \ln \left(z_{U_{t'}}^{(1-\rho)^{-1}} \right) \right].
\end{aligned}$$

Azkenik, 2.3.1 Proposiziotik, (4.7) Ekuazioarekin batera, frogapena amaitzen dugu.

4.A.3 4.4.1 Lemaren frogapena

Zero deribatuak ondokoa betetzen du

$$\left(\widetilde{W}_k^{DPS} \right)^{(0)}(0, b, x) = \mathbb{E}[\mathbf{1}[0 \geq x]] = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x > 0. \end{cases} \quad (4.36)$$

Bat deribatua lortzeko, $\left(\widetilde{W}_k^{DPS} \right)^{(1)}(0, b, x)$, (2.4) Ekuazioa erabiltzen dugu eta ondokoa kalkulatu behar dugu

$$\int_{-\infty}^{\infty} \left(\mathbb{E} \left[\mathbf{1} \left[W_k^{DPS} \left(b \middle| A(-\infty, \infty) = 1, \tau_1 = t \right) > x \right] \right] - \mathbb{E}[\mathbf{1}[0 > x]] \right) dt,$$

non $W_k^{DPS} \left(b \middle| A(-\infty, \infty) = 1, \tau_1 = t \right)$ k klaseko etiketatutako bezeroaren itxaron denbora baldintzatua den, $t \in \mathbb{R}$ unean iritsiera bat izanik, eta (4.15) betetzen duen.

Lehenik eta behin, (4.15) Ekuazioaren lehen terminoan jarriko dugu arreta, hau da, $t \leq 0 \leq t + B_{U_t}$ eta $t < \frac{g_{U_t}}{g_k} b - B_{U_t}$ kasuan. Ondokoa dugu

$$\begin{aligned}
& \int_{-\infty}^0 \mathbb{E} \left[\mathbf{1} \left[-B_{U_t} \leq t < \frac{g_{U_t}}{g_k} b - B_{U_t} \right] \mathbf{1}[t + B_{U_t} > x] \right] dt \\
&= \int_0^{\infty} \mathbb{E} \left[\mathbf{1} \left[B_{U_t} \geq t > B_{U_t} - \frac{g_{U_t}}{g_k} b \right] \mathbf{1}[-t + B_{U_t} > x] \right] dt \\
&= \mathbb{E} \left[\int_0^{\infty} \mathbf{1} \left[B_{U_t} \geq t > B_{U_t} - \frac{g_{U_t}}{g_k} b \right] \mathbf{1}[B_{U_t} - x > t] dt \right],
\end{aligned}$$

Tonelli-ren teoremagatik. Ondoren,

$$\begin{aligned}
& \int_0^\infty \mathbf{1} \left[B_{U_t} \geq t > B_{U_t} - \frac{g_{U_t} b}{g_k} \right] \mathbf{1} [B_{U_t} - x > t] dt \\
&= \int_{\left(B_{U_t} - \frac{g_{U_t} b}{g_k}\right)^+}^{\max\left\{\left(B_{U_t} - \frac{g_{U_t} b}{g_k}\right)^+, \min\{B_{U_t}, B_{U_t} - x\}\right\}} dt \\
&= \max \left\{ \left(B_{U_t} - \frac{g_{U_t} b}{g_k} \right)^+, B_{U_t} - x \right\} - \left(B_{U_t} - \frac{g_{U_t} b}{g_k} \right)^+ \\
&= \max \left\{ 0, B_{U_t} - x - \left(B_{U_t} - \frac{g_{U_t} b}{g_k} \right)^+ \right\} \\
&= \left(B_{U_t} - x - \left(B_{U_t} - \min\{B_{U_t}, \frac{g_{U_t} b}{g_k}\} \right) \right)^+ \\
&= \left(-x + \min\{B_{U_t}, \frac{g_{U_t} b}{g_k}\} \right)^+.
\end{aligned}$$

Beraz, honakoa lortzen dugu

$$\begin{aligned}
& \int_0^\infty \mathbb{E} \left[\mathbf{1} \left[B_{U_t} \geq t > B_{U_t} - \frac{g_{U_t} b}{g_k} \right] \mathbf{1} [t + B_{U_t} > x] \right] dt \\
&= \mathbb{E} \left[\left(-x + \min\{B_{U_t}, \frac{g_{U_t} b}{g_k}\} \right)^+ \right].
\end{aligned} \tag{4.37}$$

Bigarrenenez, (4.15) Ekuazioaren bigarren terminoan jarriko dugu arreta, hau da, $t \leq 0 \leq t + B_{U_t}$ eta $\frac{g_{U_t} b}{g_k} - t \leq B_{U_t}$ kasuetan. Honakoa dugu

$$\begin{aligned}
& \int_{-\infty}^0 \mathbb{E} \left[\mathbf{1} \left[\frac{g_{U_t} b}{g_k} - B_{U_t} \leq t \right] \mathbf{1} \left[\frac{g_{U_t} b}{g_k} > x \right] \right] dt \\
&= \int_0^\infty \mathbb{E} \left[\mathbf{1} \left[\frac{g_{U_t} b}{g_k} - B_{U_t} \leq -t \right] \mathbf{1} \left[\frac{g_{U_t} b}{g_k} > x \right] \right] dt \\
&= \mathbb{E} \left[\int_0^\infty \mathbf{1} \left[\frac{g_{U_t} b}{g_k} - B_{U_t} \leq -t \right] \mathbf{1} \left[\frac{g_{U_t} b}{g_k} > x \right] dt \right],
\end{aligned}$$

Tonelli-ren teoremagatik. Ondoren,

$$\begin{aligned}
& \int_0^\infty \mathbf{1} \left[t \leq B_{U_t} - \frac{g_{U_t} b}{g_k} \right] \mathbf{1} \left[\frac{g_{U_t} b}{g_k} > x \right] dt = \mathbf{1} \left[\frac{g_{U_t} b}{g_k} > x \right] \int_0^{\left(B_{U_t} - \frac{g_{U_t} b}{g_k}\right)^+} dt \\
&= \mathbf{1} \left[\frac{g_{U_t} b}{g_k} > x \right] \left(B_{U_t} - \frac{g_{U_t} b}{g_k} \right)^+ = \mathbf{1} \left[\frac{g_{U_t} b}{g_k} > x \right] \left(B_{U_t} - \min\{B_{U_t}, \frac{g_{U_t} b}{g_k}\} \right).
\end{aligned}$$

Beraz, honakoa lortzen dugu

$$\begin{aligned} & \mathbb{E} \left[\int_0^\infty \mathbf{1} \left[\frac{g_{U_t} b - B_{U_t}}{g_k} \leq -t \right] \mathbf{1} \left[\frac{g_{U_t} b}{g_k} > x \right] dt \right] \\ &= \mathbb{E} \left[\mathbf{1} \left[\frac{g_{U_t} b}{g_k} > x \right] \left(B_{U_t} - \min \left\{ B_{U_t}, \frac{g_{U_t} b}{g_k} \right\} \right) \right]. \end{aligned} \quad (4.38)$$

Hirugarrenez, (4.15) Ekuazioaren hirugarren terminoa, hau da, $t + B_{U_t} < 0$, eta $W_k^{DPS} \left(b \mid A(-\infty, \infty) = 0 \right) = b$. Honakoa dugu

$$\begin{aligned} & \int_{-\infty}^0 \mathbb{E} [\mathbf{1} [t < -B_{U_t}] \mathbf{1} [0 > x] - \mathbf{1} [0 > x]] dt \\ &= \int_0^\infty \mathbb{E} [\mathbf{1} [-t < -B_{U_t}] \mathbf{1} [0 > x] - \mathbf{1} [0 > x]] dt \\ &= \mathbb{E} \left[\mathbf{1} [0 > x] \int_0^\infty (\mathbf{1} [B_{U_t} < t] - 1) dt \right] \end{aligned}$$

Tonelli-ren teorematik. Ondoren,

$$\begin{aligned} & \mathbf{1} [0 > x] \int_0^\infty (\mathbf{1} [B_{U_t} < t] - 1) dt = \mathbf{1} [0 > x] \int_0^\infty -\mathbf{1} [B_{U_t} > t] dt = -\mathbf{1} [0 > x] \int_0^{B_{U_t}} dt \\ &= -\mathbf{1} [0 > x] B_{U_t}. \end{aligned}$$

Beraz, honakoa lortzen dugu

$$\mathbb{E} \left[\int_0^\infty (\mathbf{1} [B_{U_t} < t] \mathbf{1} [0 > x] - \mathbf{1} [0 > x]) dt \right] = -\mathbf{1} [0 > x] \mathbb{E} [B_{U_t}]. \quad (4.39)$$

Laugarrenez, (4.15) Ekuazioaren laugarren terminoan jarriko dugu arreta, hau da, $0 < t < b$ eta $\frac{b-t}{g_k} > \frac{B_{U_t}}{g_{U_t}}$ baldintzak betetzen diren kasuan. Honakoa dugu

$$\int_0^\infty \mathbb{E} \left[\mathbf{1} \left[t < b - \frac{g_k B_{U_t}}{g_{U_t}} \right] \mathbf{1} [B_{U_t} > x] \right] dt = \mathbb{E} \left[\int_0^\infty \mathbf{1} \left[t < b - \frac{g_k B_{U_t}}{g_{U_t}} \right] \mathbf{1} [B_{U_t} > x] dt \right],$$

Tonelli-ren teorematik. Ondoren,

$$\begin{aligned} & \int_0^\infty \mathbf{1} \left[t < b - \frac{g_k B_{U_t}}{g_{U_t}} \right] \mathbf{1} [B_{U_t} > x] dt = \int_0^{\left(b - \frac{g_k B_{U_t}}{g_{U_t}} \right)^+} \mathbf{1} [B_{U_t} > x] dt \\ &= \mathbf{1} [B_{U_t} > x] \left(b - \frac{g_k B_{U_t}}{g_{U_t}} \right)^+ = \mathbf{1} [B_{U_t} > x] \left(b - \min \left\{ b, \frac{g_k B_{U_t}}{g_{U_t}} \right\} \right) \end{aligned}$$

Beraz,

$$\int_0^\infty \mathbb{E} \left[\mathbf{1} \left[t < b - \frac{g_k B_{U_t}}{g_{U_t}} \right] \mathbf{1} [B_{U_t} > x] \right] dt = \mathbb{E} \left[\mathbf{1} [B_{U_t} > x] \left(b - \min \left\{ b, \frac{g_k B_{U_t}}{g_{U_t}} \right\} \right) \right]. \quad (4.40)$$

Bosgarrenez, (4.15) Ekuazioaren bosgarren terminoan jarriko dugu arreta, hau da, $0 < t < b$ eta $\frac{b-t}{g_k} \leq \frac{BU_t}{g_{U_t}}$ betetzen diren kasuan. Honakoa dugu

$$\begin{aligned} & \int_0^\infty \mathbb{E} \left[\mathbf{1} \left[b - \frac{g_k BU_t}{g_{U_t}} \leq t < b \right] \mathbf{1} \left[-t \frac{g_{U_t}}{g_k} + b \frac{g_{U_t}}{g_k} > x \right] \right] dt \\ &= \mathbb{E} \left[\int_0^\infty \mathbf{1} \left[b - \frac{g_k BU_t}{g_{U_t}} \leq t < b \right] \mathbf{1} \left[\left(b \frac{g_{U_t}}{g_k} - x \right) \frac{g_k}{g_{U_t}} > t \right] dt \right] \end{aligned}$$

Tonelli-ren teorematik. Ondoren,

$$\begin{aligned} & \int_0^\infty \mathbf{1} \left[b - \frac{g_k BU_t}{g_{U_t}} \leq t < b \right] \mathbf{1} \left[\left(b \frac{g_{U_t}}{g_k} - x \right) \frac{g_k}{g_{U_t}} > t \right] dt \\ &= \int_{\left(b - \frac{g_k BU_t}{g_{U_t}} \right)^+}^{\max\left\{ \left(b - \frac{g_k BU_t}{g_{U_t}} \right)^+, \min\left\{ b, \left(b \frac{g_{U_t}}{g_k} - x \right) \frac{g_k}{g_{U_t}} \right\} \right\}} dt \\ &= \max\left\{ \left(b - \frac{g_k BU_t}{g_{U_t}} \right)^+, \min\left\{ b, \left(b \frac{g_{U_t}}{g_k} - x \right) \frac{g_k}{g_{U_t}} \right\} \right\} - \left(b - \frac{g_k BU_t}{g_{U_t}} \right)^+ \\ &= \max\left\{ 0, b - x \frac{g_k}{g_{U_t}} - b + \min\left\{ b, \frac{g_k BU_t}{g_{U_t}} \right\} \right\}. \end{aligned}$$

Beraz, honakoa lortzen dugu

$$\begin{aligned} & \int_0^\infty \mathbb{E} \left[\mathbf{1} \left[b - \frac{g_k BU_t}{g_{U_t}} \leq t < b \right] \left(-t \frac{g_{U_t}}{g_k} + b \frac{g_{U_t}}{g_k} \right) \right] dt \\ &= \mathbb{E} \left[\left(-x \frac{g_k}{g_{U_t}} + \min\left\{ b, \frac{g_k BU_t}{g_{U_t}} \right\} \right)^+ \right]. \end{aligned} \quad (4.41)$$

Seigarrenez, (4.15) Ekuazioaren seigarren terminoa, hau da, $0 < b < t$, eta $W_k^{DPS} \left(b \middle| A(-\infty, \infty) = 0 \right) = \mathbf{1} [0 > x]$. Honakoa dugu

$$\int_0^\infty \mathbb{E} [\mathbf{1} [b < t] \mathbf{1} [0 > x] - \mathbf{1} [0 > x]] dt = -\mathbb{E} \left[\int_0^\infty \mathbf{1} [0 < t < b] \mathbf{1} [0 > x] dt \right]$$

Tonelli-ren teorematik. Ondoren,

$$-\int_0^\infty \mathbf{1} [0 < t < b] \mathbf{1} [0 > x] dt = -\int_0^b \mathbf{1} [0 > x] dt = -b \mathbf{1} [0 > x].$$

Beraz,

$$\int_0^\infty \mathbb{E} [\mathbf{1} [b < t] \mathbf{1} [0 > x] - \mathbf{1} [0 > x]] dt = -b \mathbf{1} [0 > x]. \quad (4.42)$$

Ondozioz, (4.37)-(4.42) Ekuazioak batuz, lortzen dugu

$$\begin{aligned}
& (W_k^{DPS})^{(1)}(0, b, x) \\
&= \mathbb{E} \left[\left(-x + \min\{B_{U_t}, \frac{g_{U_t}}{g_k} b\} \right)^+ \right. \\
&\quad + \mathbf{1} \left[\frac{g_{U_t}}{g_k} b > x \right] \left(B_{U_t} - \min\{B_{U_t}, \frac{g_{U_t}}{g_k} b\} \right) \\
&\quad - \mathbf{1} [0 \geq x] \mathbb{E}[B_{U_t}] \\
&\quad + \mathbf{1} [B_{U_t} > x] \left(b - \min\{b, \frac{g_k B_{U_t}}{g_{U_t}}\} \right) \\
&\quad + \left(-x \frac{g_k}{g_{U_t}} + \min\{b, \frac{g_k}{g_{U_t}} B_{U_t}\} \right)^+ \\
&\quad \left. - \mathbf{1} [0 \geq x] b \right], \tag{4.43}
\end{aligned}$$

zeina, $x > 0$ kasuan ondoko adierazpena sinplifikatzen den

$$\begin{aligned}
& (W_k^{DPS})^{(1)}(0, b, x) \\
&= \mathbb{E} \left[\left(1 + \frac{g_k}{g_{U_t}} \right) \left(-x + \min\{B_{U_t}, \frac{g_{U_t}}{g_k} b\} \right)^+ \right. \\
&\quad + \mathbf{1} \left[\frac{g_{U_t}}{g_k} b > x \right] \left(B_{U_t} - \min\{B_{U_t}, \frac{g_{U_t}}{g_k} b\} \right) \\
&\quad \left. + \mathbf{1} [B_{U_t} > x] \left(b - \min\{b, \frac{g_k B_{U_t}}{g_{U_t}}\} \right) \right]. \tag{4.44}
\end{aligned}$$

(2.1), (4.36) eta (4.44) Ekuazioekin batera frogapena amaitzen dugu.

4.A.4 4.4.1 Proposizioaren frogapena

Itxaron denbora baldintzatuaren funtzio osagarriaren interpolazioaren hurbilketa lortzen dugu DPS ereduarentzat eta $x > 0$ baldintzapean. $x > 0$ puntuan itxaron denborak atomo bat du.

2.3 Atalean azaldu bezala $G(\lambda, b, x(1-\rho)^{-1}) = W_k^{DPS}(\lambda, b, x(1-\rho)^{-1})$ ondoko polinomioarekin hurbiltzen dugu

$$\hat{G}(\lambda, b, x) = h_0(b, x) + \lambda h_1(b, x) + \lambda^2 h_2(b, x).$$

Normalizazioa deseginez, hau da, $f_\lambda^{-1} = x(1-\rho)$ funtzioa kontsideratuz, ondokoa dugu

$$\begin{aligned}
& W_k^{DPS,INT}(\lambda, b, x) = \hat{G}(\lambda, b, x(1-\rho)) \\
&= h_0(b, x(1-\rho)) + \lambda h_1(b, x(1-\rho)) + \lambda^2 h_2(b, x(1-\rho)).
\end{aligned}$$

Orduan, (2.12) trafiko-arin baldintzetatik $h_0(b, x), h_1(b, x)$ lortzen ditugu. Lehenik eta behin, ondokoa dugu

$$\hat{G}(\lambda, b, x(1 - \rho)) \Big|_{\lambda=0} = \hat{G}(0, b, x) = h_0(b, x).$$

(4.36) Ekuazioarekin batera $h_0(b, x) = \mathbf{1}[x > 0] = 0$ lortzen dugu, $x > 0$.

Bigarrenenez,

$$\begin{aligned} \frac{d\hat{G}(\lambda, b, x(1 - \rho))}{d\lambda} \Big|_{\lambda=0} &= \frac{d\hat{G}(\lambda, b, x(1 - \rho))}{d\lambda} \Big|_{\lambda=0} + \frac{d\hat{G}(\lambda, b, x(1 - \rho))}{dx} \Big|_{\lambda=0} \cdot \frac{d(x(1 - \rho))}{d\lambda} \Big|_{\lambda=0} \\ &= (h_1(b, x) + 2\lambda h_2(b, x)) \\ &+ \left(\frac{dh_0(b, x(1 - \rho))}{dx} + \lambda \frac{dh_1(b, x(1 - \rho))}{dx} + \lambda^2 \frac{dh_2(b, x(1 - \rho))}{dx} \right) \Big|_{\lambda=0} \cdot \frac{d(x(1 - \rho))}{d\lambda} \Big|_{\lambda=0} \\ &= h_1(b, x) + \frac{dh_0(b, x)}{dx} \Big|_{\lambda=0} \cdot x(-\mathbb{E}[B]) = h_1(b, x). \end{aligned}$$

(4.44) Ekuazioarekin batera ondokoa lortzen dugu, $x > 0$ -rako,

$$\begin{aligned} h_1(b, x) &= \mathbb{E} \left[\left(1 + \frac{g_k}{g_{U_t}} \right) \left(-x + \min\{B_{U_t}, \frac{g_{U_t}}{g_k} b\} \right)^+ \right. \\ &\quad \left. + \mathbf{1} \left[\frac{g_{U_t}}{g_k} b > x \right] \left(B_{U_t} - \min\{B_{U_t}, \frac{g_{U_t}}{g_k} b\} \right) \right. \\ &\quad \left. + \mathbf{1}[B_{U_t} > x] \left(b - \frac{g_k}{g_{U_t}} \min\{B_{U_t}, \frac{g_{U_t}}{g_k} b\} \right) \right]. \end{aligned}$$

Azkenik, 2.3.1 Proposiziotik, (4.9) Ekuazioarekin batera frogapena amaitzen dugu.

4.A.5 Trafiko arinpean deribatuak lortzeko beste modu bat

Atal honetan trafiko arinpeko deribatuak (4.1) Ekuazio integro-diferentzialetik ere lor daitezkeela erakutsiko dugu. Kalkulu hauek DPS ereduarentzat dira soilik baliogarriak, gure hurbilketa, aldiz, konstruktiboa eta beste ereduarentzat moldagarria da. 2.1 Atalean azaldu bezala, λ -ren balio txikietarako, $\bar{S}_k(\lambda, b)$ polinomio batekin hurbildu daiteke

$$\sum_{m=0}^{\infty} \lambda^m r_m(b), \quad (4.45)$$

$$\text{non } r_m(b) = \frac{\bar{S}_k^{(m)}(0, b)}{m!}, m = 0, 1, 2, \dots$$

2. Kapituluari, $r_m(b)$ koefizienteak nola lor daitezkeen deskribatzen dugu. Hau egia da eredu askorentzat. Orain, aldiz, DPS ereduarentzat koefiziente horiek kalkulatzeko beste modu bat azaltzen dugu. Hau, Fayolle *et al.*-en ekuazioa erabiliz lor daiteke.

Zerogarren koefizientearentzat, (4.45) Ekuaziotik $\frac{d\bar{S}_k(0, b)}{db} = \frac{dr_0(b)}{db}$ dugu eta (4.1) Ekuaziotik $\frac{d\bar{S}_k(0, b)}{db} = 1$. Honek zuzenean ematen digu

$$r_0(b) = b. \quad (4.46)$$

λ -ren balio txikientzat $\bar{S}_k(\lambda, b)$, $\sum_{m=0}^{\infty} \lambda^m r_m(b)$ adierazpenarekin hurbildu daitekeenez, $m = 1, 2, \dots$ -rako honako adierazpena dugu

$$\begin{aligned} r_m(b) &= \lim_{\lambda \rightarrow 0} \frac{1}{\lambda^m} \left(\bar{S}_k(\lambda, b) - \sum_{i=0}^{m-1} \lambda^i r_i(b) - \sum_{i=m+1}^{\infty} \lambda^i r_i(b) \right) \\ &= \lim_{\lambda \rightarrow 0} \frac{1}{\lambda^m} \left(\bar{S}_k(\lambda, b) - \sum_{i=0}^{m-1} \lambda^i r_i(b) \right) \\ &= \lim_{\lambda \rightarrow 0} \frac{1}{\lambda^m} \int_0^b \left(\frac{\partial \bar{S}_k(\lambda, \tilde{b})}{\partial \tilde{b}} - \sum_{i=0}^{m-1} \lambda^i \frac{dr_i(\tilde{b})}{d\tilde{b}} \right) d\tilde{b}. \end{aligned}$$

Orain, (4.1) Ekuazioa goiko formulaz ordezkatuz, $r_m(b)$, $m = 1, \dots$, terminoak erraz lor daitezke. Era honetan, lehenengo eta bigarren deribatuentzat honako adierazpenak lortzen ditugu

$$r_1(b) = \sum_{j=1}^K \alpha_j \frac{g_j}{g_k} \int_0^b \left(\mathbb{E}[B_j] + \left(\frac{g_k}{g_j} - 1 \right) \mathbb{E}[\min\{B_j, \frac{g_j}{g_k} \tilde{b}\}] \right) d\tilde{b} \quad (4.47)$$

eta

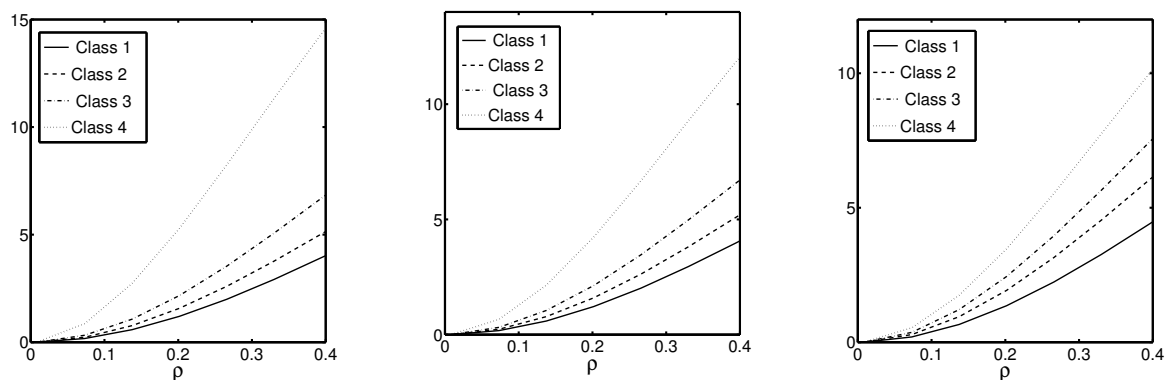
$$\begin{aligned} r_2(b) &= \sum_{j=1}^K \alpha_j \frac{g_j}{g_k} \sum_{i=1}^K \alpha_i \frac{g_i}{g_k} \int_0^b d\tilde{b} \left(\mathbb{E}[B_i] \left(\mathbb{E}[B_j] + \left(\frac{g_k}{g_j} - 1 \right) \mathbb{E}[\min\{B_j, \frac{g_j}{g_k} \tilde{b}\}] \right) \right. \\ &\quad + \left(\frac{g_k}{g_i} - 1 \right) \int_{\frac{g_j}{g_k} \tilde{b}}^{\infty} \mathbb{E}[\min\{B_i, \frac{g_i}{g_k} (x - \frac{g_j}{g_k} \tilde{b})\}] \cdot [1 - F_j(x)] dx \\ &\quad \left. + \frac{g_k}{g_j} \left(\frac{g_k}{g_i} - 1 \right) \int_{\frac{g_j}{g_k} (\tilde{b} - b)}^{\frac{g_j}{g_k} \tilde{b}} \mathbb{E}[\min\{B_i, \frac{g_i}{g_k} (\tilde{b} - \frac{g_k}{g_j} z)\}] \cdot [1 - F_j(z)] dz \right). \quad (4.48) \end{aligned}$$

$r_0(b)$ -k (4.19) Ekuazioan lorturiko zerogarren trafiko arinpeko deribatuarekin bat egiten duela ikusten dugu. Honetaz gain, (4.47) Ekuazioan lorturiko $r_1(b)$ terminoak (4.19) Ekuazioko lehenengo trafiko arinpeko deribatuarekin bat egiten duela ziurtatu dugu.

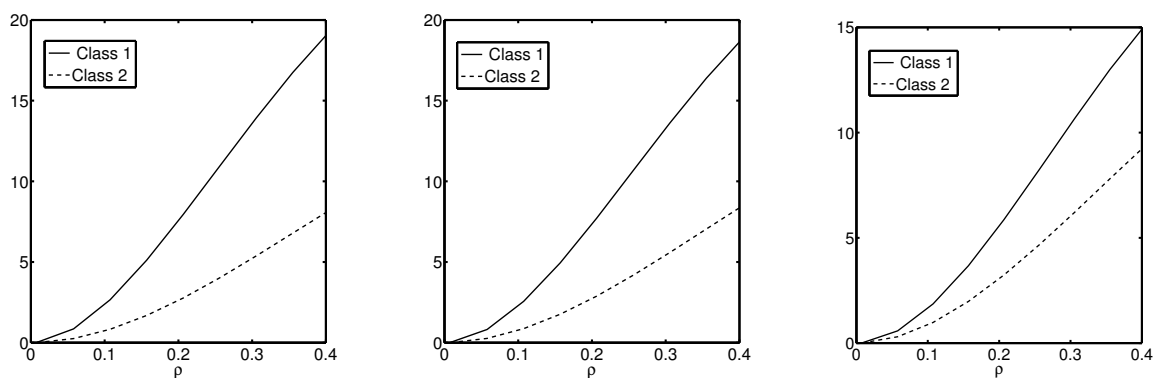
Jarraian, k klaseko etiketatutako bezeroaren sistemako egonaldiaren batezbesteko iraute denbora baldintzatuaren bigarren mailako trafiko-arin hurbilketa lortzen dugu, b beharrezko zerbitzua izanik, λ txikia denean

$$\bar{S}_k^{LT}(\lambda, b) = \sum_{m=0}^2 \lambda^m r_m(b) = r_0(b) + r_1(b)\lambda + r_2(b)\lambda^2, \quad (4.49)$$

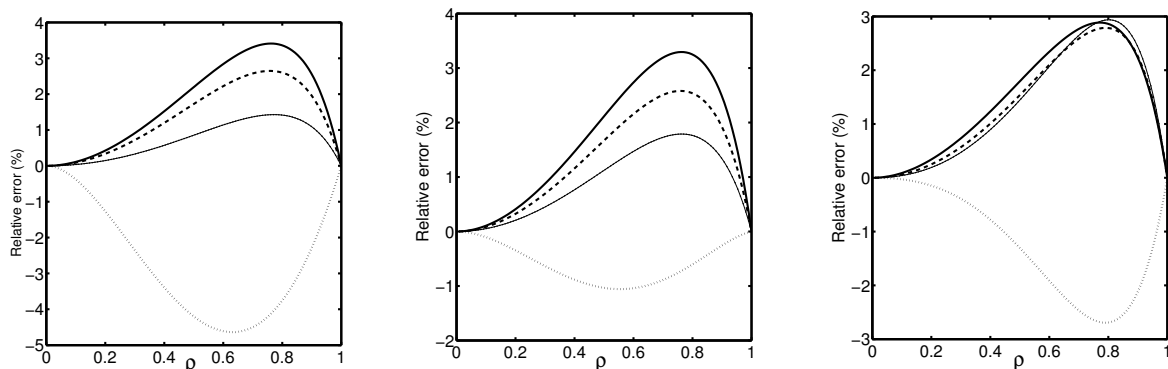
non $r_0(b)$, $r_1(b)$ eta $r_2(b)$ adierazpenak, (4.46), (4.47) eta (4.48) Ekuazioetan erakusten diren, hurrenez hurren. Emaitza hau 2.3.1 Proposizioa emaitzarekin eta trafiko geldo emaitzarekin konbinatuz, hirugarren mailako trafiko arinaren eta geldoaren interpolazioa lortzen dugu



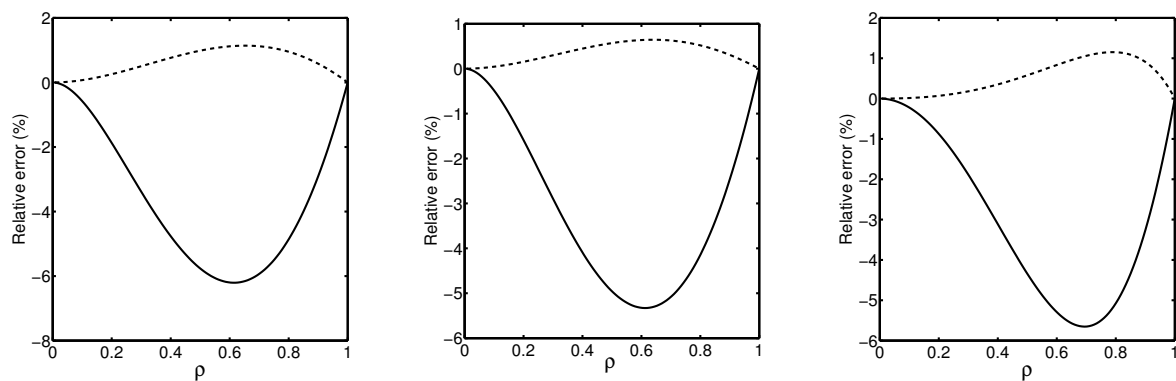
4.18. Irudia: 1 Eszenatokia: i klaseko etiketatutako bezero baten (4.19) eta (4.49) Ekuazioen errore erlatiboaren kenketa b_i , beharrezko zerbitzua, honako berdintzak bete daitezzen aukeratua izanik $\mathbb{P}(B_i \leq b_i) = 0.01$ (ezkerra), $\mathbb{P}(B_i \leq b_i) = 0.50$ (erdia), $\mathbb{P}(B_i \leq b_i) = 0.99$ (eskubia) izanik.



4.19. Irudia: 4 Eszenatokia: i klaseko etiketatutako bezero baten (4.19) eta (4.49) Ekuazioen errore erlatiboaren kenketa b_i , beharrezko zerbitzua, honako berdintzak bete daitezzen aukeratua izanik $\mathbb{P}(B_i \leq b_i) = 0.01$ (ezkerra), $\mathbb{P}(B_i \leq b_i) = 0.50$ (erdia), $\mathbb{P}(B_i \leq b_i) = 0.99$ (eskubia) izanik.



4.20. Irudia: 1 Eszenatokia: i klaseko etiketatutako bezeroaren hirugarren mailako sistemako egonaldiaren batezbesteko iraute denbora baldintzatuaren errore erlatiboa b_i , beharrezko zerbitzua, honako berdintzak bete daitezzen aukeratua izanik $\mathbb{P}(B_i \leq b_i) = 0.01$ (ezkerra), $\mathbb{P}(B_i \leq b_i) = 0.50$ (erdia), $\mathbb{P}(B_i \leq b_i) = 0.99$ (eskubia).



4.21. Irudia: 4 Eszenatokia: i klaseko etiketatutako bezeroaren hirugarren mailako sistemako egonaldiaren batezbesteko iraute denbora baldintzatuaren errore erlatiboa b_i , beharrezko zerbitzua, honako berdintzak bete daitezzen aukeratua izanik $\mathbb{P}(B_i \leq b_i) = 0.01$ (ezkerra), $\mathbb{P}(B_i \leq b_i) = 0.50$ (erdia), $\mathbb{P}(B_i \leq b_i) = 0.99$ (eskubia).

$$\overline{S_k^{DPS,INT}}(\lambda, b) = r_0(b) + r_1(b)\lambda + r_2(b)\lambda^2 + \frac{b}{g_k} \frac{\mathbb{E}[B^2]}{\sum_{j=1}^K \alpha_j \mathbb{E}[B_j^2]/g_j} \frac{(\lambda \mathbb{E}[B])^3}{1 - \lambda \mathbb{E}[B]}. \quad (4.50)$$

Jarraian, bigarren trafiko arinpeko deribatua edukitzearen eragina aztertuko dugu. Kasu honetan, bigarren mailako trafiko-arin hurbilketa eta hirugarren mailako trafiko arinaren eta geldoaren interpolazioa lortuko ditugu. Beraz, numerikoki alderatzen dugu (4.19) eta (4.49), eta (4.20) eta (4.50) Ekuazioen zehaztasun maila.

4.18 eta 4.19 Irudietan, (4.19) eta (4.49) Ekuazioen errore erlatiboaren kenketa irudikatzen dugu. Errore erlatibo guztiak positiboak direla ezagutzen dugunez, eta lorturiko funtzioak ere positiboak direnez, honek inplikutzen du maila altuagoko trafiko-arin hurbilketa zehatzagoa dela. Hala ere, 4.20 eta 4.21 Irudiak, 4.6 eta 4.14 Irudiekin alderatzen baditugu, hurrenez hurren, ondorioztatzen dugu interpolazioaren errore erlatiboa, kargaren balio guztietarako ez dela beti txikiagoa. Bi kasuetan, hirugarren mailako trafiko arinaren eta geldoaren interpolazioa edukitzeak, errore erlatibo handiena txikitzen du $\mathbb{P}(B_i \leq b_i) = 0.99$ kasurako soilik.

5. Kapitulu

Lehentasun erlatibodun trafiko geldopeko analisia

Kapitulu honetan eta hurrengoan lehentasun erlatibodun (RP) ilara aztertuko dugu, 1.1.1 Atalean deskribaturikoa. Klase anitzdun, lehentasun erlatibodun eta zerbitzu etengaitzeko M/G/1 sistema aztertzen dugu. Egoera-espazioaren kolapsoa bezala ezaguna den bat-egitea aurkezten dugu ilara luzera eskalatuarentzat trafiko geldopean, hau da, limitean, ilara luzera eskalatuaren bektorearen banaketa, banaketa esponetziala duen zorizko aldagai baten eta bektore determinista baten arteko biderkadura da. Honek ahalbidetzen digu erakustea sistemako bezero kopuru eskalatua murriztu egiten dela $c_k/\mathbb{E}[B_k]$ balio altua duten klaseei pixu altuak ezartzen zaizkienean, non c_k k klaseko bezeroei dagokien kostua den eta $\mathbb{E}[B_k]$ k klaseko bezeroei dagokien batezbesteko beharrezko zerbitzua den. Honako hau, $c\mu$ -arauaren [13] emaitzaren hedapen bat bezala ikus daiteke, lehentasun zorrotzdun diziplina da non lehentasuna $c_k/\mathbb{E}[B_k]$ balio beherakorraren arabera esleitzen den.

DROS eredurako, hau da, klase barneko planifikazio diziplina uniforme den kasurako, itxaron denbora aztertzen dugu trafiko geldo egoeran. Egoera-espazioaren kolapsoaren emaitza erabiliz, klase zehatz bati dagokion bezero baten itxaron denboraren banaketa lortzen dugu trafiko-geldo egoeran. Ikusten dugu banaketa esponetziala duten bi zorizko aldagaien biderkadura dela. Honek, [30] artikuluan ROS politikarentzat lorturiko emaitza orokortzen du. Honetaz gain, bezero arbitrario baten itxaron denboraren m -garren momentua minimizatzen duten pixuen balioak lortzen ditugu.

Amaitzeko, bi klasedun sistema bat simulatzen dugu DROS diziplinapean eta bezero kopuruaren banaketa eta bezero kopuruari eta itxaron denborari dagokien lehenengo eta bigarren momentuak erakusten ditugu trafiko-geldo egoeratik kanpo emaitza analitikoak balioztatzeko.

Ohartu kapitulu honetan trafiko geldopeko limitea kontsideratzen dugula egoera egonkor baten pean. Literaturan bezero kopuruari dagokion egoera-espazioaren kolapsoaren emaitza zuzenean trafiko geldopeko limitea aplikatuz lortu daitekeela ikus daitezke. Ikusi adibidez [7] non trafiko-geldo egoerapean aztertzen den klase anitzdun sistema bat den non bezero guztiek aldi berean jasotzen duten zerbitzua. Orokorrean, trafiko-geldo limitea eta sistema egoera egonbakorrean kontsideratzeko hartzen dugun limitea ezin dira elkar trukatu; honek azaltzen du ekarpenaren interesa. Beste desberdintasun garrantzitsu bat da gure ekarpenak ahalbidetzen duela itxaron denbora aztertzea.

Kapituluaren gainontzekoa honela antolatuta dago. 5.1. Atalean eredua aurkezten dugu eta trafiko-geldo eskalatua definitzen dugu. 5.2. eta 5.3. Ataletan bezero kopuru eskalatuaren

banaketa aurkezten dugu irteera garaian eta garai arbitrario baten, hurrenez hurren. 5.4. Atalean bezero bati dagokion itxaron denbora eskalatuaren banaketa aurkezten dugu. 5.5. Atalean lehenago aurkezturiko emaitzak erabiliz *mantentze-kostu* eskalatua eta itxaron denbora eskalatuaren momentuak nola optimizatu erakusten dugu. 5.6. Atalean emaitza numerikoak aurkezten ditugu.

5.1 Ereduaren deskribapena

Kontsidera dezagun K klase dituen eta zerbitzari bakarra duen ilara. k , $k = 1, \dots, K$, klaseko bezeroak $\lambda_k > 0$ tasako Poisson prozesu independenteen arabera iristen dira. Izan bedi iritsiera tasa totala $\lambda = \sum_{k=1}^K \lambda_k$. Onar dezagun k klaseko bezeroek banaketa orokorra duen beharrezko zerbitzua dutela, B_k , independenteak eta banaketa berekoak. $B_k(x)$ beharrezko zerbitzuari dago-kion banaketa funtzioa da eta $B_k^*(s) = \int_0^\infty e^{-sx} dB_k(x)$ Laplace-Stieltjes transformatua. Eta izan bedi $B_k^{*'}(s) = \frac{dB_k^*(s)}{ds}$. Onar dezagun $\mathbb{E}[B_k^2] < \infty$ k guztietarako. k klaseko bezeroen karga da $\rho_k = \lambda_k \mathbb{E}[B_k]$ eta karga totala

$$\rho := \sum_{k=1}^K \rho_k = \sum_{k=1}^K \lambda_k \mathbb{E}[B_k] = \lambda \sum_{k=1}^K \alpha_k \mathbb{E}[B_k],$$

non $\alpha_k = \lambda_k / \lambda$ iritsiera bat k klasekoa izateko probabilitatea den. Zerbitzua etengaitza da eta zerbitzu amaiera baten ostean zerbitzatzeko aukeratua den bezeroa k klasekoa izateko probabilitatea da

$$\frac{n_k p_k}{\sum_{j=1}^K n_j p_j}, \quad (5.1)$$

non $p_j > 0$, $j = 1, \dots, K$, pixuak diren eta n_j erabakia hartzeko unean sisteman daude j klaseko bezero kopurua den. Behin klase bat zerbitzatzeko aukeratua izan denean, klase barneko planifikazioa diziplinak erabakitzen du klase horretako zein bezero izango den zerbitzaturia. Onartzen dugu klase barneko planifikazio diziplina etengaitza dela eta ez duela bezeroen beharrezko zerbitzuaren informazioa erabiltzen.

Ilara asetasun puntuan aztertzen dugu, hau da, $\rho \uparrow 1$, trafiko-geldo egoera bezala ezagutua. Egoera hau honela lor daiteke

$$\lambda \uparrow \hat{\lambda} := \frac{1}{\sum_{k=1}^K \alpha_k \mathbb{E}[B_k]}, \quad (5.2)$$

izan ere, $\rho = \lambda \sum_{k=1}^K \alpha_k \mathbb{E}[B_k] \uparrow 1$. Trafiko-geldo egoerara igarotzean, k klaseko iritsieren frakzioa, α_k , finko mantentzen dugu eta $\hat{\lambda}_k := \alpha_k \hat{\lambda}$ definitzen dugu.

Izan bitez Q_k^{RP} eta N_k^{RP} k klaseko bezero kopurua sisteman irteera unean eta zorizko une batean, hurrenez hurren. Izan bedi W_k^{DROS} k klaseko bezero arbitrario baten itxaron denbora. Ohartu kapitulu honetan zehar notazioa ulergarriagoa izan dadin ez dugula zorizko aldagaiek ρ -rekiko duten menpekotasuna adierazteko. 5.2., 5.3. eta 5.4. Ataletan, Q_k^{RP} , N_k^{RP} eta W_k^{DROS} aztertzen ditugu, hurrenez hurren, trafiko-geldo egoeran.

5.2 Ilararen luzera irteera unean

Atal honetan egoera-espazioaren kolapsoaren emaitza aurkezten dugu irteera unean, egoera egonkorreko ilara luzeraren banaketarentzat. Hurrengo proposizioak kapitulu honetako emaitza garrantzitsuena erakusten du eta diona da, limitean, ilara luzeraren bektorea banaketa esponetziala duen zorizko aldagai baten eta bektore determinista baten arteko biderkadura dela. Frogapena 5.2.2. Atalean aurki daiteke.

5.2.1 Proposizioa *Ilara luzera irteera unean $1 - \rho$ -rekin eskalatuz gero, limitean banaketak emaitza egoki bat lortzen du. $(\lambda_1, \dots, \lambda_K) \rightarrow (\hat{\lambda}_1, \dots, \hat{\lambda}_K)$ eta $\rho \uparrow 1$ ditugunean,*

$$(1 - \rho)(Q_1^{RP}, \dots, Q_K^{RP}) \xrightarrow{d} (\widehat{Q}_1^{RP}, \dots, \widehat{Q}_K^{RP}) \stackrel{d}{=} \left(\frac{\hat{\lambda}_1}{p_1}, \dots, \frac{\hat{\lambda}_K}{p_K}\right) \cdot Y,$$

non $\xrightarrow{d}$ ikurra konbergentzia banaketan adierazten duen eta Y dimentsio bakarreko zorizko aldagaia den.

5.3. Ataleko 5.3.2 Oharrean erakutsiko dugu Y aldagaiak banaketa esponetziala duela. Hau erakutsi ahal izateko 5.3. Atalean lorturiko emaitzak behar ditugu. Honegatik, xehetasun gehiago nahi izanez gero irakurleari 5.3.2 Oharra irakurtzeko gomendatzen diogu.

Trafiko-geldo egoerarekin lanean hasi aurretik, ilara luzeraren (irteera unean) banaketaren probabilitate funtzio sortzaileak betetzen duen ekuazio sistema bat aurkeztuko dugu, Kim et al-ek lorturikoa, ikusi [27]. Izan bedi

$$\pi(q_1, \dots, q_K) := \mathbb{P}((Q_1^{RP}, \dots, Q_K^{RP}) = (q_1, \dots, q_K)),$$

eta

$$p(\vec{z}) = \mathbb{E}[z_1^{Q_1^{RP}} \dots z_K^{Q_K^{RP}}] = \sum_{q_1=0}^{\infty} \dots \sum_{q_K=0}^{\infty} z_1^{q_1} \dots z_K^{q_K} \pi(\vec{q})$$

bere probabilitate funtzio sortzailea. Izan bedi

$$r(\vec{z}) := \mathbb{E} \left[\frac{z_1^{Q_1^{RP}} \dots z_K^{Q_K^{RP}}}{\sum_{k=1}^K Q_k^{RP} p_k} \cdot \mathbf{1}_{(\sum_{k=1}^K Q_k^{RP} > 0)} \right] = \sum_{(q_1, \dots, q_K) \neq (0, \dots, 0)} \frac{\pi(\vec{q})}{q_1 p_1 + \dots + q_K p_K} z_1^{q_1} \dots z_K^{q_K}.$$

[27] artikuluan ilararen luzeraren banaketa aztertzen da, klase barneko planifikazioa diziiplina zorizkoa dela onartuz. Dena den, zerbitzu diziiplina etengaitza, aurreikusi ezina eta k klaseko bezeroak estokastikoki baliokideak direnez, ilara luzeraren banaketa bektoreak ez du menpekotasunik klase barneko politikarekiko. Hori dela eta, [27] artikulutik lana kontserbatzen duen edozein klase barneko politikarentzan honako emaitza hau dugu.

5.2.1 Teorema [27, 1 eta 2 Teorema] (a) *Oreka egoeran dagoen ilararen luzeraren (irteera unean) probabilitate funtzio sortzaileak, $p(z_1, \dots, z_K)$, honako adierazpena betetzen du*

$$p(z_1, \dots, z_K) = 1 - \rho + \sum_{i=1}^K p_i z_i \frac{\partial}{\partial z_i} r(z_1, \dots, z_K). \quad (5.3)$$

(b) $r(z_1, \dots, z_K)$ funtzioak betetzen du

$$\sum_{i=1}^K p_i \left(z_i - B_i^* \left(\lambda - \sum_{j=1}^K \lambda_j z_j \right) \right) \frac{\partial}{\partial z_i} r(z_1, \dots, z_K) = (\rho - 1) \left(1 - \sum_{i=1}^K \frac{\lambda_i}{\lambda} B_i^* \left(\lambda - \sum_{j=1}^K \lambda_j z_j \right) \right). \quad (5.4)$$

5.2.1. Atalean erakutsiko dugu (5.3) eta (5.4) Ekuazioak sinplifikatu egiten direla trafiko-geldo egoeran eta 5.2.2. Atalean erabiliko dugu 5.2.1 Proposizioa frogatzeko.

5.2.1 Trafiko-geldo eskalatua

Atal honetan 5.2.1 Proposizioaren frogapenean behar diren 3 lema aurkezten ditugu. Lehenengo leman ilararen luzera irteera unean *tight* dela erakusten dugu. Frogapena 5.A.1 eranskinean aurki daiteke.

5.2.1 Lema $(1 - \rho)(Q_1^{RP}, \dots, Q_K^{RP})$ zorizko bektorea *tight* da 1-etik gertu dagoen ρ -ren balio baterako, hau da, edozein ϵ -entzat existitzen da $\bar{\rho} \in (0, 1)$ eta $M > 0$ non $\mathbb{P}((1 - \rho)Q_k^{RP} \geq M) < \epsilon$, $k = 1, \dots, K$, guztietarako eta $\rho > \bar{\rho}$.

Komenigarria da $z_i = e^{-s_i}$ non $s_i > 0, i = 1, \dots, K$ aldagai aldaketa erabiltzea. Izan bedi $\vec{s} = (s_1, \dots, s_K)$ eta $e^{-(1-\rho)\vec{s}} = (e^{-(1-\rho)s_1}, \dots, e^{-(1-\rho)s_K})$. Honako ekuazioa

$$\lim_{\rho \uparrow 1} p(e^{-(1-\rho)\vec{s}}) = \lim_{\rho \uparrow 1} \mathbb{E}[e^{-(1-\rho)s_1 Q_1^{RP}} \dots e^{-(1-\rho)s_K Q_K^{RP}}] \quad (5.5)$$

existitzen bada orduan badago zorizko bektore bat $(\widehat{Q}_1^{RP}, \widehat{Q}_2, \dots, \widehat{Q}_K^{RP})$ non $(1 - \rho)(Q_1^{RP}, Q_2, \dots, Q_K^{RP})$ bektoreak banaketan konbergitzen duen $(\widehat{Q}_1^{RP}, \widehat{Q}_2, \dots, \widehat{Q}_K^{RP})$ bektorera, eta $(\widehat{Q}_1^{RP}, \widehat{Q}_2, \dots, \widehat{Q}_K^{RP})$ bektorearen banaketa (5.5) Ekuazioak ematen du (ikusi Feller (1971) [11]). Oraingoz, limite hori existitzen dela onartuko dugu; ohar honetara itzuliko gara 5.2.1 Proposizioaren frogapenaren amaieran.

Jarraian $\lim_{\rho \uparrow 1} p(e^{-(1-\rho)\vec{s}})$ limitearen propietateak deskribatzen dituzten bi lema ematen ditugu. Zehazki, 5.2.3 Lema 5.2.1 Proposizioaren frogapenaren giltza izango den ekuazio diferentzial partzialak lortzen ditugu.

Funtzio sortzailearen portaera deskribatzeko, defini dezagun

$$\hat{r}(\vec{s}) = \mathbb{E} \left[\frac{1 - e^{-s_1 \widehat{Q}_1^{RP}} \dots e^{-s_K \widehat{Q}_K^{RP}}}{\sum_{k=1}^K \widehat{Q}_k^{RP} p_k} 1_{(\sum_{k=1}^K \widehat{Q}_k > 0)} \right]. \quad (5.6)$$

Zenbakitzaileko “1”, $\widehat{Q}_j$ -ak zerotik hurbil daudenean parentesi arteko adierazpena bornatua mantenduko dela ziurtatzeko da.

5.2.2 Lema *Baldin eta* $\lim_{\rho \uparrow 1} p(e^{-(1-\rho)\vec{s}})$ existitzen bada, orduan honakoa betetzen da

$$\lim_{\rho \uparrow 1} p(e^{-(1-\rho)\vec{s}}) = \sum_{i=1}^K p_i \frac{\partial \hat{r}(\vec{s})}{\partial s_i}. \quad (5.7)$$

Frogapena: (5.3) Ekuaziotik dugu

$$\lim_{\rho \uparrow 1} p(e^{-(1-\rho)\vec{s}}) = \lim_{\rho \uparrow 1} \sum_{i=1}^K p_i \frac{\partial r(\vec{z})}{\partial z_i} \Big|_{\vec{z}=e^{-(1-\rho)\vec{s}}}. \quad (5.8)$$

$r(\vec{z})$ -ren definiziotik idatz dezakegu

$$\begin{aligned} \lim_{\rho \uparrow 1} \frac{\partial r(\vec{z})}{\partial z_i} \Big|_{\vec{z}=e^{-(1-\rho)\vec{s}}} &= \lim_{\rho \uparrow 1} \frac{\partial \mathbb{E} \left[\frac{z_1^{Q_1^{RP}} \cdots z_K^{Q_K^{RP}}}{\sum_{k=1}^K Q_k p_k} \cdot \mathbf{1}_{(\sum_{k=1}^K Q_k^{RP} > 0)} \right]}{\partial z_i} \Big|_{\vec{z}=e^{-(1-\rho)\vec{s}}} \\ &= \lim_{\rho \uparrow 1} \mathbb{E} \left[\frac{Q_i^{RP}}{\sum_{k=1}^K Q_k^{RP} p_k} \cdot \frac{e^{-(1-\rho)s_1 Q_1^{RP}} \cdots e^{-(1-\rho)s_K Q_K^{RP}}}{e^{-(1-\rho)s_i}} \cdot \mathbf{1}_{(\sum_{k=1}^K Q_k^{RP} > 0)} \right] \\ &= \mathbb{E} \left[\frac{\widehat{Q_i^{RP}}}{\sum_{k=1}^K \widehat{Q_k^{RP}} p_k} \cdot e^{-s_1 \widehat{Q_1^{RP}}} \cdots e^{-s_K \widehat{Q_K^{RP}}} \cdot \mathbf{1}_{(\sum_{k=1}^K \widehat{Q_k^{RP}} > 0)} \right] \\ &= \frac{\partial \hat{r}(\vec{s})}{\partial s_i}. \end{aligned} \quad (5.9)$$

Hirugarren pausoa erabiltzen dugu $\frac{Q_i}{\sum_{k=1}^K Q_k p_k} \cdot e^{-(1-\rho)s_1 Q_1^{RP}} \cdots e^{-(1-\rho)s_K Q_K^{RP}} \cdot \mathbf{1}_{(\sum_{k=1}^K Q_k > 0)}$ goi-
tik bornatua dagoela $\frac{1}{\min_j(p_j)}$ adierazpenarekin, eta banaketan konbergitzen duela $\frac{\widehat{Q_i^{RP}}}{\sum_{k=1}^K \widehat{Q_k^{RP}} p_k} \cdot$
 $e^{-s_1 \widehat{Q_1^{RP}}} \cdots e^{-s_K \widehat{Q_K^{RP}}} \cdot \mathbf{1}_{(\sum_{k=1}^K \widehat{Q_k^{RP}} > 0)}$ adierazpenera. (5.8) eta (5.9) Ekuazioetatik (5.7) lortzen
dugu. ■

Hurrengo leman erakusten dugu (5.4)-n emandako ekuazio diferentzial partziala asko sinplifikatzen dela trafiko-geldo egoeran.

5.2.3 Lema $\lim_{\rho \uparrow 1} p(e^{-(1-\rho)\vec{s}})$ existitzen bada, orduan $\hat{r}(\vec{s})$ funtzioak honako ekuazio diferentzial partziala betetzen du:

$$0 = \sum_{i=1}^K F_i(\vec{s}) \frac{\partial \hat{r}(\vec{s})}{\partial s_i} = \vec{F}(\vec{s}) \cdot \nabla \hat{r}(\vec{s}), \quad \forall \vec{s} \geq \vec{0},$$

non $\vec{F}(\vec{s}) = (F_1(\vec{s}), \dots, F_K(\vec{s}))$, eta

$$F_i(\vec{s}) = p_i(-s_i + \mathbb{E}[B_i] \sum_{k=1}^K \hat{\lambda}_k s_k) \quad i = 1, \dots, K, \quad (5.10)$$

$\hat{\lambda}_j = \alpha_j \hat{\lambda}$ eta $\hat{\lambda}$ (5.2) Ekuazioan bezala definituta egonik.

Frogapena: Har dezagun $\vec{z} = e^{-(1-\rho)\vec{s}}$ (5.4) Ekuazioan, bi aldeak $(1 - \rho)$ -rekin zatitu eta $\rho \uparrow 1$ limitea hartuz, honakoa lortzen dugu

$$\begin{aligned}
& \lim_{\rho \uparrow 1} \frac{\sum_{i=1}^K p_i (e^{-(1-\rho)s_i} - B_i^*(\lambda - \sum_{j=1}^K \lambda_j e^{-(1-\rho)s_j}))}{1-\rho} \frac{\partial}{\partial z_i} r(z_1, \dots, z_K) \Big|_{z_i = e^{-(1-\rho)s_i}} \\
&= \lim_{\rho \uparrow 1} -\left(1 - \sum_{i=1}^K \frac{\lambda_i}{\lambda} B_i^*(\lambda - \sum_{j=1}^K \lambda_j e^{-(1-\rho)s_j})\right) = 0,
\end{aligned} \tag{5.11}$$

non azken berdintza $B_i^*(0) = 1$, $\forall i$ betetzen delako ematen den. $x_i = e^{-s_i}$ aldagai aldaketa eginez lortzen dugu

$$\begin{aligned}
& \lim_{\rho \uparrow 1} \frac{\sum_{i=1}^K p_i (e^{-(1-\rho)s_i} - B_i^*(\lambda - \sum_{j=1}^K \lambda_j e^{-(1-\rho)s_j}))}{1-\rho} \frac{\partial}{\partial z_i} r(z_1, \dots, z_K) \Big|_{z_i = e^{-(1-\rho)s_i}} \\
&= \lim_{\rho \uparrow 1} \frac{\sum_{i=1}^K p_i (x_i^{1-\rho} - B_i^*(\lambda - \sum_{j=1}^K \lambda_j x_j^{1-\rho}))}{1-\rho} \frac{\partial}{\partial z_i} r(z_1, \dots, z_K) \Big|_{z_i = x_i^{1-\rho}} \\
&= \lim_{\rho \uparrow 1} \sum_{i=1}^K p_i (x_i^{1-\rho} \ln x_i \\
&\quad + \left(\frac{1}{\mathbb{E}(B)} - \left(\frac{1}{\mathbb{E}(B)} \sum_{j=1}^K \alpha_j x_j^{1-\rho} - \sum_{j=1}^K \lambda_j x_j^{1-\rho} \ln x_j \right) \right) (B_i^{*'}(\lambda - \sum_{j=1}^K \lambda_j x_j^{1-\rho}))) \frac{\partial}{\partial z_i} r(z_1, \dots, z_K) \Big|_{z_i = x_i^{1-\rho}} \\
&= \sum_{i=1}^K p_i (-s_i + \mathbb{E}(B_i) \sum_{j=1}^K \hat{\lambda}_j s_j) \frac{\partial \hat{r}(\vec{s})}{\partial s_i},
\end{aligned}$$

non bigarren pausoan l'Hopital-en erregela erabiltzen dugun eta hirugarren pausoan (5.9) eta $B_i^{*'}(0) := \frac{dB_i^*(s)}{ds} \Big|_{s=0} = -\mathbb{E}[B_i]$ edozein i -rako betetzen direla.

(5.11) Ekuazioarekin batera lortzen dugu

$$\sum_{i=1}^K p_i (-s_i + \mathbb{E}(B_i) \sum_{j=1}^K \hat{\lambda}_j s_j) \frac{\partial \hat{r}(\vec{s})}{\partial s_i} = 0.$$

■

5.2.2 5.2.1 Proposizioaren frogapena

Atal honek 5.2.1 Proposizioaren frogapena erakusten du. Frogapenaren oinarria da $\hat{r}(\vec{s})$ funtzioak 5.2.3 Leman deskribaturiko ekuazio diferentzial partziala betetzen duela. Ekuazio diferentzial partzial horretatik $\hat{r}(\cdot)$ funtzioak betetzen duen honako propietatea lortzen dugu.

5.2.4 Lema *Baldin eta $\lim_{\rho \uparrow 1} p(e^{-(1-\rho)\vec{s}})$ existitzen bada, orduan $\hat{r}(s)$ funtzioa konstantea da $(K-1)$ -dimentsiodun hiperplanoan*

$$H_c := \{\vec{s} \geq \vec{0} : \sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k = c\}, c > 0.$$

5.2.4 Lemaren frogapena 5.A.2 eranskinean aurki daiteke. Jarraian, 5.2.1 Proposizioaren frogapena emango dugu.

5.2.1 Proposizioaren frogapena: Onar dezagun $\lim_{\rho \uparrow 1} p(e^{-(1-\rho)\vec{s}})$ existitzen dela. Frogapenaren amaieran itzuliko gara honetara. $\hat{r}(\vec{s})$ konstantea denez H_c hiperplanoan, ikusi 5.2.4 Lema, $\hat{r}(\cdot)$ $\vec{s}$ bektorearen menpe dago $\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k$ baturaren baitan soilik. Beraz, existitzen da funtzio bat $\hat{r}^* : \mathbb{R} \rightarrow \mathbb{R}$ non $\hat{r}(\vec{s}) = \hat{r}^*(\sum_{k=1}^K \frac{\lambda_k}{p_k} s_k)$. 5.2.2 Lema eta $\frac{\partial \hat{r}(s)}{\partial s_i} = \frac{\hat{\lambda}_i}{p_i} \frac{d\hat{r}^*(v)}{dv} \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k}$ ekuaziotik lortzen dugu

$$\mathbb{E}[e^{-\sum_{i=1}^K s_i \widehat{Q_i^{RP}}}] = \lim_{\rho \rightarrow 1} p(e^{-(1-\rho)\vec{s}}) = \sum_{i=1}^K p_i \frac{\partial \hat{r}(s)}{\partial s_i} = \sum_{i=1}^K \hat{\lambda}_i \frac{d\hat{r}^*(v)}{dv} \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k} = \hat{\lambda} \frac{d\hat{r}^*(v)}{dv} \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k}, \quad (5.12)$$

zeina berriz ere $\vec{s}$ bektorearen menpe dagoen $\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k$ baturaren baitan soilik. Era baliokide batean, idatz dezakegu

$$\mathbb{E}[e^{-\sum_{i=1}^K s_i \widehat{Q_i^{RP}}}] = \mathbb{E}\left[e^{-\frac{p_1}{\hat{\lambda}_1} \widehat{Q_1^{RP}} - \sum_{i=1}^K \frac{\hat{\lambda}_i}{p_i} s_i - s_2 \frac{\hat{\lambda}_2}{p_2} (\frac{p_2}{\hat{\lambda}_2} \widehat{Q_2} - \frac{p_1}{\hat{\lambda}_1} \widehat{Q_1^{RP}}) - \dots - s_K \frac{\hat{\lambda}_K}{p_K} (\frac{p_K}{\hat{\lambda}_K} \widehat{Q_K^{RP}} - \frac{p_1}{\hat{\lambda}_1} \widehat{Q_1^{RP}})}\right].$$

Ekuazio hau soilik $\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k$ adierazpenaren menpe dagoenez, ikusi (5.12), honek iradokitzen du $\frac{p_i}{\hat{\lambda}_i} \widehat{Q_i^{RP}} = \frac{p_j}{\hat{\lambda}_j} \widehat{Q_j^{RP}}$ ia ziur edozein i, j -rako eta honako ekuazioa lortzen dugu

$$(\widehat{Q_1^{RP}}, \dots, \widehat{Q_K^{RP}}) = \left(\frac{\hat{\lambda}_1}{p_1}, \frac{\hat{\lambda}_2}{p_2}, \dots, \frac{\hat{\lambda}_K}{p_K}\right) \frac{p_1}{\hat{\lambda}_1} \widehat{Q_1^{RP}},$$

ia ziur. $Y \stackrel{d}{=} \frac{p_1}{\hat{\lambda}_1} \widehat{Q_1^{RP}}$ idatziz, lortzen dugu

$$(\widehat{Q_1^{RP}}, \dots, \widehat{Q_K^{RP}}) \stackrel{d}{=} \left(\frac{\hat{\lambda}_1}{p_1}, \frac{\hat{\lambda}_2}{p_2}, \dots, \frac{\hat{\lambda}_K}{p_K}\right) Y. \quad (5.13)$$

Gogorarazten dugu $\lim_{\rho \uparrow 1} p(e^{-(1-\rho)\vec{s}})$ existitzen dela onartu dugula. Horregatik, (5.13) limitea bakarria dela erakutsi behar dugu. $(1-\rho)(Q_1^{RP}, \dots, Q_K^{RP})$ *tight* denez, ikusi 5.2.1 Lema, eta ρ -ren edozein azpisegida konbergenterentzat limite bera lortzen dugunez, limitea bera existitzen dela lortzen dugu (ikusi korolaria 59. orrian, Billingsley 1999). Honek frogapena amaitutzat ematen du. ■

5.3 Ilararen luzera une arbitrarioan

Atal honetan une arbitrario batean sisteman dagoen bezero kopuruan jarriko dugu arreta, $(N_1^{RP}, \dots, N_K^{RP})$. Ondorengo emaitzak erakusten du, limitean, ilararen luzeraren bektorea banaketa esponontziala duen zorizko aldagai baten eta bektore determinista baten biderkadura dela. Gertaera hau egoera-espazioaren kolapso bezala adieraziko dugu. Frogapena 5.3.2 Atalean aurkezten dugu.

5.3.1 Oharra *Ohartu antzeko egoera-espazioaren kolapso emaitza lortu zela [49, 2.1 Proposizioa] artikuluan prozesadore-partekatze diskriminatzailea ereduarentzat. Gainera, frogapenaren teknika [49] artikuluan erabilitakoaren oso antzekoa da.*

5.3.1 Proposizioa *Ilararen luzeraren bektorea (une arbitrarioan) $1 - \rho$ -ekin eskalatuz gero limitetan banaketak emaitza egoki bat lortzen du. $(\lambda_1, \dots, \lambda_K) \rightarrow (\hat{\lambda}_1, \dots, \hat{\lambda}_K)$ eta $\rho \uparrow 1$ ditugunean,*

$$(1 - \rho)(N_1^{RP}, \dots, N_K^{RP}) \xrightarrow{d} (\widehat{N_1^{RP}}, \dots, \widehat{N_K^{RP}}) \stackrel{d}{=} \left(\frac{\hat{\lambda}_1}{p_1}, \frac{\hat{\lambda}_2}{p_2}, \dots, \frac{\hat{\lambda}_K}{p_K} \right) X, \quad (5.14)$$

non $\xrightarrow{d}$ ikurrak konbergentzia banaketan adierazten duen eta X banaketa esponentziala duen zorizko aldagai bat den $1/\nu(\vec{p})$ itxaropena duena eta

$$\nu(\vec{p}) := \frac{2 \sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k]}{\sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k^2]}. \quad (5.15)$$

Trafiko-geldo egoerarekin lanean hasi aurretik, ilara luzeraren banaketaren probabilitate funtzio sortzaileak betetzen duen ekuazio sistema bat aurkeztuko dugu, Kim et al-ek lorturikoa, ikusi [27].

Izan bedi $\psi(z_1, \dots, z_K)$, $(N_1^{RP}, \dots, N_K^{RP})$ bektorearen probabilitate funtzio sortzailea, i.e.,

$$\psi(z_1, \dots, z_K) := \mathbb{E}[z_1^{N_1^{RP}} \dots z_K^{N_K^{RP}}].$$

5.2. Atalean aipatu bezala, ilara luzeraren banaketa klase barneko diziplinaren independentea da. Hori dela eta, [27] artikuluko ondorengo emaitza erabili dezakegu.

5.3.1 Teorema [27, 3. Teorema eta 4. Teorema] *Egoera egonkorreko ilara luzeraren probabilitate funtzio sortzailea $\psi(z_1, \dots, z_K)$ honakoa da*

$$\psi(z_1, \dots, z_K) = 1 - \rho + \sum_{i=1}^K \lambda_i z_i \phi_i(z_1, \dots, z_K) \frac{1 - B_i^*(\lambda - \sum_{k=1}^K \lambda_k z_k)}{\lambda - \sum_{k=1}^K \lambda_k z_k}, \quad (5.16)$$

non $\phi_i(z_1, \dots, z_K)$ funtzioak honako adierazpena betetzen duen

$$\phi_i(z_1, \dots, z_K) = 1 - \rho + \frac{\lambda p_i}{\lambda_i} \frac{\partial}{\partial z_i} r(z_1, \dots, z_K). \quad (5.17)$$

5.3.1. Atalean erakutsiko dugu (5.16) Ekuazioa sinplifikatu egiten dela trafiko-geldo egoerapean, eta 5.3.2. Atalean hau erabiliko dugu ilara luzera eskalatuaren banaketa deskribatzeko, hau da, 5.3.1 Proposizioa frogatzeko.

5.3.1 Trafiko-geldo eskalatua

Hurrengo leman (5.16) Ekuazioa deskribatzen dugu trafiko-geldo egoeran.

5.3.1 Lema $\psi(e^{-(1-\rho)\vec{s}})$ funtzioaren limitea existitzen da $\rho \uparrow 1$ konbergitzean eta honako adierazpena betetzen du

$$\lim_{\rho \uparrow 1} \psi(e^{-(1-\rho)\vec{s}}) = \sum_{i=1}^K p_i \frac{\partial \hat{r}(s)}{\partial s_i} = \hat{\lambda} \frac{d\hat{r}^*(v)}{dv} \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k},$$

non $\hat{r}^(\cdot)$ funtzio bat den $\hat{r}^* : \mathbb{R} \rightarrow \mathbb{R}$.*

Frogapena: $(1-\rho)(Q_1^{RP}, \dots, Q_K^{RP})$ banaketan $(\widehat{Q_1^{RP}}, \dots, \widehat{Q_K^{RP}})$ bektore konbergitzen denez, badakigu $p(e^{-(1-\rho)\vec{s}})$ adierazpenaren limitea existitzen dela eta horregatik, (5.8) Ekuaziotik, $\frac{\partial r(\vec{z})}{\partial z_i} \Big|_{\vec{z}=e^{-(1-\rho)\vec{s}}}$ adierazpenaren ren limitea ere existitzen da. (5.17) Ekuaziotik zuzenean ondoriozta dezakegu $\lim_{\rho \uparrow 1} \phi_i(e^{-(1-\rho)\vec{s}})$ existitzen dela eta emaitza honakoa dela $\frac{\hat{\lambda}_i}{\hat{\lambda}_i} \frac{\partial \hat{r}(s)}{\partial s_i}$.

5.2.4 Leman ikusi dugun moduan, $\hat{r}(\vec{s})$ konstantea da H_c hiperplanoan. Horregatik, $\vec{s}$ -ren menpe dago $\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k$ -ren bidez soilik. Beraz, $\hat{r}^* : \mathbb{R} \rightarrow \mathbb{R}$ funtzio bat existitzen da non $\hat{r}(\vec{s}) = \hat{r}^*(\sum_{k=1}^K \frac{\lambda_k}{p_k} s_k)$ eta $\frac{\partial \hat{r}(s)}{\partial s_i} = \frac{\hat{\lambda}_i}{p_i} \frac{d\hat{r}^*(v)}{dv} \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k}$.

Emaitza hau eta (5.16) Ekuazioarekin batera lortzen dugu

$$\begin{aligned} \lim_{\rho \uparrow 1} \psi(e^{-(1-\rho)\vec{s}}) &= \lim_{\rho \uparrow 1} \left(1 - \rho + \sum_{i=1}^K \lambda_i e^{-(1-\rho)s_i} \phi_i(e^{-(1-\rho)\vec{s}}) \frac{1 - B_i^*(\lambda - \sum_{k=1}^K \lambda_k e^{-(1-\rho)s_k})}{\lambda - \sum_{k=1}^K \lambda_k e^{-(1-\rho)s_k}} \right) \\ &= \sum_{i=1}^K \hat{\lambda}_i \frac{\hat{\lambda}_i}{\hat{\lambda}_i} \frac{\partial \hat{r}(\vec{s})}{\partial s_i} (-B_i^*(0)) = \sum_{i=1}^K \hat{\lambda}_i \mathbb{E}[B_i] \frac{d\hat{r}^*(v)}{dv} \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k} \\ &= \hat{\lambda} \frac{d\hat{r}^*(v)}{dv} \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k} \sum_{i=1}^K \hat{\lambda}_i \mathbb{E}[B_i] = \hat{\lambda} \frac{d\hat{r}^*(v)}{dv} \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k}, \end{aligned}$$

non lehenengo pausoan l'Hopital-en erregela erabiltzen dugun eta bigarrenetan $B_i^{*'}(0) := \frac{dB_i^*(s)}{ds} \Big|_{s=0} = -\mathbb{E}[B_i], \forall i$. ■

5.3.1 Lemak inplikatzin du existitzen dela bektore bat $(\widehat{N_1^{RP}}, \dots, \widehat{N_K^{RP}})$ zeinetara ilara luzera eskalatua bektoreak banaketan konbergitzen duen.

5.3.2 5.3.1 Proposizioaren frogapena

Atal honetan 5.3.1 Proposizioaren frogapena aurkezten dugu. Bi pauso garrantzitsutan dago oinarrituta. Lehenik eta behin, ilara luzeraren bektorea zorizko aldagai baten eta bektore determinista baten arteko biderkaketa dela erakusten dugu. Eta bigarrenetaz, X zorizko aldagaiaren banaketa zehazten dugu, (5.15) Ekuazioan emandako itxaropena duen banaketa esponontziala duela ondorioztatuz.

5.3.1 Proposizioaren frogapena: $\lim_{\rho \uparrow 1} \psi(e^{-(1-\rho)\vec{s}})$ existitzen denez, ikusi 5.3.1 Lema, badakigu $(\widehat{N_1^{RP}}, \dots, \widehat{N_K^{RP}})$ bektorea existitzen dela non

$$\mathbb{E}[e^{-\sum_{k=1}^K s_k \widehat{N_k^{RP}}}] = \lim_{\rho \uparrow 1} \psi(e^{-(1-\rho)\vec{s}}) = \hat{\lambda} \frac{d\hat{r}^*(v)}{dv} \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k}. \quad (5.18)$$

5.2.1 Proposizioaren frogapenean erabilitako pausuak jarraituz lortzen dugu

$$(\widehat{N_1^{RP}}, \dots, \widehat{N_K^{RP}}) \stackrel{d}{=} \left(\frac{\hat{\lambda}_1}{p_1}, \frac{\hat{\lambda}_2}{p_2}, \dots, \frac{\hat{\lambda}_K}{p_K} \right) X, \quad (5.19)$$

non $X \stackrel{p_1}{\sim} \widehat{N_1^{RP}}$ bezala banatua dagoen.

X aldagaiaren banaketa zehazteko, sistemako lan totala kontsideratzen dugu, V^{arb} bezala adierazia. Lehenik eta behin azpimarratu nahi dugu sistemako lan totala erabiltzen den planifikazio diziplinaren independentea dela, hau kontserbakorra bada, noski. [28] artikuluan Kingman-ek FCFS ilara kontsideratzen du eta erakusten du $M/G/1$ ilaran lan total eskalatuak banaketa egoki bat duela $\rho \uparrow 1$ limitepean:

$$(1 - \rho)V^{arb} \xrightarrow{d} \hat{V}^{arb},$$

non $\hat{V}^{arb}$ -k banaketa esponentziala duen honako itxaropena duena

$$\mathbb{E}[\hat{V}^{arb}] = \frac{\sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k^2]}{2}. \quad (5.20)$$

DROS diziplinarentzat, baliokideki sistemako lan totala honela adieraz daiteke

$$V^{arb} = \sum_{k=1}^K \sum_{h=1}^{N_k^{RP}-1} B_{k,h} + \sum_{k=1}^K \tilde{B}_k,$$

non $B_{k,h}$ k klaseko h -garren bezeroaren beharrezko zerbitzua den eta $\tilde{B}_k$ k klaseko lehenengo bezeroari gelditzen zaion zerbitzu kopurua den. Alde batetik, ohartu k klaseko bezeroen beharrezko zerbitzuak independenteak eta banaketa berekoak direla, zehatzago azalduz, $B_{k,h} \stackrel{d}{=} B_k, \forall h$. Honetaz gain, $\tilde{B}_k$ B_k bezala dago banatua k klaseko N_k^{RP} -garren bezeroa ez bada zerbitzatu izaten ari, eta bestela, B_k -ren *forward-recurrence time*-ak ematen du.

$$\begin{aligned} \mathbb{E}[e^{-s\hat{V}^{arb}}] &= \lim_{\rho \uparrow 1} \mathbb{E}[e^{-(1-\rho)sV^{arb}}] \\ &= \lim_{\rho \uparrow 1} \mathbb{E}[e^{-(1-\rho)s(\sum_{k=1}^K \sum_{h=1}^{N_k^{RP}-1} B_{k,h} + \sum_{k=1}^K \tilde{B}_k)}] \\ &= \lim_{\rho \uparrow 1} \mathbb{E}[e^{-s \sum_{k=1}^K (1-\rho)(N_k^{RP}-1) \frac{\sum_{h=1}^{N_k^{RP}-1} B_{k,h}}{(N_k-1)} e^{-(1-\rho)s \sum_{k=1}^K \tilde{B}_k}}] \\ &= \mathbb{E}[e^{-s \sum_{k=1}^K \mathbb{E}[B_k] \widehat{N_k^{RP}}}], \end{aligned} \quad (5.21)$$

non azkeneko berdinketan erabiltzen dugun $e^{-s \sum_{k=1}^K (1-\rho)(N_k^{RP}-1) \frac{\sum_{h=1}^{N_k^{RP}-1} B_{k,h}}{(N_k-1)}}$ 1 balioak bokatua dagoela eta $e^{-s \sum_{k=1}^K \mathbb{E}[B_k] \widehat{N_k^{RP}}}$ adierazpena konbergitzen duela banaketan. (5.21) Ekuaziotik lortzen dugu

$$\hat{V}^{arb} \stackrel{d}{=} \sum_{k=1}^K \mathbb{E}[B_k] \widehat{N_k^{RP}}, \quad (5.22)$$

eta (5.19) Ekuazioarekin batera dugu

$$\hat{V}^{arb} \stackrel{d}{=} X \sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k]. \quad (5.23)$$

$\hat{V}^{arb}$ -k banaketa esponentziala duenez, X -k ere banaketa esponentziala izango du. Hori dela eta, itxaropena hartuz (5.23) Ekuazioan eta (5.20) aplikatuz lortzen dugu

$$\mathbb{E}[X] = \frac{\sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k^2]}{2 \sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k]},$$

zeinak 5.3.1 Proposizioaren frogapena amaitzen duen. $\square$

5.3.2 Oharra *Ohar honetan erakusten dugu bi zorizko aldagaiak Y eta X banaketan berdinak direla. Kontsidera dezagun k klaseko iritsiera arbitrario bat. PASTA propietateagatik, sistemako k klaseko bezero kopurua N_k^{RP} -ren berdina da. Lehenengo irteeraren osteko sistemako bezero kopurua $(Q_1^{RP}, \dots, Q_K^{RP})$ bezala dago banatua. ρ ordenekoa da zerbitzuan dagoen bezeroak sistema uzten duen denboran iristen diren bezeroen kopurua. Orduan,*

$$Q_k^{RP} \stackrel{d}{=} N_k^{RP} + \mathbf{O}(\rho).$$

Ekuaizio hau $(1-\rho)$ -rekin biderkatuz eta $\rho \uparrow 1$ limitea hartuz $\widehat{Q_k^{RP}} \stackrel{d}{=} \widehat{N_k^{RP}}$ lortzen dugu eta ondorioz $X \stackrel{d}{=} Y$.

5.4 Itxaron denbora

Atal honetan itxaron denbora trafiko-geldo egoeran aztertzen dugu. Klase barneko planifikazio diziplina zorizkoa kontsideratzen dugu, hau da, DROS erdua kontsideratzen dugu

Izan bedi W_l^{DROS} l klaseko bezero arbitrario baten itxaron denborari dagokion zorizko aldagaia. Bezero honi l klaseko etiketatutako bezeroa deituko diogu. Izan bedi Q_k^* k klaseko bezero kopurua sisteman (etiketatutako bezeroa izan ezik) etiketatutako bezeroa zerbitzua jasotzen hasten den unean, etiketatutako bezeroaren iritsiera unean zerbitzaria lanean ari zela suposatuz, i.e., $W_l^{DROS} > 0$. Defini dezagun honako transformatua:

$$T_l^{DROS}(u, z_1, \dots, z_K) := \mathbb{E}[e^{-uW_l^{DROS}} z_1^{Q_1^*} \dots z_K^{Q_K^*} \mathbf{1}_{\{W_l^{DROS} > 0\}}]. \quad (5.24)$$

l klaseko etiketatutako bezeroaren itxaron denboraren transformatua honakoa da

$$\mathbb{E}[e^{-uW_l^{DROS}}] = \mathbb{E}[e^{-u \cdot 0} \mathbf{1}_{\{W_l^{DROS}=0\}} + e^{-uW_l^{DROS}} \mathbf{1}_{\{W_l^{DROS}>0\}}] = 1 - \rho + T_l^{DROS}(u, \vec{1}), \quad (5.25)$$

non $1 - \rho$ l klaseko etiketatutako bezeroa iristean zerbitzaria hutsik dagoen probabilitatea den. DROS diziplinarentzat [27] artikulutik $T_l^{DROS}(u, \vec{z})$ transformatuarentzat honako emaitza dugu.

5.4.1 Teorema [27, 8 Teorema] *Klase barneko zorizko planifikazio diziplinarentzat,*

$T_l^{DROS}(u, z_1, \dots, z_K)$ transformatuak honako adierazpena du

$$\sum_{i=1}^K \frac{p_i}{p_l} \left(\frac{\partial}{\partial z_i} T_l^{DROS}(u, z_1, \dots, z_K) \right) (z_i - B_i^*(u + \lambda - \sum_{k=1}^K \lambda_k z_k)) + T_l^{DROS}(u, z_1, \dots, z_K) = W_l^1(u, z_1, \dots, z_K), \quad (5.26)$$

non $W_l^1(u, z_1, \dots, z_K)$ aldagaiak ondokoa betetzen duen

$$\begin{aligned} & W_l^1(u, z_1, \dots, z_K) \\ &= \sum_{i=1}^K ((1-\rho)\lambda_i + \lambda p_i \frac{\partial}{\partial z_i} r(z_1, \dots, z_K)) \frac{B_i^*(\lambda - \sum_{k=1}^K \lambda_k z_k) - B_i^*(u + \lambda - \sum_{k=1}^K \lambda_k z_k)}{u}. \end{aligned} \quad (5.27)$$

Itxaron denbora eskalatua aztertzeke, atal honetan $(1 - \rho)Q_k^*, \forall k$ uniformeki integragarria dela onartuko dugu. 5.6.2. Atalean aipatu bezala, emaitza numerikoek gertaera hau zuzena dela pentsatzera eramaten gaituzte.

1 Hipotesia *Klase barneko zorizko planifikazio diziplinarentzat, $\{(1 - \rho)Q_k^*\}$ zorizko aldagaien familia uniformeki integragarria da edozein k -rentzat.*

Orain aurkeztu dezakegu, limitean, l -klaseko etiketatutako bezeroaren itxaron denbora, W_l^{DROS} , banaketa esponentziala duten zorizko bi aldagaien biderkadura dela erakusten duen emaitza.

5.4.1 Proposizioa *Onar dezagun 1 Hipotesia betetzen dela eta kontsidera dezagun klase barneko zorizko planifikazio diziplina (i.e., DROS). Orduan, $\rho \uparrow 1$ limitepean honakoa betetzen da*

$$(1 - \rho)(W_l^{DROS}, Q_1^*, \dots, Q_K^*) \xrightarrow{d} (\widehat{W_l^{DROS}}, \widehat{Q_1^*}, \dots, \widehat{Q_K^*}) \stackrel{d}{=} (Z_l, \frac{\hat{\lambda}_1}{p_1}, \frac{\hat{\lambda}_2}{p_2}, \dots, \frac{\hat{\lambda}_K}{p_K})X,$$

non $\xrightarrow{d}$ ikurrak konbergentzia banaketan adierazten duen eta X eta Z_l independenteak diren zorizko aldagaia diren, banaketa esponentziala dutelarik $\mathbb{E}[X] = 1/\nu(\vec{p})$ eta $\mathbb{E}[Z_l] = 1/p_l$ itxaropenadunak.

5.4.1 Oharra 5.4.1 Proposizioa [30] artikuluan azaltzen den Kingman-en emaitzaren orokortze bat da, non trafiko-geldo egoeran itxaron denboraren banaketa lortzen den klase bakarreko DROS ilararentzat (i.e., ROS).

5.4.1 Proposizioa frogatzeko hurrengo hiru lemak beharko ditugu. Lehenengo lemak dio $(Q_1^*, \dots, Q_K^*)$ bektore eskalatuak limite bat duela.

5.4.1 Lema $1 - \rho$ -rekin eskalatzean, ilara luzeraren bektoreak, $(Q_1^*, \dots, Q_K^*)$, limite bat du banaketan $(\lambda_1, \dots, \lambda_K) \rightarrow (\hat{\lambda}_1, \dots, \hat{\lambda}_K)$. $\rho \uparrow 1$ limitepean,

$$(1 - \rho)(Q_1^*, \dots, Q_K^*) \xrightarrow{d} (\widehat{Q_1^*}, \dots, \widehat{Q_K^*}) \stackrel{d}{=} (\frac{\hat{\lambda}_1}{p_1}, \dots, \frac{\hat{\lambda}_K}{p_K}) \cdot X,$$

non $\xrightarrow{d}$ ikurrak konbergentzia banaketan adierazten duen eta X $1/\nu(\vec{p})$ itxaropena duen banaketa esponentziala duen zorizko aldagaia den.

Frogapena: Izan bedi $\tilde{Q}_i$ i klaseko bezero kopurua i klaseko etiketatutako bezeroaren zerbitzua hastean (etiketatutako bezeroa baztertuz). Definizioz, honako berdintza hau dugu:

$$\begin{aligned} \phi_l(e^{-s_1}, \dots, e^{-s_K}) &= \mathbb{E}[e^{-\sum_{i=1}^K s_i \tilde{Q}_i}] \\ &= \mathbb{E}\left[e^{-\sum_{i=1}^K s_i \tilde{Q}_i} \mathbf{1}_{\{W_i=0\}}\right] + \mathbb{E}\left[e^{-\sum_{i=1}^K s_i \tilde{Q}_i} \mathbf{1}_{\{W_i>0\}}\right] \\ &= 1 - \rho + T_l^{DROS}(0, e^{-s_1}, \dots, e^{-s_K}). \end{aligned}$$

Horregatik, (5.17) Ekuaziotik honako hau lortzen dugu

$$T_l^{DROS}(0, e^{-s_1}, \dots, e^{-s_K}) = \frac{\lambda_{p_i}}{\lambda_i} \frac{\partial}{\partial z_i} r(z_1, \dots, z_K) \Big|_{\vec{z}=e^{-\vec{s}}}. \lim_{\rho \uparrow 1} \frac{\partial}{\partial z_i} r(z_1, \dots, z_K) \Big|_{\vec{z}=e^{-(1-\rho)\vec{s}}} = \frac{\hat{\lambda}_i}{p_i} \frac{d\hat{r}^*(v)}{dv} \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k}$$

betetzen da (ikusi 5.3.1 Lemaren frogapena), hori dela eta

$$\lim_{\rho \uparrow 1} T_l^{DROS}(0, e^{-s_1}, \dots, e^{-s_K}) = \hat{\lambda} \frac{d\hat{r}^*(v)}{dv} \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k}. \quad (5.28)$$

(5.18) Ekuaziotik lortzen dugu $\mathbb{E}(e^{-\sum_{k=1}^K s_k \hat{N}_k}) = \hat{\lambda} \frac{d\hat{r}^*(v)}{dv} \Big|_{v=\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k}$. (5.28) Ekuazioa eta 5.3.1 Proposizioarekin batera frogapena amaitutzat ematen dugu. $\square$

Hurrengo lema teknikoak $W_l^1(u, z_1, \dots, z_K)$ funtzioaren balioa deskribatzen du, (5.27) Ekuazioan bezala definiturik, trafiko-geldo egoeran.

5.4.2 Lema *Klase barneko zorizko diziplina kontsideratzen dugu (i.e., DROS). Orduan, $\rho \uparrow 1$ limitepean, $W_l^1((1-\rho)u, e^{-(1-\rho)s_1}, \dots, e^{-(1-\rho)s_K})$ limitea existitzen da eta betetzen du*

$$\lim_{\rho \uparrow 1} W_l^1((1-\rho)u, e^{-(1-\rho)s_1}, \dots, e^{-(1-\rho)s_K}) = \frac{\nu(\vec{p})}{\nu(\vec{p}) + \sum_{k=1}^K s_k \frac{\hat{\lambda}_k}{p_k}},$$

non $1/\nu(\vec{p})$ (5.15) Ekuazioan emandakoa den.

5.4.2 Lemako emaitzak inplikatzan du trafiko-geldo egoeran $\lim_{\rho \uparrow 1} W_l^1((1-\rho)u, e^{-(1-\rho)s_1}, \dots, e^{-(1-\rho)s_K})$ funtzioak s aldagaiarekiko menpekotasuna soilik bere osagaien konbinazio lineal batean duela.

5.4.2 Lemaren frogapena 5.A.3 eranskinean aurki daiteke. Hurrengo lema l klaseko bezero baten itxaron denbora eskalatuak limite bat duela erakusten dugu.

5.4.3 Lema *Onar dezagun 1 Hipotesia betetzen dela klase barneko diziplina zorizkoa denean (i.e., DROS). Orduan, existitzen da $\widehat{W_l^{DROS}}$ non $(1-\rho)W_l^{DROS}$ banaketan konbergitzen duen $\widehat{W_l^{DROS}}$ -ra $\rho \uparrow 1$ limitepean.*

Frogapena: Definizioz, ondorengo bi berdintzak betetzen dira:

$$T_l^{DROS}(u, 1, \dots, 1) = \mathbb{E}[e^{-uW_l^{DROS}} \mathbf{1}_{\{W_l^{DROS} > 0\}}] \quad (5.29)$$

eta

$$\frac{\partial}{\partial z_i} T_l^{DROS}(u, z_1, \dots, z_K) \Big|_{\vec{z}=1} = \mathbb{E}[Q_i^* e^{-uW_l^{DROS}} \mathbf{1}_{\{W_l^{DROS} > 0\}}]. \quad (5.30)$$

Orain, (5.26) Ekuazioa $\vec{z} = 1$ bektorearentzat kontsideratuz trafiko-geldo egoeran lortzen dugu:

$$\lim_{\rho \uparrow 1} W_l^1((1-\rho)u, 1, \dots, 1) \quad (5.31)$$

$$\begin{aligned} &= \lim_{\rho \uparrow 1} \sum_{i=1}^K \frac{p_i}{p_l} (1-\rho) \frac{\partial}{\partial z_i} T_l^{DROS}((1-\rho)u, z_1, \dots, z_K) \Big|_{\vec{z}=1} \frac{1 - B_i^*((1-\rho)u)}{(1-\rho)} \\ &\quad + \lim_{\rho \uparrow 1} T_l^{DROS}((1-\rho)u, 1, \dots, 1) \\ &= \lim_{\rho \uparrow 1} \sum_{i=1}^K \frac{p_i}{p_l} u \mathbb{E}[B_i] \mathbb{E}[(1-\rho)Q_i^* e^{-(1-\rho)uW_l^{DROS}} \mathbf{1}_{\{W_l^{DROS} > 0\}}] + \lim_{\rho \uparrow 1} \mathbb{E}[e^{-(1-\rho)uW_l^{DROS}} \mathbf{1}_{\{W_l^{DROS} > 0\}}] \\ &= \lim_{\rho \uparrow 1} \mathbb{E} \left[\left(\sum_{i=1}^K \frac{p_i}{p_l} u \mathbb{E}[B_i] (1-\rho)Q_i^* + 1 \right) e^{-(1-\rho)uW_l^{DROS}} \mathbf{1}_{\{W_l^{DROS} > 0\}} \right] \\ &= \mathbb{E} \left[\lim_{\rho \uparrow 1} \left(\sum_{i=1}^K \frac{p_i}{p_l} u \mathbb{E}[B_i] (1-\rho)Q_i^* + 1 \right) e^{-(1-\rho)uW_l^{DROS}} \mathbf{1}_{\{W_l^{DROS} > 0\}} \right], \end{aligned} \quad (5.32)$$

non bigarren pausoan (5.29) eta (5.30) eta laugarrenean $(1 - \rho)Q_i^*$ uniformeki integragarria den hipotesia erabiltzen dugun (1 Hipotesia), [5, Teorema 3.5]. Ohartu $W_l^1((1 - \rho)u, e^{-(1-\rho)\vec{s}})$, (5.27) Ekuazioan definiturik dagoena, limitea duela $\rho \uparrow 1$ limitepean, ikusi 5.4.2 Lema. (5.31) Ekuazioak konbergitzen duenez, (5.32) Ekuazioak ere konbergitu behar du. Gainera, $\sum_{i=1}^K \frac{p_i}{p_l} u \mathbb{E}[B_i](1 - \rho)Q_i^*$ adierazpenak banaketan konbergitzen du $\sum_{i=1}^K \frac{p_i}{p_l} u \mathbb{E}[B_i] \frac{\hat{\lambda}_i}{p_i} X$ adierazpenera (ikusi 5.4.1 Lema) eta horregatik, ondorioztatzen dugu l klaseko bezero arbitrario baten itxaron denbora trafiko-geldo egoeran banaketan $\widehat{W}_l^{DROS}$ zorizko aldagai batetara konbergitzen duela. ■

5.4.2 Lematik, ohartu (5.32) u aldagaiaren independentea dela. Egiazta daiteke $(1 - \rho)(W_l^{DROS}, Q_1^*, \dots, Q_K^*)$ bektorea $X(Z_l, \frac{\hat{\lambda}_1}{p_1}, \dots, \frac{\hat{\lambda}_K}{p_K})$ bezala banatua dagoenean, ikusi 5.4.1 Proposizioa, orduan betetzen dela.

Orain 5.4.1 Proposizioa froga dezakegu eta (5.26) Ekuazioa ebatziz $T_l^{DROS}(\cdot)$ aurkituko dugu.

5.4.1 Proposizioaren frogapena: 5.4.3 Lematik ezagutzen dugu badagoela zorizko aldagai bat $\widehat{W}_l^{DROS}$ non $(1 - \rho)W_l^{DROS}$ adierazpenak banaketan konbergitzen duen $\widehat{W}_l^{DROS}$ adierazpenera. Horregatik, $\widehat{T}_l^{DROS}(u, \vec{s})$ funtzioa defini dezakegu honela:

$$\begin{aligned} \widehat{T}_l^{DROS}(u, \vec{s}) &:= \mathbb{E}[e^{-u\widehat{W}_l^{DROS}} e^{-\sum_{i=1}^K s_i \hat{Q}_i^*}] \\ &= \lim_{\rho \uparrow 1} \mathbb{E}[e^{-(1-\rho)uW_l^{DROS}} e^{-(1-\rho)s_1 Q_1^*} \dots e^{-(1-\rho)s_K Q_K^*}] \\ &= \lim_{\rho \uparrow 1} T_l^{DROS}((1 - \rho)u, e^{-(1-\rho)s_1}, \dots, e^{-(1-\rho)s_K}). \end{aligned}$$

(5.26) Ekuazioaren balioa neurtuko dugu $(u, \vec{z}) = (u(1 - \rho), e^{-(1-\rho)\vec{s}})$ puntuan $\rho \uparrow 1$ limitepean. Lehenik eta behin, lehenengo gaian zentratuko gara:

$$\begin{aligned} &\lim_{\rho \uparrow 1} (1 - \rho) \frac{\partial}{\partial z_i} T_l^{DROS}(u, \vec{z}) \Big|_{u=(1-\rho)u, \vec{z}=e^{-(1-\rho)\vec{s}}} \\ &= \lim_{\rho \uparrow 1} (1 - \rho) \mathbb{E} \left(\frac{Q_i^* e^{-(1-\rho)uW_l^{DROS}} e^{-(1-\rho)s_1 Q_1^*} \dots e^{-(1-\rho)s_i (Q_i^* - 1)} \dots e^{-(1-\rho)s_K Q_K^*}}{e^{-(1-\rho)s_i}} \right) \\ &= \mathbb{E}[\hat{Q}_i^* e^{-u\widehat{W}_l^{DROS}} e^{-s_1 \hat{Q}_1^*} \dots e^{-s_K \hat{Q}_K^*}] = -\frac{\partial}{\partial s_i} \widehat{T}_l^{DROS}(u, \vec{s}), \end{aligned} \quad (5.33)$$

non bigarren pausoan 1 Hipotesia eta [5, 3.5 Teorema] erabiltzen ditugun.

Horrez gain, ohartu $\widehat{T}_l^{DROS}(u, \vec{s}) = \mathbb{E}[e^{-u\widehat{W}_l^{DROS}} e^{-\sum_{i=1}^K s_i \hat{Q}_i^*}]$ adierazpenak $\vec{s} = (s_1, \dots, s_K)$ bektorearekiko menpekotasuna $y = \sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k$ aldagaiaren bidez duela soilik (ikusi 5.4.1 Lema). Horregatik, idatziko dugu $\widehat{T}_l^{DROS}(u, \vec{s}) = \widehat{T}_l^{DROS}(u, y)$ eta katearen arauagatik:

$$\frac{\partial}{\partial s_i} \widehat{T}_l^{DROS}(u, \vec{s}) = \frac{\partial}{\partial s_i} \widehat{T}_l^{DROS}(u, y) = \frac{\partial}{\partial y} \widehat{T}_l^{DROS}(u, y) \frac{\partial y}{\partial s_i} = \frac{\hat{\lambda}_i}{p_i} \frac{\partial}{\partial y} \widehat{T}_l^{DROS}(u, y) \Big|_{y=\sum_{k=1}^K s_k \frac{\hat{\lambda}_k}{p_k}} \quad (5.34)$$

Ondoren, (5.26) Ekuazioan trafiko geldopeko limitea hartuz eta (5.33) eta (5.34) Ekuazioak,

$$(e^{-(1-\rho)s_i} - B_i^*((1-\rho)u + \lambda - \sum_{k=1}^K \lambda_k e^{-(1-\rho)s_k}))$$

5.4.2 Lema eta $\lim_{\rho \uparrow 1} \frac{1}{1-\rho} = -s_i + \mathbb{E}[B_i] \left(u + \sum_{k=1}^K \hat{\lambda}_k s_k \right),$

l'Hopital-en erregelaren bidez lorturiko berdintza erabiliz, ondorengo ekuazio diferentzial batera iristen gara:

$$-\sum_{i=1}^K \frac{p_i}{p_l} \frac{\hat{\lambda}_i}{p_i} \frac{\partial \widehat{T_l^{DROS}}(u, y)}{\partial y} \Big|_{y=\sum_{i=1}^K s_i \frac{\hat{\lambda}_i}{p_i}} (-s_i + \mathbb{E}[B_i] (u + \sum_{k=1}^K \hat{\lambda}_k s_k)) + \widehat{T_l^{DROS}}(u, y) = \frac{\nu(\vec{p})}{\nu(\vec{p}) + y}.$$

Izan ere, $\sum_{i=1}^K \hat{\lambda}_i \left(-s_i + \mathbb{E}[B_i] \left(u + \sum_{k=1}^K \hat{\lambda}_k s_k \right) \right) = u$, eta azken hau honela idatz daiteke

$$-\frac{u}{p_l} \frac{\partial \widehat{T_l^{DROS}}(u, y)}{\partial y} \Big|_{y=\sum_{k=1}^K s_k \frac{\hat{\lambda}_k}{p_k}} + \widehat{T_l^{DROS}}(u, y) = \frac{\nu(\vec{p})}{\nu(\vec{p}) + y}. \quad (5.35)$$

(5.35) Ekuazio diferentzialaren emaitza da

$$\widehat{T_l^{DROS}}(u, y) = \frac{p_l}{u} e^{\frac{p_l}{u} y} \int_y^\infty e^{-\frac{p_l}{u} x} \frac{\nu(\vec{p})}{\nu(\vec{p}) + x} dx, \quad (5.36)$$

ikusi 5.A.4 Eranskina xehetasun gehiagorako.

Izan bitez Z_l eta X banaketa esponentziala duten bi zorizko aldagai $\mathbb{E}[Z_l] = 1/\eta$ eta $\mathbb{E}[X] = 1/\nu(\vec{p})$ itxaropenekin. Orduan, $(Z_l \cdot X, \frac{\lambda_1}{p_1} X, \dots, \frac{\lambda_K}{p_K} X)$ bektorearen Laplace Transformatua da:

$$\begin{aligned} \mathbb{E}[e^{-uZ_l \cdot X - s_1 \frac{\lambda_1}{p_1} X - \dots - s_K \frac{\lambda_K}{p_K} X}] &= \mathbb{E}[e^{-uZ_l \cdot X - \sum_{k=1}^K s_k \frac{\lambda_k}{p_k} X}] = \mathbb{E}[e^{-uZ_l \cdot X - yX}] \\ &= \mathbb{E}[\mathbb{E}[e^{-uz_l X - yX} | z_l = Z_l]] = \mathbb{E}\left[\frac{\nu(\vec{p})}{\nu(\vec{p}) + uZ_l + y}\right] = \int_0^\infty \eta e^{-\eta z_l} \frac{\nu(\vec{p})}{\nu(\vec{p}) + uz_l + y} dz_l \\ &= \frac{1}{u} \int_{\nu(\vec{p})+y}^\infty \eta e^{-\eta \frac{x - \nu(\vec{p}) - y}{u}} \frac{\nu(\vec{p})}{x} dx = \frac{\nu(\vec{p})\eta}{u} e^{\eta \frac{\nu(\vec{p})+y}{u}} \int_{\nu(\vec{p})+y}^\infty e^{-\eta \frac{x}{u}} \frac{1}{x} dx \\ &= \frac{\nu(\vec{p})\eta}{u} e^{\eta \frac{\nu(\vec{p})+y}{u}} \int_{\eta \frac{\nu(\vec{p})+y}{u}}^\infty \frac{e^{-l}}{l} dl. \end{aligned} \quad (5.37)$$

$z = p_l \frac{\nu(\vec{p})+x}{u}$ aldagai alaketa eginez (5.36) Ekuazioan ondokoa lortzen dugu

$$\hat{T}_l(u, y) = \nu(\vec{p}) \frac{p_l}{u} e^{p_l \frac{y}{u}} \int_y^\infty e^{-p_l \frac{x}{u}} \frac{1}{\nu(\vec{p}) + x} dx = p_l \frac{\nu(\vec{p})}{u} e^{p_l \frac{\nu(\vec{p})+y}{u}} \int_{p_l \frac{\nu(\vec{p})+y}{u}}^\infty e^{-z} \frac{1}{z} dz.$$

Beraz, (5.37) Ekuazioan lorturiko $(Z_l \cdot X, \frac{\lambda_1}{p_1} X, \dots, \frac{\lambda_K}{p_K} X)$ bektorearen Laplace Transformatuarekin koinziditzen du, $p_l = \eta$ izanik. Probabilitate banaketa baten Laplace Transformatua bakarrenez, (bakantasan teorema, [11]), ondorioztatzen dugu $(1-\rho)(W_l^{DROS}, Q_1^*, \dots, Q_K^*)$ bektoreak banaketan konbergitzen duela $(Z_l \cdot X, \frac{\lambda_1}{p_1} X, \dots, \frac{\lambda_K}{p_K} X)$ bektorerara, non, lehenago aipatu bezala, Z_l eta X banaketa esponentziala duten bi zorizko aldagai independente diren $\mathbb{E}[Z_l] = 1/\eta$ eta $\mathbb{E}[X] = 1/\nu(\vec{p})$ itxaropenak dituztenak. ■

5.5 Pixu optimoak

Atal honetan 5.3.1 eta 5.4.1 Proposizioan lorturiko emaitzak sistemaren errendimentua optimizatzeko nola erabili daitezkeen erakusten dugu. Zehazki, 5.5.1. Atalean mantentze-kostuan zentratzen gara eta 5.5.2. Atalean bezero arbitrario baten itxaron denboraren momentuak minimizatzen dituen pixuak aurkitzen ditugu.

5.5.1 *Mantentze-kostua*

Klase bakoitzari kostu bat esleitzen diogu $c_k \geq 0, k = 1, \dots, K$. Sistemaren errendimenduaren neur-tzaile bezala mantentze-kostua kontsideratzen dugu, $\sum_{k=1}^K c_k N_k^{RP}$. Atal honetan $N_k^{RP}(\vec{p}), \widehat{N}_k^{RP}(\vec{p})$ idatziko dugu $N_k^{RP}, \widehat{N}_k^{RP}$ -ren ordez, $\vec{p} := (p_1, \dots, p_K)$ pixuekiko duten menpekotasuna adierazteko. 5.3.1 Proposiziotik lorten dugu mantentze-kostua eskalatuak, $(1 - \rho) \sum_{k=1}^K c_k N_k^{RP}(\vec{p})$, banaketa esponenziala duen zorizko aldagai batetara konbergitzen duela banaketan honako itxaropena duena

$$\sum_{k=1}^K c_k \mathbb{E}[\hat{N}_k(\vec{p})] = \frac{\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} c_k}{2 \sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k]} \sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k^2], \quad (5.38)$$

$\rho \uparrow 1$ limitepean. Adierazpen hau erabilitz, hurrengo monotonizitate emaitza lortzen dugu trafiko-geldo egoeran: mantentze-kostua estokastikoki txikiagotzen da, $\frac{c_i}{\mathbb{E}[B_i]}$ balio handia duten bezeroei lehentasuna ematen diegun heinean. Hau $c\mu$ arauaren hedapen bat bezala ikusi genezake trafiko-geldo egoeran, [13].

5.5.1 Proposizioa *Kontsideratu $(p_1, \dots, p_K)$ eta $(q_1, \dots, q_K)$ pixkuak dituzten bi politika. Izan bedi $c_k \geq 0, k = 1, \dots, K$. Orokortasunik galdu gabe, onar dezagun klaseak honela ordenaturik daudela $\frac{c_1}{\mathbb{E}[B_1]} \geq \frac{c_2}{\mathbb{E}[B_2]} \geq \dots \geq \frac{c_K}{\mathbb{E}[B_K]}$. Baldin eta $\frac{p_k}{p_{k+1}} \leq \frac{q_k}{q_{k+1}}, \forall k = 1, \dots, K-1$ bada, orduan*

$$\lim_{\rho \uparrow 1} (1 - \rho) \sum_{k=1}^K c_k N_k^{RP}(\vec{p}) \geq_{st} \lim_{\rho \uparrow 1} (1 - \rho) \sum_{k=1}^K c_k N_k^{RP}(\vec{q}),$$

non $\geq_{st}$ -k ohizko orden estokastikoa adierazten duen, i.e., $X \geq_{st} Y$ baldin eta soilik baldin $\mathbb{P}(X \geq z) \geq \mathbb{P}(Y \geq z)$ edozion z -rentzat.

Frogapena: $(1 - \rho) \sum_{k=1}^K c_k N_k^{RP}(\vec{p})$ adierazpenak banaketan (5.38) itxaropeneko banaketa esponenzial bat duen aldagai batetara konbergitzen duela ezagutzen dugu. Banaketa esponenziala duten zorizko aldagaiak estokastikoki ordenatuak daudenez beraien itxaropenen arabera, honakoa ikustea besterik ez zaigu geratzen

$$\frac{\sum_{k=1}^K \frac{c_k \hat{\lambda}_k}{p_k}}{\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k]} \geq \frac{\sum_{k=1}^K \frac{c_k \hat{\lambda}_k}{q_k}}{\sum_{k=1}^K \frac{\hat{\lambda}_k}{q_k} \mathbb{E}[B_k]}.$$

Izan ere,

$$\begin{aligned} \left(\sum_{k=1}^K \frac{c_k \hat{\lambda}_k}{p_k}\right) \left(\sum_{k=1}^K \frac{\hat{\lambda}_k}{q_k} \mathbb{E}[B_k]\right) &= \sum_{k,i:k \neq i} \hat{\lambda}_k \hat{\lambda}_i \left(\frac{1}{p_k q_i} c_k \mathbb{E}[B_i] + \frac{1}{p_i q_k} c_i \mathbb{E}[B_k]\right) + \sum_{k=1}^K \hat{\lambda}_k^2 \frac{1}{p_k q_k} c_k \mathbb{E}[B_k] \\ &\geq \sum_{k,i:k \neq i} \hat{\lambda}_k \hat{\lambda}_i \left(\frac{1}{p_i q_k} c_k \mathbb{E}[B_i] + \frac{1}{p_k q_i} c_i \mathbb{E}[B_k]\right) + \sum_{k=1}^K \hat{\lambda}_k^2 \frac{1}{p_k q_k} c_k \mathbb{E}[B_k] = \left(\sum_{k=1}^K \frac{c_k \hat{\lambda}_k}{q_k}\right) \left(\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k]\right). \end{aligned}$$

Hemen, $c_i \mathbb{E}[B_k] \left(\frac{1}{p_i q_k} - \frac{1}{p_k q_i}\right) \geq c_k \mathbb{E}[B_i] \left(\frac{1}{p_i q_k} - \frac{1}{p_k q_i}\right)$ erabili dugu, zeina $\frac{p_i}{p_k} \leq \frac{q_i}{q_k}$ eta $\frac{c_i}{\mathbb{E}[B_i]} \geq \frac{c_k}{\mathbb{E}[B_k]}$, $i \leq k$, adierazpenetatik betetzen den. $\square$

5.5.2 Itxaron denboraren momentuak

Atal honetan k klaseko etiketatutako bezeroaren itxaron denbora eskalatuaren, $\hat{W}_k$, m -garren momentua minimizatzen duten pixuen balio optimoak erakusten ditugu. 5.4.1 Proposiziotik ezagutzen dugu

$$\widehat{W_k^{DROS}} \stackrel{d}{=} X \cdot Z_k, \quad (5.39)$$

non X eta Z_k banaketa esponentziala duten zorizko aldagaiak diren $\mathbb{E}(Z_k) = 1/p_k$ eta $\mathbb{E}(X) = 1/\nu(\vec{p})$ itxaropenak dituztenak, hurrenez hurren. Orain (5.39) adierazpena hartuz eta X eta Z_k aldagai independenteak direla erabiliz, $\hat{W}_k$ -ren m -garren momentua honako adierazpen honen bidez ematen dela ikus dezakegu

$$\mathbb{E} \left[\left(\widehat{W_k^{DROS}} \right)^m \right] = \mathbb{E}[X^m Z_k^m] = \mathbb{E}[X^m] \mathbb{E}[Z_k^m] = \frac{m!}{\nu(\vec{p})^m} \frac{m!}{p_k^m} = (m!)^2 \frac{1}{p_k^m} \left(\frac{\sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k^2]}{2 \sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k]} \right)^m.$$

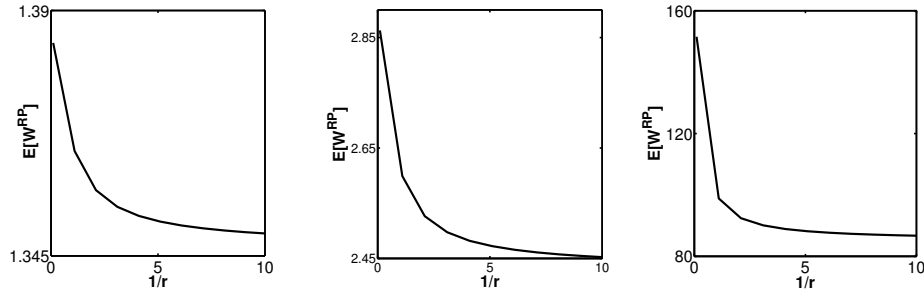
Horregatik, bezero arbitrario baten itxaron denboraren m -garren momentuak honako adierazpen hau betetzen du

$$\mathbb{E} \left[\left(\widehat{W^{DROS}} \right)^m \right] = \sum_{k=1}^K \frac{\hat{\lambda}_k}{\hat{\lambda}} \mathbb{E} \left[\left(\widehat{W_k^{DROS}} \right)^m \right] = (m!)^2 \left(\sum_{k=1}^K \frac{\hat{\lambda}_k}{\hat{\lambda} p_k^m} \right) \left(\frac{\sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k^2]}{2 \sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k]} \right)^m. \quad (5.40)$$

Jarraian $\widehat{W^{DROS}}(\vec{p})$ idatziko dugu $\widehat{W^{DROS}}$ adierazpenaren ordeztu $\vec{p}$ pixuarekiko duen menpekotasuna adierazteko.

Ohartu $\mathbb{E}[\widehat{W^{DROS}}(\vec{p})] = \frac{1}{\hat{\lambda}} \sum_{k=1}^K \mathbb{E}[\widehat{N_k^{RP}}(\vec{p})]$. Horregatik, 5.5.1 Proposizioko emaitzari Little-en legea aplikatuz, hurrengo korolarioa lortzen dugu. Korolarioan ondorioztatzen duguna zera da, itxaron denboraren itxaropena txikitu egiten dela $\mathbb{E}[B_i]$, $i = 1, \dots, K$, handia duten bezeroei lehentasuna ematen badiugu.

5.5.1 Korolarioa *Orokortasunik galdu gabe onar dezagun klaseak honela ordenaturik daudela $\mathbb{E}[B_1] \leq \dots \leq \mathbb{E}[B_K]$. Baldin eta $\frac{p_j}{p_{j+1}} \leq \frac{q_j}{q_{j+1}}$, $\forall j = 1, \dots, K-1$, bada, orduan $\mathbb{E}[\widehat{W^{DROS}}(\vec{p})] \geq \mathbb{E}[\widehat{W^{DROS}}(\vec{q})]$.*



5.1. Irudia: Sisteman hiru klase, $K = 3$, daudeneko batezbesteko itxaron denbora DROS politikapean, $\rho = 0.2610$, $\rho = 0.6083$ eta $\rho = 0.99$ kargentzat, hurrenez hurren. Ardatz horizontala $\frac{1}{r} = \frac{p_j}{p_{j+1}}$, $j = 1, \dots, K-1$ da.

5.5.1 Oharra Itxaron denboraren monotonizitate emaitza trafiko-geldo egoeran betetzen da. Bi klaseen kasuan, $K = 2$, 5.5.1 Korolario egia da edozein sistema egonkorrentzat, i.e., ρ -ren edozein balioarentzat. Hau honela ikus daiteke. Itxaron denboraren adierazpena $K = 2$ -rako honakoa da:

$$\begin{aligned} \mathbb{E}[W^{DROS}(\vec{p})] &= \sum_{i=1}^2 \frac{\lambda_i}{\lambda} \mathbb{E}[W_i^{DROS}] \\ &= \frac{\lambda_1 \mathbb{E}[B_1^2] + \lambda_2 \mathbb{E}[B_2^2]}{2\lambda} \frac{\lambda_1(1 - \rho p_1) + \lambda_2(1 - \rho p_2)}{(1 - \rho_1 - p_2 \rho_2)(1 - \rho_2 - p_1 \rho_1) - p_1 p_2 \rho_1 \rho_2}, \end{aligned} \quad (5.41)$$

non $\mathbb{E}[W_i]$, $i = 1, 2$, -ren adierazpena [27, (38) Ekuazioa] artikuluan lortu zen. Orokortasunik galdu gabe onar dezagun $p_1 + p_2 = 1$. Orduan, (5.41) Ekuazioan p_1 -rekiko deribatua hartuz 5.5.1 Korolarioan erakutsiriko monotonizitate emaitza lortzen dugu.

Horrez gain, [27] artikuluan lorturiko batezbesteko itxaron denbora kalkulatzeko kodigo bat idatzi dugu K -ren edozein balioarentzat erabili daitekena, hau da, edozein klase kopururentzat. Pixuak honela aukeratzen ditugu $\frac{p_j}{p_{j+1}} = \frac{1}{r}$, $\forall j$. Irudietan banaketa esponontziala duten beharrezko zerbitzuak aukeratzen ditugu. Dena den, monotonizitate emaitza edozein beharrezko zerbitzuren-tzat betetzen da. Lorturiko emaitzak 5.1 Irudian eta 5.2 Irudian erakusten ditugu, $K=3$ eta $K=4$ -rentzat, hurrenez hurren, trafiko intentsitatearen balio desberdinarentzat. Adibide honetan ikus daiteke beharrezko zerbitzua trikia duten bezeroei lehentasuna emanez gero (i.e., $\frac{1}{r}$ handitzen doan heinean), itxaron denboraren itxaropena txikitu egiten dela, trafiko intentsitatearen edozein balioetarako.

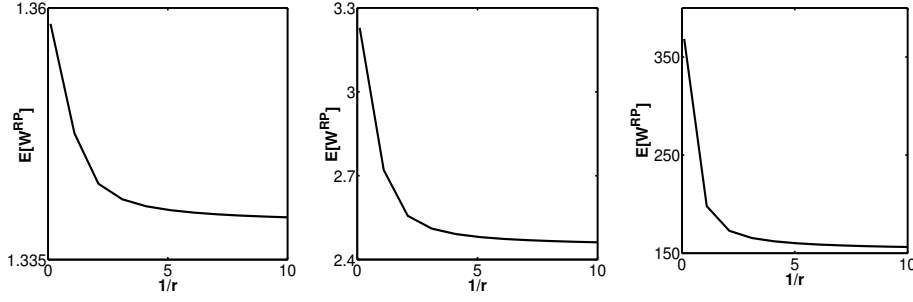
5.5.1 Korolarioan itxaron denbora eskalatuaren lehenengo momentua kontsideratzen dugu.

5.5.2 Proposizioan itxaron denbora eskalatuaren m -garren momentua aztertuko dugu eta pixuen balio optimoak kalkulatu ditugu.

5.5.2 Proposizioa Itxaron denbora eskalatuaren m -garren momentuaren limitea, $\mathbb{E} \left[\left(\widehat{W^{DROS}}(\vec{p}) \right)^m \right]$, $\vec{p}^* = (p_1^*, \dots, p_K^*)$ puntuan minimizatzen da, non

$$p_k^* := \frac{1/\mathbb{E}[B_k]^{1/m-1}}{\sum_{i=1}^K 1/\mathbb{E}[B_i]^{1/m-1}}, \quad (5.42)$$

$k \in \{1, \dots, K\}$ bakoitzeko, $m = 2, 3, \dots$



5.2. Irudia: Sisteman lau klase, $K = 4$, daudeneko batezbesteko itxaron denbora DROS politikapean, $\rho = 0.2535$, $\rho = 0.6304$ eta $\rho = 0.9951$ kargentzat, hurrenez hurren. Ardatz horizontala $\frac{1}{r} = \frac{p_j}{p_{j+1}}$, $j = 1, \dots, K-1$ da.

Frogapena: $\mathbb{E} \left[\left(\widehat{W^{DROS}}(p^*) \right)^m \right] \leq \mathbb{E} \left[\left(\widehat{W^{DROS}}(p) \right)^m \right]$ frogatu behar dugu. Hau betetzen da baldin eta soilik baldin

$$\frac{\sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k]^{m/m-1}}{\hat{\lambda} (\sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k]^{m/m-1})^m} \cdot \left(\frac{\sum_{j=1}^K \hat{\lambda}_j \mathbb{E}[B_j^2]}{2} \right)^m \leq \frac{\sum_{k=1}^K \frac{\hat{\lambda}_k}{\hat{\lambda} p_k^m}}{(\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k])^m} \cdot \left(\frac{\sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k^2]}{2} \right)^m,$$

definizioz. Hau honakoaren baliokidea da

$$\frac{1}{(\sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k]^{m/m-1})^{m-1}} \leq \frac{\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k^m}}{(\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k])^m}$$

eta berridatziz lortzen dugu

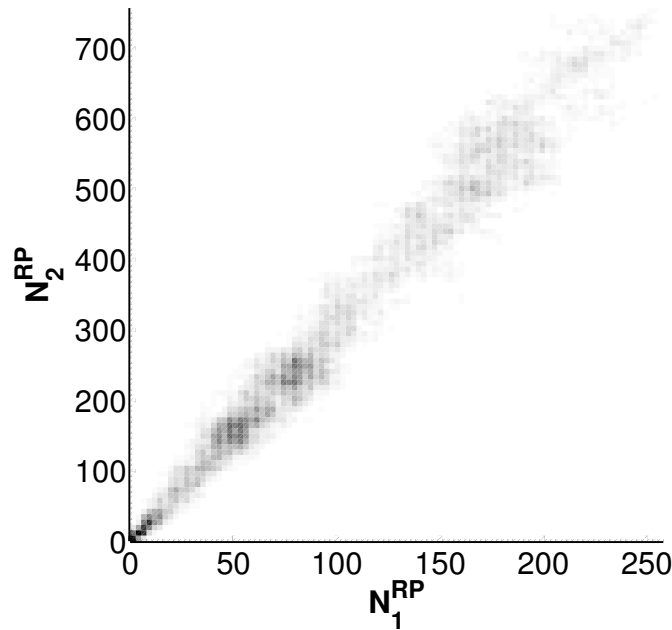
$$\left(\sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} \mathbb{E}[B_k] \right)^m \leq \sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k^m} \left(\sum_{k=1}^K \hat{\lambda}_k \mathbb{E}[B_k]^{m/m-1} \right)^{m-1}. \quad (5.43)$$

Azken hau Hölder-en desberdintzatik betetzen da. ■

5.5.2 Oharra (5.42) Ekuaziotik bi pixu optimoen arteko ratioa honakoa dela lortzen dugu:

$$\frac{p_k^*}{p_j^*} = \left(\frac{\mathbb{E}[B_j]}{\mathbb{E}[B_k]} \right)^{1/(m-1)}. \quad (5.44)$$

Orokorrean, bezeroen aukeraketa optimoa ez da tribiala. Dena den, ohartu $m \rightarrow 1$ limitepean pixu optimopean k klaseko bezero batek lehentasun osoa duela j klaseko bezero baten aurrean $\mathbb{E}[B_k] < \mathbb{E}[B_j]$ betetzen bada. Hau da, hain zuzen, $c\mu$ emaitzak dioena. Gainera, $m \rightarrow \infty$ limitepean (5.44) Ekuaziotik ikus dezakegu pixu optimoen ratioak 1-eruntz konbergitzen duela. Honek iradokitzen du m handitzen denean, optimoa dela klase guztiak berdinak kontsideratzea.



5.3. Irudia: Ilara luzeraren probailitatea. Puntuen iluntasunak puntu zehatz horretan egoteko probabilitatea adierazten du. Puntua zenbat eta ilunagoa izan, orduan eta probabilitatea handiagoa dago puntu horretan egoteko.

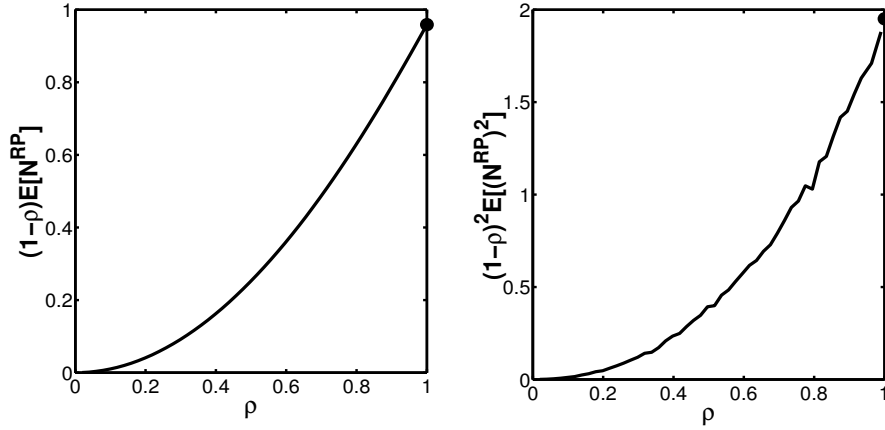
5.6 Emaiza numerikoak

Atal honetan esperimentu numerikoak aurkezten ditugu. Sistema bat kontsideratzen dugu DROS diziplinapean bi klasetako bezeroak dituen ($K = 2$) eta beharrezko zerbitzuak banaketa esponentziala dutela onartzen dugu. Esperimentu bakoitzarentzat 10^5 *busy period* simulatzen ditugu. *Busy period* bat da bi momenturen arteko denbora, non sistema hutsik egotetik berriz ere hutsik egotera igarotzen den. 5.6.1. Atalean ilarako bezero kopuruaren banaketari dagokion emaitza numerikoak aurkezten ditugu. 5.6.2. Atalean bezero kopuruaren eta itxaron denboraren momentuetan zentratzen gara eta 5.6.3. Atalean pixu optimoak aztertzen ditugu.

5.6.1 Egoera-espazioaren kolapsoa ilara luzerarentzat

Atal honetan ilara luzeraren bektorearen banaketa simulatzen dugu. Aukeraturiko parametroak hauexek dira: $\lambda_1 = 2.15$, $\lambda_2 = 2.85$, $\mathbb{E}[B_1] = 1/4$ eta $\mathbb{E}[B_2] = 1/6$, ondorioz $\rho = 0.9994$. 5.3 Irudia ilara luzeraren probabilitateak (simulazioen bitartez lortuak) erakusten ditugu $p_1 = 0.7$, $p_2 = 0.3$ pixuentzako. Ardatz horizontal eta bertikalak N_1^{RP} eta N_2^{RP} aldagaiei dagokie, hurrenez hurren. 5.3.1 Proposizioan erakutsiriko egoera-espazioaren kolapsoaren emaitzaren ondorioz, trafiko-geldo

egoeran, probabilitateek $\frac{\widehat{N_2^{RP}}}{\widehat{N_1^{RP}}} = \frac{p_1 \hat{\lambda}_2}{\hat{\lambda}_1 p_2} \approx 3.1$ malda duen lerro zuzen bat osatzen dute. Malda honek irudian lortzen dugun maldarekin bat egiten du.



5.4. Irudia: Ilara luzera eskalatuaren lehenengo eta bigarren momentuak ρ kargaren balio desberdinetarako. Bi irudietako puntu beltzak (5.14) erabiliz kalkulatu ditugu emaitzatzat $\mathbb{E}[\widehat{N^{RP}}] = 0.9589$ eta $\mathbb{E}[(\widehat{N^{RP}})^2] = 1.9510$ lortuz.

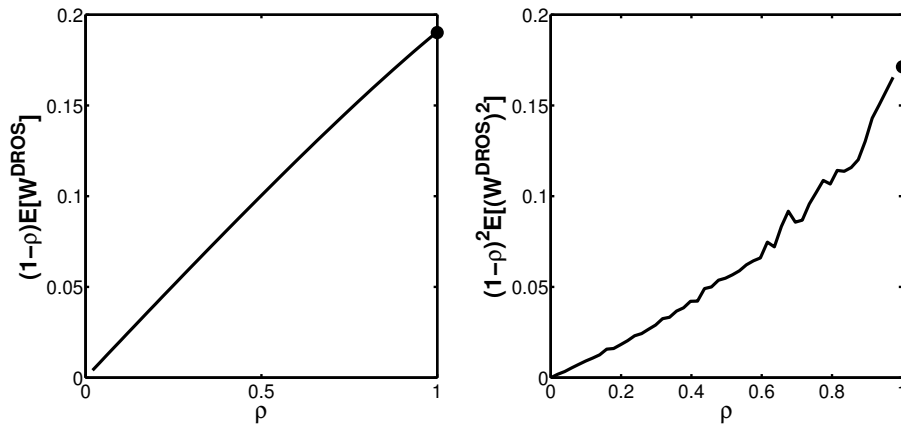
5.6.2 Ilara luzera eta itxaron denborarean momentuak

5.4 Irudian irudikatzen ditugu $(1 - \rho)\mathbb{E}[N^{RP}]$ (Little-en Legea eta (5.41) Ekuazioa erabiliz) eta $(1 - \rho)^2\mathbb{E}[(N^{RP})^2]$ (simulazioen bidez lorturikoa) ρ kargaren balio desberdinentzat. $\mathbb{E}[B_1] = 1/4$ eta $\mathbb{E}[B_2] = 1/6$ beharrezko zerbitzuak finko mantentzen ditugu eta $\lambda_2 = 1.5\lambda_1$ hartzen dugu. Honetaz gain, ilara luzera eskalatuaren lehenengo eta bigarren momentuak kalkulatzeko ditugu, i.e., $\mathbb{E}[\lim_{\rho \uparrow 1} ((1 - \rho)N^{RP})^m] = \mathbb{E}[(\widehat{N^{RP}})^m]$, $m = 1, 2$, (5.14) erabiliz, 5.4 Irudian puntu beltzez irudikaturiko balioa emanez, zeinak $\mathbb{E}[\hat{N}] = 0.9589$ eta $\mathbb{E}[\hat{N}^2] = 1.9510$ diren.

5.4 Irudian ikus daitekeen bezala, bi kasuetan, ρ bateruntz hurbiltzean, $\mathbb{E}[(1 - \rho)^m N^m]$, $m = 1, 2$ funtzioek beltzez adierazitako puntura konbergitzen dute. Honek esan nahi du, $(1 - \rho)N_k^{RP}$ zorizko aldagaiarentzat limitearen eta itxaropenaren trukaketa betetzen dela, hau da, $\lim_{\rho \uparrow 1} \mathbb{E}[(1 - \rho)^m (N_k^{RP})^m] = \mathbb{E}[\lim_{\rho \uparrow 1} (1 - \rho)^m (N_k^{RP})^m]$, $m = 1, 2$.

Ohartu limiteak trukakorrak izanik, ilara luzera eskalatuaren konbergentzia banaketan betetzen denez, honek esan nahi duela ilara luzera eskalatua uniformeki integragarria dela (ikusi [5, Teorema 3.5]), 1 Hipotesian onartu bezala.

5.5 Irudian $(1 - \rho)\mathbb{E}[W^{DROS}]$ irudikatzen dugu ((5.41) Ekuazioa erabiliz) eta $(1 - \rho)^2\mathbb{E}[(W^{DROS})^2]$ (simulazio bidez lortua) ρ kargaren balio desberdinentzat. Simulazioa gauzatzeko aukeraturiko parametroak ilara luzeraren kasuan aukeraturiko berberak dira. (5.40) Ekuazioaren balioa kalkulatzeko $m = 1$ eta $m = 2$ kasuetarako, emaitza moduan 5.5 Irudian puntu beltz batekin irudikaturiko puntuak emanez, zeinak $\mathbb{E}[\widehat{W^{DROS}}] = 0.1906$ eta $\mathbb{E}[(\widehat{W^{DROS}})^2] = 0.1713$ diren. Bi kasuetan, ρ batera hurbiltzen denean funtzioak (5.40) Ekuazioan lorturiko balioetara konbergitzen du. Honek esan nahi du, berriz ere, $(1 - \rho)W_k$ aldagaiarentzat limitearen eta itxaropenaren arteko trukaketa betetzen dela, hau da, $\lim_{\rho \uparrow 1} \mathbb{E}[(1 - \rho)W_k^{DROS}] = \mathbb{E}[\lim_{\rho \uparrow 1} (1 - \rho)W_k^{DROS}]$, $m = 1, 2$.



5.5. Irudia: Itxaron denbora eskalatuaren lehen eta bigarren momentuak ρ -ren balio desberdinetarako. Irudikaturiko bi puntu beltzak (5.40) Ekuazioa erabiliz lortu dira, $\mathbb{E}[\widehat{W^{DROS}}] = 0.1906$ eta $\mathbb{E}[(\widehat{W^{DROS}})^2] = 0.1713$ emaitzak emanez.

5.6.3 Pixuen balio optimoak

5.5.2 Proposizioan p^* pixuen aukera optimoak aurkezten ditugu helburua itxaron denbora eskalatuaren, $\widehat{W^{DROS}}$, momentuak kalkulatzeko izanik. Atal honetan numerikoki kalkulatzeko pixu horien balioak trafiko-geldoko egoeratik kanpo. Izan bitez $\mathbb{E}[B_1] = 0.2439$ eta $\mathbb{E}[B_2] = 0.1667$ eta irudikatu $(1-\rho)^2 \mathbb{E}[(W^{DROS})^2(p_1, 1-p_1)]$ ρ -ren hiru balioentzako, $\rho = 0.7, \rho = 0.8$ eta $\rho = 0.9$, ikusi 5.6 Irudia. p_1^* -ren balioa kasu honetan $p_1^* = 0.4059$ da (ikusi (5.42)). Ikus daiteke $p_1^* = 0.4059$ pixua $(1-\rho)^2 \mathbb{E}[(W^{DROS})^2(p_1, 1-p_1)]$ minimizatzeko hurbilketa egokia dela $\rho = 0.9$ baliorako. ρ txikitzean hurbilketa okerragoa bihurtzen da, baina, hala ere, funtzioaren minimotik gertu dago. $\mathbb{E}[\lim_{\rho \uparrow 1} (1-\rho)^2 (W^{DROS})^2(p_1, 1-p_1)] = \mathbb{E}[(\widehat{W^{DROS}})^2(p_1, 1-p_1)]$ ere irudikatzen dugu, zeinek $(1-\rho)^2 \mathbb{E}[(W^{DROS})^2(p_1, 1-p_1)]$ -ren hurbilketa egokia dirudi ρ bateruntz hurbiltzean.

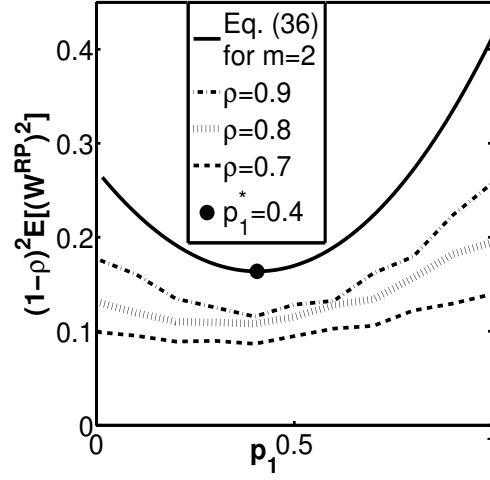
5.A Eranskina

5.A.1 5.2.1 Lemaren frogapena

Sistemako lan totala honela adieraz daiteke

$$V^{\text{dep}} = \sum_{k=1}^K \sum_{h=1}^{Q_k} B_{k,h},$$

non $B_{k,h}$ k klaseko h -garren bezeroaren beharrezko zerbitzua den. Ohartu k klaseko beharrezko zerbitzuak independenteak eta banaketa bera dutela, eta $B_{k,h} \stackrel{d}{=} B_k$, h guztientzat.



5.6. Irudia: Bigarren momentua $\mathbb{E}[(1-\rho)^2 W^2]$ pixuen balio desberdinetarako $(p_1, p_2) = (p_1, 1-p_1)$.

$\epsilon' > 0$ -rako

$$\begin{aligned}
 \mathbb{P}((1-\rho)Q_k \geq M) &= \mathbb{P}\left(Q_k \geq \frac{M}{(1-\rho)}\right) \\
 &\leq \mathbb{P}\left(\sum_{h=1}^{Q_k} B_{k,h} \geq \sum_{h=1}^{\lfloor M/(1-\rho) \rfloor} B_{k,h}\right) \\
 &\leq \mathbb{P}\left((1-\rho) \sum_{k=1}^K \sum_{h=1}^{Q_k} B_{k,h} \geq M \frac{(1-\rho)}{M} \sum_{h=1}^{\lfloor M/(1-\rho) \rfloor} B_{k,h}\right) \\
 &= \mathbb{P}\left(\frac{(1-\rho)V^{\text{dep}}}{M} - \mathbb{E}[B_k] \geq \frac{(1-\rho)}{M} \sum_{h=1}^{\lfloor M/(1-\rho) \rfloor} B_{k,h} - \mathbb{E}[B_k]\right) \\
 &= \mathbb{P}\left(\frac{(1-\rho)V^{\text{dep}}}{M} - \mathbb{E}[B_k] \geq \frac{(1-\rho)}{M} \sum_{h=1}^{\lfloor M/(1-\rho) \rfloor} B_{k,h} - \mathbb{E}[B_k], \frac{(1-\rho)}{M} \sum_{h=1}^{\lfloor M/(1-\rho) \rfloor} B_{k,h} - \mathbb{E}[B_k] > -\epsilon\right) \\
 &\quad + \mathbb{P}\left(\frac{(1-\rho)V^{\text{dep}}}{M} - \mathbb{E}[B_k] \geq \frac{(1-\rho)}{M} \sum_{h=1}^{\lfloor M/(1-\rho) \rfloor} B_{k,h} - \mathbb{E}[B_k] \mid \frac{(1-\rho)}{M} \sum_{h=1}^{\lfloor M/(1-\rho) \rfloor} B_{k,h} - \mathbb{E}[B_k] \leq -\epsilon\right) \\
 &\quad \cdot \mathbb{P}\left(\frac{(1-\rho)}{M} \sum_{h=1}^{\lfloor M/(1-\rho) \rfloor} B_{k,h} - \mathbb{E}[B_k] \leq -\epsilon\right)
 \end{aligned} \tag{5.45}$$

$$\begin{aligned}
&\leq \mathbb{P}\left(\frac{(1-\rho)V^{\text{dep}}}{M} - \mathbb{E}[B_k] > -\epsilon\right) + \tilde{\epsilon} \\
&= \mathbb{P}\left((1-\rho)V^{\text{dep}} \geq M(\mathbb{E}[B_k] - \epsilon)\right) + \tilde{\epsilon} \\
&< \bar{\epsilon} + \tilde{\epsilon} = \epsilon', \rho \text{ 1-etik gertu eta } M \text{ nahiko handia den kasurako.}
\end{aligned} \tag{5.46}$$

Bosgarren pausoa erabili dugu $\frac{(1-\rho)}{M} \sum_{h=1}^{\lfloor M/(1-\rho) \rfloor} B_{k,h}$ banaketan konbegitzen duela $\mathbb{E}[B_k]$ -ra $\rho \uparrow 1$ limitepean, beraz $\mathbb{P}\left(\frac{(1-\rho)}{M} \sum_{h=1}^{\lfloor M/(1-\rho) \rfloor} B_{k,h} - \mathbb{E}[B_k] \leq -\epsilon\right) \leq \bar{\epsilon}$, ρ batetik hurbil dagoeanean. Azken pausoa lan totala *tight* dela trafiko-geldo egoeran erabili dugu, erabilitako planifikazio diziplina kontserbakorraren independenteki, ikusi Kingman [28], hau da, $\forall \bar{\epsilon} \exists M'$ non $\mathbb{P}((1-\rho)V^{\text{dep}} \geq M') < \bar{\epsilon}$. (5.46) Ekuaziotik ondorioztatzen dugu $(1-\rho)(Q_1^{RP}, Q_2, \dots, Q_K^{RP})$ *tight* dela.

5.A.2 5.2.4 Lemaren frogapena

5.2.4 Lemaren frogapena [49] artikuluko 3 Lema oinarriturik dago. Honakoa dugu

$$\begin{aligned}
\sum_{i=1}^K \frac{\hat{\lambda}_i}{p_i} F_i(\vec{s}) &= \sum_{i=1}^K \frac{\hat{\lambda}_i}{p_i} p_i (-s_i + \mathbb{E}[B_i] \sum_{k=1}^K \hat{\lambda}_k s_k) \\
&= -\sum_{i=1}^K \hat{\lambda}_i s_i + \sum_{i=1}^K \hat{\lambda}_i \mathbb{E}[B_i] \sum_{k=1}^K \hat{\lambda}_k s_k \\
&= -\sum_{i=1}^K \hat{\lambda}_i s_i + \sum_{k=1}^K \hat{\lambda}_k s_k \\
&= 0.
\end{aligned}$$

Honek inplikatzeko du $\vec{s} \in H_c$ guztietarako, $\vec{F}(\vec{s})$ bektorea H_c hiperplanoaren paraleloa dela. $\vec{F} C^1$ denez, $\vec{s} \geq \vec{0}$ egoera bakoitzerako $\vec{f}(u) = (f_1(u), \dots, f_K(u))$ isuri bakarra existitzen da, $u \geq 0$ -k parametrizatua, hau da,

$$\vec{f}(0) = \vec{s} \quad \text{eta} \quad \frac{\partial f_i(u)}{\partial u} = F_i(\vec{f}(u)), \quad i \text{ guztientzat} \quad \text{eta} \quad u \geq 0. \tag{5.47}$$

$\vec{F}(\vec{s})$ H_c hiperplanoaren paraleloa denez $\vec{s} \in H_c$ guztietarako, H_c hasiz gero, isuria hiperplanoan mantenduko da. $\vec{f}(u)$ isuriaren beste propietate garrantzitsu bat da

$$\frac{\partial \hat{r}(\vec{f}(u))}{\partial u} = \sum_{i=1}^K \frac{\partial f_i(u)}{\partial u} \cdot \frac{\partial \hat{r}(\vec{s})}{\partial s_i} \Big|_{\vec{s}=\vec{f}(u)} = 0,$$

zeina lortzen dugun katearen araua, 5.2.3 Lema eta (5.47) Ekuazioa erabiliz. Beraz, hiperplano-ko isuri bakoitzarentzako, $\vec{f}(u)$, $\hat{r}(\vec{f}(u))$ konstantea da. Jarraian, H_c hiperplanoko isuri bakoitza puntu zehatz $c \cdot \vec{s}^* \geq 0$ batetara konbergitzen duela ikusiko dugu $u \rightarrow \infty$ limitepean.

(5.10) Ekuaziotik lortzen dugu (5.47) honela idatz daitekeela $\vec{f}(0) = \vec{s}$ eta $\vec{f}'(u)^T = A\vec{f}(u)^T$ non

$$A = \begin{pmatrix} p_1(-1 + \mathbb{E}[B_1]\hat{\lambda}_1) & p_1\mathbb{E}[B_1]\hat{\lambda}_2 & \cdots & p_1\mathbb{E}[B_1]\hat{\lambda}_K \\ p_2\mathbb{E}[B_2]\hat{\lambda}_1 & p_2(-1 + \mathbb{E}[B_2]\hat{\lambda}_2) & \cdots & p_2\mathbb{E}[B_2]\hat{\lambda}_K \\ \vdots & \vdots & \ddots & \vdots \\ p_K\mathbb{E}[B_K]\hat{\lambda}_1 & p_K\mathbb{E}[B_K]\hat{\lambda}_2 & \cdots & p_K(-1 + \mathbb{E}[B_K]\hat{\lambda}_K) \end{pmatrix}. \quad (5.48)$$

5.A.1 Leman frogatzen da A matrizearen balio propio bat 0 dela $\vec{s}^* \geq \vec{0}$, $\vec{s}^* \in H_1$ bektore propio-duna, eta beste balio propioek zati erreal negatiboa dutela. Beraz, $\vec{f}'(u)^T = A\vec{f}(u)^T$ -ren soluzioa, $\vec{f}(0) \in H_c$ izanik, honela idatz daiteke $\vec{f}(u) = c \cdot \vec{s}^* + \vec{g}(u)$, non $\lim_{u \rightarrow \infty} \vec{g}(u) = \vec{0}$ eta $\vec{s}^* \geq \vec{0}$. $\hat{r}(\vec{s})$ funtzio jarraia eta konstantea denez isuri bakoitzean zehar, eta isuri guztiek H_c hiperplanoan $c \cdot \vec{s}^* \in H_c$ -ra konbergitzen dutenez, $\hat{r}(\vec{s})$ funtzioa hiperplanoan konstantea dela ondoriozta dezakegu.

Ondoko lema teknika 5.2.4 Lemaren frogapenean erabiltzen da.

5.A.1 Lema *Kontsideratu (5.48) Ekuazioan definituriko A matrizea. A -ren balio propio bat 0 da eta beste guztiek zati erreal hertsiki negatiboa dute. Honetaz gain, $\vec{\eta} = (\eta_1, \dots, \eta_K) \geq \vec{0}$ bektore bat existitzen da non $\sum_{j=1}^K \eta_j = 1$ den eta $\vec{s}^* = (s_1^*, \dots, s_K^*)$, $s_j^* := \frac{p_j}{\hat{\lambda}_j} \eta_j$ izanik, A matrizearen bektore propio bat den, 0 balio propioari dagokiona, hain zuzen, eta $\vec{s}^* \in H_1$.*

Proof: Defini dezagun $D = \text{diag}[d_1, d_2, \dots, d_K]$ matrize diagonal bezala, non $d_i = \frac{\hat{\lambda}_i}{p_i}$, eta izan bedi S honako matrizea

$$S = DAD^{-1} = \begin{pmatrix} p_1(-1 + \mathbb{E}[B_1]\hat{\lambda}_1) & p_2\hat{\lambda}_1\mathbb{E}[B_1] & \cdots & p_K\hat{\lambda}_1\mathbb{E}[B_1] \\ p_1\hat{\lambda}_2\mathbb{E}[B_2] & p_2(-1 + \mathbb{E}[B_2]\hat{\lambda}_2) & \cdots & p_K\hat{\lambda}_2\mathbb{E}[B_2] \\ \vdots & \vdots & \ddots & \vdots \\ p_1\hat{\lambda}_K\mathbb{E}[B_K] & p_2\hat{\lambda}_K\mathbb{E}[B_K] & \cdots & p_K(-1 + \mathbb{E}[B_K]\hat{\lambda}_K) \end{pmatrix}. \quad (5.49)$$

A matrizea S -ren antzekoa da eta horregatik, A, S eta S^T matrizeek balio propio berak dituzte. S^T matrizearen ilara bakoitzaren batura 0 da $\sum_{i=1}^K \mathbb{E}[B_i]\hat{\lambda}_i = 1$ betetzen delako eta diagonaletik kanpo dauden elementuak hertsiki positiboak direlako. Honek zera esan nahi du, S^T matrizea egoera finituko, denbora jarraituko markov kate irreduzible baten sortzaile dela. Horregatik, oreka banaketa bakarra du $\vec{\eta}$, i.e., $\vec{\eta}S^T = \vec{0}$ eta $\sum_{k=1}^K \eta_k = 1$. Bereziki, 0 S^T matrizearen balio propioa da, eta $\vec{\eta}$ bektore propioari dagokio $\vec{\eta}$ eta (6.2 Proposizioa, [2]), beste balio propioen zati errealak hertsiki negatiboak dira. A eta S^T matrizeen balio propioak bat datozenez, gauza bera gertatzen da A matrizearentzat. 0 balio propioari dagokion A -ren bektore propioa $\vec{s}^{*T} = D^{-1}\vec{\eta}^T$ da, izan ere, $A\vec{s}^{*T} = D^{-1}DAD^{-1}\vec{\eta}^T = D^{-1}S\vec{\eta}^T = \vec{0}^T$

■

5.A.3 5.4.2 Lemaren frogapena

(5.27) Ekuazioan $(u, z_1, \dots, z_K) = ((1 - \rho)u, e^{-(1-\rho)s_1}, \dots, e^{-(1-\rho)s_K})$ hartuz

$$\begin{aligned} W_l^1((1 - \rho)u, e^{-(1-\rho)s_1}, \dots, e^{-(1-\rho)s_K}) = \\ \sum_{i=1}^K ((1 - \rho)\lambda_i + \lambda p_i \frac{\partial}{\partial z_i} r(z_1, \dots, z_K) \Big|_{z_i = e^{-(1-\rho)s_i}}) \cdot \\ \frac{B_i^*(\lambda - \sum_{k=1}^K \lambda_k e^{-(1-\rho)s_k}) - B_i^*((1 - \rho)u + \lambda - \sum_{k=1}^K \lambda_k e^{-(1-\rho)s_k})}{(1 - \rho)u}. \end{aligned}$$

L'Hopital-en erregela erabiliz beheko adierazpen hau lortzen dugu:

$$\begin{aligned} \lim_{\rho \uparrow 1} W_l^1((1 - \rho)u, e^{-(1-\rho)s_1}, \dots, e^{-(1-\rho)s_K}) \\ = -\frac{1}{u} \sum_{i=1}^K \left(\hat{\lambda}_i \cdot 0 + \hat{\lambda} p_i \frac{\partial}{\partial s_i} \hat{r}(s_1, \dots, s_K) \left(B_i^{*'}(0) \left(-\sum_{k=1}^K \hat{\lambda}_k s_k \right) - B_i^{*'}(0) \left(-u - \sum_{k=1}^K \hat{\lambda}_k s_k \right) \right) \right) \\ = \hat{\lambda} \sum_{i=1}^K \mathbb{E}[B_i] p_i \frac{\hat{\lambda}_i}{p_i} \frac{d}{dv} \hat{r}^*(v) \Big|_{v = \sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k} = \hat{\lambda} \frac{d}{dv} \hat{r}^*(v) \Big|_{v = \sum_{k=1}^K \frac{\hat{\lambda}_k}{p_k} s_k}, \end{aligned}$$

non $\hat{r}(\vec{s})$ (5.6) Ekuazioan definiturik dagoen eta $\hat{r}^*(\vec{s})$ 5.3.1 Leman. Orain, emaitza (5.18) Ekuaziotik lortzen dugu, azken adierazpena $\frac{\nu(\vec{p})}{\nu(\vec{p}) + \sum_{k=1}^K s_k \frac{\hat{\lambda}_k}{p_k}}$ -ren berdina izanik (izan ere, $(\widehat{N_1^{RP}}, \dots, \widehat{N_K^{RP}}) \stackrel{d}{=} X \cdot (\frac{\lambda_1}{p_1}, \dots, \frac{\lambda_K}{p_K})$, X -ek $1/\nu(\vec{p})$ batezbesteko banaketa esponentziala izanik).

5.A.4 (5.35) Ekuazio diferentzialaren ebazpena

(5.35) Ekuazio diferentzialaren emaitza soluzio homogenoaren, $T_l^{\widehat{DROS}, H}(u, y)$, eta partikularren, $T_l^{\widehat{DROS}, P}(u, y)$, batura da. Soluzio homogeneoa:

$$T_l^{\widehat{DROS}, H}(u, y) = C(u) e^{\frac{pl}{u} y}, \quad (5.50)$$

non $C(u)$ u -ren funtzio arbitrarioa den. Soluzio partikularra aurkitzeko idatz dezagun (5.35) Ekuazioa honela

$$\frac{\partial T_l^{\widehat{DROS}, P}(u, y)}{\partial y} \Big|_{y = \sum_{k=1}^K s_k \frac{\hat{\lambda}_k}{p_k}} - \frac{pl}{u} T_l^{\widehat{DROS}, P}(u, y) = -\frac{pl}{u} \frac{\nu(\vec{p})}{\nu(\vec{p}) + y}. \quad (5.51)$$

Ebatz dezagun ekuazio berria faktore integragarriaren teknika erabiliz. Horretarako $\mu(y) = e^{-\frac{pl}{u} y}$ definitzen dugu eta (5.51) Ekuazioa honekin biderkatu. $\mu(y)$ -ren deribatua da $\frac{d\mu(y)}{dy} = -\mu(y) \frac{pl}{u}$.

Orduan gure ekuazioak honako adierazpen hau du

$$\begin{aligned}
 -\mu(y) \frac{p_l}{u} \frac{\nu(\vec{p})}{\nu(\vec{p}) + y} &= \mu(y) \frac{\partial T_l^{\widehat{DROS},P}(u, y)}{\partial y} \Big|_{y=\sum_{k=1}^K s_k \frac{\hat{\lambda}_k}{p_k}} - \mu(y) \frac{p_l}{u} T_l^{\widehat{DROS},P}(u, y) \\
 &= \mu(y) \frac{\partial T_l^{\widehat{DROS},P}(u, y)}{\partial y} \Big|_{y=\sum_{k=1}^K s_k \frac{\hat{\lambda}_k}{p_k}} + T_l^{\widehat{DROS},P}(u, y) \frac{d\mu(y)}{dy} \\
 &= \frac{\partial}{\partial y} (\mu(y) T_l^{\widehat{DROS},P}(u, y)),
 \end{aligned}$$

zeina integratuz ebatz daitekeen. Zati bakoitza y aldagaiarekiko integratzen badugu jarraian erakusten den (5.35) Ekuazioaren soluzio partikularra aurkituko dugu

$$T_l^{\widehat{DROS},P}(u, y) = -\frac{p_l}{u} e^{\frac{p_l}{u} y} \int_0^y e^{-\frac{p_l}{u} x} \frac{\nu(\vec{p})}{\nu(\vec{p}) + x} dx,$$

edo gauza bera dena

$$T_l^{\widehat{DROS},P}(u, y) = \frac{p_l}{u} e^{\frac{p_l}{u} y} \int_y^\infty e^{-\frac{p_l}{u} x} \frac{\nu(\vec{p})}{\nu(\vec{p}) + x} dx. \quad (5.52)$$

Ondorioz, (5.35) Ekuazioaren soluzio orokorra da

$$T_l^{\widehat{DROS}}(u, y) = T_l^{\widehat{DROS},H}(u, y) + T_k^{\widehat{DROS},P}(u, y) = C(u) e^{\frac{p_l}{u} y} + \frac{p_l}{u} e^{\frac{p_l}{u} y} \int_y^\infty e^{-\frac{p_l}{u} x} \frac{\nu(\vec{p})}{\nu(\vec{p}) + x} dx. \quad (5.53)$$

Jarraian, $C(u)$ konstantea zero dela erakutsiko dugu. Lehenik eta behin, ohartu $T_l^{\widehat{DROS},H}(u, y) \rightarrow \infty$ betetzen dela $y \rightarrow \infty$ limitepean. Bigarrenez, $T_l^{\widehat{DROS}}(0, y) = \frac{\nu(\vec{p})}{\nu(\vec{p}) + y}$ betetzen denez ((5.24)

Ekuaziotik eta 5.4.1 Lematik), berehalakoa da $T_l^{\widehat{DROS}}(u, y)$ -k 0-ra konbergitzen duela $y = \sum_{k=1}^K s_k \frac{\hat{\lambda}_k}{p_k} \rightarrow \infty$ limitepean. Horrez gain, (5.52) soluzio partikularra hartzen badugu, l'Hopital-en erregela aplikatzen badugu $y \rightarrow \infty$ limiterako, zerora konbergitzen duela lortzen dugu, bereziki:

$$\lim_{y \rightarrow \infty} T_l^{\widehat{DROS}}(u, y) = \lim_{y \rightarrow \infty} \frac{\nu(\vec{p}) p_l}{u} \frac{\int_y^\infty e^{-\frac{p_l}{u} x} \frac{1}{\nu(\vec{p}) + x} dx}{e^{-\frac{p_l}{u} y}} = \lim_{y \rightarrow \infty} \frac{\nu(\vec{p}) p_l}{u} \frac{-e^{-\frac{p_l}{u} y} \frac{1}{\nu(\vec{p}) + y}}{-\frac{p_l}{u} e^{-\frac{p_l}{u} y}} = \lim_{y \rightarrow \infty} \frac{\nu(\vec{p})}{\nu(\vec{p}) + y} = 0.$$

Orduan, $y \rightarrow \infty$ limitepean, $C(u) e^{\frac{p_l}{u} y}$ -k zerora konbergitzeko beharrezko baldintza bakarra $C(u) = 0$ da. Ondorioz, (5.35) ekuazioaren emaitza honakoa dela ondorioztzen dugu

$$T_l^{\widehat{DROS}}(u, y) = \frac{p_l}{u} e^{\frac{p_l}{u} y} \int_y^\infty e^{-\frac{p_l}{u} x} \frac{\nu(\vec{p})}{\nu(\vec{p}) + x} dx.$$

6. Kapitulu

Interpolazio bidezko hurbilketak lehentasun erlatibodun politikarentzat

Kapitulu honetan 1.1. Atalean aurkezturiko lehentasun erlatibodun (RP) ereduaren trafiko arinaren eta geldoaren analisisa gauzatzen dugu. Interpolazioa gauzatzeko 5. Kapitulan lorturiko trafiko geldopeko emaitzak erabiliko ditugu.

5. Kapitulan aurkezturiko Kim et al.-en 5.3.1 eta 5.4.1 Teorema ezin dira orokorrean ebatzi. Hala ere, oso erabilgarriak dira sistemaren errendimendua ulertzeko. Zehazkiago, 5. Kapitulan bezero kopuruaren eta itxaron denboraren trafiko geldopeko emaitzak lortzeko oso erabilgarriak izan dira, hurrenez hurren, eta orain ere erabilgarriak izango dira bezero kopuruaren probabilitate funtzio sortzailea eta itxaron denboraren LST-ren trafiko-arin hurbilketa lortzeko. Lorturiko trafiko arinaren eta geldoaren emaitzak konbinatuz aipaturiko metriken interpolazio bidezko hurbilketak eraikiko ditugu. Guk lorturiko emaitzen abantailetakoa bat sistemako parametroekiko menpekotasuna erakusten dutela da. Horregatik, tresna erabilgarria izan daitekeela iruditzen zaigu.

Kapitulua honela antolatuta dago. 6.1. Atalean ilara luzeraren probabilitate funtzio sortzailearen interpolazio bidezko hurbilketa lortzen dugu. 6.2. Atalean itxaron denboraren interpolazio bidezko hurbilketa. 6.3. Atalean emaitza numerikoak aurkezten ditugu.

6.1 Bezero kopuruaren banaketa

5. Kapitulutik gogorarazten dugu k klaseko bezero kopurua une arbitrario batean N_k^{RP} bezala denotatuko dugula. $\vec{N}^{RP} = (N_1^{RP}, \dots, N_K^{RP})$ bektorea eta $N^{RP} := \sum_{k=1}^K N_k^{RP}$ definitzen ditugu.

Kapitulu honetan ilara luzeraren probabilitate funtzio sortzailearen interpolazio bidezko hurbilketa lortu nahi dugu. 2. Kapitulan aurkezturiko notazioarekin bat eginez $\vec{y} = \vec{z}$ finkatu eta izan bedi $G(\lambda, \vec{z}) = \psi^{RP}(\lambda, \vec{z})$, $(N_1^{RP}, \dots, N_K^{RP})$ bektorearen probabilitate funtzio sortzailea RP eredupean.

5. Kapitulan lorturiko 5.3.1 Proposizioko emaitza erabiliz ondoko trafiko geldo emaitza lortzen dugu. 5. Kapitulan aipatu bezala trafiko geldo erregimenpean bezero kopuru eskalatua-
rentzat *state-space collapse* fenomeno gertatzen da, hau da, limitean ilararen luzeraren bektorea

banaketa esponentziala duen zorizko aldagai baten eta bektore determinista baten biderkadura da.

6.1.1 Lema *Ilara luzeraren probabilitate funtzio sortzaile eskalatuak, $\psi^{RP}(\lambda, \vec{z}^{(1-\lambda\mathbb{E}[B])})$, trafiko geldo erregimenpean ondokoa betetzen du*

$$\begin{aligned}
 \lim_{\lambda \rightarrow 1/\mathbb{E}[B]} \psi^{RP}(\lambda, \vec{z}^{(1-\lambda\mathbb{E}[B])}) &= \lim_{\lambda \rightarrow 1/\mathbb{E}[B]} \mathbb{E}[z_1^{(1-\lambda\mathbb{E}[B])N_1^{RP}} \dots z_K^{(1-\lambda\mathbb{E}[B])N_K^{RP}}] \\
 &= \lim_{\lambda \rightarrow 1/\mathbb{E}[B]} \mathbb{E} \left[e^{\ln(z_1^{(1-\lambda\mathbb{E}[B])N_1^{RP}})} \dots e^{\ln(z_K^{(1-\lambda\mathbb{E}[B])N_K^{RP}})} \right] = \mathbb{E} \left[e^{\widehat{N_1^{RP}} \ln(z_1)} \dots e^{\widehat{N_K^{RP}} \ln(z_K)} \right] \\
 &= \mathbb{E} \left[e^{X \sum_{i=1}^K \frac{\hat{\lambda}_i}{p_i} \ln(z_i)} \right] = \frac{\mathbb{E}[B] \nu(\vec{p})}{\mathbb{E}[B] \nu(\vec{p}) - \sum_{i=1}^K \frac{\hat{\alpha}_i}{p_i} \ln(z_i)}. \tag{6.1}
 \end{aligned}$$

non laugarren pausoan (5.19) Ekuazioa erabili dugun eta $\nu(\vec{p})$ (5.15) Ekuazioan erakusten den.

Jarraian, 5. Kapituluan aurkezturiko 5.3.1 Teorema erabiliz, ondoko trafiko-arin hurbilketa lortzen dugu ($\lambda = 0$)

6.1.2 Lema *Bezera kopuruaren probabilitate funtzio sortzailearen trafiko-arin hurbilketa ondokoa da*

$$\begin{aligned}
 \psi^{RP,LT}(\lambda, \vec{z}) &= (\psi^{RP})^{(0)}(0, \vec{z}) + \lambda (\psi^{RP})^{(1)}(0, \vec{z}) + \frac{\lambda^2}{2} (\psi^{RP})^{(2)}(0, \vec{z}) \\
 &= 1 - \rho + \lambda \sum_{i=1}^K \alpha_i \mathbb{E}[B_i] z_i + \frac{\lambda^2}{2} \sum_{i=1}^K \alpha_i z_i \mathbb{E}[B_i^2] \left(\sum_{k=1}^K \alpha_k z_k - 1 \right).
 \end{aligned}$$

Frogapena: Ikusi frogapena 6.A.1 Eranskinean. ■

Jarraian, bezera kopuruaren interpolazio bidezko hurbilketa aurkezten dugu.

6.1.1 Proposizioa $(N_1^{RP}, \dots, N_K^{RP})$ bektorearen probabilitate funtzio sortzailearen interpolazio bidezko hurbilketa ondokoa da

$$\begin{aligned}
 \psi^{RP,INT}(\lambda, \vec{z}) &= (1 - \rho^3) + \lambda (1 - \rho^2) \left(-\mathbb{E}[B] + \sum_{i=1}^K \alpha_i \mathbb{E}[B_i] z_i^{(1-\rho)^{-1}} \right) \\
 &\quad + \frac{\lambda^2 (1 - \rho)}{2} \left(-2\mathbb{E}[B] \sum_{i=1}^K \alpha_i z_i^{(1-\rho)^{-1}} \mathbb{E}[B_i] \frac{\ln(z_i)}{1 - \rho} \right. \\
 &\quad \left. + \sum_{i=1}^K \alpha_i z_i^{(1-\rho)^{-1}} \mathbb{E}[B_i^2] \left(\sum_{k=1}^K \alpha_k z_k^{(1-\rho)^{-1}} - 1 \right) \right) \\
 &\quad + \rho^3 \frac{\mathbb{E}[B] \nu(\vec{p})}{\mathbb{E}[B] \nu(\vec{p}) - \sum_{i=1}^K \frac{\hat{\alpha}_i}{p_i} \ln(z_i^{(1-\rho)^{-1}})}. \tag{6.2}
 \end{aligned}$$

Frogapena: Frogapena 2.3.1 Proposizioa erabiliz lortzen da zuzenean, 6.1.1 Leman lorturiko trafiko geldopeko emaitza $G^{HT}(\vec{z})$ eta 6.1.2 Leman lorturiko trafiko arinpeko emaitza konbinatuz. Frogapenaren xehetasun gehiagorako ikusi 6.A.2 Eranskina. ■

(6.2) Ekuazioa bezero kopuruaren lehenengo eta bigarren momentuak lortzeko erabil daiteke. Lehenengo momenturako ondokoa dugu.

$$\begin{aligned}\mathbb{E}[N^{RP,INT}] &= \mathbb{E}[N_1^{RP,INT} + \dots + N_K^{RP,INT}] = \frac{\partial \left(\psi^{RP,INT}(\lambda, \vec{z}) \Big|_{z_i=z_j=z} \right)}{\partial z} \Big|_{z=1} \\ &= \rho + \frac{\lambda^2 \mathbb{E}[B^2]}{2} + \frac{\rho^3}{(1-\rho)} \frac{\mathbb{E}[B^2]}{2\mathbb{E}[B] \sum_{k=1}^K \frac{\alpha_k}{p_k} \mathbb{E}[B_k]} \cdot \sum_{i=1}^K \frac{\alpha_i}{p_i}.\end{aligned}\quad (6.3)$$

Sisteman klase bakarra dagoela asumituz, hau da, $\alpha_i = 0, \forall i \neq k$ eta $\alpha_k = 1$, (6.3) Ekuazioa zehatza da, hau da, aski ezaguna den $M/G/1$ ilararentzat Pollaczek-Khinchine-en formularekin bat egiten du. Ondokoa dugu

$$\mathbb{E}[N^{RP,INT}] = \rho + \frac{\lambda^2 \mathbb{E}[B^2]}{2} \left(1 + \frac{\rho}{1-\rho} \right) = \rho + \frac{\lambda^2 \mathbb{E}[B^2]}{2(1-\rho)}.$$

$\psi^{RP,INT}(\lambda, \vec{z})$ -ren deribatua z -rekiko $z = 1$ puntuan ebaluatuta da

$$\begin{aligned}\frac{\partial^2 \left(\psi^{RP,INT}(\lambda, \vec{z}) \Big|_{z_i=z_j=z} \right)}{\partial z^2} \Big|_{z=1} &= \mathbb{E} \left[(N^{RP,INT})^2 \right] - \mathbb{E}[N^{RP,INT}]^2 = \frac{\lambda^2 \mathbb{E}[B^2]}{1-\rho} \frac{2+\rho}{2} \\ &+ \frac{\rho^3}{1-\rho} \frac{2\mathbb{E}[B]}{\mathbb{E}[B^2]} \sum_{k=1}^K \frac{\alpha_k}{p_k} \mathbb{E}[B_k] \left(\frac{\mathbb{E}[B^2]}{2\mathbb{E}[B] \sum_{k=1}^K \frac{\alpha_k}{p_k} \mathbb{E}[B_k]} \right)^2 \\ &\cdot \sum_{i=1}^K \frac{\alpha_i}{p_i} \left(\frac{2}{(1-\rho)} \frac{\mathbb{E}[B^2]}{2\mathbb{E}[B] \sum_{k=1}^K \frac{\alpha_k}{p_k} \mathbb{E}[B_k]} \sum_{i=1}^K \frac{\alpha_i}{p_i} - 1 \right).\end{aligned}\quad (6.4)$$

Horregatik, bigarren momentua (6.3) eta (6.4) Ekuazioen batura da

$$\mathbb{E} \left[(N^{RP,INT})^2 \right] = \frac{\partial^2 \left(\psi^{RP,INT}(\lambda, \vec{z}) \Big|_{z_i=z_j=z} \right)}{\partial z^2} \Big|_{z=1} + \mathbb{E}[N^{RP,INT}]^2. \quad (6.5)$$

6.3. Atalean lehenengo eta bigarren momentuen adierazpenak erabiltzen ditugu, (6.3) eta (6.5), hurrenez hurren, interpolazio bidezko hurbilketaren zehaztasun maila aztertzeko.

6.2 Itxaron denboraren banaketa

Lehentasan erlatibodun ereduaren itxaron denbora klase barnean erabilitako politikaren menpe dago. 5. Kapituluaren bezala uniformeki zorizkoa den klase barneko politika hautatu dugu, DROS bezala adieraziko duguna. Gogoratu, 5. Kapituluaren adierazi bezala, k klaseko bezero arbitrario

baten itxaron denbora W_k^{DROS} bezala adieraziko dugula. Bezera honi etiketatutako bezeroa deituko diogu. Izan bedi Q_k^* k klaseko bezero kopurua sisteman (etiketatutako bezeroa bazterturik) etiketatutako bezeroa zerbitzatzen hasten den unean, etiketatutako bezeroa zerbitzaria lanean ari den unean iristen bada, i.e., $W_k^{DROS} > 0$.

Atal honetan itxaron denboraren LST lortuko dugu DROS politikapean eta horregatik, 2. Kapituluari aurkezturiko notazioarekin bat eginez $\vec{y} = u$ finkatzen dugu eta izan bedi $G(\lambda, u) = \widetilde{W}_k^{DROS}(\lambda, u)$ k klaseko bezero baten itxaron denboraren LST DROS politikapean. Orduan, 5.4.1 Proposizioa erabiliz ondoko trafiko geldopeko emaitza lortzen dugu.

6.2.1 Lema k klaseko bezero baten itxaron denbora eskalatuaren Laplace transformatua trafiko geldo erregimenpean ondokoa da

$$\begin{aligned} \lim_{\lambda \rightarrow 1/\mathbb{E}[B]} \widetilde{W}_k^{DROS}(\lambda, (1 - \lambda\mathbb{E}[B])u) &= \lim_{\lambda \rightarrow 1/\mathbb{E}[B]} \mathbb{E}[e^{-u(1 - \lambda\mathbb{E}[B])W_k^{DROS}}] \\ &= \frac{\nu(\vec{p})p_k}{u} e^{p_k \frac{\nu(\vec{p})}{u}} \int_{p_k \frac{\nu(\vec{p})}{u}}^{\infty} \frac{e^{-l}}{l} dl. \end{aligned} \quad (6.6)$$

Jarraian, (5.26) Ekuazioa λ aldagaiarekiko deribatuz ondoko trafiko-arin hurbilketa lortzen dugu.

6.2.2 Lema *Itxaron denboraren Laplace transformatuaren trafiko-arin hurbilketa DROS politikapean ondokoa da*

$$\widetilde{W}_k^{DROS,LT}(\lambda, u) = \left(\widetilde{W}_k^{DROS} \right)^{(0)}(0, u) + \lambda \left(\widetilde{W}_k^{DROS} \right)^{(1)}(0, u) = 1 - \rho + \lambda \left(\sum_{i=1}^K \alpha_i \frac{1 - B_i^*(u)}{u} \right)$$

Frogapena: Ikusi frogapena 6.A.3 Eranskinean. ■

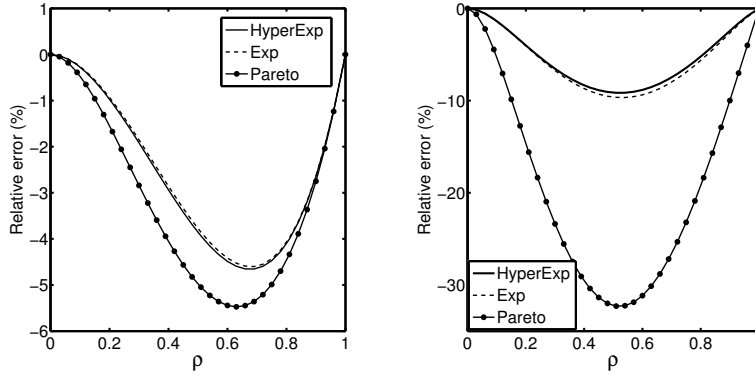
Ohartu trafiko-arin hurbilketa klasearen independentea dela. 2.1.1 Proposiziotik dakigu lehenengo mailako hurbilketa sistemara bezero bakarria iristen denean kalkulatzeko dela, eta horregatik, erabilitako politikak ez duela inongo eraginik. Bigarren mailako hurbilketa kalkulatu daiteke. Denaren, azken adierazpena oso nahasia da eta numerikoki zehaztasun maila ez da asko hobetzen. Jarraian, interpolazio bidezko hurbilketa aurkeztzen dugu, zeinak klasearen menpekotasuna duen trafiko geldo terminoan.

6.2.1 Proposizioa *Itxaron denboraren LST interpolazio bidezko hurbilketa DROS politikapean ondokoa da*

$$\begin{aligned} \widetilde{W}_k^{DROS,INT}(\lambda, u) &= (1 - \rho)^2 + \lambda(1 - \rho) \left(-\mathbb{E}[B] + \sum_{i=1}^K \alpha_i \frac{1 - B_i^*((1 - \rho)^{-1}u)}{(1 - \rho)^{-1}u} \right) \\ &\quad + \rho^2 \frac{\nu(\vec{p})p_k}{(1 - \rho)^{-1}u} e^{p_k \frac{\nu(\vec{p})}{(1 - \rho)^{-1}u}} \int_{p_k \frac{\nu(\vec{p})}{(1 - \rho)^{-1}u}}^{\infty} \frac{e^{-l}}{l} dl, \end{aligned}$$

non $\nu(\vec{p})$ (5.15) Ekuazioan ematen den.

Frogapena: Frogapena 2.3.1 Proposizioa erabiliz lortzen da zuzenean, (6.6) Ekuazioan lorturiko trafiko geldopeko emaitza $G^{HT}(u)$ eta 6.2.2 Lema lorturiko trafiko arinpeko emaitza konbinatuz. Ikusi frogapena 6.A.4 Eranskinean. ■



6.1. Irudia: 1 Eszenatokia. Errore erlatiboa RP politikapean bezero kopuruaren lehenengo eta bigarren momentuentzat, hiperesponentzial, esponentzial eta Pareto beharrezko zerbitzuen banaketentzat.

6.3 Alderaketa numerikoa

Atal honetan 6.1.1 Proposizioan lorturiko ilara luzeraren probabilitate funtzio sortzailearen interpolazio bidezko hurbilketaren zehaztasun maila neurtzen dugu.

Zehaztasun maila neurtzeko, Kim et al.-ek lorturiko (5.16) Ekuazioa erabiltzen dugu, zeina, zoritxarrez, ezin den analitikoki ebatzi edozein λ -rentzat, baina [27, 3.2. Atalean] artikuluan autoreek momentuak numerikoki nola kalkulatu azaltzen dute beharrezko zerbitzu banaketa orokorrerako. Beraien emaitzak $(*)^{KIM}$ bidez adieraziko ditugu non $*$ aztertutako metrika den.

3.7. eta 4.6. Ataletan aurkeztu ditugu erabiliko ditugun beharrezko zerbitzuen banaketak. Berriz ere, neurrak egiteko errore erlatiboa erabiliko dugu. Bezeroen lehenengo eta bigarren momentuentzat ondokoa kalkulatu dugu $100\% \times \frac{\mathbb{E}[N^{RP}] - \mathbb{E}[N^{RP,INT}]}{\mathbb{E}[N^{RP}]}$ eta $100\% \times \frac{\mathbb{E}[(N^{RP})^2] - \mathbb{E}[(N^{RP,INT})^2]}{\mathbb{E}[(N^{RP})^2]}$, hurrenez hurren. Bezero kopuruan jarriko dugu arreta interpolazio bidezko hurbilketaren errore numerikoa handiagoa baita kasu honetan itxaron denboraren kasuan baino.

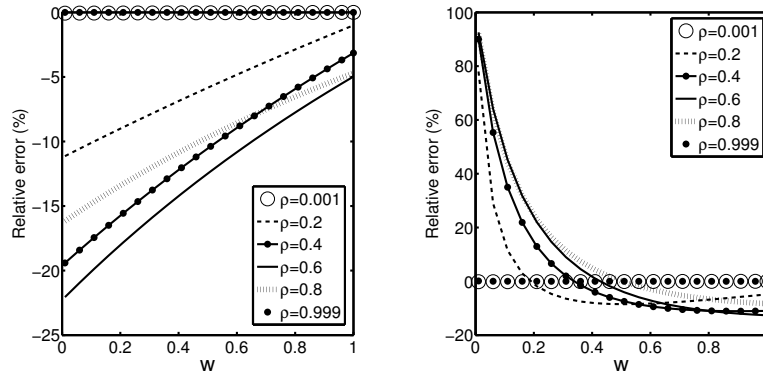
6.3.1 Interpolazio bidezko hurbilketaren zehaztasun maila aztertzen

Jarraian 6.1.1 Proposizioan lorturiko interpolazio bidezko hurbilketaren zehaztasun maila aztertuko dugu (6.3) eta (6.5) Ekuazioetan lorturiko lehenengo eta bigarren momentuak haintzakotzat hartuz.

1 *Eszenatokia*. 6.1 Irudian bezero kopuru totalaren lehenengo eta bigarren momentuen errore erlatiboak irudikatzen ditugu kargarekiko hiperesponentzial, esponentzial eta Pareto beharrezko zerbitzuen banaketentzat. Bi klase kontsideratzen ditugu $\mathbb{E}[B_1] = 11/3$, $\mathbb{E}[B_2] = 44/3$ batezbesteko beharrezko zerbitzuekin. Iritsiera bat 1 klasekoa (2 klasekoa) dela onartzen dugu $\alpha_1 = 8/12$ ($\alpha_2 = 4/12$) probabilitatearekin. $p_1 = 2$, $p_2 = 5$ pixuak finkatzen ditugu.

Lehenengo momentua nahikoa zehatza dela ikusten dugu hiru banaketentzat, baina bigarren momentuak Pareto banaketa duen beharrezko zerbitzuaren kasuan %30-eko errore erlatibo absolutua erakusten du. Pareto banaketak (2.2) Ekuazioko baldintza ez betetzeak azal dezake gertaera hau.

2 *Eszenatokia*. 6.2 Irudian bi bezero klase hartzen ditugu haintzat. 1 klaseko bezeroek μ_1



6.2. Irudia: 2 Eszenatiokia. RP ereduarentzat bezero kopuru totalaren lehenengo eta bigarren momentuen errore erlatiboa kargaren balio desberdinetarako.

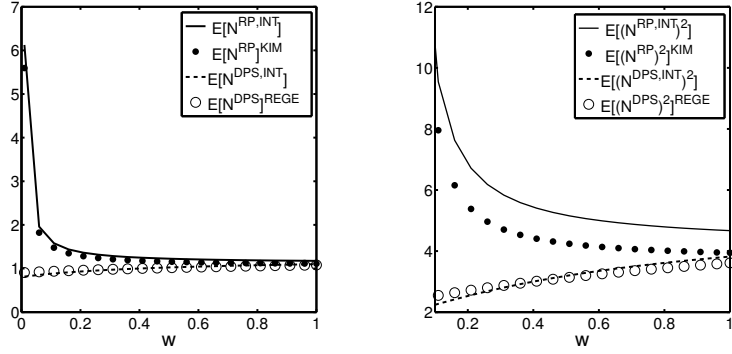
tasako banaketa esponentziala jarraitzen dute. 2 klaseko bezeroek, aldiz, (3.7) Ekuazioan definituriko banaketa hiperesponentzial degeneratibo bat ondoko parametroekin: $m_2 = 2$, $\beta_{21} = w$, $\beta_{22} = 1 - w$, $\mathbb{E}[B_{21}] = 1/(\mu_2 w)$ eta $\mathbb{E}[B_{22}] = 0$.

Kontsidera ditzagun $p_1 = 2$, $p_2 = 5$, $\alpha_1 = 7/12$, $\alpha_2 = 5/12$, $\mathbb{E}[B_1] = 11/3$, $\mathbb{E}[B_2] = 44/3$. 6.2 Irudian bezero kopuru totalaren lehenengo eta bigarren momentuen errore erlatiboak irudikatzen ditugu w parametroarekiko. Ohartu, espero bezala, interpolazio bidezko hurbilketa zehatza dela $\rho \approx 0$ eta $\rho \approx 1$ kasuetan. Errore absolutu handiena kargaren erdiko balioetarako gertatzen da w zeroruntz hurbiltzean.

6.3.2 RP versus DPS

6.3 Irudian bezero kopuru totalaren lehen eta bigarren momentuak irudikatzen ditugu RP eta DPS ereduarentzat. Gogoratu, zehaztasun maila neurtzeko, RP ereduarentzat Kim et al.-ek [27, 3.2. Atala] artikuluan aurkezturiko algoritmoa erabiltzen dugula, zeinak ahalbidetzen digun momentuak edozein beharrezko zerbitzuren banaketarentzat lortzen. Beraien emaitzak $(*)^{KIM}$ bezala adieraziko ditugu, non $*$ -k aztertutako metrika adierazten duen. DPS ereduarentzat Rege et al.-ek [40, 1. Atala] artikuluan aurkezturiko algoritmoa erabiltzen dugu, zeina beharrezko zerbitzuek banaketa esponentziala duten kasuentzat den erabilgarria soilik. Beraien emaitzak $(*)^{REGE}$ bezala adieraziko ditugu, non, berriz ere, $*$ -k aztertutako metrika adierazten duen.

3 Eszenatokia. Bi klase kontsideratzen ditugu. 1 klaseak $\mathbb{E}[B_1] = 5$ batezbesteko banaketa esponentziala jarraitzen du eta 2 klaseak, aldiz, $\mathbb{E}[B_2] = 2$ batezbesteko banaketa hiperesponentzial degeneratua. Iritsiera bat 1 klasekoa (2 klasekoa) dela onartzen dugu $\alpha_1 = 8/12$ ($\alpha_2 = 4/12$) probabilitatearekin. DPS eta RP ereduarentzat pixu berberak finkatzen ditugu, $g_1 = p_1 = 5$ eta $g_2 = p_2 = 1$. 6.3 Irudian gure hurbilketa nahikoa zehatza dela ikus daiteke. Ikus daiteke $w \rightarrow 0$ kasuan, hau da, bariazio koefizientea handia denean, DPS ereduaren errendimendua RP ereduarena baino hobea dela, gure hurbilketarentzat eta baita emaitza zehatzarentzat ere. Uste genuen gertaera bat da, izan ere, $w \rightarrow 0$ kasuan 2 klaseko bigarren momentua ∞ -runtz konbergitzen du, eta, horregatik, RP ereduaren errendimendua DPS ereduarena baino okerragoa da.



6.3. Irudia: 3 Eszenatokia. DPS eta RP ereduentzat bezero kopuru totalaren lehenengo eta bigarren momentuak, beharrezko zerbitzuen banaketa esponentzialarentzat.

6.A Eranskina

6.A.1 6.1.2 Lemaren frogapena

Ilara luzeraren probabilitate funtzio sortzailearen, $\psi^{RP}(\lambda, \vec{z})$, zero, bat eta bi deribatuak lortuko ditugu.

(5.16) Ekuaziotik lortzen dugu zero deribatua $\lambda = 0$ puntuan ondokoa dela

$$(\psi^{RP})^{(0)}(\lambda, \vec{z}) \Big|_{\lambda=0} = 1. \quad (6.7)$$

(5.16) Ekuazioa deribatuz lehenengo trafiko arinpeko deribatuaren ondoko adierazpena lortzen dugu.

$$\begin{aligned}
 (\psi^{RP})^{(1)}(0, \vec{z}) &= \frac{\partial \psi^{RP}(0, \vec{z})}{\partial \lambda} \\
 &= -\mathbb{E}[B] + \frac{1}{1 - \sum_{k=1}^K \alpha_k z_k} \sum_{i=1}^K \alpha_i z_i \left(\left(-\mathbb{E}[B] + \frac{p_i}{\alpha_i} \frac{\partial^2 r(\lambda, \vec{z})}{\partial \lambda \partial z_i} \right) \left(1 - B_i^* \left(\lambda - \lambda \sum_{k=1}^K \alpha_k z_k \right) \right) \right. \\
 &\quad \left. + \left(1 - \rho + \frac{p_i}{\alpha_i} \frac{\partial r(\lambda, \vec{z})}{\partial z_i} \right) \left(-B_i^{*'} \left(\lambda - \lambda \sum_{k=1}^K \alpha_k z_k \right) \left(1 - \sum_{k=1}^K \alpha_k z_k \right) \right) \right) \Big|_{\lambda=0} \\
 &= -\mathbb{E}[B] + \sum_{i=1}^K \alpha_i \mathbb{E}[B_i] z_i \left(1 + \frac{p_i}{\alpha_i} \frac{\partial r(\lambda, \vec{z})}{\partial z_i} \Big|_{\lambda=0} \right) \\
 &= -\mathbb{E}[B] + \sum_{i=1}^K \alpha_i \mathbb{E}[B_i] z_i. \quad (6.8)
 \end{aligned}$$

Azken pausoan ondokoa erabili dugu

$$\begin{aligned}
 \left. \frac{\partial r(\lambda, \vec{z})}{\partial z_i} \right|_{\lambda=0} &= \mathbb{E} \left[\frac{Q_i}{\sum_{k=1}^K Q_k p_k} \cdot \frac{z_1^{Q_1} \cdot \dots \cdot z_K^{Q_K}}{z_i} \cdot \mathbf{1}_{(\sum_{k=1}^K Q_k > 0)} \right] \Big|_{\lambda=0} \\
 &= \mathbb{P}(\sum Q_k > 0) \Big|_{\lambda=0} \mathbb{E} \left[\frac{Q_i}{\sum_{k=1}^K Q_k p_k} \cdot \frac{z_1^{Q_1} \cdot \dots \cdot z_K^{Q_K}}{z_i} \Big| \sum_{k=1}^K Q_k > 0 \right] \Big|_{\lambda=0} \\
 &= \rho \Big|_{\lambda=0} \mathbb{E} \left[\frac{Q_i}{\sum_{k=1}^K Q_k p_k} \cdot \frac{z_1^{Q_1} \cdot \dots \cdot z_K^{Q_K}}{z_i} \Big| \sum_{k=1}^K Q_k > 0 \right] \Big|_{\lambda=0} \\
 &= 0,
 \end{aligned} \tag{6.9}$$

RP eredua lana kontserbatzen duen politika denez, $\mathbb{P}(\sum Q_k > 0) = \rho$ zerbitzaria lanpetua ego-
tearen baliokidea da, zeina erabilitako politikaren independentea den. Beste terminoa finitua da
ondokoa betetzen baitu

$$\begin{aligned}
 &\mathbb{E} \left[\frac{Q_i}{\sum_{k=1}^K Q_k p_k} \cdot \frac{z_1^{Q_1} \cdot \dots \cdot z_K^{Q_K}}{z_i} \Big| \sum_{k=1}^K Q_k > 0 \right] \Big|_{\lambda=0} \\
 &= \sum_{\substack{q_1=0, \dots, q_K=0 \\ \sum_{k=1}^K q_k > 0}}^{\infty} \frac{q_i}{\sum_{k=1}^K q_k p_k} \cdot \frac{z_1^{q_1} \cdot \dots \cdot z_K^{q_K}}{z_i} \mathbb{P} \left(Q_1 = q_1, \dots, Q_K = q_K \Big| \sum_{k=1}^K Q_k > 0 \right) \Big|_{\lambda=0} \\
 &= \sum_{\substack{q_1=0, \dots, q_K=0 \\ \sum_{k=1}^K q_k = 1}}^{\infty} \frac{q_i}{\sum_{k=1}^K q_k p_k} \cdot \frac{z_1^{q_1} \cdot \dots \cdot z_K^{q_K}}{z_i} \mathbb{P} \left(Q_1 = q_1, \dots, Q_K = q_K \Big| \sum_{k=1}^K Q_k > 0 \right) \Big|_{\lambda=0} + o(\lambda^2) \\
 &= \sum_{\substack{q_1=0, \dots, q_K=0 \\ \sum_{k=1}^K q_k = 1}}^{\infty} \frac{q_i}{q_i p_i} \cdot 1 \cdot \dots \cdot 1 \cdot \mathbb{P} \left(\vec{Q} = e_i \Big| \sum_{k=1}^K Q_k > 0 \right) \Big|_{\lambda=0} = \frac{\alpha_i}{p_i},
 \end{aligned} \tag{6.10}$$

izan ere,

$$\begin{aligned}
 \mathbb{P} \left(\vec{Q} = e_i \Big| \sum_{k=1}^K Q_k > 0 \right) \Big|_{\lambda=0} &= \frac{\mathbb{P}(\vec{Q} = e_i \cap \sum_{k=1}^K Q_k > 0)}{\mathbb{P}(\sum_{k=1}^K Q_k > 0)} \Big|_{\lambda=0} = \frac{\mathbb{P}(\vec{Q} = e_i)}{\mathbb{P}(\sum_{k=1}^K Q_k > 0)} \Big|_{\lambda=0} \\
 &= \frac{\alpha_i \rho (1 - \rho)}{\rho} \Big|_{\lambda=0} = \alpha_i
 \end{aligned}$$

zeina betetzen den

$$A \cdot \mathbb{P}(\vec{Q} = \vec{0}) = B \cdot \mathbb{P}(\vec{Q} = e_i) + o(\lambda),$$

non $A = \alpha_i \rho$ eta $B = 1$ [26, 1 Ekuazio] artikulutik lortzen diren eta, berriz ere, RP eredua
kontserbakorra delako.

Bigarren deribatuak ondokoa betetzen du

$$\begin{aligned}
(\psi^{RP})^{(2)}(\lambda, \vec{z}) \Big|_{\lambda=0} &= \frac{\partial^2 \psi^{RP}(\lambda, \vec{z})}{\partial \lambda^2} \Big|_{\lambda=0} \\
&= \frac{1}{1 - \sum_{k=1}^K \alpha_k z_k} \sum_{i=1}^K \alpha_i z_i \left(\frac{p_i}{\alpha_i} \frac{\partial^3 r(\lambda, \vec{z})}{\partial \lambda^2 \partial z_i} \left(1 - B_i^* \left(\lambda - \lambda \sum_{k=1}^K \alpha_k z_k \right) \right) \right. \\
&\quad + 2 \left(-\mathbb{E}[B] + \frac{p_i}{\alpha_i} \frac{\partial^2 r(\lambda, \vec{z})}{\partial \lambda \partial z_i} \right) \left(-B_i^{*'} \left(\lambda - \lambda \sum_{k=1}^K \alpha_k z_k \right) \right) \left(1 - \sum_{k=1}^K \alpha_k z_k \right) \\
&\quad \left. + \left(1 - \rho + \frac{p_i}{\alpha_i} \frac{\partial r(\lambda, \vec{z})}{\partial z_i} \right) \left(-B_i^{*''} \left(\lambda - \lambda \sum_{k=1}^K \alpha_k z_k \right) \right) \left(1 - \sum_{k=1}^K \alpha_k z_k \right)^2 \right) \Big|_{\lambda=0} \\
&= \sum_{i=1}^K \alpha_i z_i \left(2 \left(-\mathbb{E}[B] + \frac{p_i}{\alpha_i} \frac{\partial^2 r(\lambda, \vec{z})}{\partial \lambda \partial z_i} \Big|_{\lambda=0} \right) \mathbb{E}[B_i] \right. \\
&\quad \left. - \left(1 + \alpha_i p_i \frac{\partial r(\lambda, \vec{z})}{\partial z_i} \Big|_{\lambda=0} \right) \mathbb{E}[B_i^2] \left(1 - \sum_{k=1}^K \alpha_k z_k \right) \right) \\
&= \sum_{i=1}^K \alpha_i z_i \mathbb{E}[B_i^2] \left(\sum_{k=1}^K \alpha_k z_k - 1 \right), \tag{6.11}
\end{aligned}$$

non azken pausoan ondokoa erabiltzen dugun

$$\begin{aligned}
\frac{\partial^2 r(\lambda, \vec{z})}{\partial \lambda \partial z_i} \Big|_{\lambda=0} &= \frac{\partial \left(\rho \cdot \mathbb{E} \left[\frac{Q_i}{\sum_{k=1}^K Q_k p_k} \cdot \frac{z_1^{Q_1} \dots z_K^{Q_K}}{z_i} \mid \sum_{k=1}^K Q_k > 0 \right] \right)}{\partial \lambda} \Big|_{\lambda=0} \\
&= \mathbb{E}[B] \mathbb{E} \left[\frac{Q_i}{\sum_{k=1}^K Q_k p_k} \cdot \frac{z_1^{Q_1} \dots z_K^{Q_K}}{z_i} \mid \sum_{k=1}^K Q_k > 0 \right] \Big|_{\lambda=0} \\
&\quad + \rho \Big|_{\lambda=0} \cdot \frac{\partial \mathbb{E} \left[\frac{Q_i}{\sum_{k=1}^K Q_k p_k} \cdot \frac{z_1^{Q_1} \dots z_K^{Q_K}}{z_i} \mid \sum_{k=1}^K Q_k > 0 \right]}{\partial \lambda} \Big|_{\lambda=0} \\
&= \mathbb{E}[B] \frac{\alpha_i}{p_i},
\end{aligned}$$

(6.10) Ekuaziotik betetzen dena.

(6.7), (6.8) eta (6.11) Ekuazioetatik 6.1.2 Leman erakutsitako emaitza lortzen dugu.

6.A.2 6.1.1 Proposizioaren frogapena

Ilara luzeraren probabilitate funtzio sortzailearen interpolazio bidezko hurbilketa lortzen dugu RP ereduarentzat.

2.3. Atalean azaldu bezala $G(\lambda, \vec{z}^{(1-\rho)}) = \psi^{RP}(\lambda, \vec{z}^{(1-\rho)})$ funtzioa ondoko polinomioaz hurbiltzen dugu

$$\hat{G}(\lambda, \vec{z}) = h_0(\vec{z}) + \lambda h_1(\vec{z}) + \lambda^2 h_2(\vec{z}) + \lambda^3 h_3(\vec{z}).$$

Normalizazioa deseginez, hau da, $f_\lambda^{-1} = \bar{z}^{(1-\rho)^{-1}}$ funtzioa kontsideratuz, ondokoa dugu

$$\begin{aligned}\psi^{RP,INT}(\lambda, \bar{z}) &= \hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}}) \\ &= h_0(\bar{z}^{(1-\rho)^{-1}}) + \lambda h_1(\bar{z}^{(1-\rho)^{-1}}) + \lambda^2 h_2(\bar{z}^{(1-\rho)^{-1}}) + \lambda^3 h_3(\bar{z}^{(1-\rho)^{-1}}).\end{aligned}$$

Orduan, (2.12) trafikoa-arin baldintzetatik $h_0(\bar{z}), h_1(\bar{z}), h_2(\bar{z})$ lortzen ditugu. Lehenik eta behin, ondokoa dugu

$$\hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}}) \Big|_{\lambda=0} = \hat{G}(0, \bar{z}) = h_0(\bar{z}).$$

(6.7) Ekuazioarekin batera $h_0(\bar{z}) = 1$ lortzen dugu.

Bigarrenez,

$$\begin{aligned}\frac{d\hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{d\lambda} \Big|_{\lambda=0} &= \frac{\partial \hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{\partial \lambda} \Big|_{\lambda=0} + \sum_{i=1}^K \frac{\partial \hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{\partial z_i} \Big|_{\lambda=0} \cdot \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \Big|_{\lambda=0} \\ &= \left(h_1(\bar{z}^{(1-\rho)^{-1}}) + 2\lambda h_2(\bar{z}^{(1-\rho)^{-1}}) + 3\lambda^2 h_3(\bar{z}^{(1-\rho)^{-1}}) \right) \Big|_{\lambda=0} \\ &\quad + \sum_{i=1}^K \left(\frac{dh_0(\bar{z}^{(1-\rho)^{-1}})}{dz_i} + \lambda \frac{dh_1(\bar{z}^{(1-\rho)^{-1}})}{dz_i} + \lambda^2 \frac{dh_2(\bar{z}^{(1-\rho)^{-1}})}{dz_i} + \lambda^3 \frac{dh_3(\bar{z}^{(1-\rho)^{-1}})}{dz_i} \right) \Big|_{\lambda=0} \\ &\quad \cdot \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \Big|_{\lambda=0} \\ &= h_1(\bar{z}) + \sum_{i=1}^K \frac{dh_0(\bar{z}^{(1-\rho)^{-1}})}{dz_i} \Big|_{\lambda=0} \cdot \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \Big|_{\lambda=0} \\ &= h_1(\bar{z}) + \sum_{i=1}^K \frac{d(1)}{dz_i} \Big|_{\lambda=0} \cdot \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \Big|_{\lambda=0} = h_1(\bar{z}).\end{aligned}$$

(6.8) Ekuazioarekin batera ondokoa lortzen dugu $h_1(\bar{z}) = -\mathbb{E}[B] + \sum_{i=1}^K \alpha_i \mathbb{E}[B_i] z_i$.

Hirugarrenez,

$$\begin{aligned}\frac{d^2 \hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{d\lambda^2} \Big|_{\lambda=0} &= \frac{\partial^2 \hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{\partial \lambda^2} \Big|_{\lambda=0} + \sum_{i=1}^K \frac{\partial (\partial \hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}}) / \partial \lambda)}{\partial z_i} \Big|_{\lambda=0} \cdot \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \Big|_{\lambda=0} \\ &\quad + \sum_{i=1}^K \left(\left(\frac{\partial (\partial \hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}}) / \partial z_i)}{\partial \lambda} \Big|_{\lambda=0} + \frac{\partial (\partial \hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}}) / \partial z_i)}{\partial z_i} \Big|_{\lambda=0} \cdot \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \Big|_{\lambda=0} \right) \right. \\ &\quad \left. \cdot \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \Big|_{\lambda=0} + \frac{\partial \hat{G}(\lambda, \bar{z}^{(1-\rho)^{-1}})}{\partial z_i} \Big|_{\lambda=0} \cdot \frac{d^2(z_i^{(1-\rho)^{-1}})}{d\lambda^2} \Big|_{\lambda=0} \right) \\ &= 2h_2(\bar{z}) + 2 \sum_{i=1}^K \frac{dh_1(\bar{z})}{dz_i} \cdot \frac{d(z_i^{(1-\rho)^{-1}})}{d\lambda} \Big|_{\lambda=0} = 2h_2(\bar{z}) + 2\mathbb{E}[B] \sum_{i=1}^K \alpha_i \mathbb{E}[B_i] z_i \ln(z_i).\end{aligned}$$

(6.11) Ekuazioarekin batera ondokoa lortzen dugu

$$h_2(\vec{z}) = \frac{1}{2} \left(\sum_{i=1}^K \alpha_i z_i \mathbb{E}[B_i^2] \left(\sum_{k=1}^K \alpha_k z_k - 1 \right) - 2\mathbb{E}[B] \sum_{i=1}^K \alpha_i z_i \mathbb{E}[B_i] \ln(z_i) \right).$$

Azkenik, 2.3.1 Proposiziotik, (6.1) Ekuazioarekin batera, frogapena amaitzen dugu.

6.A.3 6.2.2 Lemaren frogapena

l klaseko itxaron denboraren Laplace transformatuaren zero eta bat trafiko arinpeko deribatuak lortzen ditugu DROS ereduarentzat 5.4.1 Teoreman aurkezturiko emaitzak erabiliz.

Zerogarren deribatuak ondokoa betetzen du

$$\left(\widetilde{W}_k^{DROS} \right)^{(0)}(\lambda, u) \Big|_{\lambda=0} = [1 - \rho + T_l(u, \vec{1})] \Big|_{\lambda=0} = 1 + T_l(u, \vec{1}) \Big|_{\lambda=0} = 1, \quad (6.12)$$

izan ere, (5.26) Ekuaziotik ondokoa lortzen dugu

$$T_l(u, \vec{1}) \Big|_{\lambda=0} \left(1 + \sum_{i=1}^K \frac{p_i}{p_l} (1 - B_i^*(u)) \right) = 0 \Rightarrow T_l(u, \vec{1}) \Big|_{\lambda=0} = 0.$$

Eta lehenengo deribatuak ondokoa betetzen du

$$\left(\widetilde{W}_k^{DROS} \right)^{(1)}(\lambda, u) \Big|_{\lambda=0} = -\mathbb{E}[B] + \frac{\partial T_l(u, \vec{1})}{\partial \lambda} \Big|_{\lambda=0} = -\mathbb{E}[B] + \left(\sum_{i=1}^K \alpha_i \frac{1 - B_i^*(u)}{u} \right), \quad (6.13)$$

izan ere, berriz ere, (5.26) Ekuaziotik ondokoa dugu

$$\begin{aligned} & \frac{\partial T_l(u, \vec{1})}{\partial \lambda} \Big|_{\lambda=0} \left(1 + \sum_{i=1}^K \frac{p_i}{p_l} \mathbb{E}[Q_i^*] \Big|_{\lambda=0} (1 - B_i^*(u)) \right) = \sum_{i=1}^K \left(\alpha_i + p_i r(\lambda, \vec{1}) \Big|_{\lambda=0} \right) \frac{1 - B_i^*(u)}{u} \\ & \Rightarrow \frac{\partial T_l(u, \vec{1})}{\partial \lambda} \Big|_{\lambda=0} = \left(\sum_{i=1}^K \alpha_i \frac{1 - B_i^*(u)}{u} \right), \end{aligned}$$

non $\mathbb{E}[Q_i^*] \Big|_{\lambda=0} = 0$ eta $r(\lambda, \vec{1}) \Big|_{\lambda=0} = 0$, (6.9) Ekuaziotik.

6.A.4 6.2.1 Proposizioaren frogapena

Itxaron denboraren interpolazio bidezko hurbilketa RP ereduarentzat. 2.3. Atalean azaldu bezala $G(\lambda, (1 - \rho)u) = \widetilde{W}_k^{DROS}(\lambda, (1 - \rho)u)$ funtzioa ondoko polinomioarekin hurbiltzen dugu

$$\hat{G}(\lambda, u) = h_0(u) + \lambda h_1(u) + \lambda^2 h_2(u).$$

Normalizazioa deseginez, hau da, $f_\lambda^{-1} = (1 - \rho)^{-1}u$ funtzioa kontsideratuz, ondokoa dugu

$$\begin{aligned} \widetilde{W}_k^{DROS, INT}(\lambda, u) &= \hat{G}(\lambda, (1 - \rho)^{-1}u) \\ &= h_0((1 - \rho)^{-1}u) + \lambda h_1((1 - \rho)^{-1}u) + \lambda^2 h_2((1 - \rho)^{-1}u). \end{aligned}$$

Orduan, (2.12) trafiko-arin baldintzetatik $h_0(u), h_1(u)$ lortzen ditugu. Lehenik eta behin ondokoa dugu

$$\hat{G}(\lambda, (1-\rho)^{-1}u) \Big|_{\lambda=0} = \hat{G}(0, u) = h_0(u).$$

(6.12) Ekuazioaren bitartez $h_0(u) = 1$ lortzen dugu.

Bigarrenez,

$$\begin{aligned} & \frac{d\hat{G}(\lambda, (1-\rho)^{-1}u)}{d\lambda} \Big|_{\lambda=0} \\ &= \frac{d\hat{G}(\lambda, (1-\rho)^{-1}u)}{d\lambda} \Big|_{\lambda=0} + \frac{d\hat{G}(\lambda, (1-\rho)^{-1}u)}{du} \Big|_{\lambda=0} \cdot \frac{d((1-\rho)^{-1}u)}{d\lambda} \Big|_{\lambda=0} \\ &= (h_1((1-\rho)^{-1}u) + 2\lambda h_2((1-\rho)^{-1}u)) \Big|_{\lambda=0} \\ &+ \left(\frac{dh_0((1-\rho)^{-1}u)}{du} + \lambda \frac{dh_1((1-\rho)^{-1}u)}{du} + \lambda^2 \frac{dh_2((1-\rho)^{-1}u)}{du} \right) \Big|_{\lambda=0} \cdot \frac{d((1-\rho)^{-1}u)}{d\lambda} \Big|_{\lambda=0} \\ &= h_1(u) + \frac{dh_0((1-\rho)^{-1}u)}{du} \Big|_{\lambda=0} \cdot \frac{d((1-\rho)^{-1}u)}{d\lambda} \Big|_{\lambda=0} \\ &= h_1(u) + \frac{d(1)}{du} \Big|_{\lambda=0} \cdot \frac{d((1-\rho)^{-1}u)}{d\lambda} \Big|_{\lambda=0} = h_1(u). \end{aligned}$$

(6.13) Ekuazioarekin batera ondokoa lortzen dugu $h_1(u) = -\mathbb{E}[B] + \left(\sum_{i=1}^K \alpha_i \frac{1 - B_i^*(u)}{u} \right)$.

Azkenik, 2.3.1 Proposiziotik, (6.6) Ekuazioarekin batera, frogapena amaitzen dugu.

Bibliografia

- [1] E. Altman, K. Avrachenkov, and U. Ayesta. A survey on discriminatory processor sharing. *Queueing systems*, 53(1-2):53–63, 2006.
- [2] S. Asmussen. *Applied Probability and Queues*. Springer, 2003.
- [3] U. Ayesta, A. Izagirre, and I.M. Verloop. Heavy-traffic analysis of the discriminatory random-order-of-service discipline. *ACM SIGMETRICS Performance Evaluation Review*, 32(2):41–43, 2011.
- [4] M. Harchol Balter. Performance modeling and design of computer systems: Queueing theory in action. In *Cambridge University Press, Cambridge (UK)*, 2013.
- [5] P. Billingsley. *Convergence of Probability Measures*. Wiley, 1999.
- [6] S.C. Borst, R. Núñez-Queija, and A.P. Zwart. Sojourn time asymptotics in processor sharing queues. *Queueing Systems*, 53(1-2):31–51, 2006.
- [7] M. Bramson and J.G. Dai. Heavy traffic limits for some queueing networks. *Annals of Applied Probability*, 11:49–90, 2001.
- [8] D. Burman and D. Smith. An asymptotic analysis of a queueing system with markov-modulated arrivals. *Operations Research*, 34:105–119, 1986.
- [9] J.W. Cohen. The multiple phase service network with generalized processor sharing. *Acta Informatica*, 12:245–284, 1979.
- [10] G. Fayolle, I. Mitrani, and R. Iasnogorodski. Sharing a processor among many job classes. *Journal of the ACM*, 27(3):519–532, 1980.
- [11] W. Feller. *An Introduction to Probability Theory and Its Applications, Vol. II*. Wiley, New York, 1971.
- [12] P.J. Fleming and B. Simon. Interpolation approximations of sojourn time distributions. *Operations Research*, 39(2):251–260, 1991.
- [13] E. Gelenbe and I. Mitrani. *Analysis and Synthesis of Computer Systems*. Imperial College Press, 2010.
- [14] S. Grishechkin. On a relationship between processor sharing queues and Crump-Mode-Jagers branching processes. *Adv. Appl. Prob.*, 24(3):653–698, 1992.

- [15] V. Gupta, M. Harchol-Balter, K. Sigman, and W. Whitt. Analysis of join-the-shortest-queue routing for web server farms. In *Proceedings of Performance*, 2007.
- [16] M. Haviv and J. van der Wal. Equilibrium strategies for processor sharing and random queues with relative priorities. *Probability in the Engineering and Informational Sciences*, 11:403–412, 1997.
- [17] M. Haviv and J. van der Wal. Waiting times in queues with relative priorities. *Operations Research Letters*, 35:591–594, 2007.
- [18] A. Izagirre, U. Ayesta, and I.M. Verloop. Sojourn time approximations for a discriminatory-processor-sharing queue. *To appear in Transactions on Modeling and Performance Evaluation of Computing Systems*.
- [19] A. Izagirre, U. Ayesta, and I.M. Verloop. Sojourn time approximations in a multi-class time-sharing system. *IEEE INFOCOM*, 2014.
- [20] A. Izagirre, U. Ayesta, and I.M. Verloop. Heavy-traffic analysis of a multi-class queue with relative priorities. *Probability in Engineering and Informational Sciences*, 29(2):153–180, 2015.
- [21] A. Izagirre, U. Ayesta, and I.M. Verloop. Interpolation approximations for the steady-state distribution in multi-class resource-sharing systems. *Performance Evaluation*, <http://dx.doi.org/10.1016/j.peva.2015.06.005>, 2015.
- [22] A. Izagirre and A.M. Makowski. Light traffic performance under the power of two load balancing strategy: the case of server heterogeneity. *ACM SIGMETRICS Performance Evaluation Review*, 42:18–20, 2014.
- [23] F.P. Kelly. *Stochastic Networks and Reversibility*. Wiley, Chichester, 1979.
- [24] F.P. Kelly. Charging and rate control for elastic traffic. *European Transactions on Telecommunications*, 8:33–37, 1997.
- [25] D.G. Kendall. Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded markov chains. *Annals of Mathematical Statistics*, 24:338–354, 1953.
- [26] J. Kim. Queue length distribution in a queue with relative priorities. *Bull. Korean Math. Soc.*, 46:107–116, 2009.
- [27] J. Kim, J. Kim, and B. Kim. Analysis of the M/G/1 queue with discriminatory random order service policy. *Performance Evaluation*, 68(3):256–270, 2011.
- [28] J.F.C. Kingman. The single server queue in heavy traffic. *Proc. Cambridge Philos. Soc.*, 57:902–904, 1961.
- [29] J.F.C. Kingman. On queues in which customers are served in random order. *Proc. Cambridge Philos. Soc.*, 58:79–91, 1962.
- [30] J.F.C. Kingman. Queue disciplines in heavy traffic. *Mathematics of Operations Research*, 7(2):262–271, 1982.

- [31] L. Kleinrock. Analysis of a time-shared processor. *Naval Research Logistics Quarterly*, 11:59–73, 1964.
- [32] L. Kleinrock. Time-shared systems: A theoretical treatment. *Journal of the ACM*, 14(2):242–261, 1967.
- [33] J.D.C. Little. A proof for the queueing formula: $L=\lambda w$. *Operations Research*, 9(3):383–387, 1961.
- [34] A.W. Marshall and I. Olkin. Inequalities: Theory of majorization and its applications. In *Academic Press, San Diego (CA)*, 1979.
- [35] M. Mitzenmacher. The power of two choices in randomized load balancing. *IEEE Transactions on Parallel and Distributed Systems*, 12(10):1094–1104, 2001.
- [36] A. Mukhopadhyay and R.R. Mazumdar. Analysis of load balancing in large heterogeneous processor sharing systems. *Submitted on 2013*.
- [37] R.L. Dobrushin N.D. Vvedenskaya and F.I. Karpelevich. Queueing system with selection of the shortest of two queues: An asymptotic approach. *Problems of Information Transmission*, 32:20–34, 1996.
- [38] R.Ñúñez-Queija. *Processor-Sharing Models for Integrated-Services Networks*. PhD thesis, Eindhoven University of Technology, 2000.
- [39] C. Palm. Waiting times with random served queue. *Tele1 (English edition; original 1938)*, 1–107, 1957.
- [40] K.M. Rege and B. Sengupta. Queue length distribution for the discriminatory processor sharing queue. *Operations Research*, 44(4):653–657, 1996.
- [41] M.I. Reiman and B. Simon. An interpolation approximation for queueing systems with Poisson input. *Operations Research*, 36:454–469, 1988.
- [42] M.I. Reiman and B. Simon. Light traffic limits of sojourn time distributions in Markovian queueing networks. *Stochastic Models*, 4:191–233, 1988.
- [43] M.I. Reiman and B. Simon. Open queueing systems in light traffic. *Oper. Res.*, 14:26–59, 1989.
- [44] A.L. Stolyar. Pull-based load distribution in large-scale heterogeneous service systems. *Queueing Systems*, 2015.
- [45] H.C. Tijms. *A First Course in Stochastic Models*. Wiley, 2003.
- [46] G. van Kessel, R.Ñúñez-Queija, and S.C. Borst. Differentiated bandwidth sharing with disparate flow sizes. In *Proceedings of IEEE INFOCOM*, 2005.
- [47] S. Varma and A.M. Makowski. Interpolation approximations for symmetric fork-join queues. *Performance Evaluation*, 20:245–265, 1994.

- [48] I.M. Verloop, U. Ayesta, and R. Núñez-Queija. Heavy-traffic analysis of a multiple-phase network with discriminatory processor sharing. *Operations Research*, 59(3):648–660, 2011.
- [49] I.M. Verloop, U. Ayesta, and R. Núñez-Queija. Heavy-traffic analysis of a multiple-phase network with discriminatory processor sharing. *Operations Research*, 59:648–660, 2011.
- [50] J. Walrand. An introduction to queueing networks. In *Prentice-Hall, Englewood Cliffs, NJ*, 1988.
- [51] S.F. Yashkov. Processor sharing queues: Some progress in analysis. *Queueing Systems*, 2:1–17, 1987.
- [52] S.F. Yashkov. Processor sharing queues: Some progress in analysis. *Queueing Systems*, 2:1–17, 1987.
- [53] S.F. Yashkov. Mathematical problems in the theory of processor sharing queueing systems. *Journal of Soviet Mathematics*, 58:101–147, 1992.

Title: Interpolation approximations for steady-state performance measure

Abstract: The analysis of the steady-state performance in many queuing systems is complex and closed-form results are available only in particular cases. We therefore set out to develop approximations for important performance measures in steady-state such as the queue length vector, waiting time and sojourn time. We first analyse the performance in a light-traffic and heavy-traffic regime. We then show how to develop an interpolation-based approximation that is valid for any load in the system. An advantage of the approach taken is that it is not model dependent and hence could potentially be applied to other complex queuing models. We apply this technique to three widely used models in the performance evaluation of stochastic networks: The supermarket model, the Discriminatory-Processor-Sharing (DPS) queue and the Relative Priority (RP) queue. The supermarket model is a multi-server queue where upon arrival of a customer two servers are selected at random from the available pool of servers. The Join-the-Shortest-Queue policy is then used in isolation with these two servers. DPS and RP are both single-server multi-class queues that implement relative priorities among customers of the various classes. The DPS discipline serves all customers simultaneously while RP serves one customer at a time in a non-preemptive way. We show that in some instances the interpolation approximation is exact. We then use the approximation to draw structural insights onto the performance of the system, and we carry out numerical experiments that illustrate that the interpolation approximation is accurate over a wide range of parameters.

Key words: Queuing theory, heavy traffic, light traffic, interpolation approximation

Titre: Interpolation des mesures de performance à l'état stationnaire

Résumé: L'analyse de la performance à l'état stationnaire dans de nombreux systèmes de files d'attente est complexe et les résultats sous forme explicite ne sont disponibles que dans des cas particuliers. Nous avons donc développé des approximations pour des critères de performance importants à l'état stationnaire tels que la longueur de la file d'attente, le temps d'attente et le temps de traitement total. Nous analysons d'abord la performance dans des cas à faible et fort trafic. Nous montrons ensuite comment développer une approximation basée sur une interpolation qui est valable pour n'importe quelle condition de trafic. Un avantage de l'approche proposée est qu'elle n'est pas dépendante d'un modèle particulier et donc elle peut être appliquée à d'autres modèles de files d'attente complexes. Nous appliquons cette technique pour trois modèles largement utilisés dans l'évaluation des performances des réseaux stochastiques : le modèle du supermarché, la file d'attente Discriminatory-Processor-Sharing (DPS) et la file d'attente Relative Priority (RP). Le modèle du supermarché est une file d'attente à plusieurs serveurs où lorsqu'un client arrive, deux serveurs sont choisis au hasard dans un ensemble de serveurs. La politique Join-the-Shortest-Queue (JSQ) est ensuite utilisée parmi les deux serveurs sélectionnés. DPS et RP sont deux files d'attente à plusieurs classes et à serveur unique mettant en œuvre des priorités relatives entre les clients des différentes classes. La discipline DPS sert tous les clients simultanément, tandis que RP sert un seul client à la fois de manière non-préemptive. Nous montrons que dans certains cas, l'interpolation est exacte. Nous utilisons ensuite cette approximation pour déduire comment la performance dépend des paramètres des modèles, et nous effectuons des expériences numériques illustrant la précision de l'interpolation dans un grand nombre de cas de figure.

Mots-clés: Théorie des files d'attente, trafic fort, trafic léger, interpolation

Izenburua: Errendimendu metriken interpolazio bidezko hurbilketak oreka egoeran

Laburpena: Sistema askoren errendimendu metriken analisia oreka egoeran konplexua da eta karakterizazio itxiak kasu partikular batzuetan soilik daude eskuragarri. Horregatik, bezero kopurua, itxaron denbora eta sisteman igarotako denboraren hurbilketak eraiki ditugu. Lehenik eta behin, errendimendua trafiko baxu eta trafiko altu erregimenenpean aztertzen dugu. Ondoren, kargaren edozein baliorako baliagarria den interpolazio hurbilketa nola eraiki erakusten dugu. Hurbilketaren abantailetak bat ereduarekiko menpekotasunik ez duela da eta, beraz, beste edozein sistemari aplikagarria. Teknika hau hiru ilara desberdinei aplikatu diegu: Supermerkatu eredua, *Discriminatory-Processor-Sharing (DPS)* ilara eta Prioritate Erlatibodun (PE) ilara. Supermerkatu eredua zerbitzari anitzdun ilara da non iritsiera bat jazotzean bi zerbitzari zoriz aukeratuak diren. Ondoren, bi zerbitzari horietatik ilara motzena duena aukeratzen da. DPS eta PE zerbitzari bakarreko klase anitzdun ilarak dira zeinak prioritate erlatiboak ezartzen dizkiete klase desberdinetako bezeroei. DPS politikak bezero guztiak aldi berean zerbitzatzeko ditu eta, aldiz, PE politikak bezero bakarra. Kasu batzuetan interpolazio hurbilketa zehatza dela erakusten dugu. Hurbilketa erabiltzen dugu sistemako errendimenduaren inguruko ondorioak lortzeko eta interpolazio hurbilketaren zehaztasuna oso ona dela erakusten duten esperimentu numerikoak gauzatzeko ditugu.

Gako-hitzak: Ilaren toeria, trafiko geldoa, trafiko arina, interpolazio bidezko hurbilketa