



HAL
open science

Chance-Constrained Programming Approaches for Staffing and Shift-Scheduling Problems with Uncertain Forecasts: application to Call Centers

Mathilde Excoffier

► **To cite this version:**

Mathilde Excoffier. Chance-Constrained Programming Approaches for Staffing and Shift-Scheduling Problems with Uncertain Forecasts: application to Call Centers. Operations Research [math.OC]. Université Paris Sud - Paris XI, 2015. English. NNT : 2015PA112244 . tel-01321548

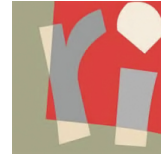
HAL Id: tel-01321548

<https://theses.hal.science/tel-01321548>

Submitted on 26 May 2016

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.



UNIVERSITÉ PARIS-SUD

ÉCOLE DOCTORALE 427
INFORMATIQUE PARIS SUD

LABORATOIRE DE RECHERCHE EN INFORMATIQUE

THÈSE DE DOCTORAT
INFORMATIQUE

par

Mathilde EXCOFFIER

**Chance-Constrained Programming Approaches
for Staffing and Shift-Scheduling Problems
with Uncertain Forecasts
Application to Call Centers**

Date de soutenance : 30/09/2015

Composition du jury :

Directeur de thèse :	Pr. Steven MARTIN	Université Paris-Sud
Président du jury :	Pr. Alain DENISE	Université Paris-Sud
Rapporteurs :	Pr. Dominique FEILLET	Ecole des Mines de Saint-Etienne
	Dr. Safia KEDAD-SIDHOUM	Université Pierre et Marie Curie
Examineurs :	Dr. Céline GICQUEL	Université Paris-Sud
	Dr. Oualid JOUINI	Ecole Centrale Paris
	Pr. Pierre L'ECUYER	Université de Montréal, Canada

Résumé

En quelques années, les centres d'appels sont devenus une structure indispensable dans de nombreuses entreprises -en particulier de service.

Ils sont souvent la principale source de contact entre les clients et les entreprises, rendant par conséquent primordiale l'évaluation de la satisfaction du client. De plus, en raison de son coût élevé, la gestion de la main d'œuvre d'un centre d'appels est un aspect critique à considérer pour assurer la rentabilité de la structure.

Le problème de dimensionnement et de planification d'agents en centre d'appels consiste à déterminer pour une période donnée le nombre d'interlocuteurs requis afin d'atteindre la qualité de service exigée et minimiser les coûts induits. Ce sujet fait l'objet d'un intérêt croissant, d'une part pour son intérêt du point de vue de la résolution théorique et d'autre part pour l'impact applicatif qu'il peut avoir. Le but de cette thèse est d'établir des approches en contraintes en probabilités en considérant l'incertitude sur la demande future. Le processus d'arrivée est un processus non stationnaire dont le taux d'arrivées est aléatoire.

La première partie de la thèse présente un modèle de résolution permettant de traiter en une seule étape la problématique complète afin de réduire l'imprécision globale de l'approche et ainsi obtenir une solution à coût réduit.

Le premier modèle décrit une approche en deux étapes que nous développons à l'aide d'une formulation d'optimisation stochastique avec contrainte en probabilité jointe afin de garantir le respect d'un niveau de risque global, contrairement à une approche en contraintes individuelles.

Le modèle en une étape se base dans sa construction sur le modèle en deux étapes en considérant une contrainte en probabilité jointe. Il est construit à l'aide de plusieurs décisions visant à diminuer les approximations : tout d'abord, les taux d'arrivées sont des variables aléatoires à loi de probabilité continue, ce qui est une considération tout à fait naturelle. De plus, une relation continue entre les taux d'arrivées et les besoins théoriques en agents est proposée afin de ne pas imposer de fortes approximations. Le problème est modélisé sous forme d'un problème d'optimisation stochastique avec contraintes jointes afin de permettre de gérer le risque sur l'horizon entier. Il est ensuite réparti sur les différentes périodes de temps. Cette répartition est alors faite pendant le processus d'optimisation plutôt qu'en pré-traitement, ce qui apporte une meilleure flexibilité dans l'approche du problème et amène à une solution au coût moins élevé qu'avec une approche à répartition décidée en avance. Cette approche est également valorisée par la considération de besoins théoriques en nombres réels plutôt qu'entiers, en permettant de réduire les effectifs sur certaines périodes tout en respectant la qualité de service imposée. Finalement, dans le but de mener à bien l'idée de continuité, les contraintes sont linéarisées de manière à obtenir un programme final rapide à résoudre par un outil d'optimisation. La linéarisation permet de proposer un encadrement (borne supérieure et borne inférieure) du coût de la solution optimale. Ces bornes respectent effectivement le niveau de risque tout en diminuant le coût de la solution par rapport à d'autres approches.

Dans la deuxième partie, le modèle en une étape est étendu pour améliorer sa représentation de la réalité. D'une part, le modèle de file d'attente est modifié afin de prendre

en compte les possibles abandons des clients qui ont atteint leur limite de patience. Ce modèle apporte une meilleure formulation mathématique ainsi qu'une meilleure considération des caractéristiques réelles de la file d'attente. D'autre part, les taux d'arrivées dans différentes périodes d'une journée sont corrélés, ce qui n'est pas considéré dans le premier modèle. Celui-ci est alors reformulé avec une nouvelle expression de l'incertitude pour prendre ce phénomène en compte dans un problème avec contrainte en probabilité jointe.

La troisième partie de la thèse aborde le problème de dimensionnement et de planification d'agents en centre d'appels en considérant une incertitude supplémentaire : la loi de probabilité des taux d'arrivées est supposée inconnue. L'approche *distributionally robust* permet de modéliser le problème en optimisant le coût final sous l'hypothèse que la loi de probabilité adéquate est inconnue et fait partie d'un ensemble de lois, défini par une moyenne et une variance données. Le problème avec l'approche *distributionally robust* est modélisé avec la même démarche que pour l'optimisation stochastique : le problème est modélisé par une contrainte en probabilité jointe, séparée en contraintes en probabilité individuelles en allouant une part du risque à chaque contrainte. Cette proportion du risque est définie par une variable à optimiser dans le programme. Les contraintes de cette approche sont reformulées et un équivalent déterministe est proposé afin de proposer une formulation rapide à résoudre. Finalement, des approximations linéaires permettent d'obtenir une formulation d'optimisation linéaire.

Abstract

In a few years, call centers have become an essential structure in numerous companies, especially in service-sector companies.

Call centers are often the main interface between the companies and their customers, making the customer satisfaction evaluation essential. Moreover, because of its high cost, workforce management in call centers is a critical issue to be considered in order to insure the structure cost-effectiveness.

The staffing and shift-scheduling problem in call centers consists in deciding how many agents handling the phone calls should be assigned to work during a given period in order to reach the required Quality of Service and minimize the costs. These problems are subjected to a growing interest, both for their interesting theoretical formulation and their possible applicative effects. This thesis aims at proposing chance-constrained approaches, considering uncertainty on demand forecasts. The arrival process is a non-stationary process with uncertain arrival rate.

The first part of this thesis proposes a model solving the staffing and shift-scheduling problems in one step so as to reduce the total imprecision errors of the approach and thus provide a cost-reducing solution.

The first model presents a two-step approach developed through a stochastic optimization formulation based on a joint chance-constrained problem. This approach guarantees that the global risk level is respected on the whole horizon whereas the individual chance-constrained approach does not.

The one-step approach is based on the two-step formulation, with the consideration of a joint chance constraint. The one-step formulation is developed on the basis of several choices aiming at reducing model approximations: first, arrival rates are random variables following continuous distributions, which is in adequation with the real case. Second, a continuous relation between the arrival rates and the theoretical real numbers of agent requirements is proposed in order to limit the approximations. The problem is then modelled as a stochastic optimization problem with a joint chance constraint so as to set the risk level on the whole time horizon. The risk is then shared among the different periods of time, for example different periods of a day. The sharing out of the risk is handled during the optimization process rather than pre-determined, which provide a better flexibility in the problem approach and leads to a cost-reduced solution compared with an approach with a pre-determined risk distribution. This approach is also enhanced by the consideration of theoretical agent requirements as real numbers rather than integers. This consideration enables to decrease the staffing on some periods while the required Quality of Service remains respected. Finally, the idea of continuity is completed with constraint linearizations in order to formulate an easily-tractable final optimization program. The linearizations enable to propose a lower bound and an upper bound of the optimal solution cost. These bounds respect the targeted risk level while reducing the solution cost compared to other approaches.

In the second part, this one-step model is extended so that it provides a better representation of real situations. First, the queuing system model is improved in order to include possible customer abandonments. Indeed, it is now considered that customers

have a limited patience. This model provides a better mathematical formulation as well as a better consideration of the real features of the queuing system. Second, the arrival rates of different periods of a day are actually correlated, which is not taken into account in the previous approach. Thus another formulation of uncertainty is proposed so that the period correlation is considered in the joint chance-constrained problem.

The third part of the thesis focuses on the staffing and shift-scheduling problems in call centers with an additional uncertainty consideration: the correct arrival rates probability distribution is supposed to be unknown. The distributionally robust approach provides a cost-optimization problem formulation while assuming that the correct probability distribution remains unknown. The distribution is supposed to belong to a set of possible distributions defined by given mean and variance. The distributionally robust optimization problem is formulated with the same approach as with the stochastic optimization problem: the problem is formulated with joint a chance constraint, split into individual chance constraints. The global risk level is dynamically shared out through the periods and thus through the individual chance constraints. The risk at each period is a decision variable to be optimized. Constraints are reformulated in order to be able to provide a deterministic equivalent problem. Finally, an easily-tractable mixed-integer linear formulation is obtained through piecewise linearization approximations.

One must still have chaos in one, to give birth
to a dancing star.

Thus Spoke Zarathustra: A book for All and None

NIETZSCHE

Acknowledgments

I would like to thank many people who contributed to this thesis. First, I am deeply indebted to Prof. Steven Martin. I would like to sincerely thank Steven for his support, his useful guidance, and his rewarding trust. I am grateful to my advisors Dr. Céline Gicquel and Dr. Oualid Jouini. Céline and Oualid provided me with help advice during these three years.

I would like to thank Prof. L'Ecuyer for his kind invitation to Montréal. I spent very rewarding months at Université de Montréal.

I would like to express my gratitude to everybody at LRI (Laboratory for Computer Science) who supported me, in particular Prof. Nicole Bidoit, Prof. Philippe Dague, Dr. Dominique Quadri and Dr. Nathann Cohen for their support. My thanks to Prof. Abdel Lisser live up to his interest in my work.

I am obliged to my thesis committee, whom I thank for their time and their advice: Dr. Safia Kedad-Sidhoum, Prof. Dominique Feillet, Prof. Alain Denise and Prof. Pierre L'Ecuyer. It is a great honor for me to have them review my work.

I would like to thank my family for their faith in me and their support during these years, but also during all the past experiences which drove me to do this thesis.

I would like to thank Guillaume who tirelessly stands by me, even after three years by the side of an stressed version of me.

Numerous friends should be thanked here for their support and their infectious good mood: Anis, Julie, Marine, Matthieu, Paul, Pierre, François, Amine, and Bruno without whom my work at the library would have been definitely gloomy. Thank you to Hélène, Philippe and Bertrand for enjoying Canadian winter with me.

I would like to thank Stéphanie, Julien, Khaleb, Xavier and Yann who had me celebrate good news when I almost forgot to do it. And of course, I cannot forget Muriel, Jean-Loup, Julien, Mayu, Gabriel, Kévin and Sam who saved my thesis from fire.

Contents

Résumé	3
Abstract	5
List of Figures	13
List of Tables	15
1 Introduction	17
1.1 Queuing Model	19
1.1.1 Queues	19
1.1.2 Modelling a Call Center	20
1.1.3 Performance Evaluation Metrics	22
1.2 Staffing and Shift-Scheduling in Call Centers	23
1.2.1 Staffing Problem	23
1.2.2 Scheduling Problem	25
1.2.3 Uncertainty	26
1.2.4 Costs	27
1.3 Uncertainty Management	27
1.3.1 Stochastic Programming and Chance Constraints	27
1.3.2 Distributionally Robust Programming	30
1.4 Main Contributions	32
2 Literature Review	37
2.1 Literature Review	37
2.1.1 Call Center Model	37
2.1.2 Stochastic Programming with Chance Constraints	41

2.1.3	Distributionally Robust Optimization	43
2.2	Work Positioning	44
3	Stochastic Approaches for Staffing and Shift-Scheduling Problems in One Step	47
3.1	Introduction	48
3.2	Problem Formulation	49
3.2.1	Queue Model	49
3.2.2	Staffing Problem	50
3.2.3	Shift-Scheduling Problem	52
3.2.4	Costs	52
3.2.5	Uncertainty Management with Chance Constraints	52
3.3	Solving the Problem in Two Steps	55
3.3.1	Uncertainty Management	55
3.3.2	Staffing Model	55
3.3.3	Agents Requirements as Random Variables	55
3.4	Staffing and Shift-Scheduling Problem in a One-Step Approach	57
3.4.1	New Staffing Model: Relation between arrival rates and requirements	58
3.4.2	Convexity of $y \mapsto F_{B_i}^{-1}(p^y)$	59
3.4.3	Final Stochastic Program	60
3.5	Approximated Solutions	61
3.5.1	Piecewise Linear Approximation	61
3.5.2	Piecewise Tangent Approximation	61
3.6	Numerical Experiments for the One-Step Approach	63
3.6.1	Instance	63
3.6.2	Comparison with Other Chance-Constrained Approaches	64
3.6.3	Results	65
3.7	Comparing Results with the Two-Step Approach	69
3.7.1	Final Deterministic Formulation	69
3.7.2	Results Comparison	69
3.8	On the Idea of Continuity	71
3.8.1	An Integer-Based One-Step Approach	73
3.8.2	Comparison of Solution Costs	75

3.9	Conclusion	76
4	Chance-Constrained Approaches with Advanced Queuing Models - Abandonments and Period Correlations	79
4.1	Introduction	80
4.2	Extension: Including Abandonments	81
4.2.1	Another Queuing Model: Erlang-A	81
4.2.2	Problem Formulation	82
4.2.3	Numerical Experiments	84
4.3	Period Correlation: Introduction of a Daily <i>Busyness</i> Factor	88
4.3.1	Arrival Rate Model	88
4.3.2	Chance-Constrained Problem with Period Correlation	90
4.3.3	Constraint Linearizations	93
4.3.4	A Tractable Approximated Formulation	99
4.4	Conclusion	99
5	Distributionally Robust Approach	101
5.1	Introduction	102
5.2	Staffing Model	103
5.3	Schedule Solution with a Wrong Assumption on Call Arrival Rates Distributions	104
5.3.1	Stochastic Program Formulation	104
5.3.2	Numerical Results in Case of a Wrong Distribution Considered	105
5.4	Distributionally Robust Model	107
5.4.1	Distributionally Robust Optimization Program	107
5.4.2	Deterministic Equivalent Problem	108
5.4.3	Linear Approximations	111
5.5	Numerical Experiments	114
5.5.1	Instance	114
5.5.2	Robustness of the Schedule Solutions	115
5.5.3	Costs of the Schedule Solutions	116
5.5.4	Comparison of costs	118
5.6	Conclusion	120

6 Conclusion	I21
6.1 Conclusion	I21
6.2 Perspectives	I22
Bibliography	I25

List of Figures

1-1	A Birth and Death process.	20
1-2	A standard queuing system.	20
1-3	A Birth and Death process for the $M/M/c$ queue.	21
1-4	A Birth and Death process for the $M/M/c + M$ queue.	22
1-5	An example of arrival rates through the day.	24
1-6	Example of a simple matrix of shifts for one day.	25
3-1	A simple $M/M/c$ queuing system.	50
3-2	Violated scenarios for each model.	67
4-1	A queue with customer abandonment.	82
4-2	Example of possible daily arrival rates scenarios.	89
5-1	An illustrative example of piecewise linear approximations of function f	112
5-2	Sharing out of the risk through the day.	118

List of Tables

3.1	Scheduling result on each shift.	66
3.2	Percentage of violated scenarios.	66
3.3	Results for different parameters.	68
3.4	Comparison of solution costs for 1-Step and 2-Step approaches.	70
3.5	Percentage of violations for the 1-Step and 2-Step approaches.	71
3.6	Values of effective Expected Waiting Time.	72
3.7	Comparison of piecewise constant-based and linear-based approaches.	75
4.1	Schedule solutions for the problem with abandonment.	86
4.2	Percentage of violated scenarios for Erlang-A Queuing Model.	87
4.3	Solution cost with or without abandonment.	87
4.4	Comparison of piecewise constant-based and real-based approaches.	88
5.1	Robustness of the stochastic approach for different wrong distributions. Targeted risk level is 5%.	106
5.2	Percentage of violations for different parameters.	115
5.3	Results for different risk levels.	117
5.4	Comparison between the stochastic solution costs and distributionally robust solution costs.	119

Chapter I

Introduction

Mathematical optimization is the discipline dedicated to the study and the formulation of optimization problems, consisting in choosing the best solution among a set of possible ones, regarding some possible constraints on the variables and some objective. When dealing with uncertain data, we focus on subfields of mathematical programming, such as stochastic programming or robust programming.

Stochastic programming approaches find an abundance of application subjects in real-life problems. Personnel capacity planning in queues and especially in call centers is one of them. A call center is an infrastructure of a company composed of working agents committed to handling customer calls. Call center structures can deal with inbound or outbound calls and can be extended around several services such as emails or IM. In [Brown et al. \[2005\]](#), it is stated that call centers employ more than 3.5 million people in the U.S. (statistics of 2002). Because of its increasing importance in industry, call centers business is a well documented subject in operations management (see [Aksin et al. \[2007\]](#) or [Gans et al. \[2003\]](#) for reviews on call centers). It has interest in various disciplines: Mathematics, OR, Industrial Engineering, Sociology, etc.

Call centers are more and more developed in many large organizations, in order to provide a convenient customer-company contact structure. According to [Brown et al. \[2005\]](#), call centers represent 70% of all business interactions in the U.S in 2002. Whether it be for emergency call centers or travel company hotlines for example, customers are to be answered within a very limited time. The Quality of Service is of prime importance in the management of call centers. Call centers are consequently a critical subject regarding customer satisfaction.

Moreover, call centers are expensive infrastructures for companies, in which the staff

agents represent 60% to 80% of the total operating budget (Aksin et al. [2007]). Thus an efficient workforce management is essential to achieve profitability in a call center.

This thesis focuses on the call center workforce short-term planning optimization.

Short-term planning fits in a global planning process: long-term view, short-term planning and real-time updating and routing.

The first step is a long-term consideration of the call center planning. Long-term forecasts enable to decide how many people should be hired. Hiring and training agents require time. Thus, the workforce need to be planned in a long-term view (usually several weeks or months).

The second step of the planning process is the short-term organization. It relies on four main aspects. First, **short-term forecasts** are to be made for a few weeks. These forecasts give a more precise representation of the upcoming call arrivals than in a long-term view. Extra agents already hired may be assigned to other tasks (such as emails handling). Second, **staffing** requirements are computed thanks to these forecasts and other considerations (Quality of Service for instance). Third, staffed agents are organized through the **shift-scheduling** step. It consists in deciding how many agents should be assigned to pre-determined shifts. Fourth, **rostering** enables to actually assign agents to the shifts. Other considerations could be included in the short-term planning, such as shifts generation. Shifts are generated according to requirements, such as employment law, typical constraints of the companies and agents requests. Among these four aspects, we focus on two of them: the staffing and shift-scheduling steps.

The last step of the planning process consists in making real-time decisions when needed, such as schedule updating or call routing for example.

The short-term staffing and shift-scheduling problems consist in deciding how many staff members handling the phone calls, i.e. "agents", should work during the forthcoming days or weeks in order to minimize workforce costs while ensuring that the required customer quality of service is reached. The Quality of Service (QoS) can be measured for instance with a maximum expected abandonment rate, ie the proportion of clients hanging up without being served, or with a maximum value of expected waiting time before entering service in the queue or abandoning it. Since this problem relies on demand forecasts, arrival rate uncertainty is an important criterion to consider, as pointed out in Gans et al. [2003]. Moreover, Aksin et al. [2007] states that considering arrival rates as random variables is a "very promising area that is just now beginning to receive attention from researchers".

Stochastic programming enables to model and solve optimization problems in which parameters are subject to uncertainty. This field is really valuable when considering real-life optimization problems with uncertainty such as ours. The goal is to optimize an objective function relatively to probable realizations while quantifying a risk. Probability distributions, either discrete or continuous, are to be integrated in the optimization programs. When lacking of information about distributions, the optimization relies on robust or distributionally robust approaches.

Stochastic programming is a huge field and the scope of this thesis is defined as follows: random variables are assumed to follow continuous probability distributions. The programs are non-linear mixed integer programs: variables of the program are integers, parameters are reals and the constraints are non linear. Chance-constrained programming approaches are developed for both stochastic and distributionally robust approach.

Next sections introduce the context and the concepts involved in the problem models proposed in this thesis. Section [I.1](#) introduces the queuing system model, Section [I.2](#) focuses on our problems, namely the staffing and shift-scheduling problems, and Section [I.3](#) presents approaches for uncertainty management and especially stochastic programming and distributionally robust programming. Finally, Section [I.4](#) gives my contributions in the subject details the structure of the document.

I.I Queuing Model

I.I.I Queues

Queuing theory was introduced by A.K. Erlang at the beginning of 20th century in order to model waiting lines in telecommunications. Queuing model relies on Markov chains and is illustrated with Birth and Death processes (Figure [I-1](#)). It consists in a chain of states with two transition processes: the *Birth* process changes the current state from n to $n + 1$ with rate λ_n ; the *Death* process changes the state from n to $n - 1$ with rate μ_n . The state value represents the number of objects in the system.

Processes in the queue are represented with Kendall notation $A/S/c$. The factor A denotes the arrival process, S the service time process and c the number of servers.

Call centers are typically represented as queuing systems, see for example [Gross et al. \[2008\]](#) or [Gans et al. \[2003\]](#). Only inbound call centers are considered here (i.e. centers receiving calls only).

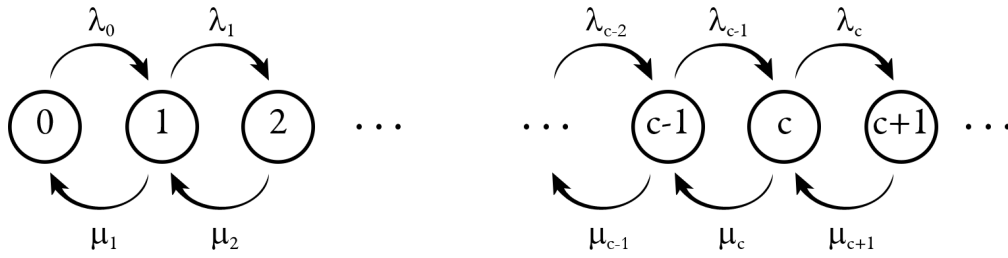


Figure 1-1: A Birth and Death process.

1.1.2 Modelling a Call Center

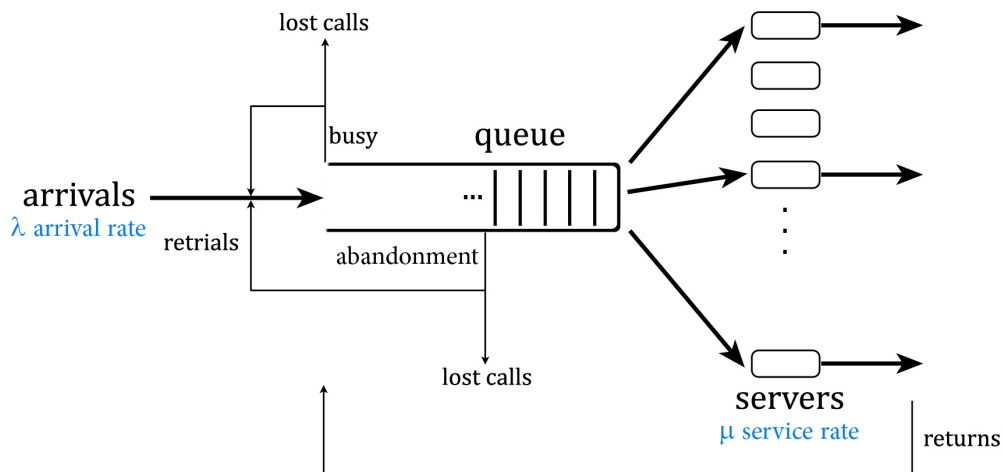


Figure 1-2: A standard queuing system.

Figure 1-2 shows a single-stage queuing system. The system is composed of telephone lines, or trunk lines, and servers, namely agents, dealing with incoming calls. Calls enter the system with arrival rate λ and are served without waiting (with service rate μ) if agents are available. If not, they join the queue and wait for an agent to be available. If a call arrives while all the trunk lines are occupied, it cannot enter the system, the call is lost and the customer receives a busy signal. While waiting, some customers may leave the queue by hanging up, purposely (when their patience limit is reached) or not. The calls are considered as abandoned and the customers are not served. Lost calls, both rejected and abandoned, may try and re-enter the system, and become retrials. Finally,

served calls, i.e. calls that went through the system and reached agents, may also re-enter the system, as returns.

In single-skill call centers, agents are assumed to be trained so that they can answer all types of calls whereas in the multi-skill call centers, agents are specialized. This tend to provide a better Quality of Service but require more training or more expensive agents, due to their specialization.

The queue discipline in a single-skill call center is usually FCFS (First Come-First Served). Other disciplines may be considered in multi-skill call centers or according to some particular criteria (priority for premium customers for example).

In this thesis, the arrival process is assumed to be a time-varying Poisson process with arrival rate $\lambda(t)$ (Aksin et al. [2007]). Customers are served by c agents. Service times are assumed to be independent and exponentially distributed with rate μ . The call center is considered single-skilled. The queue capacity is assumed to be infinite, i.e. customers never get a busy signal. Customers retrials as well as returns are ignored.

When ignoring abandoned calls, Kendall notation of this non-stationary queue is $M(t)/M/c$. The extended Kendall notation $A/S/c/K/N/D$ gives more information: the capacity of the queue K , the number of jobs N and the discipline of the queue D . Here, these parameters will always be standard: $K = \infty$, $N = \infty$ and $D = \text{FCFS}$.

Figure I-3 shows the state diagram of a $M/M/c$ queue.

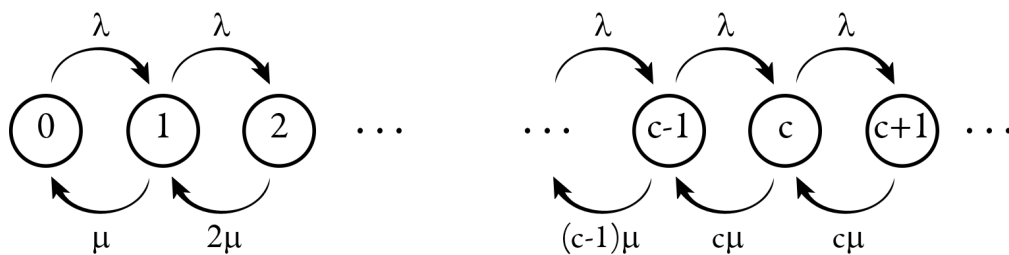


Figure I-3: A Birth and Death process for the $M/M/c$ queue.

The impatience is introduced in Baccelli and Hebuterne [1981] with the notation: $M(t)/M/c+M$. The last M denotes the impatience process is exponentially distributed with rate θ . The state diagram of the Birth and Death process is then illustrated in Figure I-4.

This queuing model is used in the staffing step in order to evaluate the performances

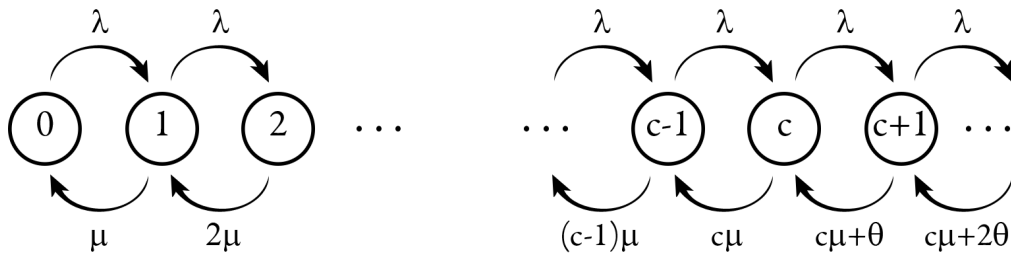


Figure I-4: A Birth and Death process for the $M/M/c + M$ queue.

of the queue (detailed in Section [I.2.1](#)).

I.1.3 Performance Evaluation Metrics

Call centers performance evaluation gives information on customer satisfaction as well as system performances. Two criteria are highly used: the Telephone Service Factor (TSF) and the Expected Waiting Time (EWT). The TSF provide a proportion of calls served in less than a fixed waiting time ; EWT gives a value of average waiting time before the call is answered. The chosen criterion for next chapters is the EWT.

The choice regarding the performance evaluation metrics is not obvious. Both of these criteria give incomplete information. Indeed, the EWT does not consider variability of waiting times. The same EWT value can describe situations with close values of waiting times or on the contrary, very distant ones. The TSF service level partially takes variability into account but does not provide information about the effective waiting time. No matter which excess waiting time, either huge or very small, it will be considered as unsatisfactory. An interesting performance evaluation metric could mix the TSF and the EWT together by introducing a limit of accepted waiting time (TSF) and computing the average of the excess waiting time (EWT). However, this type of metric will most likely be difficult to compute.

A high system performance is sought not only to provide a good Quality of Service but also to ensure its durability: by providing a satisfactory service to the customers, re-trials and returns are reduced. Then, the queue model is easier to establish. Incidentally, the return proportion could be considered in the performance evaluation.

When considering abandoned calls in the model, proportion of abandonments is an interesting criterion to be considered. It can mathematically be related to other per-

formance metrics, like the EWT. The Expected Waiting Time can be considered conditionally to the event that the call is served, or conversely unconditional. An unconditional waiting time metric would give information on global waiting time, including non-served calls. The conditional metric would give the real Average Speed of Answer without giving information on unsatisfied customers. Both considerations give interesting information, and unconditional Expected Waiting Time will be considered in next chapters.

1.2 Staffing and Shift-Scheduling in Call Centers

1.2.1 Staffing Problem

The staffing problem consists in deciding how many people handling the phone calls, that is to say agents, should be assigned to work each day at the call center. These values of agent requirements come from a calculation based on a performance evaluation and estimations of arrival rates. The objective service level is the customer Quality of Service that is to be reached in compliance with the company demand. The expected call arrival rates are forecasts computed from historical data. This data gives usually the numbers of calls for the working time horizon of past years.

As a matter of fact, the quantity of calls varies strongly in time, especially during a day (see Figure 1-5). This is why, as it is said in previous Section 1.1, the call center model is a queue with time-varying arrival rates $M(t)/M/c$.

In order to deal with this nonstationary demand and compute staffing requirements, several approaches are possible, some based on stationary approximations, some other on simulations or fluid approximation (especially when the system charge is very high). Here, the SIPP (Stationary Independent Period-by-Period) approximation is used. This approximation consists in splitting the working horizon in small periods of time during which the rate does not vary too much and the queue is assumed to be in steady state. The arrival rate at each period is considered as constant. This approximation is applied in order to represent the rate variation with good precision. The small periods are usually 15-minute or 30-minute periods. The call center is considered as a $M_t/M/c_t$ queue where stationary regime is reached and values of expected call arrival rates are obtained for each period t .

The SIPP approximation enables to use the Erlang models. We focus here on the

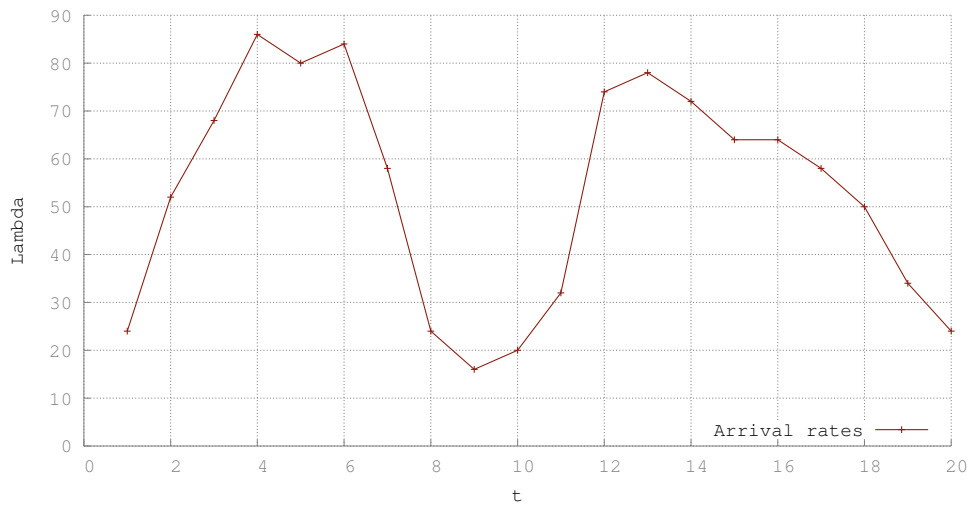


Figure 1-5: An example of arrival rates through the day.

well-known and highly used Erlang-C model which fits the simple queue model: no abandonment, no return, no retrial. Erlang-A is also considered in our model as an extended model which considers abandonments. Erlang-B considers a limited number of trunk lines. Thus, incoming calls can receive a busy signal if no line is available.

Erlang models are very convenient since they provide probabilistic formulae to model a queue. However they lack some assumptions (agents behaviour and productivity for example). Simulations tend to be more precise on a given context. However, developing and testing simulation models is complex and computation times are greater than computing Erlang formulae.

In this thesis, the objective service level is a maximum EWT value. This value is denoted EWT^* . Practically, the staffing problem consists in computing the minimum number of agents required for each period to respect the targeted value EWT^* and according to forecasts of arrival rates, with the help of Erlang formulae.

The computations of requirements will be made for each of the short periods of time constituting the whole horizon. This leads to considering a lot of values for requirements and thus to an increasing number of variables. The shift-scheduling problem handles this with an optimization program.

I.2.2 Scheduling Problem

The shift-scheduling problem is induced by the fact that variables are whole numbers of human agents working according to manpower constraints. It is indeed inconvenient to ask agents to come and work for only a few hours. They should respect typical working shifts of full-time or part-time jobs. This may lead to overstaff some periods.

Thus, several possible shifts are created in order to organize agents according to real work days, which cover the schedule of the call center. This is a multi-shift model. These shifts are made up of working hours and breaks, for lunch for example. In the problem, the shifts are predefined and cannot be changed.

A matrix of shifts describes a set of possible shifts that agents can be assigned on. One column of the matrix represents one possible shift and one row represents one period of time. The matrix of shifts $A \in \mathcal{M}_{T,S}(\{0, 1\})$ consists in 0s and 1s:

$$a_{t,s} = \begin{cases} 1 & \text{if agents assigned to shift } s \text{ are working at period } t \\ 0 & \text{if not} \end{cases}$$

Figure I-6 shows a simple example of shifts. Some shifts begins later than other, and lunch breaks are taken into consideration.

Shift	1	2	3	4	5	6	7	8
08:00 - 09:00	1	0	0	1	0	0	1	0
09:00 - 10:00	1	1	0	1	1	0	1	1
10:00 - 11:00	1	1	1	1	1	1	1	1
11:00 - 12:00	1	1	1	0	1	1	0	1
12:00 - 13:00	0	1	1	1	0	1	0	0
13:00 - 14:00	1	0	1	1	1	0	1	0
14:00 - 15:00	1	1	0	1	1	1	1	1
15:00 - 16:00	1	1	1	1	1	1	1	1
16:00 - 17:00	0	1	1	0	1	1	1	1
17:00 - 18:00	0	0	1	0	0	1	0	0

Figure I-6: Example of a simple matrix of shifts for one day.

The problem is then to decide how many working agents should be assigned to each shift in the call center in order to respect the chosen Quality of Service while minimizing the manpower cost. This computation uses data of call arrival rates previously computed and is done through an optimization program.

The deterministic formulation of the problem is the following:

$$\begin{aligned} \min \quad & c^t x \\ \text{s.t.} \quad & Ax \geq b \\ & x_i \in \mathbb{Z}^+ \end{aligned} \tag{I.1}$$

where A is the previously defined matrix of shifts and the vector x is the vector of agents assigned to each predetermined shift. Then, the vector Ax defines the number of agents working at each period t . The vector b is the vector of minimum number of agents required to ensure the chosen QoS computed in the previous step. The vector c in the objective function denotes the costs of agents according to their assignments. For this example, the costs are considered to be only dependent of their assigned shift.

The program illustrates the goal of the shift-scheduling step: minimizing the manpower cost while reaching the agents requirements computed in the staffing step at each period of time.

The nature of the matrix of shifts and the variables induce complexity in this linear program: indeed, introducing breaks, such as lunch breaks, and considering integer variables causes the program to be NP-complete. Thus, increasing the number of time periods or number of shifts can significantly increase the time computation.

1.2.3 Uncertainty

Resource deployment decisions, namely agents assignments for the staffing and shift-scheduling problems, rely on information about the load of the call center in the past. Decisions are to be made for the forthcoming days or weeks and forecasts of call arrivals are computed using historical data. It is usually the only information available.

Since the future call load is difficult to predict, the accuracy of the deterministic approach is limited. Indeed, these estimations of mean arrival rates may differ from reality. As a consequence, solutions provided by a deterministic approach is surely inadequate.

Hence, as stated previously, arrival rate uncertainty is an important criterion in the staffing and shift-scheduling problems (Aksin et al. [2007]). In this thesis, optimization under uncertainty is addressed to consider possible fluctuations.

Then, in order to take account of uncertainty, mean arrival rates are considered as random variables following probability distributions.

1.2.4 Costs

As stated previously, the staffing step relies on demand forecasts. Thus, uncertainty on mean arrival rates is expressed with random variables.

Overstaffing the call center means paying unnecessary agents and thus leads to extra operating costs. This excess is quantifiable, since the salaries are known. Conversely, understaffing the call center leads to a longer queue with high waiting time, and thus to a decreasing quality of service and customers dissatisfaction. This loss can be considered as a penalty cost as in [Liao et al. \[2012\]](#). However this quantity is not easily quantifiable. Hence, another representation of the risk would rely on the consideration of a targeted risk level. Understaffing would be tolerated up to the quantified risk with a chance-constrained formulation.

1.3 Uncertainty Management

1.3.1 Stochastic Programming and Chance Constraints

Stochastic programming was introduced in the 50's (see [Beale \[1961\]](#); [Charnes and Cooper \[1959\]](#); [Dantzig \[1955\]](#)) in order to deal with problems including uncertainty, i.e. problems for which decisions are made before some parameters are known. These parameters are expressed as random variables either in the objective function or in the constraints, or both. We focus here on single-stage stochastic programming, which describes problems where decisions are to be made before the realizations of the random variables and, once made, cannot be changed.

Consider the following deterministic program:

$$\begin{aligned} \min & f(x) \\ \text{s.t.} & \forall i \in \llbracket 1; M \rrbracket, g_i(x) \leq 0 \\ & x \in X \subset \mathbb{R}^n \end{aligned} \tag{1.2}$$

where $\forall i \in \llbracket 1; M \rrbracket$, g_i and f are functions from \mathbb{R}^n to \mathbb{R} . If they all are linear then the mathematical program is linear.

Let introduce uncertainty in the mathematical programming formulation.

$\xi: \Omega \rightarrow \mathbb{R}^k$ denotes a random vector on the probability space $(\Omega, \mathfrak{F}, \mathbb{P})$. In stochastic programming, the parameters of the probability measure is known. The decision vector

x is assumed to be deterministic.

Thus the general formulation of a stochastic program is the following:

$$\begin{aligned} \min f(x, \xi) \\ \text{s.t. } \rho(g(x, \xi)) \geq 1 - \varepsilon \\ x \in X \subset \mathbb{R}^n \end{aligned} \tag{I.3}$$

where $g(x, \xi) = {}^t(g_1(x, \xi), g_2(x, \xi), \dots, g_M(x, \xi))$, ε is the risk level, and ρ is the risk measure evaluating the random variables. Hence, the function

$$\begin{aligned} h: \mathbb{R}^n &\rightarrow \mathbb{R} \\ x &\mapsto \rho(g(x, \xi)) \end{aligned} \tag{I.4}$$

defines the evaluation function for the random vector x .

It is also possible to consider the evaluation function in the objective, by defining $f(x) + \rho(g(x, \xi))$. Moreover, the constraints can be joint or separate, so as the risk measure.

The risk measure will define the specific formulation of the stochastic program. It can be expressed with the expected values of the random variables. A standard approach involves discrete distributions and consists in generating several possible scenarios and measuring the expectation. The computational efforts depend on the number of scenarios generated.

Another risk measure evaluates random vectors by defining a subset of the feasible domain with cumulative probabilities:

$$\rho(v) = \mathbb{P}(v \leq 0) \tag{I.5}$$

where v is a random vector on $(\Omega, \mathfrak{F}, \mathbb{P})$.

Thus the constraints describe a risk level on probability violations and the problem is expressed with probabilistic constraints. Hence, a chance-constrained optimization

problem can be expressed as follows:

$$\begin{aligned} \min f(x) \\ \text{s.t. } \mathbb{P}(g(x, \xi) \leq 0) \geq 1 - \varepsilon \\ x \in X \subset \mathbb{R}^n \end{aligned} \quad (\text{I.6})$$

with ε , the risk level, typically low.

As previously explained, the evaluation function may be applied individually or jointly on the constraints. Chance constraints do not focus on the penalty cost of the non-respect of some feasible solution but provides control on how conservative the solution will be.

Chance-constrained problem are usually computationally intractable. Some details are given in the literature review in Chapter 2.

As a possible approximation, the Conditional Value-a-Risk is an expectation-based risk measure found in the literature. It is based on the Value-at-Risk (VaR) defined by:

$$VaR(v, \alpha) = \min\{z \mid [\mathbb{P}(v \leq z) \geq \alpha]\} . \quad (\text{I.7})$$

The Value-at-Risk represents the minimal loss such that the probability that the losses is greater than or equal to this value is at least α (i.e., the left endpoint of the α -quantiles of v , see [Kall and Mayer \[2011\]](#)). This is in itself a risk measure.

The Conditional Value-a-Risk (CVaR) is measuring the expected value of loss, given that the loss is greater than or equal than the Value-at-Risk (VaR). This measure is given by:

$$\begin{aligned} \rho_{CVaR}(v) &= \min_{z \in \mathbb{R}} \left[z + \frac{1}{\alpha} \mathbb{E}[(v - z)^+] \right] \\ &= \mathbb{E}[v \mid v \geq VaR(v, \alpha)] . \end{aligned} \quad (\text{I.8})$$

The CVaR could be expressed in the objective function. This measure is a conservative approximation of the chance-constrained formulation and is interesting because the function is convex. Moreover, it represents the tightest convex approximation to individual chance constraints (see [Chen et al. \[2010\]](#)). However this measure is generally computationally demanding and intractable.

In this thesis we focus on this chance-constrained formulation with a random Right-

Hand-Side -which means that $g(x, \xi) = a(x) - h(\xi)$ - and a linear function a :

$$\begin{aligned} \min \quad & c^t x \\ \text{s.t.} \quad & \mathbb{P}(Ax \geq \xi) \geq 1 - \varepsilon \\ & x \in (\mathbb{R}^+)^n \end{aligned} \tag{1.9}$$

where $A \in \mathcal{M}_{k,n}(\mathbb{R})$, ξ is a random vector in \mathbb{R}^k and $\varepsilon \in]0; 1]$ (typically small).

Under these assumptions, the problem remains NP-hard ([Luedtke et al. \[2010\]](#)).

1.3.2 Distributionally Robust Programming

Stochastic programming relies on the knowledge of the probability distribution considered in the programs. However, information about the probability distribution is not always completely available.

Robust optimization is another approach of optimization under uncertainty, initially introduced in 1973 by [Soyster \[1973\]](#) for considering linear problems where data is not exactly known. Robust optimization is then investigated anew since the end of 90's (for example [Ben-Tal and Nemirovski \[1998, 2000\]](#); [El Ghaoui and Lebret \[1997\]](#); [El Ghaoui et al. \[1998\]](#)).

A linear robust problem can be defined as follows (see [Ben-Tal et al. \[2009\]](#)):

$$\begin{aligned} \min_x \quad & \sup_{(c,A,B) \in U} c^t x \\ \text{s.t.} \quad & \forall (c,A,B) \in U, Ax \leq b \\ & x \in X \subset \mathbb{R}^n \end{aligned} \tag{1.10}$$

where $(c,A,B) \in U$, U being a given uncertainty set.

Thus, the robust optimization program aims at optimizing the cost regarding that uncertain data belongs to the uncertainty set U . This is in fact the consideration of a collection of programs. Indeed, for each data $(c,A,B) \in U$ could be associated one program. Here, the robust formulation gives a solution respecting the constraints of the problem for every possible data in U .

Between the complete knowledge required in the stochastic optimization approach and the deterministic representation of uncertainty of the robust optimization, an inter-

mediate distributionally robust is proposed.

Indeed, Scarf et al. [1958] explains that selecting a probability distribution to model the future demand may be difficult because of the unpredictability of the future. Instead of considering a definite probability distribution, it is assumed that the mean and variance are known, but not the right probability. Then, the cost is optimized regarding the worst distribution among the possible distributions, by introducing a min-max strategy.

The min-max approach is developed in the literature first by Žáčková [1966] and later by Breton and El Hachem [1995]; Calafiore and El Ghaoui [2006]; Delage and Ye [2010]. The minmax idea is now known as distributionally robust.

The distributionally robust stochastic programs may be formulated with measures previously presented in Section 1.3.1, for example the CVaR or chance constraints.

A standard chance-constrained distributionally robust formulation is given by

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & \inf_{\xi \in \Xi} \mathbb{P}(g(x, \xi) \leq 0) \geq 1 - \varepsilon \\ & x \in (\mathbb{R}^+)^n \end{aligned} \tag{1.11}$$

where ξ is a random vector in \mathbb{R}^k and its distribution is in the family Ξ .

As stated previously, we consider in this thesis linear functions in the constraints, with a random Right-Hand-Side. Thus, the distributionally robust problem is formulated as follows:

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & \inf_{\xi \in \Xi} \mathbb{P}(Ax \geq \xi) \geq 1 - \varepsilon \\ & x \in (\mathbb{R}^+)^n \end{aligned} \tag{1.12}$$

where $A \in \mathcal{M}_{k,n}(\mathbb{R})$.

This program considers the worst-case probability distributions among a known family of possible distributions. This approach stands between stochastic programming and robust programming.

1.4 Main Contributions

This thesis concentrates on the staffing and shift-scheduling problems in personnel planning under uncertain demand, especially in call centers. The call center is represented, according to numerous published works, as a single-stage queuing system ignoring re-trials and returns. Once the decision of staffing is made, real time adjustments are not allowed. This means that recourse in the optimization problem is ignored.

Problem Formulation

The call center is considered to be a single-skill and multi-period, multi-shift call center. Indeed, agents are able to answer any call. The calls do not have specific types. The time horizon is typically one day or one week split in short periods and the call center allows several possible shifts, both part-time and full-time, which cover the schedule horizon.

We base our staffing model on Erlang formulae by considering the Stationary Independent Period-by-Period approximation. Both Erlang-C (no abandonment) and, as an extension, Erlang-A (with abandonment) fit in our formulation. Erlang-A provide a better view of the real situation. The chosen Quality of Service is a maximum Expected Waiting Time (*EW*T).

Arrival rates are supposed to be subject to uncertainty, in agreement with one of the main issue of the personnel planning problem. We consider that the mean arrival rates are random variables following normal distributions with known means and variances.

The shift-scheduling problem is formulated as a joint chance-constrained optimization problem. Under uncertainty, the constraints illustrate a maximum risk we allow for not providing the truly required number of agents. Chance constraints are joint rather than individual in order to provide a guaranteed service level on the horizon.

Considering a global risk level instead of setting it period-by-period is based on several considerations. First, it may be more convenient for a manager to set a general performance goal for one day or one week rather than evaluating all periods and set different goals. Moreover, it provides a guaranteed service level for all periods. Finally, we propose in our approach to share the risk level through the different periods and optimize this sharing during the optimization process. This leads to a reduction of the total costs.

First Part: Stochastic Programming Approaches

In the light of the previous considerations, we propose and compare several approaches.

The first model considers the problem as a two-step problem: agents requirements are computed in the staffing step and the schedule is generated in the stochastic optimization step. Since agents requirements are computed using arrival rates, they are indeed random variables. We suppose they are following a continuous normal distribution.

The second model describes the problem as a one-step problem: the staffing and scheduling problems are solved together in one optimization problem. This model is proposed in order to provide a single formulation, simple to compute in short time and reducing the approximation steps between the staffing and shift-scheduling problems. General formulations are less precise than genuine simulations but provide generic tools that are easily adjustable. Here, we aim at reducing the different approximations of the model to get a better representation of the reality.

In order to provide deterministic mixed-integer linear solvable formulations of the stochastic programs, the joint chance constraint is separated into individual chance constraints. The risk levels set for the resulting constraints are proportions of the global risk level of the joint chance-constrained formulation. These proportions are dynamically attributed. Indeed, instead of splitting the risk among all the periods before solving the optimization problem, we consider that the proportion of the risk at each period is a variable to be optimized. Thus, it provides flexibility in the risk management, leading to a cost-reducing final solution. Finally, piecewise linearizations are provided to give an upper bound and a lower bound of the optimal cost.

As far as we know, previous published approaches discretize the probability distribution -or consider a discrete distribution-, providing a convenient optimization problem formulation.

In this thesis, in opposition to the standard approach, the idea of continuity leads the formulation.

First, mean arrival rates are considered to follow given continuous distributions, namely the normal distribution with known means and variances. Apart from a simpler resolution of the problem, there is no reason why the arrival rate distribution should be discrete. On the contrary, the solution obtained with a discretization-based approach leads to biased estimate of the true solution ([Robbins and Harrison \[2010\]](#)). We keep the idea of real values of random variables during all the formulation process until the

final solvable program.

Second, Erlang formulae provide tools to compute the Expected Waiting Time (EWT) for given values of arrival rate, service rate and number of agents. In our problem, the value of targeted EWT is set. Then, for given values of arrival and service rates, a discrete number of agents does not guarantee to reach exactly the required EWT . We propose a continuous function that Erlang model does not provide in order to consider real values of number of agents aiming at the exact required EWT instead of settling for a simple upper bound.

Third, the functions involved in the chance-constrained formulation are linearized, following the idea of continuity. The final solution comprises integer values of agents requirements, but the Right-Hand Side of the constraints is real.

Considering a RHS as real when agents requirements are integer is interesting when including a flexible sharing out of the risk in the approach, as we do. Indeed, it is possible to decrease the risk on some periods by considering different real values of RHS, even if the final integer value is the same. This leads to allow an increase of the risk on some other periods where this change causes a cost reduction by staffing less agents. Our approach is a mixed-integer linear problem with final agent staffed solution as integers. However, considering a real final number of agents could make sense by itself, if the problem were extended in a situation where agent available time can be a proportion of the total time.

In a nutshell, we aim at providing a high flexibility - considering a global risk level, a flexible sharing out of the risk, continuous relations - in the optimization process so that the solution cost is reduced compared with other approaches, without requiring high computation time.

The same approach stands in an extended representation of uncertainty: indeed, a positive correlation exists between different periods of a day. This dependence is introduced in the uncertainty management by multiplying the random continuous vector of arrival rates with a random *busyness* factor. This previous approach is thus adapted and leads to a mixed-integer nonlinear program.

Second Part: Distributionally Robust Approach

The second part of the thesis focuses on the distributionally robust approach of the staffing and shift-scheduling problem. Very few approaches in the call center literature were found.

In this part, the distributions of the mean arrival rates at each period are considered as unknown. However, the means and variances of the distributions are supposed to be known. The problem optimizes the cost of the call center workforce under the condition that arrival rates are random variables following unknown distributions. Hence, the problem optimizes the cost for all possible distributions with given means and variances, and thus for the most unfavorable distributions of this family. Thus, we formulate the problem as a two-step problem. Erlang-C formula is used in this part, but Erlang-A formula would also correctly fit.

The same process as for the stochastic optimization formulation occurs: we consider a joint chance-constrained optimization problem in order to consider the risk on the global scale. Then, the joint chance constraint is separated into individual chance constraints by splitting the risk through the periods. Again, the risk level is allocated during the optimization process so that its sharing guarantees to lead to a lesser cost.

A deterministic equivalent program is proposed for this distributionally robust problem. Using the convexity property of the RHS, we provide an upper bound and a lower bound of the optimal solution with piecewise tangent and linear approximations of the functions.

This approach is obviously more conservative than the stochastic approach but shows its interest in the case where the nature of the arrival rate distribution is uncertain. Indeed, in case of wrong decision on the choice of the distribution, the stochastic approach solution is not adequate.

Structure of the Thesis

Chapter 2 reviews the current literature in the relevant fields for our problem. Chapter 3 presents a stochastic programming approach and develops the formulation with joint chance constraint, dynamic allocation of the risk, consideration of theoretical real number of agents, and linearized deterministic formulations. Chapter 4 introduces an application of Erlang-A model in our approach without extra computational effort and propose another representation of the uncertainty leading to a more complex deterministic reformulation. Chapter 5 presents a distributionally robust approach and its final linearized programs. Finally Chapter 6 concludes this thesis and highlight possible perspectives for future research.

Chapter 2

Literature Review

2.1 Literature Review

The staffing and schift-scheduling problems in call centers rely on various fields and a great variety of models are present in the literature, from the call center model to the risk and uncertainty managements. The summary of literature review is threefold: first, different call center models proposed in literature are described, second, stochastic programming -and especially chance-constrained programming- models are summarized and third, the distributionally robust approach is addressed.

2.1.1 Call Center Model

In personnel planning optimization, call centers are commonly represented as queuing systems. We refer to [Gross et al. \[2008\]](#) or [Hillier and Lieberman \[1986\]](#) for general results on queuing theory.

Although simplified, **the queuing model $M/M/c$** as described in [1.1.2](#) is widely used : see for example [Green et al. \[2001\]](#); [Koole and van der Sluis \[2003\]](#); [Liao et al. \[2012, 2013\]](#). Models including customer abandonment arouse interest. Even though the model $M/M/c + M$ is more complex than a simple queue, including abandoned calls in the model provide a great improvement ([Green et al. \[2007\]](#)), since system performance and stability can be significantly affected. Usually the abandonment process is assumed to follow an exponential distribution, as in [Gans et al. \[2012\]](#); [Gurvich et al. \[2010\]](#); [Robbins and Harrison \[2010\]](#); [Saltzman and Mehrotra \[2007\]](#).

Some other extensions can be found in the literature: for example [Whitt \[1999\]](#) con-

siders a general distribution for the service times, instead of the usual exponential distribution ; [Mandelbaum and Zeltyn \[2009\]](#) consider a general distribution for the patience times. When their patience limit is reached, customers abandon the queue.

In call centers, the queuing policy is generally FCFS (for almost all published approaches) when considering only one type of call or Skill-Based in call centers considering different type of calls (see [Bassamboo et al. \[2006\]](#); [Wallace and Whitt \[2005\]](#)). Some other policies can be found, for example based on the customer priority ([Zeltyn et al. \[2011\]](#) for emergency departments).

Agents working in call centers may have specialized skills. They would be attributed calls according to their skills. This is called a multi-skill call center. Moreover, different types of calls possibly wait in different queues. [Avramidis et al. \[2010\]](#); [Cezik and L'Ecuyer \[2008\]](#); [Gurvich et al. \[2010\]](#) model the staffing problem in a multi-skill call center as a mathematical program. Multi-skill based problems are usually complex to model. These considerations are ignored in this thesis.

Evaluating performance in a call center as a non-stationary queue is complex.

Some approaches focus on stationary approximations. Two main approximations can be found in the literature. In [Bhulai et al. \[2008\]](#); [Liao et al. \[2012\]](#); [Saltzman and Mehrotra \[2007\]](#) for example, the performance measure is evaluated by considering the call center approximated as a Stationary Independent Period-by-Period model (the widely used SIPP approach). [Green et al. \[2001, 2003\]](#) propose an extension to introduce the congestion lag: it can be observed that the congestion peak appear after the arrival rate peak. In [Green and Kolesar \[1997\]](#) or [Green et al. \[2007\]](#), the Pointwise Stationary Approximation (PSA) is used. It consists in considering a stationary model at each time t , using instantaneous values of arrival rate at that time (see [Green and Kolesar \[1991\]](#) for presentation of the approximation).

These stationary approximations enable to use Erlang formulae. Erlang-C model expresses results for the $M/M/c$ queue in steady-state and is used for example in [Green et al. \[2001\]](#); [Green and Soares \[2007\]](#) or [Liao et al. \[2013\]](#). However, it is a simplified model: the queue is assumed to be in steady-state, the arrivals follow a Poisson process, the service times are exponential, and no abandonment, no return, no retrial are allowed. Erlang-A model is an extension and does take abandoned calls into consideration and is investigated in a growing number of articles, such as [Baccelli and Hebuterne \[1981\]](#); [Dietz \[2011\]](#); [Garnet et al. \[2002\]](#); [Mandelbaum and Zeltyn \[2007\]](#).

Apart from stationary approximations, other approaches can be found in the litera-

ture. Discrete-event simulation is also widely used, for example in [Avramidis et al. \[2004\]](#) or [Roubos et al. \[2011\]](#). Instead of using Erlang formula, they simulate the performance of call centers. Moreover, [Atlason et al. \[2004\]](#); [Cezik and L'Ecuyer \[2008\]](#) or [Atlason et al. \[2008\]](#) focus on simulating and estimating the service level function with a sample average approximation. In these papers, uncertainty of arrival rates is ignored. Although more precise, this approach requires longer computation times than Erlang formulae and remain specific to the data used for simulations.

Finally, fluid approximations are preferred to represent overloaded systems. This approximation gives a convenient relation between arrival rates and number of servers, particularly in large call centers, and an alternative to usual models where stationary approximations would result in an infinitely growing queue. Fluid approximation is usually used in multi-skill models, like in [Bertsimas and Doan \[2010\]](#); [Gurvich et al. \[2010\]](#); [Harrison and Zeevi \[2005\]](#).

The **staffing model** is often based on a stationary approximation and the application of an Erlang model. Under these considerations, the staffing step is usually studied in association with the forecasting step or with the shift-scheduling step. Erlang models enable to compute the minimum staffing level that satisfies the chosen performance. This heuristic seems to be prevalent (see [Bhulai et al. \[2008\]](#); [Dietz \[2011\]](#); [Ertogral and Bamuqabel \[2008\]](#); [Robbins and Harrison \[2010\]](#)).

On another consideration, the Square-Root Safety (SRS) staffing principle is suitable for large, heavily-loaded systems ([Gans et al. \[2003\]](#)) in which an asymptotic regime appear (the Quality and Efficiency Regime). The number of servers is given by $N = R + \beta\sqrt{R}$ where R is the offered load determined by the performance evaluation. For call centers, SRS staffing can be found in [Whitt \[1999\]](#) or [Borst et al. \[2004\]](#).

Mathematical programming approaches are sometimes developed for the staffing problem. In this case the staffing optimization step is considered alone, without other problems (see [Bassamboo and Zeevi \[2009\]](#); [Bertsimas and Doan \[2010\]](#)). [Atlason et al. \[2004, 2008\]](#); [Cezik and L'Ecuyer \[2008\]](#) share the same methodology, namely mathematical programming for both staffing and shift-scheduling problems.

On the other hand, as stated previously, staffing optimization in call centers is often studied in union with another capacity planning step, for example with forecasting in [Ding and Koole \[2014\]](#); [Gans et al. \[2012\]](#), or shift-scheduling in [Bhulai et al. \[2008\]](#); [Cezik and L'Ecuyer \[2008\]](#); [Ertogral and Bamuqabel \[2008\]](#).

The second step is the **shift-scheduling problem**. It is usually done once the staffing

requirements are known or computed. This two-step approach is extensively used, such as [Bhulai et al. \[2008\]](#); [Dietz \[2011\]](#); [Liao et al. \[2012\]](#); [Saltzman and Mehrotra \[2007\]](#). The independent two steps make the problems easier to study by enabling to select separated approaches independently. However, this convenience in the approach development leads to a lack of flexibility in the solution optimization and may lead to suboptimal solutions (see [Cezik and L'Ecuyer \[2008\]](#)).

The one-step approach involving solving the staffing problem and the shift-scheduling problem together is not very common because it is complex to solve. The published one-step approaches are usually based on an iterative algorithm: the staffing requirements are updated with the scheduling step results until a satisfactory solution is found. However, this approach does not guarantee the optimal solution is found. [Atlason et al. \[2004, 2008\]](#); [Cezik and L'Ecuyer \[2008\]](#) focus on a sample average approximation to evaluate the performance measure and a cutting plane method to select the optimal schedule solution. [Ingolfsson et al. \[2002\]](#) propose a genetic algorithm to generate the schedules, and simulate and evaluate the resulting service level. Finally, [Castillo et al. \[2009\]](#) generate several possible schedules in agreement with the demand, use a discrete-event simulation to evaluate the system performance and select among the efficient schedules the best in term of cost minimization. [Robbins and Harrison \[2010\]](#) model the one-step approach with a mathematical program, considering discrete distributions and piecewise approximations of the service level function.

Uncertainty on arrival rate is now considered as an important characteristic to consider since arrival rate variability highly affect the system performance ([Aksin et al. \[2007\]](#); [Chen and Henderson \[2001\]](#); [Gans et al. \[2003\]](#)). Modeling arrival rates is a point of interest and is investigated in [Aktekin and Soyer \[2011\]](#); [Aldor-Noiman et al. \[2009\]](#); [Avramidis et al. \[2004\]](#); [Ibrahim et al. \[2015\]](#) for example. Some complex approaches ignore uncertainty and consider that arrival rates are known (for example in [Cezik and L'Ecuyer \[2008\]](#); [Ertogral and Bamuqabel \[2008\]](#); [Koole and van der Sluis \[2003\]](#)).

[Gans et al. \[2012\]](#); [Liao et al. \[2012\]](#); [Luedtke et al. \[2010\]](#); [Nemirovski and Shapiro \[2006b\]](#); [Robbins and Harrison \[2010\]](#) consider discrete or discretized distributions in order to generate multiple scenarios. The computational complexity of the resulting problem as well as the solution accuracy depends on the number of generated scenarios. [Jongbloed and Koole \[2001\]](#) use a stochastic arrival rate as non-homogeneous Poisson.

The **risk management** is various through the approaches. When possible, a cost penalty is determined. This penalty illustrates the extra cost induced by understaffing

the system and is considered as a part of the objective function, see [Harrison and Zeevi \[2005\]](#); [Liao et al. \[2012\]](#); [Robbins and Harrison \[2010\]](#); [Roubos et al. \[2011\]](#). Another approach rely on the consideration of chance constraints to control the tolerated risk level. Whereas [Atlason et al. \[2004\]](#) deal with individual chance constraints, joint chance-constrained optimization problems are investigated in [Gurvich et al. \[2010\]](#) for example.

Uncertainty management is investigated in this thesis with a stochastic programming approach, namely chance constraints, and a distributionally robust approach.

2.1.2 Stochastic Programming with Chance Constraints

The origin of stochastic programming dates back to the 50's: [Dantzig \[1955\]](#) formulates linear programs with uncertain demand, [Beale \[1961\]](#) presents a quadratic programming approach to include uncertainty in a linear programming problem with recourse. They develop models for optimization problems where some parameters are uncertain. Probability distribution parameters are known and can be used as information in the programs.

Chance-constrained problem are usually intractable, mainly for two reasons. First, computing $\mathbb{P}(g(x, \xi) \leq 0)$ is usually hard, especially in the multidimensional case. Second, the feasible region defined by a chance constraint is generally not convex ([Ahmed and Shapiro \[2008\]](#); [Luedtke et al. \[2010\]](#)).

Thus, solving methods depend on the chance-constrained program characteristics.

Some programs do not present either of these difficulties. A typical example for this case defines the function as $g(x, \xi) = v - \xi^t x$ where ξ is a multivariate normal vector ([Prékopa \[2003\]](#)). A deterministic equivalent of the constraint $\mathbb{P}(v - \xi^t x \leq 0)$ and finally a deterministic equivalent of the chance-constrained program are given.

The second case represents problems in which only the second difficulty occurs. This is the case for problems where ξ has a finite distribution ([Dentcheva et al. \[2000\]](#)).

Third, problems comprising both difficulties are studied, for example in [Nemirovski and Shapiro \[2006b\]](#) with a scenarios approach or in [Nemirovski and Shapiro \[2006a\]](#) in which convex approximations are considered.

Finally, problems considered in this thesis belong to the case where only the first difficulty holds. Indeed, programs involving $g(x, \xi) = \xi - A^t x$, with A a deterministic matrix and ξ has a log-concave distribution, is a well-known example ([Prékopa \[1970\]](#)) where the feasible set of the chance constraints is convex. Uncertainty is considered in the Right-Hand Side only. When the convexity is not guaranteed, convex approxima-

tions (generally conservative approximations) are proposed (Ahmed and Shapiro [2008]; Dentcheva et al. [2000]; Nemirovski and Shapiro [2006a]; Shapiro [2009]).

As presented in 1.3.1, several deterministic formulations are proposed to solve optimization problems with random constraints. Considering in our problem a risk level on all the periods, an approach with a joint chance constraint rather than individual constraints is interesting. Both formulations are presented in Kall and Wallace [1995] and Prékopa [1995]. However, joint chance-constrained problems are significantly harder to solve, mainly because of the consideration of multivariate distributions. Calafiore and El Ghaoui [2006]; Lagoa [1999] consider individual chance-constrained linear programs and give deterministic quadratic reformulations.

Henrion and Strugarek [2008] give some convexity results for chance-constrained programs with independent random variables.

Since continuous distributions in chance constraints are complex to handle, discrete (or discretized) distributions are preferred. Ahmed and Shapiro [2008] consider a joint chance constraint with a convex feasible region and propose a scenario approach using Monte Carlo sampling of the random data. This sampling approximation is also used in Ruszczyński [2002], where uncertainty is represented with discrete distributions, or in Calafiore and Campi [2005]; Luedtke and Ahmed [2008]. However, sampling approximations can be computationally very demanding.

Considering discrete distributions, Dentcheva et al. [2000, 2002] develop a solution approximation by generating p -efficient points of distributions leading to a deterministic equivalent problem. However, the set of p -efficient points is not known and solving the problem requires numerical methods.

Joint chance-constrained optimization problems are investigated with discrete distributions in Luedtke et al. [2010], where a convex hull relaxation is proposed.

Aside from the sampling approximations, a common approximation for joint chance-constrained problems consists in decomposing the joint chance constraint into individual chance constraints. Nemirovski and Shapiro [2006a] decompose into individual chance constraints with equal distribution and Chen et al. [2010] propose a function to approximate decomposed joint chance constraint into individual constraints. Finally, bounding methods enable to give subproblems of the stochastic problem with convex functions. Kall and Wallace [1995] present piecewise linearizations of the recourse function of a two-stage program, Birge and Louveaux [2011]; van Ackooij et al. [2014] deterministic formulations of constraints or objective function using piecewise linearizations.

Instead of considering probabilistic constraints, it is possible to consider expectation of the random variables and easily provide a deterministic linear program (see [Kall and Mayer \[2011\]](#)). However, this measure lacks precision and the probability distribution is not very well represented in the problem. The chance-constrained approach focuses on controlling the probability of failure whereas the expectation-based approach minimizes the quantified loss. It is possible to consider simultaneously both probability and amount of loss by considering the Conditional Value-at-Risk as a quality of measure (derived from Value-at-Risk). Since this measure remain complex, it can be approximated as in [Nemirovski and Shapiro \[2006a\]](#) or bounded as in [Chen et al. \[2010\]](#).

2.1.3 Distributionally Robust Optimization

Evaluating the appropriate distribution to represent uncertainty on arrival rates is difficult ([Robbins \[2007\]](#); [Whitt \[2006\]](#)), and as stated previously, arrival rates variability has a high impact on the system performance. However, stochastic programming approaches suppose that information on the probability measure is completely known or can be estimated.

On the other hand, when lacking information on distribution parameters, the robust optimization selects the solution which respects the worst-case scenario when uncertain data is defined in a set. [Ben-Tal and Nemirovski \[1999\]](#) present a robust optimization approach for linear programs and [Bertsimas and Thiele \[2006\]](#) use historical data to elaborate a convex uncertainty set.

In stochastic programming, uncertainty is random whereas in robust optimization, uncertainty is represented in a deterministic set ([Ben-Tal et al. \[2009\]](#)). As an intermediate approach, the distributionally robust approach assume that uncertainty is random but its representation is not known. Hence, the goal is to optimize the objective function considering the most unfavorable distribution of a family of distributions.

[Shapiro and Ahmed \[2004\]](#) investigate distributionally robust stochastic problems with a worst-case expected cost and suggest a sample average approximation to solve the problems.

[Delage and Ye \[2010\]](#) propose a distributionally robust approach for data-driven problems by considering that the distribution support is known and the mean and second moment matrix values belong in a confidence region.

In a different approach of uncertainty, [Calafiore and El Ghaoui \[2006\]](#) propose a

distributionally robust for problems with individual chance constraints, considering a family of all distributions with known means and variances. Yue et al. [2006] consider the family of distributions with known mean and variances as well, and model the newsvendor problem constrained by a maximum expected difference cost between a wrong decision and the optimal decision under the true distribution.

Regarding the specific scheduling problem in call center, Liao et al. [2013] tackle the distributionally robust approach. The χ^2 statistic is used to build a class of possible discrete distributions around the gamma distribution, with a confidence set around estimated values and rely on scenarios generation with Monte Carlo sampling.

2.2 Work Positioning

Regarding the summary of the state of the art above, we give here a quick positioning of our contribution presented in 1.4 for the staffing and scheduling problems under uncertainty.

In this thesis, the call center model is standard: agents do not have specific skill and both $M/M/c$ and $M/M/c + M$ FCFS queues are considered. Abandoned calls are included in the second queuing model, however we ignore returned calls or retrials. The representation of the uncertainty is complex and a standard call center model makes its consideration easier.

Our approach does not rely on simulations. The call center is considered in stationary regime with the widely used SIPP model to be able to provide mathematical expressions of the performance evaluation through Erlang formulae: we focus on staffing the call center in order to respect a maximum expected waiting time. The staffing and shift-scheduling problems are solved together in a one-step approach.

In the first part of the thesis, uncertainty of input parameters is handled with a stochastic program where arrival rates are random variables following a continuous distribution. Uncertainty is only in the RHS. Unlike most published approaches, we do not discretize the random variable support and deal with continuous variables in the stochastic program.

The second part of the thesis is devoted to the distributionally robust model, considering the arrival rate distribution belongs to the set of all possible distributions with a known mean and known variance. Upper and lower approximations are given.

Both approaches of uncertainty rely on joint chance-constrained programs and an

equivalent decomposition into individual chance constraints.

Chapter 3

Stochastic Approaches for Staffing and Shift-Scheduling Problems in One Step

This chapter presents the staffing and shift-scheduling problems in call centers and propose an approach in one step, instead of solving independently the staffing problem and the shift-scheduling problem. It consists in determining the minimum-cost number of agents to be assigned to each shift on the scheduling horizon so as to reach the chosen customer Quality of Service. The one-step approach is based on the construction of a continuous relation between the call arrival rates and the number of agents required to assure the Quality of Service.

The numbers of calls considered in the problem are forecasts and thus subject to uncertainty. The time horizon is composed of several periods in which the mean call arrival rate is a random variable following a continuous distribution. We model the resulting optimization problem as a stochastic program involving a joint chance constraint. The risk level is set on the entire horizon and is dynamically distributed through the different periods in the optimization process. A linear approximation based approach gives approximated solutions providing lower and upper bound of the optimal solution cost.

Numerical experiments on a real-life instance are done in order to investigate the quality of our approach. Moreover, a comparison study with other approaches shows that our proposed approach compares well with previously published approaches both in terms of risk management and cost minimization. Similarly, one-step and two-step approaches are compared.

3.1 Introduction

Staffing problem and shift-scheduling problem are two steps of workforce planning in a call center. The staffing problem involves computing the minimum number of agents, i.e. personnel answering the phone, required to respect a desired Quality of Service. The computations are based on forecasts of historical data, in which usually the main (and often only) information available is the number of calls per period. Typical queuing system models are used in order to provide relations between the different parameters: arrival rates, number of agents and service level. The shift-scheduling problem rely on computations of the staffing process. This is an optimization problem consisting in minimizing the manpower costs by scheduling enough agents during a given period, one day or one week for example, with respect to both the staffing requirements and inherent constraints of manpower work. The goal here is to assign the working agents to established shifts.

The staffing computations are based on queuing system models. Typically, Erlang-C model describes a simple queue and provides interesting results. The model considers a stationary regime where customers are arriving in an infinite queue, waiting to be served by some agents during some time. The Erlang-C model does not consider customers leaving the queue (abandonments), or reaching a busy signal because of high influx.

First, forecasts are estimated in order to provide plausible values close to what could really occur, but they still are inherently uncertain. Thus forecasts of call arrival rates are considered to be random variables. In our approach, uncertainty management is done through continuous distributions, namely normal distributions, which we decide not to discretize.

However, since the staffing computations remain mathematically complex because of the complex queuing model formulae, uncertainty management, staffing computations and scheduling optimization are usually done independently. In this chapter, we propose a formulation enabling to solve the whole problem in one step. The formulation is based on the idea of continuous relations between the parameters of the queue in order to tend to limit approximations that may be unintentionally introduced at each resolution step and are difficult to quantify.

Second, due to the queuing model, time horizon is split in several time periods. We consider here that the random variables at each period are independent. The risk management is done through joint chance constraint: the stochastic program is formulated

with chance constraints joined with a global risk level. This risk level is dynamically shared among the different chance constraints: its sharing is not predefined but is done during the optimization process.

This approach results in a mixed integer non-linear program for the staffing and shift-scheduling problem.

Finally, linear approximations of this program are computed in order to propose easily solvable mixed integer linear programs which provide schedule solutions giving an upper bound and a lower bound of the optimal cost.

Numerical experiments are done in order to compare our approach with several other approaches: the two-step approach, different chance-constrained approaches, and simpler approaches.

Section 3.2 introduces the detailed context of the problem. Section 3.3 describes the two-step approach based on chance constraints. Then, Section 3.4 develops our one-step approach and Section 3.5 gives linear approximations of the resulting program. Sections 3.6 and 3.7 give numerical experiments and show the qualities of our approach. Finally, Section 4.2 presents the extended formulation of the problem using Erlang-A and gives some numerical results.

3.2 Problem Formulation

We propose here to describe and solve a mixed integer non-linear stochastic program with joint chance constraint able to solve a joint staffing and shift-scheduling problem.

The call center is single-skill multi-shift and no call abandonments are considered. Agents in the call center are capable of answering every type of calls. The values of call arrival rates are considered to be forecasts obtained from historical data and are subject to uncertainty. This uncertainty is taken into consideration with a chance-constrained optimization program and the risk management is carried by joint constraints.

The optimization process does not allow recourse. The resulting formulation is a one-stage program.

3.2.1 Queue Model

The call center is represented as a simple queue:

In the queue, several parameters are to be considered. The arrival process is assumed

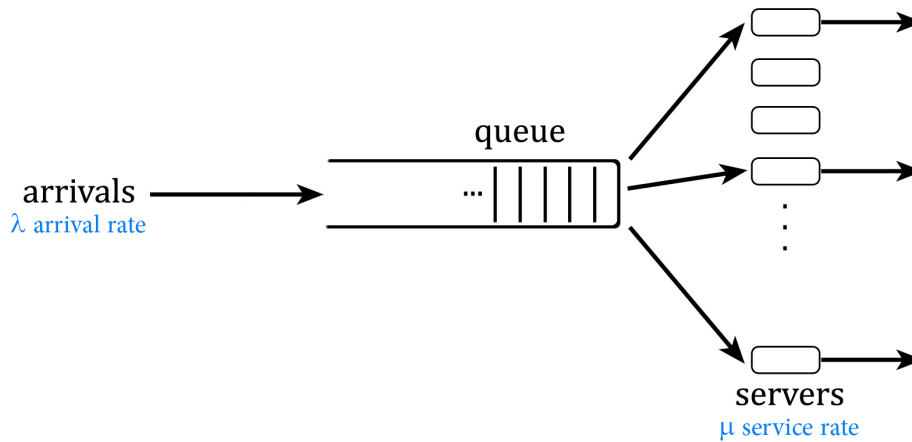


Figure 3-1: A simple $M/M/c$ queuing system.

to be Poisson with known mean arrival rate λ . Clients are served by c servers, which are here agents. Service times are assumed to be independent and exponentially distributed with rate μ . Arrival rates and servers numbers are varying in time, then the queue is a non-stationary queue $M(t)/M/c(t)$.

Calls are served in the order of their arrivals, i.e. under the First Come-First Served (FCFS) discipline of service. Calls are not sorted and agents are able to answer to any call. This is a single-skill call center.

The queue capacity is assumed to be infinite, customers never get a busy signal. Finally, customer abandonments, retrials as well as returns are ignored.

3.2.2 Staffing Problem

The problem consists in deciding how many agents should be assigned to work each day at the call center. As explained in previous Section 3.2.1, the call center model is a queue with varying arrival rates. The call center is approximated with SIPP (Stationary Independent Period-by-Period) model in order to use computations based on considerations of a Quality of Service, characterized here by the time a customer spends waiting in the queue, which requires to consider the queue in steady state. Hence, the stationary regime is considered to be reached so that we are able to use the probability distribution of the waiting time given by Erlang-C model, which is a random variable.

In order to represent the variation of rate with the best precision possible, the work-

ing horizon is split in small periods of time t during which the queue is assumed to be in steady state and the arrival rate is considered as known. These small periods are usually 15-minute or 30-minute periods. Finally the call center is considered as $M_t/M/c_t$ where stationary regime is reached and values of expected call arrival rates are obtained for each period t . This describes the well-known Erlang-C model.

Expectancy of the waiting time random variable is given by (see [Gans et al. \[2003\]](#) for details):

$$\begin{aligned}
 EWT(c, \lambda, \mu) &= \mathbb{E}[\text{Wait}] \\
 &= \mathbb{P}(\text{Wait} > 0) \mathbb{E}[\text{Wait} | \text{Wait} > 0] \\
 &= \frac{1}{c\mu(1 - \frac{\lambda}{c\mu}) \left(1 + (1 - \frac{\lambda}{c\mu}) \sum_{m=0}^{c-1} \frac{c!}{m!} (\frac{\mu}{\lambda})^{c-m} \right)}. \tag{3.1}
 \end{aligned}$$

The quantity is denoted EWT for Expected Waiting Time. This formula gives the average waiting time for a call before being answered according to the parameters of the queue: the service rate μ , the arrival rate λ and the number of servers c .

Note In this relation λ and μ are real numbers whereas c is an integer.

We denote EWT^* our objective service level, i.e. the maximum accepted EWT value. The EWT formula will be used in next sections in order to find a relation between the forecasted values of arrival rates and values of agents requirements that should be determined. Practically, minimum values of requirements for each period will be computed in order to respect the targeted value EWT^* and according to forecasts of arrival rates.

It is considered that once staffing is done for the time horizon, it is not possible to change it even if the actual situation differs from the forecasts. Moreover, updating the workforce load after decision can not be allowed. In terms of optimization in the scheduling problem, recourse is not allowed.

Next section presents the shift-scheduling which consists in organizing the call center workforce through different possible shifts.

3.2.3 Shift-Scheduling Problem

The problem consists in minimizing the workforce cost while ensuring the staffing requirements are respected for each period of time.

The shift-scheduling problem is presented as a mixed-integer linear program as introduced in Section 1.2.2

Agents are assigned to pre-determined possible shifts and follow typical full-time or part-time schedules. This states a multi-shift call center model. These shifts are made up of working hours and breaks, for lunch for example.

Under parameters uncertainty, the deterministic formulation is not enough anymore, and consideration of random variables will be introduced in the linear problem in next sections.

3.2.4 Costs

Considering the call center requirements are subject to uncertainty, a risk level ϵ is set. This risk level indicates how conservative the solution should be and is expressed as a probability on the constraints.

Instead of considering penalty costs, the optimization program presented in next sections focuses on controlling the risk of understaffing the call center while limiting the salary costs.

Agents are considered to be equally paid according to the number of worked hours. The cost of one agent is therefore proportional to the number of periods worked. Thus, it depends on the shift the agent works on.

3.2.5 Uncertainty Management with Chance Constraints

As explained previously, mean arrival rates are forecasts computed with historical data. Thus, in order to take account of uncertainty, they are considered as random variables following continuous probability distributions. The random variables computed at different periods of the forecast time horizon are considered to be independent.

Let the problem be on one period. The chance constraint makes sure that the staffed agents x_t meet our requirement condition b_t under uncertainty. The probability of understaffing should not exceed the targeted risk level ϵ . In other words, the probability of staffing enough agents should satisfy $\mathbb{P}(x_t \geq b_t) \geq 1 - \epsilon$. The term $1 - \epsilon$ is the qual-

ity interval. The value of staffing x_t will be determined in order to respect this chance constraint, where b_t is the random variable of requirements.

The risk level represents the portion of the set of all possible situations for which understaffing is tolerated. The lower the risk level, the more conservative the solution and the greater the cost.

In the program, one period of time makes one constraint. Periods are considered as independent from each other. Using the previous chance-constrained formulation, it is possible to write an optimization program. As said previously in Section 3.2.3, the schedule is multi-shift. Then agents will be assigned to possible shifts through the shifts matrix A . The term $A_t x$ in the constraint takes over the simple term x_t . The term x is the vector of agents assigned to each shift and A_t is the row vector of assignments for period t . The vector c is the cost vector.

This leads to an individual chance-constrained model:

$$\begin{aligned} \min \quad & c^t x \\ \text{s.t.} \quad & \forall t \in \llbracket 1; T \rrbracket, \mathbb{P}(A_t x \geq b_t) \geq 1 - \varepsilon \\ & x \in (\mathbb{Z}^+)^S. \end{aligned} \quad (3.2)$$

This program minimizes the manpower costs under the constraints that probabilities of reaching the requirements at each period is greater than the quality interval $1 - \varepsilon$, with $\varepsilon \in]0; 1]$. The number of shifts is S .

This program states an equal risk on each period of time. It is possible to consider different risks on the different periods, by considering different probability values. Uncertainty can be expressed on each period with a given quality interval p_t for period t :

$$\begin{aligned} \min \quad & c^t x \\ \text{s.t.} \quad & \forall t \in \llbracket 1; T \rrbracket, \mathbb{P}(A_t x \geq b_t) \geq p_t \\ & x \in (\mathbb{Z}^+)^S. \end{aligned} \quad (3.3)$$

However, it will be required to know the risk level for each period. Considering the periods are very short, typically 15 or 30 minutes, it is laborious to evaluate the proper risk for each of them.

In order to deal with a global problem, this risk will be set for the entire horizon (for example one week or one month). Practically, it is more convenient to decide the risk

level for the whole horizon rather than period by period.

The global risk level is expressed through a joint chance-constrained approach:

$$\mathbb{P}(Ax \geq B) \geq 1 - \varepsilon . \quad (3.4)$$

This one constraint deals with uncertainty on the whole horizon of study. The shifts matrix A is introduced in the formulation, $A \in \mathcal{M}_{T,S}(\{0,1\})$. The value S is the number of shifts and T is the number of periods. The term x is the vector of agents assigned to each shift. Thus Ax is the vector defining the number of agents working at each period. The vector B is the vector of the staffed agents random variables.

The joint chance-constrained program is then:

$$\begin{aligned} \min \quad & c^t x \\ \text{s.t.} \quad & \mathbb{P}(Ax \geq B) \geq 1 - \varepsilon \\ & x \in (\mathbb{Z}^+)^S . \end{aligned} \quad (3.5)$$

This program optimizes the cost of hired agents under the constraint that the probability of reaching the requirements for the whole schedule is greater than the quality interval $1 - \varepsilon$.

Since the random variables B_t are supposed to be independent variables, the chance constraint can be considered as a product of the T probabilities:

$$F_B(Ax) = \prod_{t=1}^T F_{B_t}(A_t x) . \quad (3.6)$$

Finally

$$\begin{aligned} \min \quad & c^t x \\ \text{s.t.} \quad & \prod_{t=1}^T F_{B_t}(A_t x) \geq 1 - \varepsilon \\ & x \in (\mathbb{Z}^+)^S . \end{aligned} \quad (3.7)$$

Next Sections 3.3 and 3.4 propose two approaches in order to express F_{B_t} function. Moreover the formulations aim at handling the product of probabilities in order to turn the one constraint into several independent constraints.

3.3 Solving the Problem in Two Steps

This section develops the previous stochastic program with a two-step approach of the problem:

- First step: the staffing problem solved with computations based on Erlang-C model formulae.
- Second step: the shift-scheduling problem expressed through the stochastic optimization program using data of the first step.

3.3.1 Uncertainty Management

As explained in Section 1.2.3, values of agents requirements are computed with the help of arrival rates forecasted values. They are thus random variables. However, deducing the probability distribution of agents requirements from the arrival rates distribution is obviously complex because of the queuing model. Indeed, no analytical expression between agents requirements and arrival rates is known.

Then in this section, agents requirements are assumed to follow normal distributions whose means are the computed value of requirements. This allows to express the constraints with a known cumulative distribution function.

3.3.2 Staffing Model

Section 3.2.2 defines EWT function based on Erlang-C model. The staffing load is the resulted minimum value of c which respects our service level for given arrival rate λ and service rate μ computed through the EWT function. This minimum value at period t is denoted by c_t :

$$c_t = \min\{c \in |EWT(c, \lambda_t, \mu) \leq EWT^*\} . \quad (3.8)$$

Note In this relation EWT^* defines the maximum value of Expected Waiting Time accepted.

3.3.3 Agents Requirements as Random Variables

Agents requirements are random variables following normal distributions whose means are computed through Erlang-C model. At each period, c_t denotes the expected value

and σ_t^2 the variance of the distribution. Let β denote the standard normal deviate. Using the standardized normal distribution, the previous program (3.7) presented in Section 3.2.5 is equivalent to

$$\begin{aligned} \min \quad & c' x \\ \text{s.t.} \quad & \prod_{t=1}^T F_{\beta} \left(\frac{A_t x - c_t}{\sigma_t} \right) \geq 1 - \varepsilon \\ & x \in (\mathbb{Z}^+)^S. \end{aligned} \quad (3.9)$$

Under the assumption of periods independence, the global risk level can be equally split through the periods, according to the Bonferroni approximation, so as to get back to the individual chance-constrained program (3.3), where $p_t = (1 - \varepsilon)^{\frac{1}{T}}$.

The program would be

$$\begin{aligned} \min \quad & c' x \\ \text{s.t.} \quad & \forall t \in \llbracket 1; T \rrbracket, F_{\beta} \left(\frac{A_t x - c_t}{\sigma_t} \right) \geq (1 - \varepsilon)^{\frac{1}{T}} \\ & x \in (\mathbb{Z}^+)^S. \end{aligned} \quad (3.10)$$

However, unlike the individual probabilities program, this approach considers a global risk level which is split among the periods. The practical interest of this approach is the consideration of only one risk level for the whole horizon. Nevertheless, even with non-equal proportions, this approach requires to choose in advance the sharing out of the risk through the different periods. Hence, the same drawback appears.

It is possible to propose another approach which allow to consider this sharing out of the risk as a part of the optimization process. Indeed, instead of considering the fixed value $\frac{1}{T}$, the proportion of the risk for each period will be a variable of the optimization program. Let the variable y_t denote this proportion at period t and consider its value is between 0 and 1. The sum of all the variables y_t should be 1 in order to respect our global risk level.

The program can be expressed as follows:

$$\begin{aligned}
& \min c'x \\
& \text{s.t. } \forall t \in \llbracket 1; T \rrbracket, \frac{A_t x - c_t}{\sigma_t} \geq F_{\beta}^{-1}((1 - \varepsilon)^{y_t}) \\
& \sum_{t=1}^T y_t = 1 \\
& x \in (\mathbb{Z}^+)^S, \forall t \in \llbracket 1; T \rrbracket, y_t \in]0; 1[.
\end{aligned} \tag{3.II}$$

In this program, the right-hand side is a non-linear function of y_t .

This program can be approximated in order to be solved. However, this approach relies on the assumption that agents requirements are random variables following normal distributions. Still, it is difficult to ascertain an adequate distribution for the agent requirement variables, for they are the result of a complex computation through Erlang-C model using call arrival rates as random variables. Making an assumption on the distribution of the arrival rates would be easier to comprehend. As a consequence, the distribution of agent requirements becomes unknown. Moreover, this two-step approach may introduce some approximations between the staffing resolution and the shift-scheduling optimization. Solving the staffing problem and the shift-scheduling problem altogether enables to consider uncertainty on the arrival rates without making assumption on the agents requirements. This new approach is developed in next section.

3.4 Staffing and Shift-Scheduling Problem in a One-Step Approach

In this section, the staffing problem and shift-scheduling problem are solved together in one step, as an optimization program dealing with staffing computations. Assumption of uncertainty is made on arrival rates and information of the distribution parameters are known. The distribution of agents requirements random variable is expressed through a composition of the distribution of arrival rate and a continuous relation between the two random variables. The continuous approach enables to minimize unvoluntary approximations and thus ensure a better precision in the uncertainty management.

Back to the general approach introduced by (3.7), we apply again the flexible sharing out of the risk:

$$\begin{aligned}
 & \min c^t x \\
 & \text{s.t. } \forall t \in \llbracket 1; T \rrbracket, F_{B_t}(A_t x) \geq (1 - \epsilon)^{y_t} \\
 & \sum_{t=1}^T y_t = 1 \\
 & x \in (\mathbb{Z}^+)^S, \forall t \in \llbracket 1; T \rrbracket, y_t \in]0; 1[.
 \end{aligned} \tag{3.12}$$

Next sections focus on giving an expression for F_{B_t} .

3.4.1 New Staffing Model: Relation between arrival rates and requirements

Arrival rates are now considered as random variables following normal distributions. In order to keep the best precision possible in the process, a continuous relation between the agent requirement values and the arrival rates is introduced. The distribution of agent requirement variables remains unknown because of the complexity of the computation.

The following function ψ gives a continuous relation between b and λ :

$$\begin{aligned}
 \psi: \mathbb{R}^+ & \rightarrow \mathbb{R}^+ \\
 \lambda & \mapsto \psi(\lambda) = b(\lambda, EWT^*, \mu) .
 \end{aligned} \tag{3.13}$$

The function ψ gives the minimum number of agents b required to ensure that the targeted QoS EWT^* is reached when the call arrival rate is λ and the expected service rate is μ . The chosen QoS is the Expected Waiting Time (EWT). The computed value of b is a real number and not an integer, which lead the linear approximations in the next sections.

To the best of our knowledge, no analytical expression computing ψ is known. However, for a given value of λ , we propose the following algorithm in order to compute $b = \psi(\lambda)$.

Erlang-C model's function is reminded here:

$$EWT(c, \lambda, \mu) = \mathbb{E}[\text{Wait}] = \frac{1}{c\mu(1 - \frac{\lambda}{c\mu}) \left(1 + (1 - \frac{\lambda}{c\mu}) \sum_{m=0}^{c-1} \frac{c!}{m!} (\frac{\mu}{\lambda})^{c-m} \right)} .$$

This formula gives the expectation of waiting time (EWT) given the arrival rate λ , the service rate μ and the number of servers c which is an integer. In order to consider ψ as function of a positive real value of b , the algorithm below is formulated:

- Computation of $EWT(c, \lambda, \mu)$ and $EWT(c + 1, \lambda, \mu)$ such that

$$EWT(c, \lambda, \mu) \geq EWT^* \text{ and } EWT(c + 1, \lambda, \mu) < EWT^* .$$

$EWT(c, \lambda, \mu)$ is denoted as $EWT_{c, \lambda}$.

- The real value of c is computed by a linearization in the $[EWT_{c, \lambda}; EWT_{c+1, \lambda}]$ segment. The affine function is:

$$EWT^* = (EWT_{c+1, \lambda} - EWT_{c, \lambda})b + (c + 1)EWT_{c, \lambda} - cEWT_{c+1, \lambda}$$

and the value b solving this equation is the real value $\psi(\lambda)$ that is looked for. \square

Using this algorithm for the value of λ which is considered in the ψ function, the requirements b is finally obtained:

$$\psi(\lambda) = b = \frac{EWT^* + cEWT_{c+1, \lambda} - (c + 1)EWT_{c, \lambda}}{EWT_{c+1, \lambda} - EWT_{c, \lambda}} . \quad (3.14)$$

The function ψ determines the values of b as a function of λ , μ and the targeted EWT^* . In a nutshell, it determines the number of agents required to deal with the arrival rates λ with respecting a Quality of Service previously defined. This function is strictly increasing.

The cumulative distribution function of B can then be denoted by

$$F_B(b) = F_\Lambda(\psi^{-1}(b)) . \quad (3.15)$$

3.4.2 Convexity of $y \mapsto F_B^{-1}(p^y)$

The cumulative distribution function was previously defined by $F_B(b) = F_\Lambda(\psi^{-1}(b))$.

Then with this notation, it follows that $F_B(b) = F_\Lambda(\psi^{-1}(b)) = 1 - \varepsilon$ and so $F_B^{-1}(1 - \varepsilon) = \psi(F_\Lambda^{-1}(1 - \varepsilon))$.

In the problem, the risk $1 - \varepsilon$ is split. Since $1 - \varepsilon$ represents a probability, let it be called p in this part. It is obvious that a high quality interval is wanted in the problem,

since the practical interest of uncertainty management is to provide control on the situation. Thus a small value of ε is considered. From here the quality interval is considered so that $p > 0.5$, which is necessary for convexity.

In the optimization program, the inverse of the cumulative distribution function is applied on $p^y, y \in]0; 1[$. So with $y \in]0; 1[$, the following equality holds:

$$F_B^{-1}(p^y) = \Psi(F_\Lambda^{-1}(p^y)) .$$

Lemma $y \mapsto F_\Lambda^{-1}(p^y)$ is convex.

Proof Since $f : y \mapsto p^y$ is convex decreasing and $g : p \mapsto F_\Lambda^{-1}(p)$ is convex for $p > 0.5$ and increasing, thus $y \mapsto F_\Lambda^{-1}(p^y)$ is convex. \square

This previous result and the strictly increasing function $\lambda \mapsto \Psi(\lambda)$ helped to note that $y \mapsto F_B^{-1}(p^y)$ is a decreasing quasiconvex function. Thereafter, an approximated convex function of $y \mapsto F_B^{-1}(p^y)$ is considered.

3.4.3 Final Stochastic Program

The final mixed integer non-linear stochastic program is the following:

$$\begin{aligned} \min \quad & c^t x \\ \text{s.t.} \quad & \forall t \in \llbracket 1; T \rrbracket, A_t x \geq \Psi(F_\Lambda^{-1}((1 - \varepsilon)^{y_t})) \\ & \sum_{t=1}^T y_t = 1 \\ & x \in (\mathbb{Z}^+)^S, \forall t \in \llbracket 1; T \rrbracket, y_t \in]0; 1[. \end{aligned} \tag{3.16}$$

This chance-constrained program minimizes the costs in term of agents under the constraint of respecting a global risk level which is dynamically shared through all the periods.

In order to solve it, a piecewise linearization of the Right-Hand Side is proposed in next section.

3.5 Approximated Solutions

3.5.1 Piecewise Linear Approximation

In this section we give an upper approximation of $y \mapsto F_{B_t}^{-1}(p^y)$.

Let $y_j \in]0; 1[$, $j \in \llbracket 1; n \rrbracket$ be n points such that $y_1 < y_2 < \dots < y_n$.

Let us denote $\hat{F}_{B,j}^{-1}$ the linearized secant approximation of F_B^{-1} between y_j and y_{j+1} .

$$\begin{aligned} \forall j \in \llbracket 1; n-1 \rrbracket, \hat{F}_{B,j}^{-1}(p^y) &= F_B^{-1}(p^{y_j}) + \frac{y - y_j}{y_{j+1} - y_j} (F_B^{-1}(p^{y_{j+1}}) - F_B^{-1}(p^{y_j})) \\ &= \delta_j y + \alpha_j. \end{aligned} \quad (3.17)$$

Since $F_B(b) = F_\Lambda(\Psi^{-1}(b))$, we have $\forall p \in]\frac{1}{2}; 1[$, $F_{B,j}^{-1}(p) = \Psi(F_{\Lambda,j}^{-1}(p))$, thus the coefficients are:

$$\begin{cases} \delta_j = \frac{\Psi(F_{\Lambda,j}^{-1}(p^{y_{j+1}})) - \Psi(F_{\Lambda,j}^{-1}(p^{y_j}))}{y_{j+1} - y_j} \\ \alpha_j = \Psi(F_{\Lambda}^{-1}(p^{y_j})) - y_j \delta_j. \end{cases} \quad (3.18)$$

Owing to the convexity of the approximation, it is guaranteed that each segment is the greatest one on its own interval.

Thus our approximated program is:

$$\begin{aligned} \min \quad & c^t x \\ \text{s.t.} \quad & \forall t \in \llbracket 1; T \rrbracket, \forall j \in \llbracket 1; n-1 \rrbracket, A_t x \geq \alpha_{u,t,j} + \delta_{u,t,j} y_t \\ & \sum_{t=1}^T y_t = 1 \\ & \forall i \in \llbracket 1; S \rrbracket, x_i \in \mathbb{Z}^+, \\ & \forall t \in \llbracket 1; T \rrbracket, y_t \in]\alpha_1; 1[\end{aligned} \quad (3.19)$$

with n points for linear secant approximation of $F_{B_t}^{-1}$ with $(\alpha_{u,t,j}, \delta_{u,t,j})$ coordinates (u stands for *upper*). S is the number of shifts and T the total number of periods. The risk is given by $\varepsilon \in]0; \frac{1}{2}[$.

3.5.2 Piecewise Tangent Approximation

Let us now develop a lower approximation of $y \mapsto F_{B_t}^{-1}(p^y)$.

Let $y_j \in]0; 1$, $j \in \llbracket 1; n \rrbracket$ be n points such that $y_1 < y_2 < \dots < y_n$.

We apply a first-order Taylor series expansion around these n tangents points. Let us denote $\hat{F}_{B,j}^{-1}$ the linearized approximation of F_B^{-1} around y_j . Then

$$\begin{aligned} \forall j \in \llbracket 1; n \rrbracket, \hat{F}_{B,j}^{-1}(p^y) &= F_B^{-1}(p^{y_j}) + (y - y_j)(F_B^{-1})'(p^{y_j}) \ln(p) p^{y_j} \\ &= \delta_j y + \alpha_j \end{aligned} \quad (3.20)$$

$$\text{with } (F_B^{-1})'(p^{y_j}) = \frac{1}{F_B'(F_B^{-1}(p^{y_j}))} = \frac{1}{f_B(F_B^{-1}(p^{y_j}))}$$

And since

$$f_b(b) = \frac{f_\Lambda(\Psi^{-1}(b))}{\Psi'(\Psi^{-1}(b))},$$

according to the definition of the composite function of random variables :

$$\begin{aligned} f_B(F_B^{-1}(p^{y_j})) &= \frac{f_\Lambda(\Psi^{-1}(F_B^{-1}(p^{y_j})))}{\Psi'(\Psi^{-1}(F_B^{-1}(p^{y_j})))} = \frac{f_\Lambda(\Psi^{-1}(\Psi(F_\Lambda^{-1}(p^{y_j}))))}{\Psi'(\Psi^{-1}(\Psi(F_\Lambda^{-1}(p^{y_j}))))} \\ &= \frac{f_\Lambda(F_\Lambda^{-1}(p^{y_j}))}{\Psi'(F_\Lambda^{-1}(p^{y_j}))}. \end{aligned}$$

Hence, the coefficients are:

$$\begin{cases} \delta_j = \ln(p) p^{y_j} \frac{\Psi'(F_\Lambda^{-1}(p^{y_j}))}{f_\Lambda(F_\Lambda^{-1}(p^{y_j}))} \\ \alpha_j = \Psi(F_\Lambda^{-1}(p^{y_j})) - y_j \delta_j. \end{cases} \quad (3.21)$$

Like previously, we can assure that these piecewise tangent approximations are always below the curve.

Similarly as before, each linearized straight line has the greatest values on its interval of interest, thus only one constraint for each segment is required:

$$\hat{F}_B^{-1}(p^y) = \max_{j \in \llbracket 1; n \rrbracket} \{F_{B,j}^{-1}(p^y)\}. \quad (3.22)$$

Finally the piecewise tangent approximated program is:

$$\begin{aligned}
& \min c^t x \\
& \text{s.t. } \forall t \in \llbracket 1; T \rrbracket, \forall j \in \llbracket 1; n \rrbracket, A_t x \geq \alpha_{l,t,j} + \delta_{l,t,j} y_t \\
& \quad \sum_{t=1}^T y_t = 1 \\
& \quad \forall i \in \llbracket 1; S \rrbracket, x_i \in \mathbb{Z}^+, \forall t \in \llbracket 1; T \rrbracket, y_t \in]0; 1[.
\end{aligned} \tag{3.23}$$

with n points for tangent approximation with $(\alpha_{l,t,j}, \delta_{l,t,j})$ coordinates (l stands for *lower*). Again, $\varepsilon \in]0; \frac{1}{2}[$.

In this section, deterministic approximations of the initial stochastic problem are developed through piecewise tangent and linear secant approximations. This leads to consider upper bound and lower bound programs as mixed integer linear programs.

These two programs are easily computed with an optimization software (CPLEX for example). Next section gives results on numerical experiments.

Note Computations of $p \mapsto F_{\Lambda}^{-1}(p)$ were done with the coefficients of Gauss error function according to Takuya Ooura (see [Ooura \[2006\]](#)).

3.6 Numerical Experiments for the One-Step Approach

3.6.1 Instance

The previously developed approach gives an upper bound and a lower bound of the problem's objective function. In this section, the programs will be applied to an instance from a health insurance call center. A matrix of 19 different shifts is considered, with both full-time and part-time work hours and the scheduling is made for one week (5.5 days, from Monday to Saturday midday).

The time horizon is split into 30-minute periods, considering 10 hours a day from Monday to Friday and 3.5 hours for Saturday morning, which gives 107 periods.

The agents are assumed to be paid according to the number of worked hours. The cost of one agent is proportional to the number of periods worked. Thus it depends on the shift the agent works on. Here the cost is set to 1 for the fullest shifts (with the greatest number of periods) and the costs of other shifts are a strict proportionality of the number of worked periods.

The data used to staff and schedule are arrival rates varying between 6 calls/min and 86 calls/min, following a typical daily seasonality, as described in Gans et al. [2003]. Let $\forall t \in \llbracket 1; T \rrbracket, \lambda_t$ denote the mean of the T random variables, which are the given data. The variances σ_t^2 are random values generated in $[\frac{\lambda_t}{4}; \frac{\lambda_t}{2}]$.

3.6.2 Comparison with Other Chance-Constrained Approaches

The previous instance is applied to the upper and lower programs (3.19) and (3.23) and, as a comparison, to the following approaches.

Joint Chance Constraint with Equal Sharing out of the Risk

As presented in Section 3.3.3, here is the program in which we equally divided the risk through the periods in a pre-treatment:

$$\begin{aligned} \min c^t x \\ \text{s.t. } \forall t \in \llbracket 1; T \rrbracket, A_t x \geq F_{B_t}^{-1} \left((1 - \varepsilon)^{\frac{1}{T}} \right) \\ x \in (\mathbb{Z}^+)^S. \end{aligned} \quad (3.24)$$

Individual Chance-Constrained Approach

The individual chance-constrained approach focuses on reaching the given risk level at each period, not through the whole horizon:

$$\begin{aligned} \min c^t x \\ \text{s.t. } \forall t \in \llbracket 1; T \rrbracket, \mathbb{P}(A_t x \geq b_t) \geq 1 - \varepsilon \\ x \in (\mathbb{Z}^+)^S. \end{aligned} \quad (3.25)$$

Thus we got this final program:

$$\begin{aligned} \min c^t x \\ \text{s.t. } \forall t \in \llbracket 1; T \rrbracket, (A_t x) \geq F_{B_t}^{-1} ((1 - \varepsilon)) \\ x \in (\mathbb{Z}^+)^S. \end{aligned} \quad (3.26)$$

Deterministic Approach

Finally, the simplest approach is the deterministic approach.

$$\begin{aligned}
 & \min c^t x \\
 & \text{s.t. } \forall t \in \llbracket 1; T \rrbracket, A_t x \geq b_t \\
 & \quad x \in (\mathbb{Z}^+)^S.
 \end{aligned} \tag{3.27}$$

The values of b_t here are computed with the Erlang-C formula using the mean forecasted values, considered as certain.

3.6.3 Results

All the programs are tested on the instance with the following parameters:

- $\mu = 1$
- $EWT^* = 0.4$
- $\varepsilon = 0.05$

Table 3.1 shows the solutions for staffing and shift-scheduling of this instance for 5 programs: column 1 gives the shift, column 2 (Deter) presents the x vector obtained with the deterministic model (3.27), column 3 (Indiv) gives the results with the individual chance constraints model (3.26) and column 4 (Fixed) with the fixed-risk model (3.24). Finally, columns 5 (LowerB) and 6 (UpperB) present the results obtained with the lower bound (3.23) and the upper bound (3.19) approximations.

Computations of both lower and upper bounds were done with 5 interpolation points. The cost gap between the two bounds is $D = 3.3\%$.

In order to check the efficiency of these solutions, 100 scenarios were randomly generated according to the historical data we previously used. The feasibility of the 5 solutions was investigated. If the number of agents scheduled in at least one period of a scenario is insufficient, the latter is considered as violated.

In Table 3.2 are the results for a batch of 100 scenarios. JCC stands for *Joint Chance Constraint*.

Figure 3-2 plots for each model and each scenario the difference between the scheduled number of agents by the precited solutions and the number of agents actually needed

Table 3.1: Scheduling result on each shift.

Shift	Deter	Indiv	Fixed	LowerB	UpperB
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
6	2	5	6	3	4
7	0	0	0	0	0
8	0	0	0	0	0
9	0	0	0	0	0
10	0	0	0	0	0
11	26	29	25	31	36
12	20	21	22	18	21
13	9	12	15	14	16
14	0	0	2	3	0
15	12	12	13	12	10
16	11	13	14	14	15
17	0	0	6	0	0
18	8	9	7	9	6
19	7	8	12	9	9
Total	95	109	122	113	117
Cost function	94.44	107.6	120.3	112.2	115.9

Table 3.2: Percentage of violated scenarios.

Model	% of violation
Deterministic model	100%
Individual chance constraint model	42%
JCC model with fixed risk level	0%
JCC lower bound and flexible risk	4%
JCC upper bound and flexible risk	2%
Targeted maximal risk	5%

$b_{present} - b_{needed}$. When this value is negative for at least one period, the model is invalidated for the scenario.

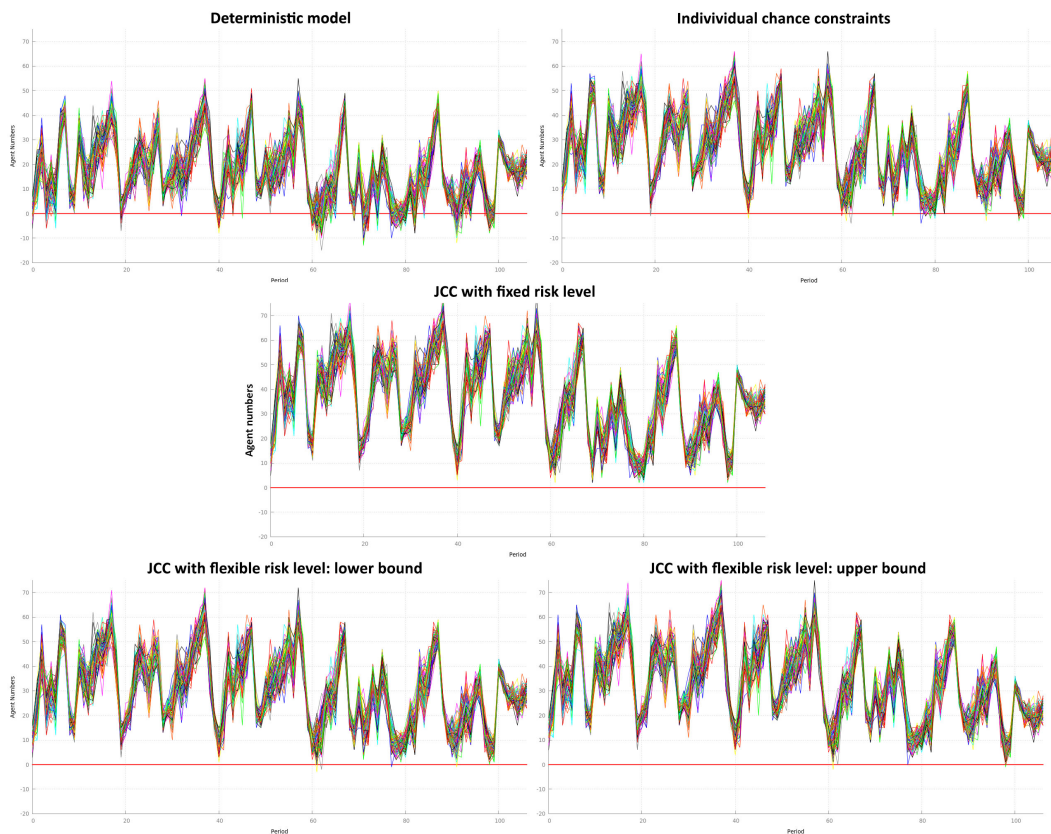


Figure 3-2: Violated scenarios for each model.

First, the schedule solutions obtained by using the various models should be compared in term of efficiency based on the total staffing costs.

Table 3.1 shows that the deterministic (3.27) and individual (3.26) models provide less expensive schedules than the joint chance constraint models (3.19), (3.23) and (3.24).

However the approximated programs where the risk is dynamically divided through the periods provide less expensive schedules than the joint chance constraint model where the risk is a priori equally divided between the scheduling periods (3.24). This shows the interest of allowing some flexibility in the way the risk is allocated between the scheduling periods.

In small instances, other predetermined proportions of the risk (by analysing wisely the risky periods) could have been chosen but it remains too complex, especially on instances considering many periods.

Second, Table 3.2 shows that all the solutions cheaper than the two bound solutions of the new model of joint chance constraint with a flexible sharing out of the risk do not validate the condition of the targeted QoS. These approaches are the deterministic

and the individual chance constraints ones. Thus they cannot be considered as possible alternatives. The robustness, i.e. the capability of providing the required QoS over the whole scheduling horizon within the maximum allowable risk level, is an essential criterion and its validation is mandatory to approve the model.

The last 3 programs which are joint chance-constrained models are the only models respecting the objective service level. The bound programs are cheaper than the joint chance constrained model with a fixed sharing out of the risk. This approach allows to save between 3.8% (upper bound) and 7.2% (lower bound) compared to this latter program.

This shows the interest of the proposed modelling and solution approach for it provides robust schedules at a lesser cost than previously joint chance-constrained models.

Table 3.3: Results for different parameters.

Parameters				Results	
λ range	μ	EWT^*	Risk ϵ	Gap	Violations
3 – 43	1	1	10%	0.05	3 – 1
3 – 43	1	1	05%	0.04	2 – 1
3 – 43	2	1	30%	0.0	13 – 5
3 – 43	1	2	30%	0.03	18 – 12
6 – 86	1	2	10%	0.004	6 – 5
6 – 86	1	1	10%	0.005	6 – 5
6 – 86	1	1	01%	0.003	0 – 0

In Table 3.3, results for different values of λ , μ , EWT^* and ϵ (illustrated in the table as percentages) are presented.

The *Gap* column gives the cost gap between the lower bound and the upper bound solution costs. The *Violations* column gives the numbers of violated scenarios for the lower and the upper bounds. We note that for greater values of λ , the bounds are really close and give good results.

In the scope of the one-step approaches, the approximated joint chance-constrained program with flexible sharing out of the risk shows good qualities both for costs minimizing and robustness. Next section compares this approach with the two-step approach previously defined in Section 3.3.

3.7 Comparing Results with the Two-Step Approach

3.7.1 Final Deterministic Formulation

The program presented in 3.3.3 is reminded here:

$$\begin{aligned}
 & \min c' x \\
 & \text{s.t. } \forall t \in \llbracket 1; T \rrbracket, \frac{A_t x - c_t}{\sigma_t} \geq F_{\beta}^{-1}((1 - \varepsilon)^{y_t}) \\
 & \sum_{t=1}^T y_t = 1 \\
 & x \in (\mathbb{Z}^+)^S, \forall t \in \llbracket 1; T \rrbracket, y_t \in]0; 1[.
 \end{aligned} \tag{3.28}$$

In this program, the right-hand side is a non-linear function of y_t .

Since this approach is considering the common normal distribution, the convexity of the RHS is known for a probability $p \geq 0.5$. Hence, the same approximations as in Section 3.5 are possible. Similarly, the coefficients of the approximations are obtained through piecewise linearizations and the final programs have the same form.

The coefficients for the upper bound program are

$$\begin{aligned}
 \forall j \in \llbracket 1; n-1 \rrbracket, \hat{F}_{u,\beta,j}^{-1}(p^y) &= F_{\beta}^{-1}(p^{y_j}) + \frac{y - y_j}{y_{j+1} - y_j} (F_{\beta}^{-1}(p^{y_{j+1}}) - F_{\beta}^{-1}(p^{y_j})) \\
 &= \delta_j y + \alpha_j.
 \end{aligned} \tag{3.29}$$

Again, the coefficients for the lower bound program are

$$\begin{aligned}
 \forall j \in \llbracket 1; n \rrbracket, \hat{F}_{l,\beta,j}^{-1}(p^y) &= F_{\beta}^{-1}(p^{y_j}) + (y - y_j)(F_{\beta}^{-1})'(p^{y_j}) p^{y_j} \ln(p) \\
 &= \delta_j \times y + \alpha_j
 \end{aligned} \tag{3.30}$$

where

$$(F_{\beta}^{-1})'(p^{y_j}) = \frac{1}{f_{\beta}(F_{\beta}^{-1}(p^{y_j}))}.$$

3.7.2 Results Comparison

The costs for the one-step programs and two-step programs are compared with varying values of parameters. The targeted EWT^* was set to 1. Some results are given in Table 3.4. Several other computations were made, with similar results.

Table 3.4: Comparison of solution costs for 1-Step and 2-Step approaches.

Resolution	Parameters		Cost functions		Gap
	λ range	Risk ϵ	Lower	Upper	
JCC 1Step	3 – 43	10%	60.44	61.72	2.1%
JCC 2Step	3 – 43	10%	59.72	64.44	7.9%
JCC 1Step	3 – 43	05%	61.44	63.44	3.3%
JCC 2Step	3 – 43	05%	60.44	64.72	7.1%
JCC 1Step	3 – 43	20%	59.00	60.00	1.7%
JCC 2Step	3 – 43	20%	57.44	62.44	8.7%
JCC 1Step	6 – 86	05%	112.2	113.9	1.5%
JCC 2Step	6 – 86	05%	110.7	115.6	4.4%

Table 3.4 shows the gap between the bounds is really tight for the one-step approach and is less satisfactory for the two-step approach. The closer the bounds, the less confusing it is for the manager to decide for a final schedule solution. We also notice that cost-wise, the lower bound of the two-step approach is a bit better than the lower solution of the one-step approach, while this is the contrary for the upper bound. On condition that all these solutions are robust, it's safer to consider the set of solutions with the cheapest upper bound.

Furthermore, robustness of the solutions was evaluated like in Section 3.6.3. The same process as in previous section was used, and the solutions of both approaches were tested. Table 3.5 gives some results of our simulations, for which the percentage of violated scenarios was extracted. 3 sets of simulations were run, each consisting in 100 of generated scenarios.

According to Table 3.5, the lower bound of the simplified two-step approach does not always respect the accepted risk level. It is in adequation with the idea of proposing a lower bound of a problem, but it may be confusing when it is needed to actually build a schedule that will indeed respect the set risk level. This test however did not enable us to show a strong similarity between our unknown random distribution for B_t (introduced in Section 3.2.5) and the normal distribution.

An interesting feature on one simulation can be highlighted: the costs of the upper bounds for the two different approaches were very close (118.2 for the two-step and 118.4 for the one-step for example), but the number of violated scenarios were very different. The one-step solution did respect the risk level whereas the two-step solution

Table 3.5: Percentage of violations for the 1-Step and 2-Step approaches.

Model	Parameters		% Violations	
	λ range	Risk ϵ	Lower	Upper
JCC 1Step	6 – 86	01%	0.67	0.67
JCC 2Step	6 – 86	01%	3	2
JCC 1Step	6 – 86	05%	3.67	2.67
JCC 2Step	6 – 86	05%	4.33	2.67
JCC 1Step	6 – 86	20%	14	9.33
JCC 2Step	6 – 86	20%	20.33	8.67
JCC 1Step	6 – 86	30%	20.33	19
JCC 2Step	6 – 86	30%	37.33	16.67

was often jeopardised. This is actually the consequence of a different distribution of the workforce through the possible shifts. The two-step solution seemed to have chosen a worse arrangement than the one-step solution.

3.8 On the Idea of Continuity

Our stochastic programming approach is based on several ideas. We compared the schedule solutions with applying these ideas. First joint chance constraint and individual chance constraints were compared. Second, the flexible sharing out of the risk and the fixed separation of the risk approaches were compared. Third, the one-step approach and the two-step approach were compared. Now we investigate another position of our approach: the idea of continuity, i.e. considering function ψ are functions from \mathbb{R} to \mathbb{R} .

The continuous expression of $F_B: \lambda \mapsto F_B(b)$ proposed in Section 3.4.1 is opposed to a more standard design considering a piecewise constant function based on server numbers as integers. Indeed, ψ function proposes real numbers of agents as an intermediary result even if the final solution obtained through the program is an integer vector. Constructing the ψ function as a continuous function is proposed since it is part of a whole process of computation considering continuous distributions and flexible sharing out of the risk. This idea is based on the observation that the computation of integer agent requirements as the minimum load as presented in Section 3.3.2 provides number of agents which give in fact a better objective service level than required. Thus, the continuity-based solution aims at a cheaper final solution through a continuous intermediary function

while respecting the objective service level.

Indeed, computing the minimum agents numbers to reach the objective service level for given arrival rate λ and service rate μ using

$$c = \min\{c \in |EWT(c, \lambda, \mu) \leq EWT^*\} \tag{3.31}$$

may provide a better service level than required. Table 3.6 shows that effective EWT are significantly better than EWT^* for different parameter values. The value of μ was set to 1 without loss of generality. Column 3 gives the minimum value c given by equation (3.31) for the given parameters and column 4 gives the results of the service level for the selected number of agents $EWT(c, \lambda, \mu)$.

Table 3.6: Values of effective Expected Waiting Time.

Parameters		Results	
EWT^*	λ	c	$EWT(c, \lambda, \mu)$
0.3	5	7	0.16
0.3	20	22	0.28
0.3	30	33	0.16
0.3	100	103	0.23
0.6	5	6	0.59
0.6	20	22	0.28
0.6	30	32	0.32
0.6	100	102	0.39
0.8	5	6	0.59
0.8	20	21	0.76
0.8	30	31	0.80
0.8	100	102	0.39

However, this idea of continuity needs to be validated by comparing with a more intuitive approach based on integers.

Thus, the following sections introduce a new model in one-step and the schedule costs between integer and continuous-based approaches are compared.

3.8.1 An Integer-Based One-Step Approach

This section introduces another expression for $F_B: b \mapsto F_B(b)$ in order to describe an integer-based approach. The optimization problem considered in this section is expressed as introduced in 3.2.5, considering $1 - \varepsilon \in]0; 1[$:

$$\begin{aligned}
 & \min c' t x \\
 & \text{s.t. } \forall t \in \llbracket 1; T \rrbracket, A_t x \geq F_{C_t}^{-1}((1 - \varepsilon)^{y_t}) \\
 & \quad \sum_{t=1}^T y_t = 1 \\
 & \quad x \in (\mathbb{Z}^+)^S, \forall t \in \llbracket 1; T \rrbracket, y_t \in]0; 1[.
 \end{aligned} \tag{3.32}$$

where $F_{C_t}^{-1}$ denotes the inverse of the cumulative distribution function of the C_t integer random variable.

Section 3.4.1 presents the ψ function to express $F_B(b) = F_\Lambda(\psi^{-1}(b))$ with $b \in \mathbb{R}^+$. This section introduces the φ function defined by

$$\begin{aligned}
 & \varphi: \mathbb{R}^+ \rightarrow \mathbb{N} \\
 & \lambda \mapsto \varphi(\lambda) = \min\{c \in \mathbb{N} \mid EWT(c, \lambda, \mu) \leq EWT^*\} .
 \end{aligned} \tag{3.33}$$

This function is piecewise constant. Let $0 = \lambda_0 < \lambda_1 < \dots < \lambda_m$ be $m + 1$ variables so that $\forall c \in \llbracket 1; m \rrbracket, \varphi(\lambda \in]\lambda_{c-1}; \lambda_c]) = c$ and $\varphi(\lambda_0) = \varphi(0) = 0$.

The values $\lambda_1, \lambda_2, \dots, \lambda_m$ are determined by computing the φ function for several values $k \times \Delta\lambda$, $k \in \llbracket 1; k_{\max} \rrbracket$, with a given precision level $\Delta\lambda$, in a chosen range of values $]0; \lambda_m]$.

Considering $c > 0$, we focus on expressing y_t so that

$$F_{C_t}^{-1}((1 - \varepsilon)^{y_t}) = \varphi(F_{\Lambda_t}^{-1}((1 - \varepsilon)^{y_t})) = c . \tag{3.34}$$

Since $F_{C_t}^{-1}$ is piecewise constant, we consider

$$\lambda_{c-1} < F_{\Lambda_t}^{-1}((1 - \varepsilon)^{y_t}) \leq \lambda_c , \tag{3.35}$$

and the bounds are the following:

$$\frac{\ln(F_{\Lambda_t}(\lambda_{c-1}))}{\ln(1-\varepsilon)} > y_t \geq \frac{\ln(F_{\Lambda_t}(\lambda_c))}{\ln(1-\varepsilon)}. \quad (3.36)$$

Thus, the Right-Hand Side has value c is $y_t \in \left[\frac{\ln(F_{\Lambda_t}(\lambda_c))}{\ln(1-\varepsilon)}; \frac{\ln(F_{\Lambda_t}(\lambda_{c-1}))}{\ln(1-\varepsilon)} \right[$.

Hence,

$$y_t \mapsto F_{C_t}^{-1}((1-\varepsilon)^{y_t}) = \sum_{c \in \mathbb{N}} c \mathbf{1}_{\left[\frac{\ln(F_{\Lambda_t}(\lambda_c))}{\ln(1-\varepsilon)}; \frac{\ln(F_{\Lambda_t}(\lambda_{c-1}))}{\ln(1-\varepsilon)} \right[} \quad (3.37)$$

with

$$F_{C_t}^{-1}\left(p \in]F_{\Lambda_t}(\lambda_{c-1}); F_{\Lambda_t}(\lambda_c)]\right) = c. \quad (3.38)$$

In this expression, $\left[\frac{\ln(F_{\Lambda_t}(\lambda_c))}{\ln(1-\varepsilon)}; \frac{\ln(F_{\Lambda_t}(\lambda_{c-1}))}{\ln(1-\varepsilon)} \right[$ describes each interval of risk for which an additional agent is needed at period t . Thus, the number of agents can be described as the amount of intervals whose infimum is exceeded by y_t . This is expressed in the mixed-integer linear program by a binary variable $\kappa_{t,c}$ which is 1 if $y_t \geq \frac{\ln(F_{\Lambda_t}(\lambda_c))}{\ln(1-\varepsilon)}$. The resulting MILP is finally:

$$\min c' t x$$

$$\text{s.t. } \forall t \in \llbracket 1; T \rrbracket, A_t x \geq F_{C_t}^{-1}(1-\varepsilon) + \sum_{c=F_{C_t}^{-1}(1-\varepsilon)}^m \kappa_{t,c} \quad (3.39a)$$

$$\forall t \in \llbracket 1; T \rrbracket, y_t \geq \frac{\ln(F_{\Lambda_t}(\lambda_{F_{C_t}^{-1}(1-\varepsilon)}))}{\ln(1-\varepsilon)} + \sum_{c=F_{C_t}^{-1}(1-\varepsilon)}^m \kappa_{t,c} (\alpha_{t,c} - \alpha_{t,c-1}) \quad (3.39b)$$

$$\forall t \in \llbracket 1; T \rrbracket, \forall c \in \llbracket 1; m \rrbracket, \kappa_{t,c} \geq \kappa_{t,c-1} \quad (3.39c)$$

$$\sum_{t=1}^T y_t = 1 \quad (3.39d)$$

$$x \in (\mathbb{Z}^+)^S, \forall t \in \llbracket 1; T \rrbracket, y_t \in]0; 1[$$

$$\forall t \in \llbracket 1; T \rrbracket, \forall c \in \llbracket 1; m \rrbracket, \kappa_{t,c} \in \{0; 1\}$$

where $\alpha_{t,c} = \frac{\ln(F_{\Lambda_t}(\lambda_c))}{\ln(1-\varepsilon)}$ and $\alpha_{t,0} = 0$.

Constraints (3.39a) are the reformulation of constraints in (3.32) by considering that each interval of risk $\left[\frac{\ln(F_{\Lambda_t}(\lambda_c))}{\ln(1-\varepsilon)}; \frac{\ln(F_{\Lambda_t}(\lambda_{c-1}))}{\ln(1-\varepsilon)} \right[$ considered in addition to the base risk level

$1 - \varepsilon$ requires to add an extra agent.

Constraints (3.39b) show the relation between y_t variables and the number of additional agents. In order to decrease the value of y_t from one interval to the previous one, it is required to add one extra agent.

Finally, constraints (3.39c) state that if y_t belongs to one interval, then one extra agent is required for each following interval representing greater values of y_t .

In next section the solution cost provided by this integer-based approach are compared with cost of the one-step real-based approach.

3.8.2 Comparison of Solution Costs

Similar instances as in Section 3.6 are used in this section in order to compare the solution costs of the integer-based and real-based approaches.

The schedule horizon is one-week long with 107 periods of 30 minutes ; 19 different shifts compose the shifts matrix. The variances σ_t^2 are random values generated in $[\frac{\lambda_t}{4}; \frac{\lambda_t}{2}]$ and $\mu = 1$ without loss of generality.

Table 3.7: Comparison of piecewise constant-based and linear-based approaches.

Parameters			Cost functions		
Risk ε	EWT^*	λ range	Lower Real	Upper Real	Approx Int
0.05	0.4	3 – 43	61.44	61.72	63.44
0.05	0.7	3 – 43	60.44	60.72	62.44
0.1	0.4	3 – 43	59.72	60.44	62
0.05	0.4	6 – 86	112.2	112.9	114.4
0.05	0.7	6 – 86	111.2	111.9	113.2
0.05	1	6 – 86	111.2	111.4	112.9
0.1	0.4	6 – 86	109.9	110.9	112.4
0.1	0.7	6 – 86	109.2	109.9	111.2

Table 3.7 shows the solution costs for the upper (column *Upper Real*) and lower (column *Lower Real*) bounds of the linearized real-based approach as well as for the approximated integer-based approach (column *Approx Int*) for different values of the risk ε , the targeted Expected Waiting Time EWT^* and arrival rates λ range. Both are based on a one-step formulation. It can be noticed that the integer-based solutions are more expensive than the cost bounds of the real-based approach. These results tend to confirm

the idea that the integer-based one-step approach is a conservative approximation of the real-based one-step approach.

3.9 Conclusion

This chapter introduced a complete resolution approach for solving the staffing and shift-scheduling problem in one step with uncertain arrival rates. By solving the whole problem in one single step, the procedure aimed at restricting the addition of several approximations which could occur at different steps of the process.

The problem was formulated with a joint chance-constrained program. We introduced the representation of mean arrival rates as continuous normal distributions and proposed the construction of the ψ function in order to solve the staffing problem and the shift-scheduling problem altogether in one step. This function was expressed with our chosen Quality of Service, the Expected Waiting Time for a customer in the queue. The risk level was set for the whole horizon of study, and was expressed as a composition of risks on each period of staffing. Instead of deciding this composition a priori, the sharing out was handled during the optimization process, represented with variables to be optimized. Finally, classical piecewise linear approximations were proposed in order to provide upper and lower bounds of the objective function's optimal solution for the staffing and shift-scheduling problem.

This approach was successfully applied to Erlang-C model. Computational results show that the proposed approach compares well with previously published approaches both in terms of risk management and cost minimization. The results show the interest of the joint chance-constrained approach in term of robustness and the advantage of the dynamic sharing out of the risk in term of cost reduction. The lower bound and upper bound give quite close results and both propose cheaper solutions than other chance-constrained program, while respecting the targeted Quality of Service.

Moreover, the one-step and the two-step approaches were compared in order to check similarities between the unknown distribution of B_t and the normal distribution. Even if the two-step approach gives some interesting results, the one-step approach appears to be more precise and averagely cheaper. Furthermore, it was observed that its robustness was a bit better.

Finally an approach with a piecewise constant representation of $y \mapsto F_C^{-1}$ was proposed in order to highlight the interest of our continuous approach with real values of

agents.

This work can be extended theoretically and experimentally.

First, the shape of $y \mapsto F_B^{-1}$ is complex and further analyses need to be done. Second, as it could be seen in our results, both upper and lower bounds give a better QoS than expected and thus, probably illustrates an overstaffing. This overstaffing is costly and would rather be avoided. Improving the queuing system model may help solving this drawback. Skill-based call centers can be considered, assuming that agents are specialized in specific fields and will answer the appropriate calls according to these skills. This implies the creation of multiple queues which can be connected (see [Cezik and L'Ecuyer \[2008\]](#) for example). Finally, periods were supposed to be independent, which is not really the case. This consideration can be taken into account in the chance-constrained approach.

Experimentally, a study with advanced simulations of a call center could show the qualities and lacunes of our model, which could help improving it and choosing more realistic assumptions if needed.

Chapter 4

Chance-Constrained Approaches with Advanced Queuing Models - Abandonments and Period Correlations

This chapter presents an approach for the staffing and shift-scheduling problem with uncertain demand forecasts. Compared to previous chapter, we consider here important extra features: first, customers have a limited patience and may leave the queue before being served. This is considered as call abandonments. Second, correlation between different periods of time within a day is included in the uncertainty model with the consideration of a *bussyness* factor. This factor represents the random load of a day and is a random variable considered in addition to the random arrival rates. This chapter shows that the one-step stochastic approach developed previously is still practicable with an extended call center model and its global formulation can be adapted to improve the uncertainty management. We propose here a deterministic approximation of this problem.

4.1 Introduction

This chapter investigates some perspectives of previous Chapter 3. First, Erlang-C formulae have some constraints when the system is over-loaded. The queue is not stable anymore and the Erlang-C formula of Expected Waiting Time cannot be applied. Moreover, the $M/M/c$ queue is standard but simple by assumption. Erlang-A model solves these two drawbacks: first, the system considering abandonment is stable (confirmed in Green et al. [2007]). Second, customer abandonment is an important aspect of call center operations (see Mandelbaum and Zeltyn [2007]) and affect the Quality of Service. Hence, we propose to consider Erlang-A instead of Erlang-C in a new formulation of our ψ function in order to consider call abandonments.

The second perspective investigated in this chapter is related to period independence. Indeed, arrival rates at each period were considered as independent. These periods were introduced by the SIPP (Stationary Independent Period-by-Period) approximation, required to use queuing results. These periods usually represent one single day during which the numbers of calls at different times are actually not independent, as precised in Avramidis et al. [2004]. Hence, random variables in this chapter represent both random arrival rates and the random load of a day. This load represents the dependence between periods : when a period is heavily-loaded, other periods of the same day tend to handle the same consideration, and conversely.

In this chapter, we use some results of previous Chapter 3: the staffing and shift-scheduling problem is solved in one step by considering a continuous relation between queuing parameters. The problem is formulated with a joint chance-constrained approach, and the global risk level is dynamically shared among the periods.

However, the staffing model is now based on Erlang-A, which advantageously represents a stable queue with a better representation of real situations. Moreover, arrival rates random variables are not following normal distributions anymore. Arrival rates are indeed a product of a random variables: a continuous variable represents the arrival rate random seasonality and a discrete variable represents the random load of one day. Linear approximations lead to an approximated mixed-integer program.

The extended queuing model including abandonments is presented in Section 4.2. This section gives the new formulation of the staffing model with Erlang-A and the resulting new ψ_a function. Then, experimental results are given. Section 4.3 presents a new consideration of uncertainty in order to include period correlations in the formu-

lation. A new stochastic program is presented. A deterministic approximation based on linearizations is proposed and reformulations of products of variables are introduced in order to obtain a mixed-integer program. The final program is composed of a high number of non-linear constraints and a two-step approach is proposed to consider linear programs.

4.2 Extension: Including Abandonments

The call center model presented in Section 3.2.1 and used for solving the problem ignores call abandonment. The customer patience is considered as infinite. Section 3.2.2 presented Erlang-C model, which describes the call center under this assumption. However considering that some customers may hang up on purpose before being served is consistent with a human behaviour: they reached their patience limit and decided to end the call. This is considered as call abandonments.

Erlang-C model is commonly used for its complexity is handable. However, customer abandoning influences significantly the performances of call centers. Thus, abandonment is an important criterion to be included in the call center model. Targeted values of probability of abandonment are often set in call centers as performance measures (Mandelbaum and Zeltyn [2009]). By preventing congestion, models including abandonments show a better system stability (see Green et al. [2007]) whereas the simple Erlang-C model is not satisfactory when the system is overloaded.

In this section, the previous one-step joint chance-constrained approach is applied with another queuing system model, namely Erlang-A. The resolution presented in the previous section is still valid.

4.2.1 Another Queuing Model: Erlang-A

In this Section, abandonments are taken into account in the queuing system model in order to propose a more precise representation of the reality. Erlang-C describes a model where customer patience is supposed to be infinite. Hence, Erlang-C model is replaced by Erlang-A model: Erlang-A model is an extended model in which customers are considered to have a limited patience. They accept to wait some time before service, but beyond this limit, they abandon the queue. The patience time is considered to be exponentially distributed. The abandonment rate is denoted by θ (and θ^{-1} represents the

mean patience time).

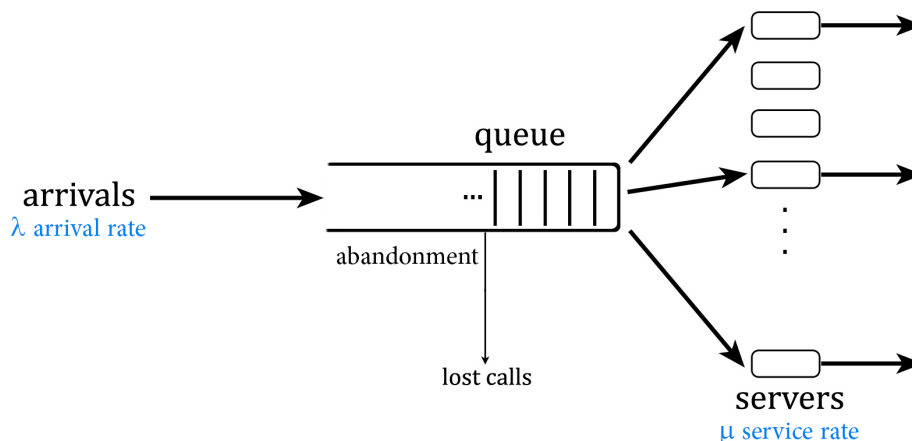


Figure 4-1: A queue with customer abandonment.

In the following computations, we decide to keep the same Quality of Service as previously: a targeted EWT^* value, i.e. a targeted value of Expected Waiting Time. This decision is made to keep consistency with the previous sections and does not omit the useful information given by Erlang-A model, particularly the fraction abandoning, since our chosen QoS can be related to the probability of abandonment.

In [Mandelbaum and Zeltyn \[2007\]](#), the authors give the following relation between the probability of abandonment $\mathbb{P}(Ab)$ and EWT :

$$\mathbb{P}(Ab) = \theta \mathbb{E}[W]. \quad (4.1)$$

Hence, the results presented in this section with the Expected Waiting Time as Quality of Service are qualitatively consistent with results we would have with the probability of abandonment as QoS.

4.2.2 Problem Formulation

The problem formulation relies on the expression of the chosen Quality of Service, EWT . Here are the detailed formulae associated to the Erlang-A model. The queue taking into account abandonment is denoted by $M_t/M/c_t + M$ according to [Baccelli and Hebuterne \[1981\]](#), where $+M$ denotes the patience time distribution. The queue is, as

the previous Erlang-C queue, assumed to be in stationary regime at each period. Note that, because of abandonments, the Erlang-A queue is unconditionally stable.

From [Mandelbaum and Zeltyn \[2005\]](#), the probability of waiting and the expected waiting time for delayed calls in Erlang-A model are given respectively by

$$\mathbb{P}(\text{Wait} > 0) = \frac{A\left(\frac{c\mu}{\theta}, \frac{\lambda}{\theta}\right) B_c}{1 + (A\left(\frac{c\mu}{\theta}, \frac{\lambda}{\theta}\right) - 1)B_c} \quad (4.2)$$

and

$$\mathbb{E}[\text{Wait} | \text{Wait} > 0] = \frac{c\mu + A\left(\frac{c\mu}{\theta}, \frac{\lambda}{\theta}\right) (\lambda - c\mu)}{\lambda\theta A\left(\frac{c\mu}{\theta}, \frac{\lambda}{\theta}\right)}, \quad (4.3)$$

where λ is the call arrival rate, μ the service rate, c the number of agents, θ the abandonment rate, and

$$A\left(\frac{c\mu}{\theta}, \frac{\lambda}{\theta}\right) = \frac{c\mu}{\theta} e^{\lambda/\theta} \left(\frac{\lambda}{\theta}\right)^{-\frac{c\mu}{\theta}} \gamma\left(\frac{c\mu}{\theta}, \frac{\lambda}{\theta}\right), \quad (4.4)$$

where $\gamma(x, y)$ is the incomplete lower gamma function

$$\gamma(x, y) = \int_0^y t^{x-1} e^{-t} dt, \quad (4.5)$$

and finally B_c represents the blocking probability for an $M/M/c/c$ queue:

$$B_c = \frac{\left(\frac{\lambda}{\mu}\right)^c}{c! \sum_{i=0}^c \frac{1}{i!} \left(\frac{\lambda}{\mu}\right)^i}. \quad (4.6)$$

The blocking probability represents the probability that a call meets a busy signal in a system without a queue where there are n agents and n telephone lines.

Finally, the Quality of Service is given by

$$EWT(c, \lambda, \mu, \theta) = \frac{c\mu}{\lambda\theta} \left(\frac{\frac{\lambda}{c\mu} A\left(\frac{c\mu}{\theta}, \frac{\lambda}{\theta}\right) B_c + 1}{1 + (A\left(\frac{c\mu}{\theta}, \frac{\lambda}{\theta}\right) - 1)B_c} - 1 \right), \quad (4.7)$$

defined on \mathbb{R}^+ .

The ψ_a function, is the same function as the previously introduced ψ function (Sec-

tion 3.4) but uses the new EWT which includes abandonment:

$$\begin{aligned} \Psi_a: \mathbb{R}^+ &\rightarrow \mathbb{R}^+ \\ \lambda \mapsto \Psi_a(\lambda) &= b(\lambda, EWT^*, \mu) = \frac{EWT^* + cEWT_{c+1,\lambda} - (c+1)EWT_{c,\lambda}}{EWT_{c+1,\lambda} - EWT_{c,\lambda}}. \end{aligned} \quad (4.8)$$

The convexity of $\Psi_a(F_\Lambda^{-1})$ is not proven yet but computations are partially done. However it is numerically observed to be convex.

The stochastic problem is the following:

$$\begin{aligned} \min c^t x \\ \text{s.t. } \forall t \in \llbracket 1; T \rrbracket, A_t x &\geq \Psi_a(F_\Lambda^{-1}((1-\varepsilon)^{y_t})) \\ \sum_{t=1}^T y_t &= 1 \\ x \in (\mathbb{Z}^+)^S, \forall t \in \llbracket 1; T \rrbracket, y_t &\in]0; 1]. \end{aligned} \quad (4.9)$$

The approximated solutions are obtained with the same method as previously.

Note Computations of $A\left(\frac{c\mu}{\theta}, \frac{\lambda}{\theta}\right)$ were done using the Kummer's confluent hypergeometric function, with a 10^{-15} precision.

The next section gives some numerical results based on the same experiments: we compare the deterministic, stochastic individual chance-constrained, and joint chance-constrained approaches. Then, the scheduled solutions are tested on 100 simulated scenarios.

4.2.3 Numerical Experiments

Similarly as in previous chapter in Section 3.6, our approach is applied on some instances. These results enable to evaluate the performance of the approach.

Data of a French health call center are used for these numerical experiments. The following parameters were used:

- $\varepsilon = 0.1$
- $EWT^* = 0.7$

- $\mu = 1$
- $\theta = 0.1$
- 15 points for linear approximations

The matrix of shifts consists of 24 shifts, both part-time and full-time. The schedule horizon is one typical 10-hour weekday. The day is composed of 30-minute periods.

Table 4.1 gives the schedule solutions for different approaches. Column *Determ* shows the schedule solution for the deterministic approach, column *Indiv* stand for the individual chance-constrained approach, column *Fixed* represents the solution for the joint chance-constrained approach with a fixed sharing out of the risk and finally the two last columns *LowerB* and *UpperB* give the lower bound and upper bound of our approach. The costs of the different approaches show the same trends: the deterministic and individual chance-constrained approaches give cheaper solutions than the joint chance-constrained approaches. Moreover, among the joint chance-constrained approaches, the approximated lower and upper bounds are both cheaper than the stochastic approach with a sharing out of the risk done a priori.

In order to evaluate the robustness of the schedule solutions, scenarios are generated with the same parameters as previously. A scenario represents a possible realisation of the call arrival rate values during a day, following a typical daily seasonality. A violated scenario brings out the inability of the scheduled agents to manage the load during at least one period of time in this realisation.

Table 4.2 compares the different schedules solutions quality wise: the percentage of violated scenarios can be compared to the targeted percentage of violations, namely the risk level. Similarly to the results with Erlang-C approach, the deterministic and individual chance-constrained models highly exceed the target of 10%. These approaches can not be considered as usable. The three joint chance-constrained approaches respect the risk level, and our new approach is cheaper than the fixed risk-based approach.

The cost gap between the lower and upper bounds on this instance is $CG = 0.005$ which is good. It is explained by the number of linearization points.

Indeed, $y \mapsto F_{B_{a,t}}^{-1}(p^y)$ is convex. Then, it is possible to highly increase the number of points and propose closer bounds than with the Erlang-C approach.

Our approach shows good results in term of cost saving as well as robustness for the Erlang-A queue. This observation matches the previous evaluation for Erlang-C queue,

Table 4.I: Schedule solutions for the problem with abandonment.

Shift	Deter	Indiv	Fixed	LowerB	UpperB
1	18	20	21	19	20
2	8	6	6	6	4
3	0	0	0	0	0
4	9	8	3	3	1
5	0	0	0	0	0
6	0	0	0	0	0
7	0	0	0	0	0
8	0	0	0	0	0
9	0	0	0	0	0
10	1	2	4	6	6
11	0	2	8	8	11
12	19	24	29	27	28
13	0	0	0	0	0
14	0	0	0	0	0
15	0	0	0	0	0
16	0	0	0	0	0
17	4	5	5	5	4
18	0	0	0	0	0
19	4	3	2	3	2
20	4	5	0	0	0
21	18	20	17	15	15
22	4	0	0	0	0
23	0	0	0	0	0
24	10	9	12	11	11
Total	99	104	107	103	102
Cost function	70.62	76.94	83.44	80.88	81.25

but the Erlang-A based approach may bring more advantages since the Erlang-A model is closer to reality and more interesting to consider.

Table 4.2: Percentage of violated scenarios for Erlang-A Queuing Model.

Model	% of violation
Deterministic model	100%
Individual chance constraint model	59%
JCC model with fixed risk level	2%
JCC lower bound and flexible risk	9%
JCC upper bound and flexible risk	8%
Targeted maximal risk	10%

Finally, Table 4.3 gives the solution costs for a given instance with the following parameters:

- $\varepsilon = 0.5$
- $EWT^* = 0.4$
- $\mu = 1$

The table shows the results for different values of abandonment rate θ . Unsurprisingly, Erlang-A-based model provides cheaper solutions than Erlang-C-based model for the same instance. By greatly reducing the abandonment rate, Erlang-A-based solution tends to meet the Erlang-C-based solution.

Table 4.3: Solution cost with or without abandonment.

	θ	Lower Cost	Upper Cost
Erlang-A	0.15	79.25	79.5
	0.1	81.19	81.44
	0.05	83.31	83.56
	0.01	85.44	85.69
	0.005	85.75	86
Erlang-C	0	86.06	87.75

Finally, the ψ_a function was introduced in the integer-based one-step approach. Table 4.4 shows the cost solutions for the integer-based and real-based one-step approach. The results are in adequation with observations made with the Erlang-C model. The one-step continuous real-based approach provide cheaper bounds than the integer-based approach. Thanks to the convexity of ψ_a function, by increasing the number of interpolation points, the gap between the upper and lower bounds can be decreased to reach

a predetermined gap. However we did not examine this issue. Comparing the computational efforts according to the gap precision could show the limit of the model in term of tractability.

Table 4.4: Comparison of piecewise constant-based and real-based approaches.

Parameters				Cost functions		
Risk ϵ	EWT^*	λ range	θ	Lower Reel	Upper Reel	Approx Int
0.05	0.4	3 – 43	0.05	59.15	59.44	61.15
0.05	0.7	3 – 43	0.05	57.15	57.44	59.15
0.1	0.4	3 – 43	0.1	56.15	56.44	58.15
0.1	0.4	3 – 43	0.05	58	58.44	59.72
0.05	1	3 – 43	0.05	55.72	56.15	57.72
0.05	0.4	6 – 86	0.01	111.9	112.2	113.6
0.05	0.4	6 – 86	0.05	108.9	109.4	110.9
0.05	0.7	6 – 86	0.05	106.2	106.4	107.9
0.05	1	6 – 86	0.05	104.2	104.4	105.9
0.1	0.4	6 – 86	0.05	106.9	107.4	108.9
0.1	0.7	6 – 86	0.05	104.2	104.6	106.2
0.05	0.4	6 – 86	0.1	106.9	107.4	108.9

4.3 Period Correlation: Introduction of a Daily *Busyness* Factor

This section focuses on improving the current uncertainty management, presented in Section 3.2.5.

Until now, periods of the staffing model and thus, the random variables, are assumed to be independent. In this section, an improved representation of uncertainty is proposed. We now consider a product of random variables, the continuous arrival rates and a discrete *busyness* factor.

4.3.1 Arrival Rate Model

The SIPP approximation, used to model the queue in stationary regime, consists in splitting the time horizon into shorts periods of time in which the arrival rate does not vary

too quickly. Thus, it is considered as constant at each period. Up to now, these mean arrival rates were considered as independent. However practically this is not the case. Indeed, [Avramidis et al. \[2004\]](#) state that there is a positive correlation between arrivals at different periods of a day. Thus, call volumes in the morning and in the afternoon tend to be related.

In this section, we propose to develop a formulation taking the period dependence into consideration.

This dependence is illustrated with a random *busyness* factor which represents the load of a day. This busyness factor was proposed in [Whitt \[1999\]](#), but without considering uncertainty on the arrival rates. This random variable Π is a multiplicative factor of the random arrival rates and does not depend on the period.

Hence, the expression of the arrival rate at period t is now the following:

$$\Lambda_{\text{corr},t} = \Pi \Lambda_t . \quad (4.10)$$

The resulting random variable is the mean random arrival rate considered so far with a random weighting according to the possible load for the day. The arrival rates Λ_t at each period are still assumed to be independent. Moreover, the *busyness* factor and the arrival rates are assumed to be independent.

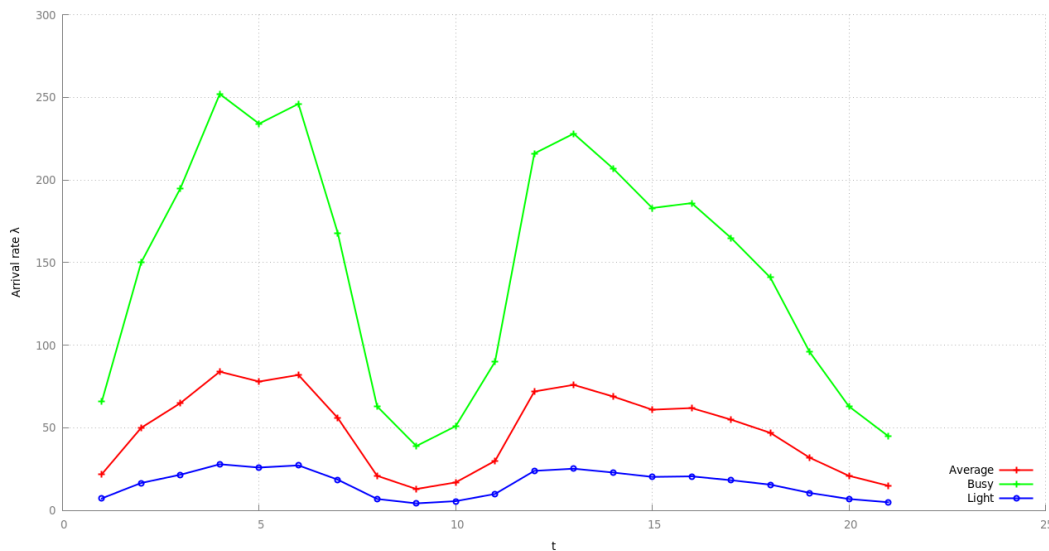


Figure 4-2: Example of possible daily arrival rates scenarios.

Figure 4-2 shows possible different daily loads. The typical seasonality of arrival rates through the day is visible. The load is influenced by random event such as particular

weather conditions, political announcements, etc.

4.3.2 Chance-Constrained Problem with Period Correlation

The problem is formulated as a one-step problem, solving the staffing and shift-scheduling problems together. The chance-constrained approach presented in Section 3.4 is used in this section: the joint chance-constrained problem expressed the choice of a risk level set on the whole time horizon. Since the correlation is defined on a daily basis, the joint chance constraint of the problem described above is stated for one day. However, it is possible to add one chance constraint for each additional day to be considered. In this case, the model proposed in this section is still valid.

In this section, ψ_a is the function giving a real number of agents requirements b for given values of λ , μ and EWT^* according to Erlang-A model. We define the following function for a vector $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_T)$:

$$\begin{aligned} \psi_{a,T} : (\mathbb{R}^+)^T &\rightarrow (\mathbb{R}^+)^T \\ \lambda &\mapsto \psi_{a,T}(\lambda) = (\psi_a(\lambda_1), \psi_a(\lambda_2), \dots, \psi_a(\lambda_T)) . \end{aligned} \quad (4.11)$$

Hence, the joint chance-constrained formulation of the problem is the following:

$$\begin{aligned} \min c^T x \\ \text{s.t. } \mathbb{P} \left(\psi_{a,T}^{-1}(Ax) \geq \Lambda_{\text{corr}} \right) &\geq 1 - \varepsilon \\ x &\in (\mathbb{Z}^+)^S . \end{aligned} \quad (4.12)$$

Λ_{corr} is the vector of $\Lambda_{\text{corr},t}$ random variables.

The cumulative distribution function of Λ_{corr} is expressed as follows:

$$\mathbb{P}(\Lambda_{\text{corr}} \leq \lambda) = \mathbb{P}(\Pi\Lambda \leq \lambda) . \quad (4.13)$$

Then, we consider:

$$\begin{aligned} \min c^T x \\ \text{s.t. } \mathbb{P} \left(\psi_{a,T}^{-1}(Ax) \geq \Pi\Lambda \right) &\geq 1 - \varepsilon \\ x &\in (\mathbb{Z}^+)^S . \end{aligned} \quad (4.14)$$

In the following formulations, Ψ_a is denoted by Ψ for easier reading.

The *busyness* factor is assumed to be a discrete random variable. Let $\pi_1, \pi_2, \dots, \pi_J$ be the J possible realizations and p_1, p_2, \dots, p_J be the associated probabilities.

Since Π is a discrete random vector, the cumulative distribution function of the product of random variables is expressed with the possible outcomes of the variable:

$$\begin{aligned} \mathbb{P}(\Pi\Lambda \leq \lambda) &= \mathbb{P}\left(\bigcap_{t=1}^T \{\Lambda_t \Pi \leq \lambda_t\}\right) \\ &= \sum_{j=1}^J p_j \mathbb{P}\left(\bigcap_{t=1}^T \left\{\frac{\lambda}{\pi_j} \geq \Lambda_t\right\}\right) = \sum_{j=1}^J p_j \prod_{t=1}^T \mathbb{P}\left(\frac{\lambda}{\pi_j} \geq \Lambda_t\right). \end{aligned} \quad (4.15)$$

Finally, the joint chance constraint of the problem (4.14) is given by

$$\mathbb{P}\left(\Psi_T^{-1}(Ax) \geq \Pi\Lambda\right) \geq 1 - \varepsilon \Leftrightarrow \sum_{j=1}^J p_j \prod_{t=1}^T \mathbb{P}\left(\frac{\Psi^{-1}(A_t x)}{\pi_j} \geq \Lambda_t\right) \geq 1 - \varepsilon. \quad (4.16)$$

Unlike in Chapter 3, the constraint here is expressed with a sum and products of cumulative distribution function of Λ_t . The product inside the sum is similar to the result obtained in the previous chapter, by considering independency. The idea here is to reformulate the constraint in order to propose linearizations similar as before.

First, the sum of elements is decomposed into individual elements, each representing a scenario. Then, the products are split into individual constraints, each representing a period. Like previously, the decomposition is proposed by considering variables to be optimized. This approach provides a reduced solution cost compared to a fixed sharing out of the risk.

The reformulation introduces a vector f of J variables defined on $]0; 1[$ and a vector y of T variables on $]0; 1[$:

$$\begin{aligned} \sum_{j=1}^n p_j \prod_{t=1}^T \mathbb{P}\left(\frac{\Psi^{-1}(A_t x)}{\pi_j} \geq \Lambda_t\right) \geq 1 - \varepsilon \\ \Leftrightarrow \begin{cases} \forall j \in \llbracket 1; J \rrbracket, p_j \prod_{t=1}^T \mathbb{P}\left(\Lambda_t \leq \frac{\Psi^{-1}(A_t x)}{\pi_j}\right) \geq f_j(1 - \varepsilon) \\ \sum_{j=1}^n f_j = 1 \\ \forall j \in \llbracket 1; J \rrbracket, f_j < \frac{p_j}{1 - \varepsilon}. \end{cases} \end{aligned} \quad (4.17)$$

Equations (4.17) give the separation of the sum into several constraints. The constraint $\sum_{j=1}^n f_j = 1$ describes the fact that $1 - \varepsilon$ is the targeted global risk level. The constraint $\forall j \in \llbracket 1; J \rrbracket, f_j < \frac{p_j}{1-\varepsilon}$ insures the probability level is still below 1.

And then,

$$\sum_{j=1}^n p_j \prod_{t=1}^T \mathbb{P} \left(\frac{\Psi^{-1}(A_t x)}{\pi_j} \geq \Lambda_t \right) \geq 1 - \varepsilon$$

$$\Leftrightarrow \begin{cases} \forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, \mathbb{P} \left(\Lambda_t \leq \frac{\Psi^{-1}(A_t x)}{\pi_j} \right) \geq \left(\frac{f_j(1-\varepsilon)}{p_j} \right)^{y_t} \\ \sum_{j=1}^n f_j = 1 \\ \forall j \in \llbracket 1; J \rrbracket, f_j < \frac{p_j}{1-\varepsilon} \\ \sum_{t=1}^T y_t = 1 . \end{cases} \quad (4.18)$$

Equations (4.18) give the separation of the products into several constraints. Again, the constraint $\sum_{t=1}^T y_t = 1$ is added. Here, the variables y_t do not depend on the scenario j . This illustrates the assumption that the risk should have the same distribution through the periods for every scenario, which is quite consistent. It is possible to change the formulation and introduce $y_{j,t}$ variables, allowing any values for the variables, without consideration of consistency between the scenarios.

The joint chance constraint is now separated into individual chance constraints. The number of constraints is now $J \times T$ (in addition to the constraints on f and y). The resulting mixed-integer non-linear program is the following

$$\min c^t x$$

$$\text{s.t. } \forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, \mathbb{P} \left(\Lambda_t \leq \frac{\Psi^{-1}(A_t x)}{\pi_j} \right) \geq \left(\frac{f_j(1-\varepsilon)}{p_j} \right)^{y_t} \quad (4.19a)$$

$$\sum_{j=1}^n f_j = 1 \quad (4.19b)$$

$$\sum_{t=1}^T y_t = 1 \quad (4.19c)$$

$$\forall j \in \llbracket 1; J \rrbracket, f_j < \frac{p_j}{1-\varepsilon} \quad (4.19d)$$

$$x \in (\mathbb{Z}^+)^S, f \in]0; 1[^J, y \in]0; 1[^T .$$

The RHS of the constraints (4.19a) is reformulated according to [Liberti et al. \[2009\]](#):

$$\forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, \mathbb{P} \left(\Lambda_t \leq \frac{\Psi^{-1}(A_t x)}{\pi_j} \right) \geq \left(\frac{f_j(1-\varepsilon)}{p_j} \right)^{y_t}$$

$$\Leftrightarrow \begin{cases} \forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, \mathbb{P} \left(\Lambda_t \leq \frac{\Psi^{-1}(A_t x)}{\pi_j} \right) \geq w_{j,t} & (4.20a) \\ \forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, y_t \ln \left(\frac{f_j(1-\varepsilon)}{p_j} \right) = \ln(w_{j,t}) & (4.20b) \end{cases}$$

The set of variables $w_{j,t}$, $(j,t) \in \llbracket 1; J \rrbracket \times \llbracket 1; T \rrbracket$ defines the Right-Hand-Side of constraints (4.19a), as expressed in constraint (4.20a). The second set of constraints (4.20b) is the simple expression of the equality defining variables $w_{j,t}$ for which the logarithm function is applied.

The resulting individual constraints are non-linear. The next section proposes a approximation of the MINLP through linearizations.

4.3.3 Constraint Linearizations

The problem considered here is the following:

$$\min c^t x$$

$$\text{s.t. } \forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, A_t x \geq \Psi \left(\pi_j F_{\Lambda_t}^{-1}(w_{j,t}) \right) \quad (4.21a)$$

$$\forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, y_t \ln \left(\frac{f_j(1-\varepsilon)}{p_j} \right) = \ln(w_{j,t}) \quad (4.21b)$$

$$\sum_{j=1}^n f_j = 1$$

$$\sum_{t=1}^T y_t = 1$$

$$\forall j \in \llbracket 1; J \rrbracket, f_j < \frac{p_j}{1-\varepsilon}$$

$$x \in (\mathbb{Z}^+)^S, f \in]0; 1[^n, y \in]0; 1[^T .$$

Linearizations of reformulated constraints defined by (4.21a) and (4.21b) are proposed in order to provide an approximation of the problem. The linearizations are computed in order to propose a lower approximation of the initial MINLP problem

(4.21).

Piecewise linearization of Constraints (4.21a)

The constraints defined (4.21a) are approximated through a piecewise tangent approximation in order to provide a lower bound of the Right-Hand Side. Indeed, the functions

$$\forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, F_{B_{t,j}}^{-1} : w \mapsto \Psi \left(\pi_j F_{\Lambda_t}^{-1}(w) \right) \quad (4.22)$$

are increasing by composition of increasing functions Ψ and $F_{\Lambda_t}^{-1}$. Previous Section 4.2 notes that F_B^{-1} is convex.

In order to keep the formulae clear, it is considered that the number of linearization points n does not depend on j neither t . Since the constraints have the same formulation, this simplification makes sense. It is still possible with our approach to choose different point numbers for each constraint of (4.22) if needed.

Let $w_n \in]0; 1[$, $j \in \llbracket 1; N \rrbracket$ be N points such that $w_1 < w_2 < \dots < w_N$. The tangent approximated function $\hat{F}_{B_{t,j,n}}^{-1}$ of a function $F_{B_{t,j}}^{-1}$ around point w_n is given by

$$\begin{aligned} \forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, \hat{F}_{B_{t,j,n}}^{-1}(w) &= F_{B_{t,j}}^{-1}(w_n) + (w - w_n)(F_{B_{t,j}}^{-1})'(w_n) \\ &= \frac{\Psi' \left(\pi_j F_{\Lambda_t}^{-1}(w_n) \right)}{f_{\Lambda_t} \left(\pi_j F_{\Lambda_t}^{-1}(w_n) \right)} w \\ &\quad + \Psi(\pi_j F_{\Lambda_t}^{-1}(w_n)) - w_n \frac{\Psi' \left(\pi_j F_{\Lambda_t}^{-1}(w_n) \right)}{f_{\Lambda_t} \left(\pi_j F_{\Lambda_t}^{-1}(w_n) \right)} \\ &= \delta_{t,j,n} w + \alpha_{t,j,n} . \end{aligned} \quad (4.23)$$

Equation (4.23) gives the tangent approximation of a constraint of the set (4.21a) around point w_n .

Thus, the constraints (4.21a) are reformulated as following:

$$\forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, \forall n \in \llbracket 1; N \rrbracket, A_t x \geq \delta_{t,j,n} \times w_{j,t} + \alpha_{t,j,n} . \quad (4.24)$$

Piecewise linearizations of the LHS of Constraints (4.21b)

Second, the Left-Hand Side is linearized by giving piecewise tangent and secant approximations of \ln function. These upper and lower approximations are applied to give bounds of the Right-Hand Side.

The piecewise linear secant approximation is the same formulation as (4.29) above: considering K points $f_1 < f_2 < \dots < f_K$, the secant approximation between f_k and f_{k+1} is given by

$$\hat{\ln}\left(\frac{f(1-\varepsilon)}{p_j}\right) = d_{1,j,k}f + r_{1,j,k}. \quad (4.25)$$

These linearizations provide a lower bound with a piecewise secant approximations and require to introduce binary variables: one binary variable for each section defined by to interpolation points f_k and f_{k+1} , $k \in \llbracket 1; K-1 \rrbracket$. These variables are denoted $\forall j \in \llbracket 1; J \rrbracket$, $\forall t \in \llbracket 1; T \rrbracket$, $\forall k \in \llbracket 1; K-1 \rrbracket$, $\kappa_{j,t,k}$. Hence, the resulting constraints are:

$$\begin{aligned} \forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, \forall k \in \llbracket 1; K-1 \rrbracket, \\ \kappa_{j,t,k} \times y_t (d_{1,j,k}f_j + r_{1,j,k}) \leq p_m w_{j,t} + q_m. \end{aligned} \quad (4.26)$$

where $\forall j \in \llbracket 1; J \rrbracket$, $\forall t \in \llbracket 1; T \rrbracket$, $\sum_{k=1}^{K-1} \kappa_{j,t,k} = 1$.

The product of binary variable $\kappa_{j,t,k}$ and continuous y_t variables is reformulated according to [Liberti et al. \[2009\]](#): we introduce $z_{j,t,k}$ a continuous variable in $]0; 1[$ and replace $\kappa_{j,t,k} \times y_t$ by $z_{j,t,k}$. Then, new constraints are added:

$$z_{j,t,k} < \kappa_{j,t,k} \quad (4.27a)$$

$$z_{j,t,k} \leq y_t \quad (4.27b)$$

$$\kappa_{j,t,k} + y_t - z_{j,t,k} < 1. \quad (4.27c)$$

Finally, with L chosen points $f_1 < f_2 < \dots < f_L$ the piecewise tangent approximation of \ln around f_l is given by:

$$\begin{aligned} \hat{\ln}\left(\frac{f(1-\varepsilon)}{p_j}\right) &= \frac{f}{f_l} + \ln\left(\frac{f_l(1-\varepsilon)}{p_j}\right) - 1 \\ &= d_{2,j,l}f_j + r_{2,j,l}. \end{aligned} \quad (4.28)$$

Piecewise linearizations of the RHS of Constraints (4.21b)

The set of equality constraints described by (4.21b) is approximated by introducing bounds of the linear approximation of function \ln .

First, we give a linearization of the Right-Hand Side which is a simple \ln function.

Given M chosen points such that $w_1 < w_2 < \dots < w_M$, the piecewise secant approximation of \ln function between points w_m and w_{m+1} is

$$\begin{aligned}\hat{\ln}(w) &= \frac{\ln w_{m+1} - \ln w_m}{w_{m+1} - w_m} w + \frac{w_{m+1} \ln w_m - w_m \ln w_{m+1}}{w_{m+1} - w_m} \\ &= p_m w + q_m.\end{aligned}\quad (4.29)$$

Like previously, binary variables $\forall j \in \llbracket 1; J \rrbracket$, $\forall t \in \llbracket 1; T \rrbracket$, $\forall m \in \llbracket 1; M-1 \rrbracket$, $\tau_{j,t,m}$ are introduced. For each linearization, only one τ is equal to 1:

$$\forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, \sum_{m=1}^{M-1} \tau_{j,t,m} = 1$$

Again, the products of binary variable $\tau_{j,t,m}$ and continuous variable $w_{j,t}$ is reformulated by introducing a continuous variable $u_{j,t,m} \in]0; 1[$:

$$u_{j,t,m} < \tau_{j,t,m} \quad (4.30a)$$

$$u_{j,t,m} \leq w_{j,t} \quad (4.30b)$$

$$\tau_{j,t,m} + w_{j,t} - u_{j,t,m} < 1. \quad (4.30c)$$

By gathering these linearizations, the constraints (4.21b) are approximated as follows:

$$\begin{aligned}\forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, \forall (k, m, l) \in \llbracket 1; K-1 \rrbracket \times \llbracket 1; M-1 \rrbracket \times \llbracket 1; L \rrbracket, \\ z_{j,t,k}(d_{1,j,k} f_j + r_{1,j,k}) \leq p_m u_{j,t,m} + q_m \tau_{j,t,k} \leq y_t (d_{2,j,l} f_j + r_{2,j,l})\end{aligned}\quad (4.31)$$

and

$$\begin{aligned} \forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, \forall k \in \llbracket 1; K-1 \rrbracket, z_{j,t,k} < \kappa_{j,t,k} \\ z_{j,t,k} \leq y_t \\ \kappa_{j,t,k} + y_t - z_{j,t,k} < 1 \end{aligned}$$

$$\begin{aligned} \forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, \forall k \in \llbracket 1; M-1 \rrbracket, u_{j,t,m} < \tau_{j,t,m} \\ u_{j,t,m} \leq w_{j,t} \\ \tau_{j,t,m} + w_{j,t} - u_{j,t,m} < 1 \end{aligned}$$

$$\begin{aligned} \forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, \sum_{k=1}^{K-1} \kappa_{j,t,k} = 1 \\ \sum_{m=1}^{M-1} \tau_{j,t,m} = 1 . \end{aligned}$$

Final Mixed-Integer Formulation

Finally, the mixed-integer program is composed of the previous approximated constraint:

$$\min c^t x \quad (4.32a)$$

$$\text{s.t. } \forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, \forall n \in \llbracket 1; N-1 \rrbracket, A_t x \geq \delta_{t,j,n} \times w_{j,t} + \alpha_{t,j,n} \quad (4.32b)$$

$$\forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, \forall (k, m, l) \in \llbracket 1; K-1 \rrbracket \times \llbracket 1; M-1 \rrbracket \times \llbracket 1; L \rrbracket,$$

$$z_{j,t,k}(d_{1,j,k}f_j + r_{1,j,k}) \leq p_m u_{j,t,m} + q_m \tau_{j,t,k} \leq y_t(d_{2,j,l}f_j + r_{2,j,l}) \quad (4.32c)$$

$$\sum_{j=1}^n f_j = 1 \quad (4.32d)$$

$$\sum_{t=1}^T y_t = 1 \quad (4.32e)$$

$$\forall j \in \llbracket 1; J \rrbracket, f_j < \frac{p_j}{1 - \varepsilon} \quad (4.32f)$$

$$\forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, \sum_{k=1}^{K-1} \kappa_{j,t,k} = 1 \quad (4.32g)$$

$$\sum_{l=1}^{M-1} \tau_{j,t,m} = 1 \quad (4.32h)$$

$$\forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, \forall k \in \llbracket 1; K-1 \rrbracket, z_{j,t,k} < \kappa_{j,t,k} \quad (4.32i)$$

$$z_{j,t,k} \leq y_t \quad (4.32j)$$

$$\kappa_{j,t,k} + y_t - z_{j,t,k} < 1 \quad (4.32k)$$

$$\forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, \forall k \in \llbracket 1; M-1 \rrbracket, u_{j,t,m} < \tau_{j,t,m} \quad (4.32l)$$

$$u_{j,t,m} \leq w_{j,t} \quad (4.32m)$$

$$\tau_{j,t,m} + w_{j,t} - u_{j,t,m} < 1 \quad (4.32n)$$

$$x \in (\mathbb{Z}^+)^S, \quad (4.32o)$$

$$\forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, (f_j, y_t) \in]0; 1[^2 \quad (4.32p)$$

$$\forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, \forall k \in \llbracket 1; K-1 \rrbracket, (\kappa_{j,t,k}, \tau_{j,t,k}) \in \{0; 1\}^2 \quad (4.32q)$$

$$\forall j \in \llbracket 1; J \rrbracket, \forall t \in \llbracket 1; T \rrbracket, \forall k \in \llbracket 1; K-1 \rrbracket, (z_{j,t,k}, u_{j,t,k}) \in]0; 1[^2. \quad (4.32r)$$

The constraints (4.32b) is the linearization of previous set of constraints (4.21a). Constraints defined by (4.32d, 4.32e, 4.32f) are introduced by the flexible sharing out of the risk through the different scenarios and different periods of time.

Constraints (4.32c) show the piecewise linear approximations, providing lower and upper bounds of the linearized \ln function. The lower bounds required to introduce binary variables for each linearization interval. Only one function should be active for each variable $w_{j,t}$. This is expressed in constraints (4.32g).

The reformulation of the product of continuous variables introduced by the flexible sharing out of the risk and binary variables required for the lower bounds of these constraints led to the introduction of $z_{j,t,k}$ and $u_{j,t,k}$ variables in (4.32c) and the introduction of new constraints (4.32i, 4.32j, 4.32k) as well as (4.32l, 4.32m, 4.32n).

However this program is not linear and is usually intractable. A two-step approach may enable to find an alternative tractable program.

4.3.4 A Tractable Approximated Formulation

In order to propose a tractable program, the previous program (4.32) may be solved in two steps: first, the problem is solved for one scenario. It is consistent to consider the existence of the scenario of *business* factor equal to 1. Thus, the problem is solved for this scenario. The values of y_t are then used for solving the full program. The product of binary and continuous variables can be reformulated with the same approach as previously. Thus, a linear program would be obtained.

This would lead to an intermediate approach between the flexible sharing out of the risk and the Bonferroni approximation (with an equal splitting): risk levels are partially pre-determined, but their values are optimized on one scenario. However, we let this reformulation for further work.

4.4 Conclusion

This chapter proposed two extensions for the staffing and shift-scheduling problem formulation.

First, the previous one-step approach with uncertain arrival rates was evaluated on an advanced queuing system model. Instead of modeling a $M_t/M/c_t$ with Erlang-C, we introduced a limited customer patience and abandonments by considering an $M_t/M/c_t +$

M queue with Erlang-A model. Erlang-A requires more computationally efforts than Erlang-C. However, it fits in our approach without involving tractability issue. This one-step approach was successfully applied to both Erlang-C and Erlang-A models and the results show the interest of the joint chance-constrained approach in term of robustness and the advantage of the dynamic sharing out of the risk in term of cost reduction.

Numerical results were given for this advanced queuing model. First, an instance was given in order to illustrate the robustness of the approach regarding the Quality of Service. Second, Erlang-C and Erlang-A solution costs were compared and showed the cost interest of Erlang-A compared to an Erlang-C-based approach. Third, the real-based one-step approach was compared to the integer-based one-step approach. Again, our model was adapted to Erlang-A and provided similar qualitative results as with Erlang-C. Moreover, the convexity of the new ψ_a function enables to reduce the gap between the upper and lower bounds.

Second, we focused on an widely-used assumption when approximating non-stationary queues: independence of the periods. This assumption is indeed the basis of the SIPP (Stationary Independent Period-by-Period) approximation. Thus, arrival rates random variables at each period were multiplied by a *busyness* factor. This factor is a discrete random variable representing the random load of a day. A joint chance-constrained approach was proposed. The reformulation into individual chance constraints led to a complex program. Piecewise linear approximations and reformulations were proposed in order to consider a flexible sharing out of the risk, and thus pursue our previous original idea. A final mixed-integer non-linear program was obtained. Because of its intractability, a two-step approach could be used to approximate the program into two mixed-integer linear programs. Unfortunately, numerical experiments on this latter are not done yet.

For future research, an implementation of this two-step mixed-integer linear program would be very interesting. Indeed, except with simulation-based approaches, daily period dependence is hardly considered in the scheduling optimization literature. The model would show the interest of including the period correlation by comparing the solution costs and robustness.

Chapter 5

Distributionally Robust Approach

This chapter focuses on the staffing and shift-scheduling problem in call centers under the assumption that the call arrival rates are subject to uncertainty and are following unknown probability distributions. We assume that we only know the means and variances of the distributions and thus propose to model this optimization problem as a distributionally robust program with joint chance constraint. Moreover, the risk level is dynamically shared throughout the entire scheduling horizon during the optimization process. We propose a non-linear deterministic equivalent of the problem and solve linear approximations of the constraints to provide upper and lower bounds of the optimal solution cost. We apply our approach on a real-life instance and give numerical results. Finally, we showed the practical interest of this approach compared to a stochastic approach in which the choice of the distribution is incorrect.

5.1 Introduction

In this chapter, the call center model is similar to the model presented in previous Chapter 3: the call center model is based on an Erlang model -both Erlang-C and Erlang-A are fit our approach- and forecasts of arrival rates are computed from historical data. Thus, they are subject to uncertainty. The Expected Waiting Time is the Quality of Service and the risk level is set for the entire working horizon.

However in previous chapters, we considered that we knew the right distribution to represent the uncertainty. This distribution depends in fact on the call center type and on the period of the year. The adequate distributions are actually difficult to assess, and sometimes they are varying in time. Then, what if we do not know the distribution? What are the consequences if the assumed distribution is not the right one? In our approach, we consider that the respect of the constraint under the given risk level is an essential criterion. The choice of the distribution can be crucial. The stochastic programming approach presented in previous chapters may not be enough anymore.

As a consequence, we investigate a way of solving the problem for unknown distributions and propose in this chapter another approach, the distributionally robust approach, in which the distribution is considered as unknown. The goal is then to optimize the call center costs assuming that the distribution belongs to a set of possible distributions. This set includes all the possible distributions with given mean and variance. Very few publications present the distributionally approach for the scheduling problem in queues.

In this chapter, the staffing and shift-scheduling problems are solved in two steps. The problem is formulated with uncertain mean arrival rates and a joint chance-constrained mixed-integer nonlinear program.

First, this program is written with the distributionally robust assumption, and the constraints are to be respected for a given set of probability distributions instead of a given distribution. We assume that the first two moments of the distribution are known. The periods are supposed to be independent.

Second the risk level is set on the whole horizon of study with joint chance constraint and the dynamic sharing out of the risk through the periods is applied. Finally, the linear approximations lead to solutions giving bounds of the optimal solution cost.

Numerical experiments give an overview of the robustness and the cost of this approach.

The staffing model is briefly presented in Section 5.2. Section 5.3 investigates the

impact in term of robustness of a wrong assumption on the choice of the probability distribution. This example takes the assumption of the previous chapters, that is to say the random variables are following normal distributions. Section 5.4 presents the distributionally robust chance-constrained approach and propose computations leading to the deterministic equivalent of the program. We also describe the piecewise linear approximations leading to the final programs whose solution costs are lower and upper bounds of the initial optimal solution cost. Finally, some numerical results and comparisons are done in Section 5.5.

5.2 Staffing Model

In this chapter, the staffing model is the same than in the two-step approach introduced in Chapter 3. An approach based on the ψ function as presented in 3.4.1 could be used instead of this staffing solution. The model establishment for the distributionally robust approach in next sections would still be valid.

The call center is considered to be a $M_t/M/c_t$ queue where stationary regime is reached. Expected Waiting Time in Erlang-C model is reminded here:

$$\begin{aligned} EWT(c, \lambda, \mu) &= \mathbb{E}[Wait] = \mathbb{P}(Wait > 0) \mathbb{E}[Wait | Wait > 0] \\ &= \frac{1}{c\mu(1 - \frac{\lambda}{c\mu}) \left(1 + (1 - \frac{\lambda}{c\mu}) \sum_{m=0}^{c-1} \frac{c!}{m!} (\frac{\mu}{\lambda})^{c-m} \right)}. \end{aligned} \quad (5.1)$$

The staffing load is the minimum value of c which respects the targeted service level for given arrival rate λ and service rate μ computed through the EWT function:

$$c_t = \min\{c \in \mathbb{N} | EWT(c, \lambda_t, \mu) \leq EWT^*\}. \quad (5.2)$$

In next sections, c_t is denoted by μ_t , considering that it is the mean value of the random variable of the agent requirements.

5.3 Schedule Solution with a Wrong Assumption on Call Arrival Rates Distributions

5.3.1 Stochastic Program Formulation

This section investigates the possibility of a wrong assumption on the chosen distribution in the stochastic optimization approach. The model is based on the standard chance-constrained approach under the assumption that we know the distribution of the random variables. What if the problem is developed under the assumption of the normal distribution which actually appears to be the wrong distribution?

Agent requirements are random variables following a normal distribution with known means and variances.

The two-step stochastic program approach previously presented in Section 3.3.3 is considered in this part:

$$\begin{aligned}
 & \min c^t x \\
 & \text{s.t. } \forall t \in \llbracket 1; T \rrbracket, \frac{A_t x - \mu_t}{\sigma_t} \geq F_{\beta}^{-1}((1 - \varepsilon)^{y_t}) \\
 & \sum_{t=1}^T y_t = 1 \\
 & x \in (\mathbb{Z}^+)^S, \quad \forall t \in \llbracket 1; T \rrbracket, y_t \in]0; 1[.
 \end{aligned} \tag{5.3}$$

with $\varepsilon \in]0; 1]$, and μ_t and σ_t^2 the expected value and variance of the random variable B_t at period t respectively. Indeed, the distribution is supposed to be known and is namely the normal distribution. Random variables are supposed to be independent from each other. The variable β denotes the standard normal deviate and the function $F_{\beta \sim \mathcal{N}(0,1)}^{-1}$ denotes the inverse of its cumulative distribution function.

Piecewise tangent and linear approximations give upper and lower bound programs. Since we are considering the normal distribution, the convexity of $y_t \mapsto F_{\beta \sim \mathcal{N}(0,1)}^{-1}((1 - \varepsilon)^{y_t})$ is known for a probability $1 - \varepsilon \geq \frac{1}{2}$. Hence, piecewise linearizations described in Section 3.5 can be applied and are guaranteed to give upper and lower bounds.

The constraints have the following expression:

$$\frac{A_t x - \mu_t}{\sigma_t} \geq \delta_{t,j} \times y_t + \alpha_{t,j}. \tag{5.4}$$

With $p = 1 - \epsilon$, the coefficients for the lower bound program are

$$\begin{aligned} \forall j \in \llbracket 1; n \rrbracket, \hat{F}_{l,\beta,j}^{-1}(p^y) &= F_{\beta}^{-1}(p^{y_j}) + (y - y_j)(F_{\beta}^{-1})'(p^{y_j})p^{y_j} \ln(p) \\ &= \delta_{l,j} \times y + \alpha_{l,j} \end{aligned} \quad (5.5)$$

where

$$(F_{\beta}^{-1})'(p^{y_j}) = \frac{1}{F_{\beta}'(F_{\beta}^{-1})(p^{y_j})} = \frac{1}{f_{\beta}(F_{\beta}^{-1}(p^{y_j}))}$$

Likewise, the coefficients for the upper bound program are

$$\begin{aligned} \forall j \in \llbracket 1; n-1 \rrbracket, \hat{F}_{u,\beta,j}^{-1}(p^y) &= F_{\beta}^{-1}(p^{y_j}) + \frac{y - y_j}{y_{j+1} - y_j} (F_{\beta}^{-1}(p^{y_{j+1}}) - F_{\beta}^{-1}(p^{y_j})) \\ &= \delta_{u,j} \times y + \alpha_{u,j} . \end{aligned} \quad (5.6)$$

These programs are to give bounds of the optimal solution cost considering that call arrival rates are following normal distributions with given means and variances. We would like to evaluate the significance of this assumption, in the case it is wrong.

5.3.2 Numerical Results in Case of a Wrong Distribution Considered

As shown in the previous chapter, even if the one-step approach is better, this approach is satisfactory in terms of robustness when considering that the assumption of normal distribution is right. The following numerical experiments evaluate the robustness of the solutions in the event of the right distribution being not the normal distribution, but another one.

This experiment comprises two steps:

- First, upper x_u and lower x_l schedule solutions are computed through the mixed integer linear programs. The data used for the instance is taken from a health insurance call center. These solutions are based on the assumption that the requirements are following normal distributions.
- Second, scenarios are elaborated with another assumption: instead of considering the normal distribution, other distributions are taken into account. A scenario is a simulation of possible realizations of agents requirements in the call centers on the whole schedule horizon. The same data and parameters as previously are used

and one of these distributions is considered for the random variable realizations: gamma distribution, uniform distribution, Pareto distribution, and variations of normal distributions (log-normal, folded normal).

For each distribution, 100 scenarios are computed. The values of simulated requirements are compared with the scheduled requirements of our approximated solutions. A scenario is considered as violated when for at least one period of the horizon, the realization of simulated requirements is higher than what the schedule solution proposes. It means that our schedule solutions provide insufficient number of agents compared to what is needed in the scenario.

The result of the second step on an instance are seen in Table 5.1.

The parameters of the instance are the following:

- $EWT^* = 1$
- $\varepsilon = 0.05$
- λ range: 6 – 16 calls/min
- σ^2 range: $[\frac{\lambda_l}{4}; \frac{\lambda_l}{2}]$

The scheduling horizon is a 10-hour day of 20 30-minute periods.

Table 5.1: Robustness of the stochastic approach for different wrong distributions. Targeted risk level is 5%.

Distribution	Percentage of Violations	
	For Lower Bound	For Upper Bound
Normal	6%	3%
Gamma	15%	7%
Folded Normal	5%	2%
Log-normal	10%	6%
Pareto	3%	1%

Table 5.1 presents the percentages of violations of our solutions if the real distributions are the ones indicated. The first line shows the quality of the stochastic approach when the assumption of normal distribution is right.

The second line of the table shows that if the right distribution is the gamma distribution, both the lower and the upper staffing solutions do not give a satisfactory risk

management. The solutions were computed in order to respect a risk of 5% but 7% to 15% of scenarios were violated.

It can be noticed that if we wrongly choose the normal distribution instead of the Gamma or the Log-normal distribution, the quality of the solution is not satisfactory anymore. The other distributions still respect the targeted risk level on this batch of scenarios.

Note It has been noticed some rare batches where the scenarios generated with the Pareto distribution did not respect the risk level for both upper and lower solutions. As far as we tested, the Folded normal distribution scenarios showed violations only for upper solutions, but the percentage was high compared to the allowed risk level (between 3 and 4 times).

This result shows that if the choice of the distribution is wrong, the resulted staffing solutions can not longer be considered as satisfactory.

Hence the following sections focus on proposing an approach when the exact probability distribution is assumed to be unknown. The mean and variance are the only information about the distribution of the random variable.

5.4 Distributionally Robust Model

5.4.1 Distributionally Robust Optimization Program

It is assumed that the distributions of the random vector B are not exactly known. However means μ and variances σ^2 vectors are supposed to be known.

The distributionally robust approach consists in optimizing the agents costs considering that we want to respect our constraints for all the possible distributions. The correct distribution is not known but the problem is to be optimized for all the possible distributions and thus for the most unfavourable distribution with known expected values and variances.

Hence, the joint chance constraint

$$\mathbb{P}(Ax \geq B) \geq 1 - \varepsilon$$

becomes

$$\inf_{B \sim (\mu, \sigma^2)} \mathbb{P}(Ax \geq B) \geq 1 - \varepsilon$$

where $B \sim (\mu, \sigma^2)$ denotes the vector of variables B_t , with means μ_t and variances σ_t^2 .

The resulting joint chance-constrained distributionally robust optimization program is

$$\begin{aligned} \min \quad & c^t x \\ \text{s.t.} \quad & \inf_{B \sim (\mu, \sigma^2)} \mathbb{P}(Ax \geq B) \geq 1 - \varepsilon \\ & x \in (\mathbb{Z}^+)^S \end{aligned} \tag{5.7}$$

where c is the cost vector, x is the agents vector, B is the vector of agents requirements b_t and A is the matrix of shifts. The variable ε is the risk level.

This program minimizes the cost of working agents while respecting the chosen objective service level for the time horizon under the risk level ε .

In order to propose a resolution of this problem, the next section provides a deterministic equivalent of the constraint.

5.4.2 Deterministic Equivalent Problem

[Bertsimas and Popescu \[2005\]](#) (Theorem 6.1) give a tight bound based on Chebyshev's inequality for a random vector X with mean vector $M = \mathbb{E}[X]$ and covariance matrix $\Sigma = \mathbb{E}[(X - M)^t(X - M)]$ and a convex event S :

$$\sup_{X \sim (M, \Sigma)} \mathbb{P}(X \in S) = \frac{1}{1 + d^2} \tag{5.8}$$

where $d^2 = \inf_{x \in S^c} (x - M)^t \Sigma^{-1} (x - M)$.

Since our goal is the expression of the infimum, we first split our chance constraint with the same method as in Chapter 3: a dynamic sharing out of the global risk level through the periods. Since the random variables are supposed to be independent, the same process occurs.

The constraint

$$\inf_{B \sim (\mu, \sigma^2)} \mathbb{P}(Ax \geq B) \geq 1 - \varepsilon \tag{5.9}$$

is then equivalent to

$$\inf_{B \sim (\mu, \sigma^2)} \prod_{t=1}^T \mathbb{P}(Ax \geq B) \geq 1 - \varepsilon, \quad (5.10)$$

and finally

$$\prod_{t=1}^T \inf_{B \sim (\mu, \sigma^2)} \mathbb{P}(Ax \geq B) \geq 1 - \varepsilon. \quad (5.11)$$

Finally, the program is the following:

$$\begin{aligned} & \min c^t x \\ & \text{s.t. } \forall t \in \llbracket 1; T \rrbracket, \quad \inf_{B_t \sim (\mu, \sigma_t^2)} \mathbb{P}(A_t x \geq B_t) \geq (1 - \varepsilon)^{y_t} \\ & \quad \sum_{t=1}^T y_t = 1 \\ & \quad x \in (\mathbb{Z}^+)^S, \forall t \in \llbracket 1; T \rrbracket, y_t \in]0; 1[\end{aligned} \quad (5.12)$$

where the variables y_t represent the proportion of risk allocated to each period t .

Let us focus on the expression of one constraint. For a given period t , we have:

$$\inf_{B_t \sim (\mu, \sigma_t^2)} \mathbb{P}(A_t x \geq B_t) \geq (1 - \varepsilon)^{y_t}. \quad (5.13)$$

Then, noticing that

$$\inf_{B_t \sim (\mu_t, \sigma_t^2)} \mathbb{P}(A_t x \geq B_t) = 1 - \sup_{B_t \sim (\mu_t, \sigma_t^2)} \mathbb{P}(A_t x < B_t), \quad (5.14)$$

it is sufficient to solve the following optimization problem in order to have an equivalent expression of our constraint through (5.8):

$$\begin{aligned} & \min (b_t - \mu_t)^2 \sigma^{-2} \\ & \text{s.t. } A_t x - b_t \leq 0 \\ & \quad b_t \in \mathbb{Z}^+. \end{aligned} \quad (5.15)$$

The result of the objective function at the optimal value b_t^* will give the value of d^2 in (5.8) we are looking for.

The optimal solution is obviously

$$A_t x \geq \mu_t, \quad b_t^* = A_t x. \quad (5.16)$$

Finally,

$$d^2 = \frac{\sigma^2}{\sigma^2 + (A_t x - \mu_t)^2} \quad (5.17)$$

and we obtain the following expression for our constraint:

$$\sup_{B_t \sim (\mu_t, \sigma_t^2)} \mathbb{P}(A_t x < B_t) = \begin{cases} \frac{\sigma_t^2}{\sigma_t^2 + (A_t x - \mu_t)^2} & \text{if } A_t x \geq \mu_t \\ 1 & \text{otherwise.} \end{cases} \quad (5.18)$$

The constraint (5.13) is respected if and only if

$$A_t x - \mu_t \geq 0, \quad \frac{(A_t x - \mu_t)^2}{\sigma_t^2 + (A_t x - \mu_t)^2} \geq (1 - \varepsilon)^{y_t}.$$

Then, an equivalent of the constraint can be given:

$$\begin{aligned} \inf_{B_t \sim (\mu_t, \sigma_t^2)} \mathbb{P}(A_t x \geq B_t) \geq (1 - \varepsilon)^{y_t} &\Leftrightarrow \frac{(A_t x - \mu_t)^2}{\sigma_t^2 + (A_t x - \mu_t)^2} \geq (1 - \varepsilon)^{y_t} \\ &\Leftrightarrow \frac{(A_t x - \mu_t)^2}{\sigma_t^2} \geq \frac{(1 - \varepsilon)^{y_t}}{1 - (1 - \varepsilon)^{y_t}}. \end{aligned} \quad (5.19)$$

The probability is denoted $p = 1 - \varepsilon$ and, since $A_t x - \mu_t \geq 0$, we have the following result:

$$\frac{A_t x - \mu_t}{\sigma_t} \geq \sqrt{\frac{p^{y_t}}{1 - p^{y_t}}}. \quad (5.20)$$

We now have a deterministic equivalent of our distributionally robust chance-constrained approach. In the next section, the Right-Hand Side of the equation is linearized in order to obtain bounds of the optimal solution cost through a lower bound program and an upper bound program.

5.4.3 Linear Approximations

The lower and upper bounds of the problem cost are proposed by linearizing the Right-Hand Side.

The Right-Hand Side is expressed by the following function:

$$f:]0; 1[\rightarrow \mathbb{R}^+ \quad (5.21)$$

$$y \mapsto \sqrt{\frac{p^y}{1-p^y}}$$

with $\varepsilon \in]0; 1]$ and $p = 1 - \varepsilon$.

By deriving this function twice, the convexity of the function is proven.

Function f is C^∞ , so the second derivative of function f can be computed.

First,

$$\begin{aligned} \frac{df}{dy} &= \frac{\frac{\ln p}{2} p^{\frac{y}{2}} (1-p^y)^{\frac{1}{2}} + \frac{\ln p}{2} p^y (1-p^y)^{-\frac{1}{2}} p^{\frac{y}{2}}}{1-p^y} \\ &= \frac{\ln(p)(1-p^y)^{-\frac{1}{2}}(p^{\frac{y}{2}}(1-p^y) + p^{\frac{3}{2}y})}{2(1-p^y)} \\ &= f(y) \frac{\ln p}{2(1-p^y)}. \end{aligned}$$

Then,

$$\begin{aligned} \frac{d^2f}{dy^2} &= \frac{\ln p f'(y)(1-p^y) + \ln(p)p^y f(y)}{2(1-p^y)^2} \\ &= \frac{\ln^2(p)(1+2p^y)}{4(1-p^y)^2} f(y) \\ &= \frac{\ln^2(p)(1+2p^y)}{4(1-p^y)^2} \frac{p^{\frac{y}{2}}}{(1-p)^{\frac{y}{2}}}. \end{aligned}$$

Since every term of the second derivative is positive, we conclude that $\frac{d^2f}{dy^2}$ is positive and then, f is convex. \square

This result guarantees that the following linearizations are upper and lower bounds of the functions, that is to say, linearizations are always above or below the function's curve respectively.

The graph 5-I shows an illustration of the piecewise approximations of function f .

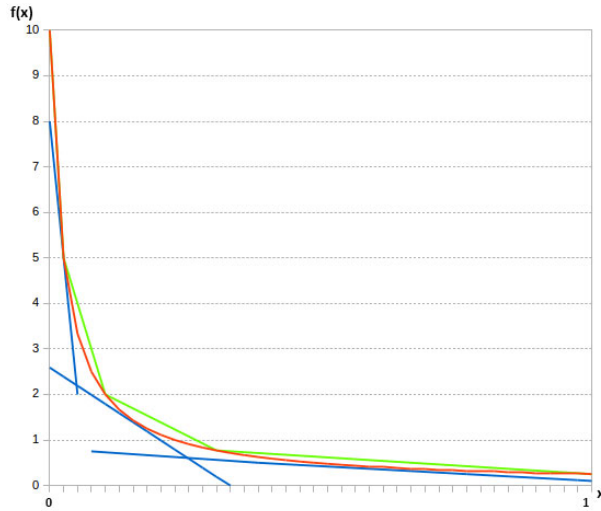


Figure 5-1: An illustrative example of piecewise linear approximations of function f .

Piecewise Tangent Approximation

The secant linearization of $f: y \mapsto \sqrt{\frac{p^y}{1-p^y}}$ gives a lower bound of the initial problem's objective function.

As in Chapter 3, let $y_j \in]0; 1]$, $j \in \llbracket 1; n \rrbracket$ be n points such that $y_1 < y_2 < \dots < y_n$.

The piecewise tangent approximation is given by a linearization around these n points. The linearisation $\hat{f}_{l,j}$ around the point y_j (the subscript l stands for *lower*) is given by a first-order Taylor series expansion:

$$\begin{aligned}
 \forall j \in \llbracket 1; n \rrbracket, \hat{f}_{l,j}(y) &= f(y_j) + (y - y_j)f'(y_j) \\
 &= f(y_j) + (y - y_j)f(y_j) \frac{\ln p}{2(1 - p^{y_j})} \\
 &= \delta_{l,j} \times y + \alpha_{l,j}.
 \end{aligned} \tag{5.22}$$

In the program, the condition is expressed in each period with one constraint for each approximation point. Finally, a lower bound of the solution cost of the initial problem

is given by the following program:

$$\begin{aligned}
& \min c^t x \\
& \text{s.t. } \forall t \in \llbracket 1; T \rrbracket, \forall j \in \llbracket 1; n \rrbracket, \frac{A_t x - b_t}{\sigma_t} \geq \delta_{l,t,j} y_t + \alpha_{l,t,j} \\
& \sum_{t=1}^T y_t = 1 \\
& x \in (\mathbb{Z}^+)^S, \forall t \in \llbracket 1; T \rrbracket, y_t \in]0; 1[
\end{aligned} \tag{5.23}$$

where S is the number of shifts and T the number of periods.

Piecewise Linear Approximation

Similarly, an upper bound of the Right-Hand Side is given with a piecewise linear approximation.

Let $y_j \in]0; 1]$, $j \in \llbracket 1; n \rrbracket$ be n points such that $y_1 < y_2 < \dots < y_n$.

The piecewise linear approximated function is denoted by $\hat{f}_{u,j}$ between the points y_j and y_{j+1} (the subscript u stands for *upper*) is given by secant linearization between them:

$$\begin{aligned}
\forall j \in \llbracket 1; n-1 \rrbracket, \hat{f}_{u,j}(y) &= f(y_j) \\
&+ \frac{y - y_j}{y_{j+1} - y_j} (f(y_{j+1}) - f(y_j)) \\
&= \delta_{u,j} \times y + \alpha_{u,j} .
\end{aligned} \tag{5.24}$$

Finally, the following program gives an upper bound of our optimal solution cost:

$$\begin{aligned}
& \min c^t x \\
& \text{s.t. } \forall t \in \llbracket 1; T \rrbracket, \forall j \in \llbracket 1; n-1 \rrbracket, \\
& A_t x - b_t \geq (\delta_{u,t,j} \times y_t + \alpha_{u,t,j}) \sigma_t \\
& \sum_{t=1}^T y_t = 1 \\
& x \in (\mathbb{Z}^+)^S, \forall t \in \llbracket 1; T \rrbracket, y_t \in]0; 1[
\end{aligned} \tag{5.25}$$

where S is the number of shifts and T the number of periods.

This section presented a distributionally robust approach aiming to solve an chance-constrained optimization program where random variables are following unknown dis-

tributions with fixed means and variances. A deterministic equivalent to the initial distributionally robust stochastic problem was proposed, using an extension of Chebyshev's inequality. We had to deal with a mixed-integer nonlinear program.

Similarly as in previous chapters, the dynamic sharing out of the risk was taking into account and upper and lower bounds of the optimal solution cost were provided by computing piecewise tangent and secant approximations. This was possible because of the convexity of the RHS of the constraints. This led to two mixed-integer linear programs whose numbers of integer variables remains acceptable. This enables to give schedule solutions assuring bounds of the optimal solution cost of the initial complex problem.

Next section gives numerical experiments on a scheduling problem in call centers.

5.5 Numerical Experiments

5.5.1 Instance

In order to evaluate the quality and the robustness of our model, we applied our approach to instances based on the same data as previously in Section 5.3, from a health insurance call center.

The forecast horizon is one typical 10-hour day of a week. The horizon is split in 30-minute periods. 24 different shifts, from both full-time and part-time schedules, make up the matrix of shifts. The service rate μ is set to 1 without loss of generality, since it simply represents a multiplicity factor. The agents costs are supposed to be only proportional to the number of periods worked. Hence, agents have the same hourly wage.

The vectors of scheduled agents x_l and x_u are computed for one week with the two MILP programs (5.23) and (5.25), providing an upper bound and a lower bound of the optimal solution cost. The piecewise tangent and secant approximations were computed with 17 interpolation points. We noticed that the order of magnitude of variables y_l is between 10^{-2} and 10^{-1} , thus the gap between the upper and lower bounds was reduced by gathering most of the points around this area.

In order to evaluate the quality of our solutions x_l and x_u , possible realizations of arrival rates were simulated according to different distributions with the same data as previously.

As in Section 5.3, the simulation process is the following. Different possible distri-

butions are considered: gamma distributions, uniform distribution, Pareto distribution, and variations of normal distributions (log-normal, folded normal).

A scenario is elaborated as following: for each period of time a call arrival rate is simulated according a given probability distribution selected among the mentioned ones previously. Then we compute the number of effective required agents for each period. A scenario covers requirements for the whole time horizon. Finally we compare these values of requirements with our schedule solutions of the problem (lower solution x_l and upper solution x_u). A scenario is considered as violated if the scheduled solution x_l or x_u at one period at least is not enough in comparison of what the realization requires.

We computed between 100 and 500 scenarios for each probability distributions. The percentage of violations gives us an idea of the robustness of our approach for several chosen distributions.

5.5.2 Robustness of the Schedule Solutions

Table 5.2: Percentage of violations for different parameters.

Parameters				Results	
Variance range	λ range	EWT^*	Risk ϵ	Cost Gap	% Violations
0.3 – 1	16 – 86	1	15%	0.0045	5 – 3
0.3 – 5	16 – 86	1	15%	0.011	1 – 0
0.1 – 1	4 – 20	1	15%	0.027	13 – 9
0.1 – 1	4 – 20	1	10%	0.034	7 – 4
2.5 – 9	4 – 20	0.3	15%	0.052	1 – 1

Table 5.2 gives the percentage of violated scenarios for various ranges of values of means and variances, and risk level. The first column gives the range of values of the variances through the day. The second column gives the range of values of the means through the day, following a typical seasonality.

The variable Cost Gap (CG) of the 5th column is given by the relative difference between the cost of the upper solution and the cost of the lower solution:

$$CG = \frac{c^t x_u - c^t x_l}{c^t x_l} . \quad (5.26)$$

The last column gives the number of violated scenarios for the lower solution and for the upper solution.

In Table 5.2 we can notice that both upper and lower solutions respect the set risk level. The variations of the parameters show that the bigger the variances, the better the model. The distributionally robust model deals very well with increasing of variances. We notice that even if we allow 15% risk, only a few scenarios are violated when the variances are higher (second and last lines of Table 5.2). In these cases the call center is overstaffed and the given solutions seem too conservative.

However, it is important to note that all the observations are based on simulations of only a few examples of distributions. These very low percentages only show that if the arrival rates follow in reality one of the studied distributions, it may be overstaffed. Nevertheless the distributionally robust model indeed consists in taking all possible distributions with given mean and variance into account. Thus it may be possible to reach the maximum risk level with other particular distributions.

These results show that our approach is robust, considering the numbers of violations never exceed the targeted risk level. The values of Cost Gap show that the two bounds are close enough to select and provide a very close solution to optimal solution.

We can notice that even if the solutions costs are very close, the number of violations is different between the upper solution and the lower solution. This is due to the fact that the distribution of the agents through the different shifts is different according to the programs.

5.5.3 Costs of the Schedule Solutions

The costs of the solutions give an idea of the quality of the minimization.

Table 5.3 focuses on comparing, for different risk levels and values of EWT^* , the results in term of solution costs and violations. The simulations were made with the following parameters:

- $\mu = 1$
- λ follows a daily seasonality, varying between 4 calls/min and 21 calls/min
- σ^2 varies through the periods, between 0.25 and 1.

These parameters show well the performance of the model. Table 5.3 gives the costs of the two approximated solutions and the Cost Gap. Like previously 200 simulations

were executed and the number of violated scenarios was evaluated, which is given in the last column of the table.

Table 5.3: Results for different risk levels.

Parameters		Results				
EWT^*	Risk ϵ	Lower Cost	Upper Cost	Cost Gap	% Violations Upper Solution	% Violations Lower Solution
5	15%	27	27.75	0.028	4.5	3
5	10%	29	30	0.034	2.5	1
5	05%	33.81	35.31	0.044	0.5	0.5
1	15%	27.5	28.31	0.029	6	4
1	10%	29.44	30.5	0.036	3	2
1	05%	34.25	35.81	0.046	0.5	0.5
0.3	15%	28.88	29.75	0.030	5.5	3.5
0.3	10%	30.94	32	0.034	4	1
0.3	05%	35.69	37.31	0.045	0.5	0.5

The first two columns of Table 5.3 gives the chosen parameters for the problem. Columns 3 and 4 give the solution costs of the two programs and column 5 gives the Cost Gap. Finally, the two last columns give the number of violated scenarios for the two solutions.

Unsurprisingly, the cost of the solution increases when the risk level decreases. The Cost Gap seems to remain in a small range, even if we notice a small increase of the gap when the risk is lowered.

We can also see an increasing of the cost when EWT^* (the objective Expected Waiting Time) decreases.

Like previously, the violation results show that our model respects the initial risk conditions, for both upper and lower solutions. However, the solutions are overstaffing the call center. The higher the variance, the more conservative the solution. For high variances (indicating very bad forecasts), this distributionally robust approach seems to be too conservative and the violation rate tends to 0. However this overstaffing may be understandable since the main consideration is that we do not know what is the right distribution. Thus it would be interesting to consider a smaller set of distributions. This would be possible with a statistical analysis of call center historical data.

These results show that the costs stay consistent with the variations of parameters

and no unusual deviations was observed.

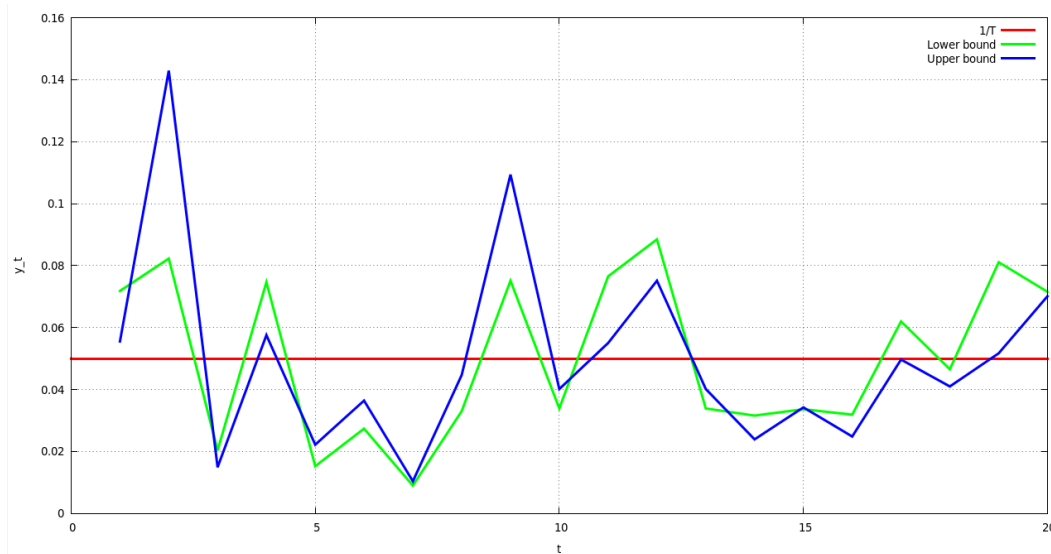


Figure 5-2: Sharing out of the risk through the day.

Figure 5-2 shows the values of y_t variables through the horizon for the upper bound (in blue) and the lower bound (in green). The red line shows the equal sharing out of the risk through the day. This figure brings out the interest of dynamically sharing out the risk: optimization of the variable y_t shows their value are different from the simple equal proportion through the periods. Thus our approach is more complicated but leads to cheaper solutions than a simpler approach with fixed risk levels.

5.5.4 Comparison of costs

Since it might be difficult to ascertain a cost for non-respect of the constraint, our approach focuses on respecting the constraint with a set risk level and does not fix a value on violated constraints. Hence, it is difficult to compare the stochastic approach with a wrong known distribution and the distributionally robust approach, since the respect of the QoS is de facto mandatory in our approach.

However, we compared the costs of the two solutions for various parameters anyway in order to have an idea of the range difference between the two approaches. This result can be useful in case that we have information on having a wrong assumption on the chosen distribution or on penalty costs.

In this numerical experiments, $EWT^* = 1$.

Table 5.4: Comparison between the stochastic solution costs and distributionally robust solution costs.

Risk ϵ	λ range	σ^2 range	LS Cost	US Cost	LDR Cost	UDR Cost	Average CG
0.15	4 – 20	$\frac{\lambda}{10}$	22.12	22.5	27.5	28.31	0.25
0.1	4 – 20	$\frac{\lambda}{10}$	22.31	22.62	29.44	30.5	0.33
0.3	16 – 80	$\frac{\lambda}{2}$	86.62	91.38	114.7	119.7	0.32
0.15	16 – 80	$\frac{\lambda}{10}$	81.5	82.94	97.12	99.69	0.20
0.1	16 – 80	$\frac{\lambda}{4}$	85.81	88.62	128.5	134.3	0.51
0.1	16 – 80	$\frac{\lambda}{3}$	87.38	90.94	139.5	145.7	0.60
0.1	16 – 80	$\frac{\lambda}{2}$	90.12	94.31	153.4	160.2	0.70
0.05	16 – 80	$\frac{\lambda}{2}$	91.88	95.75	190.4	199	1.08

Table 5.4 shows the results of schedule solution costs for the stochastic approach with normal distribution and the distributionally robust approach. Columns 1, 2 and 3 define the parameters of the simulation. Columns 4 and 5 show the lower bound and upper bound costs of the stochastic approach. *LS Cost* stands for the Lower bound Stochastic solution Cost and *US Cost* stands for the Upper bound Stochastic solution Cost. Columns 6 and 7 show the Lower bound Distributionally Robust solution Cost (*LDR Cost*) and the Upper bound Distributionally Robust solution Cost (*UDR Cost*). The last column shows the average gap between the solution cost of the two approaches, namely

$$CG_a = \frac{1}{2} \left(\frac{c_{ldr} - c_{us}}{c_{us}} + \frac{c_{ldr} - c_{ls}}{c_{ls}} \right). \quad (5.27)$$

The superior cost of the distributionally robust approach is easily predictable. Table 5.4 helps us notice that the values of variances have a significant impact on the cost difference. As we previously noticed for the distributionally robust approach, the higher the variance, the more conservative the solution. This table corroborate this observation: indeed, costs difference is higher for bigger variances (see rows 5, 6 and 7).

Moreover, the decrease of the accepted risk level is also increasing the cost difference (compare rows 3, 7 and 8). This is the consequence of the consideration of leptokurtic distributions with heavy tail in the distributionally robust approach.

5.6 Conclusion

We considered the staffing and shift-scheduling problem under uncertainty arising in call center. The distributionally robust approach was introduced, considering that the call arrival rates are following unknown distributions. Moreover, instead of considering the risk level on a period-by-period basis, we decided to set this risk level for the whole horizon of study and thus consider a joint chance-constrained program. Then, a deterministic equivalent of the distributionally robust approach was proposed, based on a tight bound of Chebyshev's inequality, including a dynamic sharing out of the risk. We were thus able to propose solutions with reduced costs compared to less flexible approaches. Finally lower and upper bound of the problem objective function were computed with piecewise linear approximations. Computational results show that both upper and lower solutions respect the objective risk level for a given set of distributions. This shows that our approach proposes robust solutions in term of risk management. The Cost Gap was small enough to be able to bring out a valid solution for the initial problem, which is eventually useful for the managers.

In the simulations, we noticed that mainly the Pareto distribution and Gamma distribution are the ones with violated scenarios. The solutions of the model show that for other distributions, the call center may be overstaffed. Thus, we could further study the call center model in order to evaluate what are the interesting distributions to consider. This can lead, as an improvement for our work in the future, to the study of a reduced set of distributions, according to some conditions (in addition to the known means and variances). A way to reduce the set of distributions could be to integrate the statistical Kolmogorov-Smirnov test into the consideration of the constraints.

The distributionally robust approach showed an advantage compared to a stochastic approach with a wrong assumption on the distribution. Indeed, if the assumption of normal distribution turns out to be incorrect, the staffing solutions are not satisfactory whereas the distributionally robust approach considers this possibility per se.

Chapter 6

Conclusion

6.1 Conclusion

This thesis investigated two steps composing the personnel planning optimization problem arising in call centers, as well as numerous systems with queues and random customers arrivals, namely the staffing and shift-scheduling problems with uncertain demand forecasts. Customer-company interactions are essential for the company strategy, but the staff agent cost represents the main component of the total operating budget. Personnel planning intends to organize the staffing for forthcoming days, and thus making decisions according to uncertain forecasts. Thus, the idea of optimizing this infrastructure arises.

The objective of this thesis was then to develop optimization approaches with uncertainty management for the staffing and shift-scheduling problems. The purpose was to propose easily-tractable cost-reducing optimization programs, providing general formulations which could be transposed to various systems.

Different characteristics of the call center model were considered in the thesis: after focusing on a standard simple queuing model, we proposed other call center models. Erlang-A was introduced instead of Erlang-C in order to consider the customer patience limit, and thus possible abandonments. Moreover the positive correlation between call arrivals during the different periods of a day was introduced in the formulation with a multiplicative random factor. All these models were including uncertainty management by considering arrival rates as random variables.

Optimization under uncertainty was considered through two subfields of mathematical programming: stochastic programming and distributionally robust programming.

Stochastic programming models were developed by considering that random variables are following continuous distributions with known means and variances, especially the normal distribution. Our formulation is not highly specific to the choice of the distribution and others could be considered without jeopardizing the model. Distributionally robust programming was used to propose a program optimizing the staffing cost when the distribution of random variables considered in the problem is assumed to be unknown. We thus formulated a problem where chance constraints were defined considering the right distribution was unknown and belonged to the set of all possible probability distributions with known means and variances.

The optimization problems were formulated with joint chance-constrained programs. This representation set a risk level for the whole time horizon. We proposed equivalent reformulations into individual chance constraints by allocating a proportion of the risk on each period of the staffing model. Instead of considering Bonferroni approximation, these proportions were set as variables of the optimization problems and computed while solving the optimization program. Deterministic reformulations were finally proposed through linearizations and led to tractable mixed-integer programs.

In opposition to the more standard two-step approach, the staffing and shift-scheduling problem was presented in a one-step formulation. This approach was based on the introduction of a relation between the queue parameters into the optimization problem. This one-step approach is based on a continuous relation between the parameters but is also presented with a piecewise constant formulation based on integer values of agents number.

6.2 Perspectives

The work presented in this thesis brings out possible extensions for future research. Some possible points of interest and perspectives are presented here.

- In our approach, we introduced a function ψ which is convex for Erlang-A model and quasiconvex for Erlang-C model. Unfortunately, this statement is not entirely proven and further computations need to be done. This result is visible in abundant observations and intermediate calculations are already made, but the proof is not complete.
- We compared our approach with other various chance-constrained approaches but

we did not focus on other formulations, such as the penalty-cost based approach. It would be interesting to compare performances and costs between these two methods since they focus on different considerations. This comparison could serve as a validation of our approach. This analysis is not risky since we already carried out numerical experiments with satisfactory results and may provide useful information.

- Comparing our model solutions with results given by simulating an entire call center would show the qualities and lacunes of our assumptions. Moreover, testing our models with other scenario data from other areas, such as emergency services or any FCFS queue with uncertain customer (or vehicle) arrivals, would enable to evaluate its adaptability.
- A non-linear program was proposed to model the staffing and shift-scheduling problem with correlated periods. However, the final linear reformulation was not implemented. Thus, this model lacks numerical experiments, as a support of the theoretical formulation. In order to assess the quality of the model, this should be done in a future work.
- The distributionally robust approach as presented in our work may lead to over-staffing. In order to improve this model, the set of distributions can be refined. Some new information about the distribution (additional moments) could be added, or the reduction of the set of probabilities with the consideration of the statistical Kolmogorov-Smirnov test could be considered.
- In this thesis, the call center was considered as single skilled, i.e. agents are able to answer any call. However, agents often have specific skills, for example specific spoken languages. An agent cost depends on the skills he/she has. Considering a skill-based call center in the model is thus an interesting feature. The call center would then be represented with multiple queues, possibly connected. The system would be more complex than the one considered here for a single-skill structure.
- In the problem formulation for this thesis, the objective service level is stated period by period. It would be interesting to consider an objective for the whole horizon of study in order to provide flexibility in the service level management. This implies adding a consideration on the global service level in the constraints of

the program. Moreover, it would highly change the constraint formulation for the targeted service level would not be a fixed value anymore.

Bibliography

- Ahmed, S. and Shapiro, A. (2008). Solving chance-constrained stochastic programs via sampling and integer programming. *Tutorials in Operations Research*, pages 261–269. 41, 42
- Aksin, Z., Armony, M., and Mehrotra, V. (2007). The modern call center: A multi-disciplinary perspective on operations management research. *Production and Operations Management*, 16:665–688. 17, 18, 21, 26, 40
- Aktekin, T. and Soyer, R. (2011). Call center arrival modeling: A bayesian state-space approach. *Naval Research Logistics*, 58(1):28–42. 40
- Aldor-Noiman, S., Feigin, P. D., and Mandelbaum, A. (2009). Workload forecasting for a call center: Methodology and a case study. *The Annals of Applied Statistics*, 3(4):1403–1447. 40
- Atlason, J., Epelman, M. A., and Henderson, S. G. (2004). Call center staffing with simulation and cutting plane methods. *Annals of Operations Research*, 127(1–4):333–358. 39, 40, 41
- Atlason, J., Epelman, M. A., and Henderson, S. G. (2008). Optimizing call center staffing using simulation and analytic center cutting-plane methods. *Management Science*, 54(2):295–309. 39, 40
- Avramidis, A. N., Chan, W., Gendreau, M., L’Ecuyer, P., and Pisacane, O. (2010). Optimizing daily agent scheduling in a multiskill call center. *European Journal of Operational Research*, 200(3). 38
- Avramidis, A. N., Deslauriers, A., and L’Ecuyer, P. (2004). Modeling daily arrivals to a telephone call center. *Management Science*, 50(7):896–907. 39, 40, 80, 89

- Baccelli, F. and Hebuterne, G. (1981). On queues with impatient customers. Technical report, Inria. RR-0094. 21, 38, 82
- Bassamboo, A., Harrison, J. M., and Zeevi, A. (2006). Design and control of a large call center: Asymptotic analysis of an lp-based method. *Operations Research*, 54(3):419–435. 38
- Bassamboo, A. and Zeevi, A. (2009). On a data-driven method for staffing large call centers. *Operations Research*, 57(3):714–726. 39
- Beale, E. (1961). The use of quadratic programming in stochastic linear programming. Technical report, RAND Corporation, Santa Monica, CA. P-2404. 27, 41
- Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. (2009). *Robust Optimization*. Princeton University Press. 30, 43
- Ben-Tal, A. and Nemirovski, A. (1998). Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805. 30
- Ben-Tal, A. and Nemirovski, A. (1999). Robust solutions of uncertain linear programs. *Operations Research Letters*, 25(1):1–13. 43
- Ben-Tal, A. and Nemirovski, A. (2000). Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 88(3):411–424. 30
- Bertsimas, D. and Doan, X. V. (2010). Robust and data-driven approaches to call centers. *European Journal of Operational Research*, 207(2):1072–1085. 39
- Bertsimas, D. and Popescu, I. (2005). Optimal inequalities in probability theory: A convex optimization approach. *SIAM Jour. on Optimization*, 15:780–804. 108
- Bertsimas, D. and Thiele, A. (2006). Robust and data-driven optimization: Modern decision-making under uncertainty. *INFORMS tutorials in operations research: models, methods, and applications for innovative decision making*, page 137. 43
- Bhulai, S., Koole, G., and Pot, A. (2008). Simple methods for shift scheduling in multi-skill call centers. *Manufacturing & Service Operations Management*, 10(3):411–420. 38, 39, 40

- Birge, J. R. and Louveaux, F. (2011). *Introduction to Stochastic Programming*. Springer Series in Operations Research and Financial Engineering. Springer New York. 42
- Borst, S., Mandelbaum, A., and Reiman, M. I. (2004). Dimensioning large call centers. *Operations Research*, 52(1):17–34. 39
- Breton, M. and El Hachem, S. (1995). Algorithms for the solution of stochastic dynamic minimax problems. *Computational Optimization and Applications*, 4(4):317–245. 31
- Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S., and Zhao, L. (2005). Statistical analysis of a telephone call center: A queueing-science perspective. *Journal of the American Statistical Association*, 100:36–50. 17
- Calafiore, G. and Campi, M. (2005). Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46. 42
- Calafiore, G. C. and El Ghaoui, L. (2006). On distributionally robust chance-constrained linear programs. *Jour. of Optimization Theory and Applications*, 130:1–22. 31, 42, 43
- Castillo, I., Joro, T., and Li, Y. Y. (2009). Workforce scheduling with multiple objectives. *European Journal of Operational Research*, 196(1):162–170. 40
- Cezik, M. T. and L'Ecuyer, P. (2008). Staffing multiskill call centers via linear programming and simulation. *Management Science*, 54:310–323. 38, 39, 40, 77
- Charnes, A. and Cooper, W. W. (1959). Chance-constrained programming. *Management Science*, 6(1):73–79. 27
- Chen, B. P. and Henderson, S. G. (2001). Two issues in setting call centre staffing levels. *Annals of Operations Research*, 108(1-4):175–192. 40
- Chen, W., Sim, M., Sun, J., and Teo, C.-P. (2010). From cvar to uncertainty set: Implications in joint chance-constrained optimization. *Operations Research*, 58(2):470–485. 29, 42, 43
- Dantzig, G. B. (1955). Linear programming under uncertainty. *Management Science*, 1(3/4):197–206. 27, 41

- Delage, E. and Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612. 31, 43
- Dentcheva, D., Prékopa, A., and Ruszczyński, A. (2000). Concavity and efficient points of discrete distributions in probabilistic programming. *Mathematical Programming*, 89(1):55–77. 41, 42
- Dentcheva, D., Prékopa, A., and Ruszczyński, A. (2002). Bounds for probabilistic integer programming problems. *Discrete Applied Mathematics*, 124(1-3):55–65. 42
- Dietz, D. C. (2011). Practical scheduling for call center operations. *Omega*, 39(5):550–557. 38, 39, 40
- Ding, S. and Koole, G. (2014). Optimal call center forecasting and staffing under arrival rate uncertainty. Working paper. 39
- El Ghaoui, L. and Lebret, H. (1997). Robust solutions to least-squares problems with uncertain data. *SIAM Journal on Matrix Analysis and Applications*, 18(4):1035–1064. 30
- El Ghaoui, L., Oustry, F., and Lebret, H. (1998). Robust solutions to uncertain semidefinite programs. *SIAM Jour. on Optimization*, 9(1):33–52. 30
- Ertogral, K. and Bamuqabel, B. (2008). Developing staff schedules for a bilingual telecommunication call center with flexible workers. *Computers & Industrial Engineering*, 54(1):118–127. 39, 40
- Gans, N., Koole, G., and Mandelbaum, A. (2003). Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management*, 5:79–141. 17, 18, 19, 39, 40, 51, 64
- Gans, N., Shen, H., and Zhou, Y.-P. (2012). Parametric stochastic programming models for call-center workforce scheduling. Working paper. 37, 39, 40
- Garnet, O., Mandelbaum, A., and Reiman, M. (2002). Designing a call center with impatient customers. *Manufacturing & Service Operations Management*, 4(3):208–227. 38
- Green, L. V. and Kolesar, P. J. (1991). The pointwise stationary approximation for queues with nonstationary arrivals. *Management Science*, 37:84–97. 38

- Green, L. V. and Kolesar, P. J. (1997). The lagged psa for estimating peak congestion in multiserver markovian queues with periodic arrival rates. *Management Science*, 43(1):80–87. 38
- Green, L. V., Kolesar, P. J., and Soares, J. (2001). Improving the sipp approach for staffing service systems that have cyclic demands. *Operations Research*, 49(4):549–564. 37, 38
- Green, L. V., Kolesar, P. J., and Soares, J. (2003). An improved heuristic for staffing telephone call centers with limited operating hours. *Production and Operations Management*, 12(1):46–61. 38
- Green, L. V., Kolesar, P. J., and Whitt, W. (2007). Coping with time-varying demand when setting staffing requirements for a service system. *Production and Operations Management*, 16(1):13–39. 37, 38, 80, 81
- Green, L. V. and Soares, J. (2007). Computing time-dependent waiting time probabilities in $m(t)/m/s(t)$ queueing systems. *Manufacturing & Service Operations Management*, 9(1):54–61. 38
- Gross, D., Shortle, J. F., Thompson, J. M., and Harris, C. M. (2008). *Fundamentals of Queueing Theory*. Wiley Series. 19, 37
- Gurvich, I., Luedtke, J., and Tezcan, T. (2010). Staffing call centers with uncertain demand forecasts: A chance-constrained optimization approach. *Management Science*, 56:1093–1115. 37, 38, 39, 41
- Harrison, J. M. and Zeevi, A. (2005). A method for staffing large call centers based on stochastic fluid models. *Manufacturing & Service Operations Management*, 7(1):20–36. 39, 41
- Henrion, R. and Strugarek, C. (2008). Convexity of chance constraints with independent random variables. *Computational Optimization and Applications*. 42
- Hillier, F. S. and Lieberman, G. J. (1986). *Introduction to Operations Research*, 4th Ed. Holden-Day, Inc. 37
- Ibrahim, R., Ye, H., L'Ecuyer, P., and Shen, H. (2015). Modeling and forecasting call center arrivals: A literature survey. Working paper. 40

- Ingolfsson, A., Haque, M. A., and Umnikov, A. (2002). Accounting for time-varying queueing effects in workforce scheduling. *European Journal of Operational Research*, 139(3):585–597. 40
- Jongbloed, G. and Koole, G. (2001). Managing uncertainty in call centers using poisson mixtures. *Applied Stochastic Models in Business and Industry*, 17:307–318. 40
- Kall, P. and Mayer, J. (2011). *Stochastic Linear Programming*. Springer. 29, 43
- Kall, P. and Wallace, S. W. (1995). *Stochastic Programming*. Wiley. 42
- Koole, G. and van der Sluis, E. (2003). Optimal shift scheduling with a global service level constraint. *IIE Transactions*, 35(11):1049–1055. 37, 40
- Lagoa, C. (1999). On the convexity of probabilistically constrained linear programs. In *Proceedings of the 38th IEEE Conference on Decision and Control*, volume 1, pages 516–521. 42
- Liao, S., Koole, G., van Delft, C., and Jouini, O. (2012). Staffing a call center with uncertain non-stationary arrival rate and flexibility. *OR Spectrum*, 34:691–721. 27, 37, 38, 40, 41
- Liao, S., van Delft, C., and Vial, J.-P. (2013). Distributionally robust workforce scheduling in call centers with uncertain arrival rates. *Optimization Methods and Software*, 28:501–522. 37, 38, 44
- Liberti, L., Cafieri, S., and Tarissan, F. (2009). Reformulations in mathematical programming: A computational approach. In Abraham, A., Hassanien, A.-E., Siarry, P., and Engelbrecht, A., editors, *Foundations of Computational Intelligence Volume 3*, volume 203 of *Studies in Computational Intelligence*, pages 153–234. Springer Berlin Heidelberg. 93, 95
- Luedtke, J. and Ahmed, S. (2008). A sample approximation approach for optimization with probabilistic constraints. *SIAM Jour. on Optimization*, 19(2):674–699. 42
- Luedtke, J., Ahmed, S., and Nemhauser, G. (2010). An integer programming approach for linear programs with probabilistic constraints. *Mathematical Programming*, 122(2):247–272. 30, 40, 41, 42

- Mandelbaum, A. and Zeltyn, S. (2005). Service engineering in action: The palm/erlang-a queue, with applications to call centers. In *Advances in Services Innovations*, pages 17–48. Springer-Verlag. 83
- Mandelbaum, A. and Zeltyn, S. (2007). Service engineering in action: The palm/erlang-a queue, with applications to call centers. In Spath, D. and Fähnrich, K.-P., editors, *Advances in Services Innovations*, pages 17–45. Springer Berlin Heidelberg. 38, 80, 82
- Mandelbaum, A. and Zeltyn, S. (2009). Staffing many-server queues with impatient customers: Constraint satisfaction in call centers. *Operations Research*, 57(5):1189 – 1205. 38, 81
- Nemirovski, A. and Shapiro, A. (2006a). Convex approximations of chance constrained programs. *SIAM Jour. on Optimization*, 17(4):969–996. 41, 42, 43
- Nemirovski, A. and Shapiro, A. (2006b). Scenario approximations of chance constraints. In Calafiore, G. and Dabbene, F., editors, *Probabilistic and Randomized Methods for Design under Uncertainty*, pages 3–47. Springer London. 40, 41
- Ooura, T. (2006). <http://www.kurims.kyoto-u.ac.jp/ooura/index.html>. 63
- Prékopa, A. (1970). On probabilistic constrained programming. In *Proceedings of the Princeton symposium on mathematical programming*, pages 113–138. Princeton, New Jersey: Princeton University Press. 41
- Prékopa, A. (1995). *Stochastic Programming*, volume 324 of *Mathematics and Its Applications*. Springer Netherlands. 42
- Prékopa, A. (2003). Probabilistic programming. In Ruszczyński, A. and Shapiro, A., editors, *Handbooks in operations research and management science*, volume 10, chapter 5, pages 267–351. Elsevier. 41
- Robbins, T. R. (2007). *Managing Service Capacity under Uncertainty*. PhD thesis, The Pennsylvania State University. 43
- Robbins, T. R. and Harrison, T. P. (2010). A stochastic programming model for scheduling call centers with global service level agreements. *European Journal of Operational Research*, 207:1608–1619. 33, 37, 39, 40, 41

- Roubos, A., Bhulai, S., and Koole, G. (2011). Flexible staffing for call centers with non-stationary arrival rates. Working paper. 39, 41
- Ruszczynski, A. (2002). Probabilistic programming with discrete distributions and precedence constrained knapsack polyhedra. *Mathematical Programming*. 42
- Saltzman, R. M. and Mehrotra, V. (2007). Managing trade-offs in call center agent scheduling: Methodology and case study. *Proceedings of the 2007 Summer Computer Simulation Conference*, pages 643–651. 37, 38, 40
- Scarf, H., Arrow, K., and Karlin, S. (1958). A min-max solution of an inventory problem. *Studies in the mathematical theory of inventory and production*, 10:201–209. 31
- Shapiro, A. (2009). Statistical inference. In Shapiro, A., Dentcheva, D., and Ruszczyński, A., editors, *Lectures on Stochastic Programming: Modeling and Theory*. 42
- Shapiro, A. and Ahmed, S. (2004). On a class of minimax stochastic programs. *SIAM Jour. on Optimization*, 14(4):1237–1249. 43
- Soyster, A. L. (1973). Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21(5):1154–1157. 30
- van Ackooij, W., Henrion, R., Möller, A., and Zorgati, R. (2014). Joint chance constrained programming for hydro reservoir management. *Optimization and Engineering*, 15(2):509–531. 42
- Wallace, R. B. and Whitt, W. (2005). A staffing algorithm for call centers with skill-based routing. *Manufacturing & Service Operations Management*, 7(4):276–294. 38
- Whitt, W. (1999). Dynamic staffing in a telephone call center aiming to immediately answer all calls. *Operations Research Letters*, 24:205–212. 37, 39, 89
- Whitt, W. (2006). Staffing a call center with uncertain arrival rate and absenteeism. *Production and Operations Management*, 15(1):88–102. 43
- Yue, J., Chen, B., and Wang, M.-C. (2006). Expected value of distribution information for the newsvendor problem. *Operations Research*, 54(6):1128–1136. 44
- Žáčková, J. (1966). On minimax solutions of stochastic linear programming problems. *Časopis pro pěstování matematiky*, 91(4):423–430. 31

Zeltyn, S., Marmor, Y. N., Mandelbaum, A., Carmeli, B., Greenshpan, O., Mesika, Y., Wasserkrug, S., Vortman, P., Shtub, A., Lauterman, T., Schwartz, D., Moskovitch, K., Tzafrir, S., and Basis, F. (2011). Simulation-based models of emergency departments: Operational, tactical, and strategic staffing. *ACM Trans. Model. Comput. Simul.*, 21(4):24:1–24:25. 38