



**HAL**  
open science

# Estimation of the environment distribution of a random walk in random environment

Antoine Havet

► **To cite this version:**

Antoine Havet. Estimation of the environment distribution of a random walk in random environment. Statistics [math.ST]. Université Paris Saclay (COmUE), 2019. English. NNT: 2019SACLX033 . tel-02273106

**HAL Id: tel-02273106**

**<https://theses.hal.science/tel-02273106>**

Submitted on 28 Aug 2019

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

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

# THÈSE DE DOCTORAT

de

L'UNIVERSITÉ PARIS-SACLAY

École doctorale de mathématiques Hadamard (EDMH, ED 574)

*Établissement d'inscription* : École polytechnique

*Laboratoire d'accueil* : Centre de mathématiques appliquées de polytechnique,  
UMR 7641 CNRS

*Spécialité de doctorat* : Mathématiques appliquées

**Antoine HAVET-MOREL**

Estimation de la loi du milieu  
d'une marche aléatoire en milieu aléatoire

*Date de soutenance* : 19 août 2019

*Lieu de soutenance* : Palaiseau

*Après avis des rapporteurs* : ISMAËL CASTILLO (Sorbonne Université)  
VINCENT RIVOIRARD (Université Paris Dauphine)

*Jury de soutenance* :

EMMANUEL GOBET	(Professeur, École polytechnique, UMR 7641)	Examineur
MATTHIEU LERASLE	(Chargé de recherche, Université Paris-Sud, UMR 8628)	Invité
ÉRIC MOULINES	(Professeur, École polytechnique, UMR 7641)	Directeur de thèse
VINCENT RIVOIRARD	(Professeur, Université Paris Dauphine, UMR 7534)	Rapporteur
JUDITH ROUSSEAU	(Professeur, Université Paris Dauphine, UMR 7534)	Présidente du jury



Estimation de la loi du milieu  
d'une marche aléatoire en milieu aléatoire sur  $\mathbb{Z}$



# Résumé

**Estimation de la loi du milieu d'une marche aléatoire en milieu aléatoire.** Introduit dans les années 1960, le modèle de la marche aléatoire en milieu aléatoire i.i.d. sur les entiers relatifs (ou MAMA) a récemment été l'objet d'un regain d'intérêt dans la communauté statistique. Divers travaux se sont en particulier intéressés à la question de l'estimation de la loi du milieu à partir de l'observation d'une unique trajectoire de la MAMA. Cette thèse s'inscrit dans cette dynamique. Dans un premier temps, nous considérons le problème d'estimation d'un point de vue fréquentiste. Lorsque la MAMA est transiente à droite ou récurrente, nous construisons le premier estimateur non paramétrique de la densité de la loi du milieu et obtenons une majoration du risque associé mesuré en norme infinie. Dans un deuxième temps, nous envisageons le problème d'estimation sous un angle Bayésien. Lorsque la MAMA est transiente à droite, nous démontrons la consistance à posteriori de l'estimateur Bayésien de la loi du milieu. La principale difficulté mathématique de la thèse a été l'élaboration des outils nécessaires à la preuve du résultat de consistance bayésienne. Nous démontrons pour cela une version quantitative de l'inégalité de concentration de type Mac Diarmid pour chaînes de Markov. Nous étudions également le temps de retour en 0 d'un processus de branchement en milieu aléatoire avec immigration. Nous montrons l'existence d'un moment exponentiel fini uniformément valable sur une classe de processus de branchement en milieu aléatoire. Le processus de branchement en milieu aléatoire constituant une chaîne de Markov, ce résultat permet alors d'explicitier la dépendance des constantes de l'inégalité de concentration en fonction des caractéristiques de ce processus.

**Mots Clefs :** Milieu aléatoire, Chaînes de Markov, Statistiques bayésiennes, Estimation non-paramétrique.

# Abstract

**Estimation of the environment distribution of a random walk in random environment.** Introduced in the 1960s, the model of random walk in i.i.d. environment on integers (or RWRE) raised only recently interest in the statistical community. Various works have in particular focused on the estimation of the environment distribution from a single trajectory of the RWRE. This thesis extends the advances made in those works and offers new approaches to the problem. First, we consider the estimation problem from a frequentist point of view. When the RWRE is transient to the right or recurrent, we build the first non-parametric estimator of the density of the environment distribution and obtain an upper-bound of the associated risk in infinite norm. Then, we consider the estimation problem from a Bayesian perspective. When the RWRE is transient to the right, we prove the posterior consistency of the Bayesian estimator of the environment distribution. The main difficulty of the thesis was to develop the tools necessary to the proof of Bayesian consistency. For this purpose, we demonstrate a quantitative version of a Mac Diarmid's type concentration inequality for Markov chains. We also study the return time to 0 of a branching process with immigration in random environment (or BPIRE). We show the existence of a finite exponential moment uniformly valid on a class of BPIRE. The BPIRE being a Markov chain, this result enables then to make explicit the dependence of the constants of the concentration inequality with respect to the characteristics of the BPIRE.

**Keys words :** Random environment, Markov chains, Bayesian statistics, Non-parametric estimation.



# Remerciements

La préparation de cette thèse de doctorat n'a pas été un long fleuve tranquille : ce fut une période pleine de doutes, de questionnements, de découvertes, de révélations, de joies et bien évidemment ... de travail. Somme toute, ces trois dernières années ont profondément et heureusement contribué à ma formation intellectuelle et humaine et je tiens donc à remercier dans ces quelques lignes toutes les personnes qui m'ont soutenu, supporté et accompagné sur le Chemin.

Pour commencer, j'adresse de sincères remerciements à mes trois encadrants de thèse : Éric Moulines, Matthieu Lerasle et Élodie Vernet. Je leur suis en particulier reconnaissant de m'avoir permis de travailler sur un sujet riche en questions et développements et de m'avoir partagé leurs connaissances et expériences : découvrir et expérimenter la recherche à leurs côtés fut une chance. Je remercie chacun pour tout le temps et toute l'attention qu'il a pu m'accorder en fonction de ses disponibilités.

C'est un grand honneur pour moi d'avoir Vincent Rivoirard et Ismaël Castillo comme rapporteurs ainsi que Judith Rousseau, Emmanuel Gobet et Vincent Rivoirard comme membres du jury de thèse. Deux d'entre eux furent mes enseignants en Master Recherche et m'ont notamment fait découvrir les statistiques non paramétriques tandis que les deux autres ont par leurs travaux occupé mon quotidien ces deux dernières années. Je remercie les rapporteurs d'avoir accepté d'évaluer ce manuscrit en un temps inhabituellement court et le jury d'avoir accepté de se rassembler au beau milieu de l'été pour la soutenance de mes travaux de thèse.

Je me dois de remercier les institutions qui m'ont donné les moyens et les conditions pour réaliser ce travail. En premier lieu, l'État français qui grâce à une bourse publique du LabEx Mathématique Hadamard porté par le *Programme d'Investissements d'Avenir* a donné les moyens financiers. Ensuite, l'École polytechnique qui m'a offert la chance de faire l'expérience de la recherche et de l'enseignement dans un environnement à maints égards exceptionnel. Finalement, le CMAP (*Centre de Mathématiques Appliquées de Polytechnique*) dont le personnel administratif et scientifique m'a accueilli et épaulé dans de nombreuses tâches administratives, informatiques et scientifiques.

Une grande reconnaissance également envers tous les mathématiciens qui ont marqué mon parcours d'études secondaires et supérieures : mes enseignants au lycée, Ludovic Vermoyal et Chantal Demetz; mes enseignants en classes préparatoires, Christian de Moliner et Jean Zurek; les enseignants du Magistère de mathématiques d'Orsay qui m'ont aussi bien donné le goût des Probabilités et des Statistiques qu'aiguillé dans mes choix d'orientation, Frédéric Paulin, Christophe Giraud, Arvind Singh, Edouard Maurel-Segala, Dominique Hulin...



Pour conclure, je souhaite remercier les personnes qui ont partagé mon quotidien depuis de nombreuses années (les trois dernières mais pas seulement...).

Tout d'abord, mes camarades de galère, les doctorants du CMAP (anciens et actuels) qui non contents de souffrir mon "*comique*" de répétition, m'ont apporté soutien et réconfort, chacun à leur manière.

Les membres et amis de la "CCX", le père Nicolas R. s.j. et les amis du groupe Magis qui sont des guides, des modèles et des compagnons dans l'accomplissement de mon devoir d'état pour la plus grande gloire du Tout Autre. Sans oublier, les amis d'*A Bras Ouverts* qui m'ont accompagné dans ma recherche du sens de la Vie et sont ainsi indissociables de mon travail.

Mes amis et anciens colocataires, Thomas, Étienne, Fedor et Marie-Liesse avec lesquels j'ai partagé une vie des plus bucoliques au pied des marches de Lozère.

Également, mes fidèles amis, Camille, Clotilde, Lucie, Antoine... qui m'ont souvent écouté, rassuré, encouragé et distrait dès que je les sollicitais.

Finalement, un tendre et non moins immense merci à toute ma famille ! Ma sœur, Anne-Fleur, dont l'amour fraternel et la confiance me sont si chers. Mes parents sans lesquels ce travail n'aurait jamais été réalisé : je leur dois énormément et les qualités (et défauts...) qu'ils m'ont transmis ont été extrêmement précieux. Mon épouse Hélène pour sa patience, ses conseils, son optimisme, ses encouragements, son amour : elle fut une alliée de chaque instant pour la conception, la gestation et la mise au monde de cette thèse.

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>11</b>
1.1	Le modèle de la MAMA en environnements i.i.d. sur $\mathbb{Z}$	11
1.1.1	La marche aléatoire simple sur $\mathbb{Z}$	11
1.1.2	Un modèle intermédiaire	12
1.1.3	Le modèle de la MAMA en milieu i.i.d.	13
1.2	Quelques repères sur les MAMA	13
1.2.1	Une définition plus formelle de la MAMA	13
1.2.2	Quelques résultats probabilistes	14
1.2.3	Le processus des sauts à gauche	16
1.2.4	Processus de branchement en milieu aléatoire	17
1.2.5	Le lien entre les marches aléatoires en milieu aléatoire et les processus de branchement en milieu aléatoire	18
1.2.6	Un estimateur des moments par [AE04]	18
1.2.7	Un estimateur de type maximum de vraisemblance par [CFL16, CFL <sup>+</sup> 14, FLM14, FGL14]	19
1.2.8	Un premier estimateur non paramétrique par [DL18]	21
1.3	Contributions	22
1.3.1	Estimation de la densité de la loi du milieu : Chapitre 2	23
1.3.2	Brève introduction aux statistiques bayésiennes	25
1.3.3	Consistance a posteriori de l'estimateur Bayésien : Chapitre 5	29
1.3.4	Inégalité de concentration pour transformation de chaînes de Markov avec la propriété de différences bornées : Chapitre 4	30
1.3.5	Contrôle uniforme de la queue de distribution du temps de retour en 0 du processus de branchement en milieu aléatoire : Chapitre 3	31
1.4	Conclusion et perspectives	32
<b>2</b>	<b>Nonparametric density estimation of the RWRE</b>	<b>33</b>
2.1	Introduction	33
2.2	Random walks in random environment (RWRE)	35
2.3	Estimator construction	38
2.4	Main results	39
2.5	Simulation Study	41
2.5.1	Influence of the regularity	42
2.5.2	Influence of the regime	42
2.5.3	Goldenshluger-Lepski estimator	45
2.6	Proof	46
2.6.1	Proof of Proposition 1	46

2.6.2	Bounding $\widehat{f}_n^M$ and $f^M$ in sup-norm . . . . .	46
2.6.3	Proof of Theorem 2 . . . . .	48
2.6.4	Proof of Theorem 5 . . . . .	54
2.6.5	Proof of Proposition 6 . . . . .	55
<b>3</b>	<b>First return time of a branching process in random environment</b>	<b>57</b>
3.1	Setting and main result . . . . .	59
3.1.1	Assumptions . . . . .	59
3.1.2	The case of RWRE . . . . .	60
3.1.3	Main result . . . . .	61
3.2	Proofs . . . . .	61
3.2.1	Theorem 16 . . . . .	62
3.2.2	Sketch of proof of Theorem 16 . . . . .	62
3.2.3	Detailed proof of Theorem 16 . . . . .	63
3.2.4	A different interpretation of the BPIRE . . . . .	69
3.2.5	Theorem 22 . . . . .	70
3.2.6	Sketch of proof of Theorem 22 . . . . .	70
3.2.7	Detailed proof of Theorem 22 . . . . .	73
3.2.8	Proof of Theorem 14 through Theorems 16 and 22 . . . . .	86
3.2.9	Proof of the existence of exponential moments for $\text{BPIREG}(\nu, 0)$ . . . . .	87
<b>4</b>	<b>Concentration inequality for geometrically ergodic Markov chains</b>	<b>89</b>
4.1	Framework . . . . .	91
4.1.1	Markovian setting . . . . .	91
4.1.2	Assumptions in Markovian framework . . . . .	92
4.2	Main results : Theorems 39 and 40 . . . . .	92
4.3	Proof of Theorem 39 . . . . .	94
4.3.1	Sketch of proof of Theorem 39 . . . . .	94
4.3.2	Intermediate results for proof of Theorem 39 . . . . .	95
4.4	Proof of Theorem 40 . . . . .	109
4.4.1	$V$ -geometric ergodicity . . . . .	109
4.4.2	Satisfaction of Assumptions <b>M1</b> , <b>M2</b> and <b>M3</b> . . . . .	111
<b>5</b>	<b>Posterior consistency of Bayes estimator of the environment</b>	<b>121</b>
5.1	RWRE, BPIRE and the Bayesian setting . . . . .	122
5.1.1	RWRE . . . . .	123
5.1.2	BPIRE . . . . .	124
5.1.3	Towards posterior consistency . . . . .	125
5.2	Bayesian setting and main results . . . . .	126
5.2.1	Assumptions on the prior . . . . .	126
5.2.2	Bayesian framework for RWRE . . . . .	127
5.2.3	Bayesian framework for BPIRE . . . . .	128
5.2.4	Main results : posterior consistency for RWRE and $\text{BPIREG}(\nu, 0)$ . . . . .	128
5.3	Proof of the main results . . . . .	129
5.3.1	Analysis : a look into Bayesian techniques . . . . .	129
5.3.2	Sketch of proof of Theorem 64 . . . . .	131
5.3.3	Minoration of the denominator $D_n$ : proof of Proposition 70 . . . . .	134
5.3.4	Some properties of $d^\nu$ and $d_n$ . . . . .	136

5.3.5	Link between $d^v$ and $d_n$ : proof of Proposition 66 . . . . .	139
5.3.6	Proof of Proposition 67 . . . . .	139
5.3.7	Proof of Proposition 68 . . . . .	140
5.3.8	Proof of Proposition 69 . . . . .	141
5.3.9	Control of the first and second kind risks . . . . .	143
5.3.10	Final step of the proof of Theorem 64 . . . . .	146
5.3.11	Proof of Theorem 63 . . . . .	146

## Appendices

<b>A</b>	<b>Reminder on transition kernels and Markov chains</b>	<b>149</b>
<b>B</b>	<b>Reminder on covering numbers</b>	<b>153</b>
<b>C</b>	<b>Reminder on stopping times</b>	<b>155</b>
	<b>Bibliography</b>	<b>161</b>



# Notations

Dans cette thèse, nous adopterons les notations suivantes

- Pour tout entier  $k$  dans  $\mathbb{Z}$ , on note  $\mathcal{S}_k$  l'ensemble des suites de  $\mathbb{Z}^{\mathbb{N}}$  issues de  $k$  dont les termes successifs diffèrent de 1 exactement

$$(1) \quad \mathcal{S}_k = \left\{ x = (x_t)_{t \in \mathbb{N}} \in \mathbb{Z}^{\mathbb{N}} : x_0 = k; \forall t \in \mathbb{N}, |x_{t+1} - x_t| = 1 \right\}.$$

- Si  $f$  est une fonction à valeurs réelles définie sur un intervalle (ouvert)  $I \subset \mathbb{R}$  dérivable  $k$ -fois avec  $k$  dans  $\mathbb{N}^*$ , on note  $f^{(i)}$  la dérivée d'ordre  $i$  de  $f$ , pour tout entier  $1 \leq i \leq k$ .
- Si  $\mathcal{X}$  est un ensemble et  $(x_0, \dots, x_{n-1}) \in \mathcal{X}^n$ , on note pour tous  $0 \leq i \leq j \leq n-1$

$$x_{i:j} = (x_i, \dots, x_j).$$

- L'ensemble de toutes les distributions de probabilité sur une tribu  $\mathcal{X}$  est noté  $\mathbb{M}^1(\mathcal{X})$ .
- Pour tout  $p$  dans  $(0, 1)$ ,  $\mathcal{G}_{\mathbb{N}}(p)$  désigne la distribution de probabilité sur  $\mathbb{N}$  dite *loi géométrique de paramètre  $p$*  définie de la manière suivante : une variable aléatoire  $X$  est distribuée suivant  $\mathcal{G}_{\mathbb{N}}(p)$  sous  $\mathbb{P}$ , si et seulement si, pour tout  $k$  dans  $\mathbb{N}$

$$(2) \quad \mathbb{P}(X = k) = p(1 - p)^k.$$

- Soit  $\Omega$  un ensemble muni d'une topologie  $\mathcal{T}$  et de la tribu  $\sigma(\mathcal{T})$  engendrée par  $\mathcal{T}$  (i.e. la plus petite tribu contenant  $\mathcal{T}$ ). Soit une mesure  $\Pi$  sur  $\sigma(\mathcal{T})$ .

Le *support* de  $\Pi$ , noté  $\text{supp}(\Pi)$ , est défini comme le complémentaire de l'union des éléments de  $\mathcal{T}$  de mesure nulle sous  $\Pi$  (ou de manière équivalente comme l'ensemble des éléments de  $\Omega$  dont chaque voisinage pour la topologie  $\mathcal{T}$  est de mesure strictement positive sous  $\Pi$ ).

- Soit  $(\Omega, \mathcal{A})$  un espace mesurable muni d'une mesure de probabilité  $\mathbb{P}$  et  $(E, d)$  un espace métrique. Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires définies sur  $(\Omega, \mathcal{A})$  à valeur dans  $(E, d)$  et  $X$  une variable aléatoire également définie sur  $(\Omega, \mathcal{A})$  et à valeur dans  $(E, d)$ . On dit que la suite  $(X_n)_{n \in \mathbb{N}}$  converge vers  $X$  en probabilité sous  $\mathbb{P}$ , si et seulement si, pour tout  $\varepsilon > 0$ ,  $\mathbb{P}(d(X_n, X) \geq \varepsilon) \xrightarrow[n \rightarrow +\infty]{} 0$  et on note

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X.$$



# Chapter 1

## Introduction

Introduite par le biologiste Chernov [Che67] dans l'étude de la réplication de l'ADN puis reprise peu après par le physicien Temkin [Tem72] pour l'étude de la transition de phase dans les alliages, la marche aléatoire en milieu aléatoire (abrégé en *MAMA*) constitue un modèle simple pour divers phénomènes de transport comme par exemple la diffusion de chaleur ou le déplacement de matière au sein d'un milieu physique.

Plus récemment, lors de la modélisation du phénomène de combinaison-recombinaison observé lors du dégrafage d'une molécule d'ADN, des travaux tels que [AMJR12, HFR09, BBC<sup>+</sup>07, BBC<sup>+</sup>06, KSJW02, LN02] ont eu recours à la MAMA sur  $\mathbb{Z}$  pour tenter de déterminer la nature de chaque liaison nucléotidique (A-T ou G-C) à partir de la seule observation du déroulé de l'expérience.

Dans de tels phénomènes, un "déplacement" se produit au sein d'un "milieu" dont les propriétés sont inconnues ou très irrégulières : deux niveaux d'aléa sont ainsi superposés. D'une part, le milieu est imprévisible localement mais possède une certaine régularité statistique : cela conduit à modéliser le milieu comme aléatoire. D'autre part, le déplacement en un milieu donné est variable : cela amène à modéliser le déplacement comme également aléatoire.

Depuis son introduction dans [Che67], le modèle de la MAMA a été l'objet de nombreuses études dans la communauté probabiliste, que ce soit sur  $\mathbb{Z}$ ,  $\mathbb{Z}^d$  ou plus généralement sur des graphes (le cours [Zei12] expose l'état de l'art dans ce domaine). De nombreux outils probabilistes ont été développés pour étudier la MAMA sur  $\mathbb{Z}$  et de féconds parallèles ont pu être établis avec d'autres domaines des probabilités. Notre travail se concentre ainsi principalement sur les MAMA sur  $\mathbb{Z}$  et s'inspire du lien avec les *processus de branchement en milieu aléatoire* mis en évidence par [KKS75] (voir Chapitre 3).

Nous allons maintenant définir de manière plus précise la marche aléatoire en milieux aléatoires indépendants et identiquement distribués (en abrégé *i.i.d.*) sur  $\mathbb{Z}$ .

### 1.1 Le modèle de la MAMA en environnements i.i.d. sur $\mathbb{Z}$

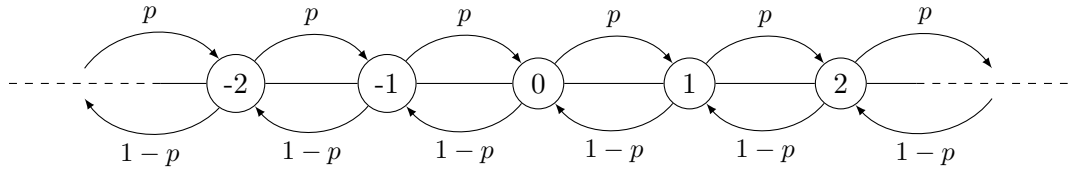
Décrivons tout d'abord un modèle plus simple qui n'est pas en environnement aléatoire.

#### 1.1.1 La marche aléatoire simple sur $\mathbb{Z}$

Le modèle de la marche aléatoire simple sur  $\mathbb{Z}$  de paramètre  $p \in [0, 1]$  permet de modéliser un déplacement aléatoire sur  $\mathbb{Z}$ . Notant  $X_t$  la position à l'instant  $t$  (élément de  $\mathbb{N}$ ) d'une

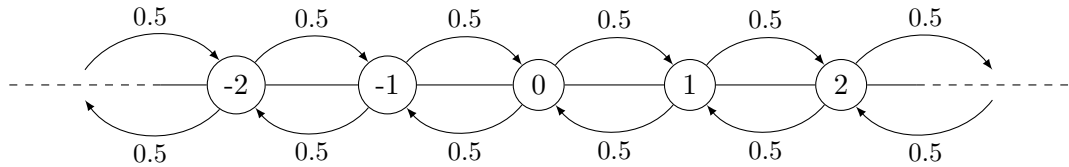


particule sur  $\mathbb{Z}$  qui se déplace aléatoirement selon la règle suivante : à chaque instant, la particule se déplace soit de 1 pas en avant avec probabilité  $p$  (et donc  $X_{t+1} = X_t + 1$ ) soit de 1 pas en arrière avec probabilité  $1 - p$  (et donc  $X_{t+1} = X_t - 1$ ).



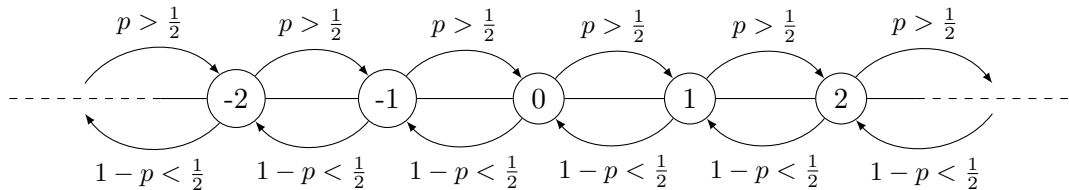
Le comportement de  $(X_t)_{t \in \mathbb{N}}$  au cours du temps est alors fonction de la valeur de  $p$  par rapport à 0.5 :

- Si  $p = 0.5$ , alors,  $1 - p = 0.5$  et la particule a autant de chance de se déplacer en avant qu'en arrière à chaque instant.



N'ayant pas de sens de déplacement privilégié, la particule fait des allers et retours incessants. On montre alors que *la marche est récurrente*, i.e. la particule passe en tout endroit une infinité de fois.

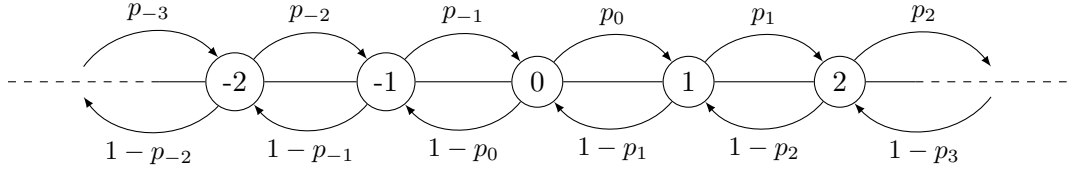
- Si  $p > 0.5$ , alors,  $1 - p < 0.5$  et la particule se déplace préférentiellement en avant : *la marche est transiente à droite*, i.e. la position de la particule tend vers  $+\infty$ .



- Si  $p < 0.5$ , alors,  $1 - p > 0.5$  et la particule se déplace préférentiellement en arrière à chaque instant : *la marche est transiente à gauche*, i.e. la position de la particule tend vers  $-\infty$ .

### 1.1.2 Un modèle intermédiaire

Implicitement, le modèle de la marche aléatoire simple considère que le milieu dans lequel a lieu le déplacement est homogène : la chance d'avancer (ou de reculer) est la même en chaque position  $k$ . Introduisons de l'hétérogénéité en supposant que la probabilité d'avancer de 1 pas en avant (ou de reculer de 1 pas en arrière) est fonction de la position  $k$ . Ainsi, notant  $X_t$  la position à l'instant discret  $t$  d'une particule sur  $\mathbb{Z}$ , on considère qu'elle se déplace aléatoirement selon la règle suivante : si la particule se trouve en  $k$  à l'instant  $t$  (i.e.  $X_t = k$ ), alors en l'instant  $t + 1$ , la particule s'est déplacée soit de 1 pas en avant avec probabilité  $p_k$  (et donc  $X_{t+1} = X_t + 1$ ), soit de 1 pas en arrière avec probabilité  $1 - p_k$  (et donc  $X_{t+1} = X_t - 1$ ).

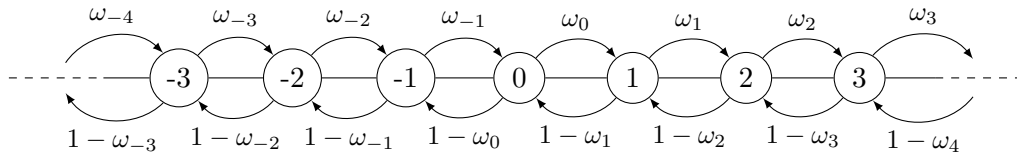


### 1.1.3 Le modèle de la MAMA en milieux i.i.d.

Le modèle de la MAMA impose que les  $p_k$  du modèle proposé en Section 1.1.2 présentent une régularité statistique : ce sont les tirages indépendants de même loi  $\nu$  sur  $[0, 1]$  fixée. Nous allons ainsi décrire le modèle de la MAMA comme la superposition de deux niveaux successifs d'aléa.

Dans un premier temps, nous tirons  $\omega = (\omega_k)_{k \in \mathbb{Z}}$  des nombres de l'intervalle  $[0, 1]$  indépendamment selon la même loi  $\nu$  (i.e. *i.i.d.* de loi  $\nu$ ).

Dans un second temps, nous considérons le déplacement aléatoire d'une particule sur  $\mathbb{Z}$  suivant un modèle de marche aléatoire sur  $\mathbb{Z}$  de probabilités de transition  $(\omega_k)_{k \in \mathbb{Z}}$ . Le graphique ci-dessous reprend les règles de déplacement de la particule une fois la réalisation des milieux  $(\omega_k)_{k \in \mathbb{Z}}$  fixée :



Une première notion propre à la théorie des processus aléatoires en milieu aléatoire, prenant en compte cette superposition de deux couches d'aléa, apparaît ainsi : la notion de *loi quenched*  $\mathbf{P}_\omega$  qui est définie comme la loi des trajectoires de la particule une fois l'environnement  $\omega$  connu.

Bien que cette vision hiérarchisée soit indispensable à la description, à la simulation ou à l'étude théorique des MAMA, la réalité statistique est bien différente : les réalisations  $(\omega_k)_{k \in \mathbb{Z}}$  ne sont pas connues et seule la trajectoire de la particule est observée.

Une seconde notion propre à la théorie des processus aléatoires en milieu aléatoire, prenant en compte ce constat, apparaît alors naturellement : la notion de *loi annealed*  $\mathbf{P}^\nu$  définie comme loi de probabilité des trajectoires de la particule (sans information sur les réalisations  $(\omega_k)_{k \in \mathbb{Z}}$  de l'environnement).

Avant d'exposer notre travail, nous définissons un peu plus formellement la MAMA et donnons quelques repères fondamentaux pour les développements ultérieurs.

## 1.2 Quelques repères sur les MAMA

### 1.2.1 Une définition plus formelle de la MAMA

Notons  $E = (0, 1)^\mathbb{Z}$  l'ensemble des environnements muni de la tribu  $\mathcal{E} = \mathcal{B}([0, 1])^\mathbb{Z}$  engendrée par les cylindres et considérons le processus canonique  $(X_t)_{t \in \mathbb{N}}$  de l'espace  $\mathbb{Z}^\mathbb{N}$  muni de la tribu

$\mathcal{S}$  engendrée par les cylindres , i.e. pour tout  $t$  dans  $\mathbb{N}$  et tout  $x = (x_t)_{t \in \mathbb{N}}$  dans  $\mathbb{Z}^{\mathbb{N}}$

$$(1.1) \quad X_t(x) = x_t .$$

Pour tout environnement fixé  $\omega$  dans  $E$ , on définit la mesure de probabilité  $\mathbf{P}_\omega$  dite *quenched* dans le milieu  $\omega$  comme l'unique mesure de probabilité sur  $\mathcal{S}$  sous laquelle le processus canonique  $X = (X_t)_{t \in \mathbb{N}}$  constitue une chaîne de Markov homogène partant de 0 et de noyau de transition défini pour  $x$  et  $y$  dans  $\mathbb{Z}$  par

$$p_\omega(x, y) = \begin{cases} \omega_x & \text{si } y = x + 1 \\ 1 - \omega_x & \text{si } y = x - 1 \\ 0 & \text{sinon} \end{cases} .$$

En particulier, pour tout environnement  $\omega$  dans  $E$  fixé,  $(X_t)_{t \in \mathbb{Z}}$  appartient  $\mathbf{P}_\omega$  presque sûrement à l'ensemble  $\mathcal{S}_0$  (défini par l'Equation (1) des Notations).

Considérant alors une mesure de probabilité fixée  $\nu$  sur  $([0, 1], \mathcal{B}([0, 1]))$ , on munit l'espace des environnements  $(E, \mathcal{E})$  de la mesure de probabilité  $\mathbb{P}^\nu = \nu^{\otimes \mathbb{Z}}$ . Intégrant la loi quenched  $\mathbf{P}_\omega$  (fonction de l'environnement) sous la loi  $\mathbb{P}^\nu$ , on définit la loi de probabilité  $\tilde{\mathbf{P}}^\nu$  sur  $(E \times \mathbb{Z}^{\mathbb{N}}, \mathcal{E} \otimes \mathcal{S})$ , i.e. pour  $A$  dans  $\mathcal{E}$  et  $B$  dans  $\mathcal{S}$

$$\tilde{\mathbf{P}}^\nu(A \times B) = \int_E \mathbb{1}_A(\omega) \mathbf{P}_\omega(B) \mathbb{P}^\nu(d\omega) = \int_A \mathbf{P}_\omega(B) \mathbb{P}^\nu(d\omega) .$$

La seconde marginale de  $\tilde{\mathbf{P}}^\nu$  notée  $\mathbf{P}^\nu$  est appelée *loi annealed*.

En particulier, pour n'importe quelle distribution de probabilité  $\nu$  sur  $[0, 1]$  fixée, on déduit également que  $(X_t)_{t \in \mathbb{Z}}$  appartient  $\mathbf{P}^\nu$  presque sûrement à  $\mathcal{S}_0$ .

## 1.2.2 Quelques résultats probabilistes

Après leur utilisation pour la modélisation de phénomènes concrets dans [Che67, Tem72], les MAMA sont étudiées pour la première fois sous un angle probabiliste par [Sol75] qui établit que le comportement des trajectoires sous la loi annealed  $\mathbf{P}^\nu$  dépend de la distribution de la variable aléatoire

$$\rho_0 = \frac{1 - \omega_0}{\omega_0} \quad \text{où} \quad \omega_0 \sim \nu .$$

Si la quantité  $\mathbb{E}^\nu [|\log \rho_0|]$  est finie, [Sol75, Theorem 1.7] dresse une première classification du comportement en temps long de la MAMA:

- (i) Si  $\mathbb{E}^\nu [\log \rho_0] \neq 0$ , alors
  - (a) soit  $\mathbb{E}^\nu [\log \rho_0] < 0$  et  $X_n$  converge  $\mathbf{P}^\nu$  presque sûrement vers  $+\infty$  quand  $n$  tend vers  $+\infty$ , on dit alors que  $(X_n)_{n \in \mathbb{N}}$  est *transiente à droite*,
  - (b) soit  $\mathbb{E}^\nu [\log \rho_0] > 0$  et  $X_n$  converge  $\mathbf{P}^\nu$  presque sûrement vers  $-\infty$  quand  $n$  tend vers  $+\infty$ , on dit alors que  $(X_n)_{n \in \mathbb{N}}$  est *transiente à gauche*.
- (ii) Si  $\mathbb{E}^\nu [\log \rho_0] = 0$ , alors  $\mathbf{P}^\nu$  presque sûrement

$$\limsup_{n \rightarrow +\infty} (X_n) = +\infty \quad \text{et} \quad \liminf_{n \rightarrow +\infty} (X_n) = -\infty ,$$

et on dit alors que  $(X_n)_{n \in \mathbb{N}}$  est *récurrente*.

Pour tout  $n$  dans  $\mathbb{N}$ , nous définissons la fonction  $T_n : \mathbb{Z}^{\mathbb{N}} \rightarrow \mathbb{N} \cup \{+\infty\}$  en posant pour tout  $x = (x_t)_{t \in \mathbb{N}}$  dans  $\mathbb{Z}^{\mathbb{N}}$

$$(1.2) \quad T_n(x) = \inf \{t \in \mathbb{N} : x_t = n\},$$

et nous convenons de noter  $T_n = T_n(X)$  où  $X$  est défini par l'Equation (1.1).

Cette caractérisation est alors précisée par [Sol75, Theorem 1.16] qui établit un premier résultat quantitatif sur la vitesse asymptotique en régime transient à droite :

(i) soit  $\mathbb{E}^\nu[\rho_0] < 1$  et  $\mathbf{P}^\nu$  presque sûrement

$$\frac{T_n}{n} \xrightarrow[n \rightarrow +\infty]{} \frac{1 + \mathbb{E}^\nu[\rho_0]}{1 - \mathbb{E}^\nu[\rho_0]},$$

et la MAMA est alors qualifiée de *balistique*,

(ii) soit  $\mathbb{E}^\nu[\rho_0] \geq 1$  et  $\mathbf{P}^\nu$  presque sûrement

$$\frac{T_n}{n} \xrightarrow[n \rightarrow +\infty]{} +\infty,$$

et la MAMA est alors qualifiée de *sous-balistique*.

Constatant que le modèle de la marche aléatoire simple de paramètre  $p$  introduit en Section 1.1.1 est une MAMA où  $\nu$  est une masse de Dirac en  $p$ , il est alors intéressant de mettre les caractérisations de [Sol75] en regard de celles établies auparavant.

D'une part, comparer la position de la quantité  $\mathbb{E}^\nu[\log \rho_0] = \log((1-p)/p)$  par rapport à 0 revient exactement à comparer la position par rapport à 0.5 de la quantité  $p$  et on retrouve alors les trois types de comportement déjà mis en évidence.

D'autre part, en régime transient à droite, la quantité  $\mathbb{E}^\nu[\log \rho_0] = \log((1-p)/p)$  est toujours strictement inférieure à 1 puisque  $p < 0.5$ . La MAMA s'avère donc être un modèle au comportement généralement plus complexe.

Avec l'hypothèse que le sous-groupe de  $(\mathbb{R}, +)$  engendré par le support de  $\log \rho_0$  sous  $\mathbb{P}^\nu$  est dense dans  $\mathbb{R}$  et qu'il existe  $\kappa > 0$  tel que  $\mathbb{E}^\nu[\rho_0^\kappa] = 1$  et  $\mathbb{E}^\nu[\rho_0^\kappa \log^+(\rho_0)] < +\infty$ , [KKS75] caractérise plus finement les fluctuations de  $T_n$  et le comportement asymptotique des trajectoires de la MAMA sous  $\mathbf{P}^\nu$

- (i) si  $\kappa < 1$ ,  $T_n/n^{1/\kappa}$  et  $X_t/t^\kappa$  convergent en loi vers une distribution non triviale,
- (ii) si  $\kappa = 1$ ,  $T_n/(n \log n)$  et  $(\log t/t)X_t$  convergent en probabilité vers une constante  $c_\kappa \neq 0$ ,
- (iii) si  $\kappa > 1$ ,  $T_n/n$  et  $X_t/t$  convergent en probabilité vers une constante  $\tilde{c}_\kappa \neq 0$ .

Avec l'hypothèse qu'il existe  $0 < \alpha < 0.5$  tel que  $\nu([\alpha, 1-\alpha]) = 1$  (on dit que le milieu est *elliptique*), et l'hypothèse que  $0 < \text{Var}(\rho_0) = \mathbb{E}^\nu[\log^2(\rho_0)]$ , [Sin82] donne alors le comportement de la marche en régime récurrent en établissant que la variable aléatoire

$$\frac{\text{Var}(\rho_0)}{\log^2(t)} X_t,$$

converge en loi sous  $\mathbf{P}^\nu$  vers une variable  $b_\infty$  de loi symétrique. Ce résultat est ensuite précisé par [Gol83] et [Kes86] qui déterminent que la transformée de Laplace de  $b_\infty$  est donnée pour tout  $\lambda > 0$  par

$$\mathbb{E}^\nu \left[ e^{-\lambda |b_\infty|} \right] = \frac{\cosh(\sqrt{2\lambda}) - 1}{\lambda \cosh(\sqrt{2\lambda})}.$$

La technique de preuve initiée par [Sin82] repose principalement sur l'introduction d'une marche aléatoire simple, appelée *potentiel*, dont les retombées vont bien au delà de ce résultat : elle ouvre le champ à la plupart des études probabilistes ultérieures sur la MAMA en milieu i.i.d. sur  $\mathbb{Z}$  ([GPS10, PZ09, GS02, Ali99a]). Avec cette interprétation d'un milieu comme un potentiel aléatoire, [ESZ09] précise notamment les quantités limites du résultat principal de [KKS75].

Pour la suite de notre propos, nous nous attardons cependant sur un lien entre MAMA et processus de branchement en milieu aléatoire. Mis en évidence pour la première fois dans la preuve de [KKS75], il s'est déjà avéré fécond dans des travaux tels que [MWRZ04, Ali99b].

### 1.2.3 Le processus des sauts à gauche

Les résultats de [Sol75] sur le comportement asymptotique en régimes récurrent et transient à droite de la MAMA garantissent que pour tout entier  $n$ , le temps  $T_n$  de premier passage en  $n$  est une quantité finie  $\mathbf{P}^\nu$  presque sûrement.

D'un point de vue pratique,  $T_n$  correspond à la réalité expérimentale. Par exemple,  $T_n$  peut désigner le temps nécessaire au dégrafage complet d'une molécule d'ADN de longueur  $n$ . D'un point de vue théorique,  $T_n$  possède de bonnes propriétés : il définit un *temps d'arrêt* (voir l'Appendice C pour un rappel sur la définition et des propriétés de base).

Ces deux approches motivent alors l'étude du modèle statistique où l'observation est  $X_{0:T_n}$ .

Pour tout  $k$  dans  $\mathbb{Z}$ , pour tout entier  $n$  dans  $\mathbb{N}$  et pour toute suite  $x = (x_t)_{t \in \mathbb{N}}$  dans  $\mathbb{Z}^{\mathbb{N}}$  telle que  $T_n(x) < +\infty$ , nous définissons les nombres de sauts à gauche et à droite de  $k$  de la suite  $x$  avant d'atteindre  $n$  par

$$L(k, x, n) = \sum_{t=0}^{T_n(x)-1} \mathbb{1}_{\{x_t=k, x_{t+1}=k-1\}} \quad \text{et} \quad R(k, x, n) = \sum_{t=0}^{T_n(x)-1} \mathbb{1}_{\{x_t=k, x_{t+1}=k+1\}}.$$

Pour tout  $n$  dans  $\mathbb{N}$  et tout  $k$  dans  $\mathbb{Z}$ , les propriétés combinatoires des éléments de  $\mathcal{S}_0$  et la définition même de  $T_n$  nous assurent alors que pour tout  $x$  dans  $\{x \in \mathcal{S}_0 : T_n(x) < +\infty\}$

$$(1.3) \quad R(k, x, n) = \begin{cases} L(k+1, x, n) & \text{si } k < 0 \\ L(k+1, x, n) + 1 & \text{si } 0 \leq k \leq n-1 \\ L(k+1, x, n) = 0 & \text{si } n-1 < k \end{cases}.$$

Avec la même convention de notation  $T_n = T_n(X)$ , pour tout  $n$  dans  $\mathbb{N}$  et tout  $k$  dans  $\mathbb{Z}$ , nous définissons  $L_k^n$  sur le sous-ensemble  $\{T_n < +\infty\} \subset \mathbb{Z}^{\mathbb{N}}$  par

$$L_k^n = L(k, X, n) = \sum_{t=0}^{T_n-1} \mathbb{1}_{\{X_t=k, X_{t+1}=k-1\}}.$$

La famille  $(L_k^n)_{0 \leq k \leq n}$  constitue le processus des sauts à gauche de la MAMA.

### 1.2.4 Processus de branchement en milieu aléatoire

Comme dans Section 1.2.1, nous notons  $E = (0, 1)^{\mathbb{Z}}$  l'ensemble des environnements muni de la tribu  $\mathcal{E} = \mathcal{B}([0, 1])^{\mathbb{Z}}$  engendrée par les cylindres. Nous considérons le processus canonique  $(Z_t)_{t \in \mathbb{N}}$  de l'espace  $\mathbb{N}^{\mathbb{N}}$  muni de la tribu  $\mathcal{Z}$  engendrée par les cylindres, i.e. pour tout  $t$  dans  $\mathbb{N}$  et tout  $z = (z_t)_{t \in \mathbb{N}}$  dans  $\mathbb{N}^{\mathbb{N}}$

$$(1.4) \quad Z_t(z) = z_t.$$

Pour tout environnement fixé  $\omega$  dans  $E$  et toute mesure de probabilité fixée  $\mu$  sur  $\mathbb{N}$ , on définit la mesure de probabilité  $\mathbf{Q}_{\mu, \omega}$  comme l'unique mesure de probabilité sur  $\mathcal{Z}$  sous laquelle le processus canonique  $Z = (Z_t)_{t \in \mathbb{N}}$  constitue une chaîne de Markov non-homogène partant de  $\mu$  dont les noyaux de transition  $(K_{\omega, k})_{k \in \mathbb{N}}$  sont définis pour tout  $k, i$  et  $j$  dans  $\mathbb{N}$  par

$$(1.5) \quad K_{\omega, k}(i, j) = \binom{i+j}{i} \omega_k^{i+1} (1 - \omega_k)^j.$$

Pour toute mesure de probabilité  $\nu$  sur  $([0, 1], \mathcal{B}([0, 1]))$ , on munit l'espace des environnements  $(E, \mathcal{E})$  de la même mesure de probabilité  $\mathbb{P}^\nu = \nu^{\otimes \mathbb{Z}}$  qu'en Section 1.2.1. Considérant alors une mesure de probabilité fixée  $\mu$  sur  $\mathbb{N}$ , on intègre la loi  $\mathbf{Q}_{\mu, \omega}$  (fonction de l'environnement) sous la loi  $\mathbb{P}^\nu$ , on définit la loi de probabilité  $\tilde{\mathbf{Q}}_\mu^\nu$  sur  $(E \times \mathbb{N}^{\mathbb{N}}, \mathcal{E} \otimes \mathcal{Z})$ , i.e. pour  $A$  dans  $\mathcal{E}$  et  $B$  dans  $\mathcal{Z}$

$$\tilde{\mathbf{Q}}_\mu^\nu(A \times B) = \int_E \mathbb{1}_A(\omega) \mathbf{Q}_{\mu, \omega}(B) \mathbb{P}^\nu(d\omega) = \int_A \mathbf{Q}_{\mu, \omega}(B) \mathbb{P}^\nu(d\omega).$$

La seconde marginale de  $\tilde{\mathbf{Q}}_\mu^\nu$ , notée  $\mathbf{Q}_\mu^\nu$ , est la loi du *processus de branchement avec immigration en milieu aléatoire* (abrégé en *PBMA*).

Pour tout  $\nu$  dans  $\mathbb{M}^1([0, 1])$ , [CFL<sup>+</sup>14, Proposition 4.3] montre alors que  $\mathbf{Q}_\mu^\nu$  est l'unique mesure de probabilité sur  $\mathbb{N}$  sous laquelle le processus canonique  $(Z_t)_{t \in \mathbb{N}}$  constitue une chaîne de Markov homogène partant de  $\mu$  et de noyau de transition  $K^\nu$  défini pour tous  $i$  et  $j$  dans  $\mathbb{N}$  par

$$(1.6) \quad K^\nu(i, j) = \binom{i+j}{i} \int_0^1 t^{i+1} (1-t)^j \nu(dt).$$

Considérons des mesures de probabilité  $\mu$  et  $\nu$  respectivement sur  $\mathbb{N}$  et  $[0, 1]$ . Nous donnons une autre définition de la loi  $\mathbf{Q}_\mu^\nu$  du processus de branchement milieu aléatoire qui indique notamment une manière de simuler un tel processus. Soit une suite de variables aléatoires i.i.d.  $e = (e_k)_{k \in \mathbb{Z}}$  de loi  $\nu$  sur un espace  $(\Omega, \mathcal{A}, \mathbb{P})$ . Soit une suite de variables aléatoires  $(Y_k)_{k \in \mathbb{N}}$  sur  $\Omega$  telle que la variable  $Y_0$  est de loi  $\mu$  et est indépendante de  $e$  et telle que pour tout  $k$  dans  $\mathbb{N}$

$$(1.7) \quad Y_{k+1} = \sum_{i=0}^{Y_k} \xi_{k,i},$$

où les  $(\xi_{k,i})_{(k,i) \in \mathbb{N}^2}$  sont des variables indépendantes conditionnellement à la tribu engendrée par  $e$  de loi géométrique donnée pour tous entiers  $k, i$  et  $n$

$$(1.8) \quad \mathbb{P}(\xi_{k,i} = n | e) = e_k (1 - e_k)^n.$$

La loi image de  $\mathbb{P}$  par  $(Y_k)_{k \in \mathbb{N}}$  définie sur  $\mathbb{N}^{\mathbb{N}}$  est alors la loi  $\mathbf{Q}_\mu^\nu$ .

### 1.2.5 Le lien entre les marches aléatoires en milieu aléatoire et les processus de branchement en milieu aléatoire

Dans le cadre de la Section 1.2.1, pour toute suite  $x = (x_t)_{t \in \mathbb{N}}$  dans  $\{x \in \mathcal{S}_0 : T_n(x) < +\infty\}$ , nous déduisons une identité fondamentale à partir des relations (1.3)

$$\begin{aligned}
(1.9) \quad \mathbf{P}^\nu (X_{0:T_n} = x_{0:T_n(x)}) &= \int_E \prod_{t=0}^{T_n(x)-1} (\omega_{x_t} \mathbb{1}_{x_{t+1}=x_t+1} + (1 - \omega_{x_t}) \mathbb{1}_{x_{t+1}=x_t-1}) \nu^{\otimes \mathbb{Z}}(d\omega) \\
&= \prod_{k \in \mathbb{Z}} \int_{[0,1]} t^{R(k,x,n)} (1-t)^{L(k,x,n)} \nu(dt) \\
&= \prod_{k=-\infty}^{-1} \int_0^1 t^{L(k+1,x,n)} (1-t)^{L(k,x,n)} \nu(dt) \times \prod_{k=0}^{n-1} \int_0^1 t^{L(k+1,x,n)+1} (1-t)^{L(k,x,n)} \nu(dt).
\end{aligned}$$

D'après l'Equation (1.9), la famille  $(L_k^n)_{k \leq n, k \in \mathbb{Z}}$  constitue une statistique exhaustive.

Mais la propriété de la suite  $(L_k^n)_{k \leq n, k \in \mathbb{Z}}$  la plus importante pour notre travail est celle mise en évidence pour la première fois par [KKS75] et déjà exploitée par [CFLL16, CFL<sup>+</sup>14, FLM14, FGL14, DL18] : la distribution du processus des sauts à gauche  $(L_{n-k}^n)_{0 \leq k \leq n}$  sous  $\mathbf{P}^\nu$  est identique à la distribution de  $(Z_t)_{0 \leq t \leq n}$  sous  $\mathbf{Q}'_0$ .

### 1.2.6 Un estimateur des moments par [AE04]

[AE04] est le premier à adopter une approche statistique pour la MAMA et se pose la question de l'inférence directe sur la loi du milieu  $\nu$ . Ne se restreignant pas à la MAMA sur  $\mathbb{Z}$ , il travaille sur des groupes abéliens (donc en particulier  $(\mathbb{Z}, +)$ ). Il propose alors un estimateur des moments qui est construit non pas sur tous les sites visités par la marche mais seulement sur ceux ayant une *histoire* spécifique. Nous proposons ici une exposition rapide de cet estimateur dans le cadre de la MAMA sur  $\mathbb{Z}$ .

Définissant pour tout  $k$  dans  $\mathbb{Z}$  et tout  $t$  dans  $\mathbb{N}$ , l'histoire du site  $k$  au temps  $t$  par

$$H(t, k) = (L(k, X_{0:t}), R(k, X_{0:t})) ,$$

on peut alors considérer l'histoire du site courant  $X_t$  à n'importe quel temps  $t$

$$H(t) = H(t, X_t) = (L(X_t, X_{0:t}), R(X_t, X_{0:t})) .$$

On définit alors pour tout couple d'entiers naturels  $h = (h_-, h_+)$ , la suite  $(K_i^h)_{i \in \mathbb{N}}$  des temps successifs où l'histoire du site courant est  $h$

$$K_0^h = \inf\{t \in \mathbb{N} : H(t) = h\} \quad \text{et} \quad K_{i+1}^h = \inf\{t > K_i^h : H(t) = h\} ,$$

et  $\Delta_i^h$  le mouvement de la marche la  $i$ -ème fois où l'histoire du site courant est  $h$

$$\Delta_i^h = X_{K_{i+1}^h} - X_{K_i^h} .$$

D'après [AE04, Proposition 4, Corollary 2], les variables  $(\Delta_i^h)_{i \in \mathbb{N}}$  sont i.i.d. et si  $\omega_0 \sim \nu$ , alors  $\mathbf{P}^\nu$  presque sûrement pour tout  $\varepsilon$  dans  $\{-1, 1\}$

$$\frac{1}{m} \sum_{i=1}^m \mathbb{1}_{\Delta_i^h = \varepsilon} \xrightarrow[n \rightarrow +\infty]{} V_\varepsilon(h) = \frac{\mathbb{E}^\nu \left[ \omega_0^{\frac{1+\varepsilon}{2} + h_+} (1 - \omega_0)^{\frac{1-\varepsilon}{2} + h_-} \right]}{\mathbb{E}^\nu \left[ \omega_0^{h_+} (1 - \omega_0)^{h_-} \right]}.$$

Définissant pour tout  $\varepsilon$  dans  $\{-1, 1\}$ , les estimateurs  $\widehat{V}_\varepsilon^n(h)$  par

$$\widehat{V}_\varepsilon^n(h) = \frac{1}{M_n^h} \sum_{i=1}^{M_n^h} \mathbb{1}_{\Delta_i^h = \varepsilon} \quad \text{où} \quad M_n^h = \sup\{K_i^h < T_n : i \geq 1\},$$

le fait que  $\mathbf{P}^\nu$  presque sûrement  $T_n \xrightarrow[n \rightarrow +\infty]{} +\infty$  assure alors que  $\mathbf{P}^\nu$  presque sûrement

$$\widehat{V}_\varepsilon^n(h) \xrightarrow[n \rightarrow +\infty]{} V_\varepsilon(h),$$

et on peut ainsi estimer tous les moments de la loi  $\nu$  en faisant varier  $h$ .

[AE04, Section 4.3] illustre cette procédure d'estimation dans un cadre paramétrique particulier et alerte sur la non-efficacité d'un tel algorithme dans des situations plus générales.

### 1.2.7 Un estimateur de type maximum de vraisemblance par [CFLL16, CFL<sup>+</sup>14, FLM14, FGL14]

Dans le prolongement de [AE04], les travaux [CFLL16, CFL<sup>+</sup>14, FLM14, FGL14] reprennent la question de l'inférence directe de la loi du milieu dans un cadre paramétrique.

Dans un premier temps, [CFL<sup>+</sup>14, FLM14, FGL14] exploitent l'égalité en loi entre le processus des sauts à gauche de la MAMA et le processus de branchement avec immigration en milieu aléatoire initialement mise en évidence par [KKS75] et explicitée en Section 1.2.5.

Le premier mérite de cette égalité en loi est de réduire le problème statistique de l'estimation directe de la loi du milieu de la MAMA à partir de l'observation d'une unique trajectoire  $X_{0:T_n}$  au problème d'estimation du paramètre  $\nu$  définissant le PBMA à partir de l'observation d'une réalisation de ses  $n + 1$  premières générations.

Le second mérite de cette égalité en loi provient de [CFL<sup>+</sup>14, Proposition 4.3] : le PBMA défini par la loi  $\nu$  est une chaîne de Markov homogène de noyau  $K^\nu$  défini dans l'Equation (1.6).

En particulier, lorsque  $\nu$  dans  $\mathbb{M}^1([0, 1])$  vérifie  $\int_0^1 \log(\frac{1-t}{t}) \nu(dt) < 0$ , [CFL<sup>+</sup>14, Proposition 4.5] établit que la chaîne de Markov de noyau de transition  $K^\nu$  est récurrente positive et admet une unique loi de probabilité invariante  $\mu^\nu$ . S'appuyant sur les propriétés d'ergodicité des chaînes de Markov, [CFL<sup>+</sup>14, Proposition 4.6] établit pour tout  $\nu'$  dans  $\mathbb{M}^1([0, 1])$

$$\frac{1}{n} \sum_{t=0}^{n-1} \log \left( \int_0^1 t^{Z_t+1} (1-t)^{Z_{t+1}} \nu'(dt) \right) \xrightarrow[n \rightarrow +\infty]{\mathbf{Q}_0^\nu} \mathbf{E}_{\mu^\nu}^\nu \left[ \log \left( \int_0^1 t^{Z_0+1} (1-t)^{Z_1} \nu'(dt) \right) \right],$$

où le processus canonique  $(Z_t)_{t \in \mathbb{N}}$  et la loi  $\mathbf{Q}_{\mu^\nu}^\nu$  sont définis en Section 1.2.4.



Ainsi, se fixant un modèle paramétrique  $(\nu_\theta)_{\theta \in \Theta}$  où  $\Theta \subset \mathbb{R}^d$ , [CFL<sup>+</sup>14, FGL14] proposent pour les régimes transients sous-balistique et balistique un  $M$ -estimateur  $\widehat{\theta}_n$  défini par

$$\widehat{\theta}_n \in \underset{\theta \in \Theta}{\operatorname{Argmax}}(l_n(\theta) - l_n(\theta_0)),$$

où  $\theta_0$  dans  $\Theta$  est une valeur fixée du paramètre et  $l_n(\theta)$  est définie pour tout  $\theta$  dans  $\Theta$  par

$$l_n(\theta) = \sum_{k=0}^{n-1} \log \left( \int_0^1 t^{L_n^{k+1}+1} (1-t)^{L_n^k} \nu_\theta(dt) \right).$$

Supposant que l'application  $\theta \mapsto \int_0^1 t^{x+1} (1-t)^y \nu_\theta(dt)$  est continue sur  $\Theta$  pour tous  $x$  et  $y$  dans  $\mathbb{N}$ , que le modèle est identifiable (i.e.  $\theta \mapsto \nu_\theta$  est injective) et sous une autre hypothèse plus technique différente selon le régime, [CFL<sup>+</sup>14] et [FGL14] montrent notamment la consistance de l'estimateur  $\widehat{\theta}_n$  en régimes transients balistique et sous-balistique.

Précisant l'étude en régime balistique, sous des hypothèses techniques de régularité relativement au paramètre, [FLM14] établit la normalité asymptotique et même l'efficacité asymptotique (i.e. il réalise asymptotiquement la borne de Cramer-Rao) de l'estimateur  $\widehat{\theta}_n$ . Des résultats semblables en découlent dans [FGL14] pour le régime sous-balistique.

Finalement, [FGL14, Section 4], [FLM14, Section 5] et [CFL<sup>+</sup>14, Section 5] illustrent numériquement les performances de leur estimateur sur des modèles paramétriques particuliers et les comparent à celles obtenues pour l'estimateur issu de [AE04]. En plus d'y souligner le caractère plus systématique de leur procédure d'estimation, ils montrent également que les performances sont bien meilleures que ce soit en terme de biais ou en terme de variance.

Dans un second temps, [CFLL16] se base sur le phénomène de *forte localisation* révélé par [Sin82] en régime récurrent. Il choisit pour modèle la famille  $(\nu_\theta)_{\theta \in \Theta}$  où  $\Theta \subset (0, 1)^d \times (0, 1)^d$  avec  $d \geq 2$  entier et où pour tout  $\theta = (a, p)$  dans  $\Theta$

$$\nu_\theta = \sum_{i=1}^d p_i \delta_{a_i},$$

il propose deux  $M$ -estimateurs. Le premier estimateur  $\widehat{\theta}_n$  est basé sur une méthode de type maximum de vraisemblance et procède à l'estimation des deux composantes du paramètre simultanément. L'autre  $\bar{\theta}_n$  est obtenu par un développement asymptotique de la log-vraisemblance du modèle. Il identifie d'abord la première composante du paramètre (i.e. le support de la loi  $\nu_\theta$ ) à l'aide du terme de premier ordre du développement. Puis il identifie à partir de cette première estimation la seconde composante du paramètre (i.e. le vecteur de probabilité de la loi  $\nu_\theta$ ) à partir du terme de second ordre. Sous les seules hypothèses de récurrence de la MAMA et d'identifiabilité du modèle, la consistance des estimateurs  $\widehat{\theta}_n$  et  $\bar{\theta}_n$  est obtenue.

Finalement, [CFLL16, Section 6] illustre les performances numériques de  $\widehat{\theta}_n$  et  $\bar{\theta}_n$ . Il les compare également à celles de l'estimateur issu de [AE04] sur des modèles particuliers. Ne tenant pas compte du phénomène de *pièges* introduit par la localisation dans les *vallées de potentiel* de [Sin82], l'estimateur issu de [AE04] présente en particulier un biais et une variance plus importants que  $\widehat{\theta}_n$  et  $\bar{\theta}_n$ .

### 1.2.8 Un premier estimateur non paramétrique par [DL18]

Après les résultats asymptotiques et paramétriques de [CFL16, CFL<sup>+</sup>14, FLM14, FGL14], l'article [DL18] s'affranchit simultanément de ces deux contraintes. En effet, ce dernier se place dans un cadre non paramétrique et propose un estimateur de la fonction de répartition de la loi du milieu dont il étudie les propriétés de concentration.

Dans un premier temps, [DL18] considère le problème d'estimation de la famille des  $\beta$ -moments  $(m^{a,b}(\nu))_{(a,b) \in \mathbb{M}([0,1]) \times \mathbb{N}^2}$  de la loi  $\nu$  où pour tous entiers  $a$  et  $b$

$$m^{a,b}(\nu) = \int_0^1 u^a (1-u)^b \nu(du).$$

S'appuyant sur l'égalité en loi entre processus des sauts à gauche de la MAMA rappelée en Section 1.2.5 et sur le caractère markovien du processus de branchement en milieu aléatoire, il propose d'estimer  $m^{a,b}(\nu)$  par

$$\widehat{m}_n^{a,b} = \frac{1}{N_n^a} \sum_{k=1}^n \phi^{a,b} \left( L_{n-(k-1)}^n, L_{n-k}^n \right),$$

où  $(L_{n-k}^n)_{0 \leq k \leq n}$  désigne le processus des sauts à gauches introduit en Section 1.2.3 et

$$\phi^{a,b}(i, j) = \frac{\binom{i+j-(a+b)}{i-a}}{\binom{i+j}{i}} \mathbb{1}_{\{i \geq a, j \geq b\}} \quad \text{et} \quad N_n^a = \sum_{k=1}^n \mathbb{1}_{\{L_{n-(k-1)}^n \geq a\}}.$$

Par un argument de martingale, [DL18, Theorem 4] établit qu'en régime récurrent ou transien, pour tout entier  $n \geq 1$  et tout réel positif  $z$

$$\mathbf{P}^\nu \left( \left| \widehat{m}_n^{a,b} - m^{a,b}(\nu) \right| \geq \frac{n}{N_n^a} \binom{a+b}{a}^{-1} \frac{z}{\sqrt{2n}} \right) \leq 2e^{-z^2}.$$

[Mna08] établit que la famille de fonctions  $(F^M)_{M \in \mathbb{N}^*}$  construite à partir des  $\beta$ -moments de la loi  $\nu$  et définie pour tout  $M$  dans  $\mathbb{N}^*$  et tout  $u$  dans  $[0, 1]$  par

$$(1.10) \quad F^M(u) = \sum_{k=0}^{\lfloor (M+1)u \rfloor - 1} \binom{M}{k} m^{k, M-k}(\nu),$$

permet d'approcher de manière déterministe la fonction de répartition  $F$  de la loi  $\nu$ .

[DL18] substitue alors les estimateurs  $\widehat{m}_n^{k, M-k}$  aux  $\beta$ -moments  $m^{k, M-k}(\nu)$  dans l'Equation (1.10). Après renormalisation, il propose une famille d'estimateurs  $(\widehat{F}_n^M)_{M \in \mathbb{N}^*}$  de la fonction de répartition de la loi de la MAMA en posant pour tout  $M$  dans  $\mathbb{N}^*$  et tout  $u$  dans  $[0, 1]$

$$\widehat{F}_n^M(u) = \frac{1}{N_n^M} \sum_{k=1}^n \psi_M^{\lfloor (M+1)u \rfloor} \left( L_{n-(k-1)}^n, L_{n-k}^n \right),$$

avec

$$\psi_M^l(i, j) = \frac{\mathbb{1}_{i \geq M}}{\binom{i+j}{M}} \sum_{k=0}^{l-1} \binom{i}{k} \binom{j}{M-k} \quad \text{et} \quad N_n^M = \sum_{k=1}^n \mathbb{1}_{\{L_{n-(k-1)}^n \geq M\}}.$$

La procédure de Goldenshluger-Lepski [GL08] permet alors de choisir automatiquement et de manière optimale le paramètre  $M$  en posant pour tout  $z > 0$  et tout entier  $M \geq 1$

$$\widehat{M}_n^z = \underset{M \geq 1}{\operatorname{Argmin}} \left( \Delta(M) + \frac{2n}{N_n^M} \sqrt{\frac{z + 3 \log M}{2n}} \right)$$

où

$$\Delta(M) = \sup_{M' \geq 1} \left( \left\| \widehat{F}_n^{M'} - \widehat{F}_n^{M \wedge M'} \right\|_\infty - \frac{2n}{N_n^{M'}} \sqrt{\frac{z + 3 \log M'}{2n}} \right).$$

Pour tous réels  $\beta > 0$  et  $L > 0$  fixés, notons  $m = \sup \{ \ell \in \mathbb{N} : \ell < \beta \}$  et désignons par  $\Sigma(\beta, L)$  l'ensemble des fonctions  $\beta$ -Hölder, i.e. les fonctions  $m$ -fois différentiables  $g : [0, 1] \rightarrow \mathbb{R}$  telles que pour  $x$  et  $x'$  dans  $[0, 1]$

$$(1.11) \quad \left| g^{(m)}(x) - g^{(m)}(x') \right| \leq L |x - x'|^{\beta - m}.$$

Ainsi, [DL18, Theorem 1] assure que si la fonction de répartition de la loi du milieu  $F$  est  $\gamma$ -Hölder pour  $\gamma$  dans  $(0, 2]$  et que la MAMA est transiente ou récurrente, alors, il existe une constante  $C_\nu$  telle que pour tout entier  $n \geq 2$ , l'estimateur  $\widehat{F}_n = \widehat{F}_n^{\widehat{M}_n^{\log(n)}(X_{0:T_n})}(X_{0:T_n})$  vérifie

- (i) si  $\mathbb{E}^\nu[\log \rho_0] = 0$  et  $\mathbb{E}^\nu[\log^2 \rho_0] > 0$  et s'il existe  $a$  tel que  $\mathbb{E}^\nu[\rho_0^a] + \mathbb{E}^\nu[\rho_0^{-a}] < \infty$

$$\mathbf{E}^\nu \left[ \left\| \widehat{F}_n - F \right\|_\infty \right] \leq C_\nu \frac{\log n}{\sqrt{n}},$$

- (ii) si  $\mathbb{E}^\nu[\log \rho_0] < 0$  et s'il existe  $0 < \kappa < \infty$  tel que  $\mathbb{E}^\nu[\rho_0^\kappa] = 1$  et  $\mathbb{E}^\nu[\rho_0^\kappa \log^+ \rho_0] < \infty$

$$\mathbf{E}^\nu \left[ \left\| \widehat{F}_n - F \right\|_\infty \right] \leq C_\nu \left( \frac{\log n}{n} \right)^{\frac{\gamma}{2\gamma + 4\kappa}}.$$

Contrairement aux estimateurs proposés dans les approches paramétriques de [CFL16, CFL<sup>+</sup>14, FLM14, FGL14], cet estimateur est défini indépendamment du régime. Par ailleurs, il est adaptatif vis-à-vis de la régularité Hölderienne du régime.

Hormis leur approche statistique, le point commun essentiel des travaux exposés jusqu'alors est leur mise à profit du parallèle entre les deux processus aléatoires en milieu aléatoire sur  $\mathbb{Z}$  que constituent la MAMA et le PBMA.

Exploitant de nouveau ce lien fondamental, nous allons maintenant exposer nos contributions au problème de l'estimation de la loi du milieu de la MAMA en environnements i.i.d. sur  $\mathbb{Z}$ .

### 1.3 Contributions

L'objet de cette thèse étant l'estimation de la loi du milieu, nous y proposons des solutions au problème d'apprentissage de la loi  $\nu$  à partir de l'observation d'une unique trajectoire que nous considérons sous l'angle de la loi annealed. Nous envisageons principalement deux approches.

La première approche, non paramétrique, fait l'hypothèse que la loi  $\nu$  possède une densité régulière. Nous proposons un estimateur de cette densité et étudions ses propriétés d'un

point de vue fréquentiste à la suite de [DL18]. C'est l'objet du Chapitre 2 qui reprend notre publication [HLM19].

La seconde approche, également non paramétrique, envisage quant à elle pour la première fois sous un angle bayésien le problème d'estimation de la loi du milieu de la MAMA. Nous munissons l'espace des probabilités sur  $[0, 1]$  d'une loi a priori  $\Pi$ . Nous étudions l'estimateur Bayésien et montrons la consistance de la loi a posteriori de cet estimateur.

Cette approche fait l'objet du Chapitre 5. Toutefois, elle requiert plusieurs résultats intermédiaires ne portant pas directement sur la MAMA. Au Chapitre 3, nous faisons ainsi l'étude de la distribution du temps de retour en 0 du PBMA présenté en Section 1.2.5. Et au Chapitre 4, reprenant plusieurs résultats de [DMPS18], nous proposons une inégalité de concentration des transformations ayant la propriété de différences bornées qui soit uniformément valable pour une classe de chaînes de Markov. Puis nous appliquons cette inégalité au PBMA.

### 1.3.1 Estimation de la densité de la loi du milieu : Chapitre 2

Dans le prolongement du travail d'estimation de la fonction de répartition de la loi du milieu par [DL18] exposé en Section 1.2.8, nous nous posons la question de l'estimation non paramétrique et non asymptotique de la densité de la loi du milieu  $\nu$ .

Nous adoptons alors la même démarche en deux étapes que [DL18] : d'abord trouver une approximation déterministe de la densité  $f$  de la loi  $\nu$ , ensuite estimer cette approximation.

Les ensembles  $\Sigma(\beta, L)$  étant défini par l'Equation (1.11), nous imposons que la densité  $f$  de la loi  $\nu$  appartienne à  $\Sigma(\beta, L)$  avec  $\beta$  dans  $(0, 2]$ .

La loi  $\nu$  du milieu étant à support borné dans  $[0, 1]$ , elle est entièrement déterminée par la suite de ses moments  $\mu_j(\nu) = \int_0^1 t^j \nu(dt) = \int_0^1 t^j f(t) dt$ . Dans un tel contexte, [Mna08] montre que la suite de fonctions  $(\nu_M)_{M \in \mathbb{N}}$  définies pour  $M$  dans  $\mathbb{N}$  et  $x$  dans  $[0, 1]$  par

$$\nu_M(x) = \sum_{k=0}^{\lfloor Mx \rfloor} \sum_{j=k}^M \binom{M}{j} \binom{j}{k} (-1)^{j-k} \mu_j[\nu],$$

converge simplement vers la fonction de répartition de  $\nu$  quand  $M$  tend vers  $+\infty$ .

Le lien entre fonction de répartition et densité nous amène alors à définir une suite de densités de probabilité  $(f_M)_{M \in \mathbb{N}}$  en posant pour tout  $M$  dans  $\mathbb{N}$  et tout  $x$  dans  $[0, 1]$

$$f^M(x) = (M+1) (\nu_M([0, x]) - \nu_M([0, x - M^{-1}])) = (M+1) \binom{M}{\lfloor Mx \rfloor} m^{\lfloor Mx \rfloor, M - \lfloor Mx \rfloor}.$$

Lorsque la densité  $f$  de la loi  $\nu$  est continue, [Mna08, Theorem 1] garantit la convergence uniforme de la suite de densités de probabilité  $(f_M)_{M \in \mathbb{N}}$  vers  $f$ . Nous précisons ce résultat en garantissant l'existence d'une constante  $C_{\beta, L}$  telle que

$$\|f - f^M\|_{\infty} \leq C_{\beta, L} M^{-\frac{\beta}{2}}.$$

Nous substituons l'unique  $\beta$ -moment figurant dans l'expression de cette approximation déterministe de la densité  $f$  par l'estimateur correspondant de [DL18]. Nous définissons ainsi une famille d'estimateurs  $(\hat{f}_n^M)_{M \in \mathbb{N}}$  de la densité par

$$\hat{f}_n^M(x) = (M+1) \binom{M}{\lfloor Mx \rfloor} \hat{m}_n^{\lfloor Mx \rfloor, M - \lfloor Mx \rfloor}.$$

Grâce aux propriétés de concentration des  $\beta$ -moments par [DL18] et aux contrôles déterministes de [BR03, Mna08, BK10], nous contrôlons respectivement les erreurs d'estimation et d'approximation de ces estimateurs. Nous obtenons alors les premières bornes de risque non asymptotiques non paramétriques : il existe des constantes  $C_\nu^r > 0$  et  $C_\nu^t > 0$  telles que

(i) si  $\mathbb{E}^\nu[\log \rho_0] = 0$  et  $\mathbb{E}^\nu[\log^2 \rho_0] > 0$  et s'il existe  $a$  tel que  $\mathbb{E}^\nu[\rho_0^a] + \mathbb{E}^\nu[\rho_0^{-a}] < \infty$

$$\mathbf{E}^\nu \left[ \left\| \widehat{f}_n^M - f \right\|_\infty \right] \leq C_\nu^r \left( LM^{-\frac{\beta}{2}} + \frac{\log n}{\sqrt{n}} \right),$$

et pour  $M_n = \left\lfloor (\sqrt{n}/\log(n))^{2/(\beta+2)} \right\rfloor$

$$\mathbf{E}^\nu \left[ \left\| f - \widehat{f}_n^{M_n} \right\|_\infty \right] \leq C_\nu^t \left( \frac{\log(n)}{\sqrt{n}} \right)^{\beta/(\beta+2)},$$

(ii) si  $\mathbb{E}^\nu[\log \rho_0] < 0$  et s'il existe  $0 < \kappa < \infty$  tel que  $\mathbb{E}^\nu[\rho_0^\kappa] = 1$  et  $\mathbb{E}^\nu[\rho_0^\kappa \log^+ \rho_0] < \infty$

$$\mathbf{E}^\nu \left[ \left\| \widehat{f}_n^M - f \right\|_\infty \right] \leq C_\nu^t \left( LM^{-\frac{\beta}{2}} + M^{1+\kappa} \sqrt{\frac{\log n}{n}} \right),$$

et pour  $M_n = \left\lfloor (n/\log(n))^{\frac{1}{\beta+2(1+\kappa)}} \right\rfloor$

$$\mathbf{E}^\nu \left[ \left\| f - \widehat{f}_n^{M_n} \right\|_\infty \right] \leq C_\nu^t \left( \frac{\log(n)}{n} \right)^{\frac{\beta}{2(\beta+2(1+\kappa))}}.$$

Au même titre que [DL18], la méthode de Goldenshluger-Lepski [GL08] nous fournit une procédure de sélection automatique du paramètre de régularisation  $M$  dont notre estimateur dépend. Cela permet de réaliser les optima des bornes de risque sans connaître ni le régime de la MAMA, ni la classe de régularité hölderienne de la densité. Nous définissons ainsi pour tout entier  $n \geq 2$  et tout entier  $1 \leq M \leq n-1$

$$\widehat{M}_n \in \underset{M \in \{1, \dots, n-1\}}{\operatorname{Argmin}} \left\{ C_n(M) + \frac{M+1}{N_n^M} 2\sqrt{3n \log n} \right\}$$

où  $C_n(M) = \sup_{M' \in \{1, \dots, n-1\}} \left\{ \left\| \widehat{f}_n^M - \widehat{f}_n^{M \vee M'} \right\|_\infty - \frac{M'+1}{N_n^{M'}} 2\sqrt{3n \log n} \right\}.$

Considérant alors les estimateurs  $\widehat{f}_n = \widehat{f}_n^{\widehat{M}_n(X_{0:T_n})}(X_{0:T_n})$  et  $\widetilde{f}_n = \min(\widehat{f}_n, \log n)$ , nous obtenons alors l'existence de deux constantes  $C_\nu^r > 0$  et  $C_\nu^t > 0$  telles que

(i) si  $\mathbb{E}^\nu[\log \rho_0] = 0$  et  $\mathbb{E}^\nu[\log^2 \rho_0] > 0$  et s'il existe  $a$  tel que  $\mathbb{E}^\nu[\rho_0^a] + \mathbb{E}^\nu[\rho_0^{-a}] < \infty$

$$\mathbf{P}^\nu \left( \left\| f - \widehat{f}_n \right\|_\infty \leq C_\nu^r \left( \frac{\log n}{\sqrt{n}} \right)^{\beta/(\beta+2)} \right) \geq 1 - C_\nu^r \frac{\log n}{\sqrt{n}},$$

et

$$\mathbf{E}^\nu \left[ \left\| f - \widetilde{f}_n \right\|_\infty \right] \leq C_\nu^t \left( \frac{\log n}{\sqrt{n}} \right)^{\beta/(\beta+2)},$$

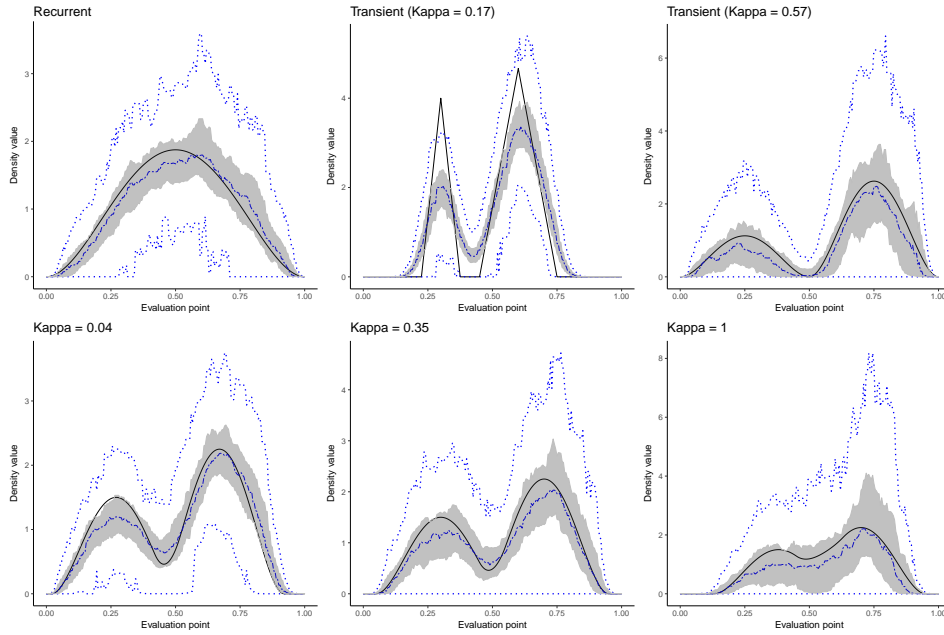


Figure 1.1: La véritable densité (trait plein noir) et en chaque point  $x$  de  $(0, 1)$  la médiane (alternance tirets-pointillés bleus), l'écart interquartile (zone grisée) et les *hinges* inférieur et supérieur (pointillés bleus) de l'estimateur  $\hat{f}_{100}$  pour les 6 milieux considérés.

(ii) si  $\mathbb{E}^\nu[\log \rho_0] < 0$  et s'il existe  $0 < \kappa < \infty$  tel que  $\mathbb{E}^\nu[\rho_0^\kappa] = 1$  et  $\mathbb{E}^\nu[\rho_0^\kappa \log^+ \rho_0] < \infty$

$$\max \left( \mathbf{E}^\nu \left[ \left\| f - \hat{f}_n \right\|_\infty \right], \mathbf{E}^\nu \left[ \left\| f - \tilde{f}_n \right\|_\infty \right] \right) \leq C_\nu^t \left( \frac{\log n}{n} \right)^{\frac{\beta}{2(\beta+2(1+\kappa))}}.$$

Nous illustrons finalement par des simulations numériques les performances de notre estimateur issu de la procédure de sélection de Goldenshluger-Lepski. Pour 6 milieux différents, nous simulons 100 trajectoires de la MAMA jusqu'à ce qu'elles atteignent le site  $n = 100$ . Pour chacune d'elles et pour toutes les valeurs du paramètre  $M$  allant de 1 à 100, nous calculons l'estimateur  $f_n^M$ . La méthode de sélection de Goldenshluger-Lepski permet le choix du paramètre optimal  $\hat{M}_n$  et nous obtenons ainsi l'estimateur  $\hat{f}_n$ . Sur la Figure 1.1, nous donnons un aperçu des performances de cet estimateur  $\hat{f}_n$  de la densité en traçant différentes grandeurs statistiques pour chacun des milieux considérés.

Conformément à l'intuition et étant donné notre schéma d'observation, plus la chaîne atteint rapidement le site  $n$  ou plus la densité est irrégulière, moins l'estimation de la densité est précise.

### 1.3.2 Brève introduction aux statistiques bayésiennes

Le paradigme bayésien consiste à considérer comme aléatoire aussi bien l'observation que le paramètre d'intérêt. L'espace des paramètres  $\Theta$  est muni d'une probabilité  $\Pi$ , appelée *prior* ou *loi a priori*. Supposons que les observations  $(X_n)_{n \in \mathbb{N}}$  sont i.i.d. à valeurs dans un espace mesurable  $\mathcal{X}$  et de loi commune une loi dénotée  $\mathbf{P}^\theta$  dépendant d'un paramètre  $\theta$  dans  $\Theta$ . En interprétant  $\mathbf{P}^\theta$  comme la distribution conditionnelle de l'observation sachant la variable aléatoire  $\theta$  de loi  $\Pi$ , la loi du couple  $((X_n)_{n \in \mathbb{N}}, \theta)$  est entièrement déterminée.

La suite de lois *a posteriori* ou *posteriors* est alors définie comme la suite de loi  $(\Pi(\cdot|X_{0:n}))_{n \in \mathbb{N}}$  où pour tout  $n$  entier,  $\Pi(\cdot|X_{0:n})$  désigne la loi conditionnelle de la variable  $\theta$  sachant l'observation  $X_{0:n}$ . Ainsi, l'étude du comportement de la suite des lois *a posteriori*  $(\Pi(\cdot|X_{0:n}))_{n \in \mathbb{N}}$  est naturelle. Nous n'entrerons pas dans cette introduction dans les considérations théoriques justifiant l'existence d'une telle loi conditionnelle et renvoyons par exemple à [Dud02] pour une exposition exhaustive.

Dans le cadre bayésien, la question du choix de la loi a priori est importante. Dans certains problèmes, nous disposons d'une information préalable sur le paramètre qu'il est donc judicieux d'exploiter : l'approche bayésienne est alors parfaitement adaptée à la prise en compte d'une telle information. Dans d'autres situations, il n'existe pas de façon naturelle de spécifier la loi a priori. Le caractère arbitraire de ce choix est souvent avancé pour critiquer l'approche bayésienne. Nous ne pouvons exclure la possibilité que des praticiens qui choisiraient des lois a priori différentes aboutiraient à partir d'observations identiques à des conclusions différentes... Dès lors, une réponse positive quant à l'indifférence du choix du prior est souhaitable car elle signifie que l'arbitraire du choix de la loi a posteriori disparaît asymptotiquement.

L'étude fréquentiste de la distribution a posteriori donne une réponse à ce type de questions. Faisant l'hypothèse fréquentiste que les données sont une réalisation du modèle statistique pour une valeur  $\theta^*$  du paramètre (inconnue), cette approche recherche quelles conditions imposer sur le modèle (i.e. sur la loi a priori) pour que la distribution a posteriori se *concentre* autour du paramètre  $\theta^*$ . Soulignons que la justification ainsi apportée de l'approche bayésienne n'est pas le seul intérêt de l'étude fréquentiste dans un cadre bayésien ; voir par exemple [DF86] et [GvdV17].

Dans le cadre d'une étude fréquentiste, nous entendons généralement par *concentration* les propriétés de consistance et de vitesse de convergence de la distribution a posteriori. La consistance de la suite des distributions a posteriori en  $\nu^*$  pour une topologie  $\mathcal{T}$  (sur l'espace  $\Theta$  des paramètres) signifie que pour tout ouvert  $U$  de  $\mathcal{T}$  contenant  $\theta^*$ , la probabilité a posteriori  $\Pi(U|X_{0:n})$  converge vers 1 quand  $n$  tend vers  $+\infty$ . Si cette convergence a lieu presque sûrement, on parle de *consistance forte a posteriori* tandis que si elle a seulement lieu en probabilité, on parle de *consistance faible a posteriori* ou tout simplement de *consistance a posteriori*. La question faisant naturellement suite à la consistance est celle de la vitesse de convergence a posteriori et a d'ailleurs donné lieu à de très nombreux travaux, entre autres [GvdV07a, GvdV07b, SvdV01, GGvdV00].

Nous nous concentrons dans ce travail sur la consistance a posteriori : le Chapitre 5 est dédié à la démonstration de la consistance a posteriori pour un modèle de MAMA; les résultats intermédiaires nécessaires à cette preuve font l'objet des Chapitres 3 et 4.

Un résultat très fort concernant la consistance a posteriori remonte à [Doo49] et semble résoudre complètement la question dans le cadre d'observations i.i.d. : si les observations sont i.i.d. à valeurs dans l'espace euclidien  $\mathcal{X}$ , si le modèle est identifiable et si l'espace des paramètres est séparable complet pour la topologie  $\mathcal{T}$ , alors, la suite des distributions a posteriori est consistante pour  $\mathcal{T}$  en  $\Pi$ -presque tout  $\theta^*$ .

Quand l'espace des paramètres  $\Theta$  est de dimension finie (i.e. que le modèle est appelé *paramétrique*), il existe même des résultats nettement plus fins sur le comportement des distributions a posteriori comme par exemple le théorème de Bernstein-von Mises énonçant des conditions suffisantes pour que la suite des lois conditionnelles de la suite  $\sqrt{n}(\theta - \theta_*)$  sachant les observations  $X_{0:n}$ , converge en variation totale.

Quand l'espace des paramètres  $\Theta$  est de dimension infinie (i.e. que le modèle est appelé *non paramétrique*), la consistance a posteriori n'est pas automatique. En effet, bien que la consistance de la suite des distributions a posteriori soit assurée sur un ensemble de probabilité 1 pour  $\Pi$ , il existe dans des modèles infini-dimensionnels des ensembles de probabilité 0 pour  $\Pi$  qui sont "très gros" dans un sens topologique (voir par exemple [Fre63, Fre65, GvdV17] pour de telles situations). [DF86] montre que de tels problèmes sont typiques pour "la plupart des lois a priori".

Pour résoudre ce problème, de nombreux auteurs ont développé des démonstrations de consistance pour des modèles statistique non paramétriques, i.i.d. et dominés : voir par exemple [Sch65, Bar88, BSW99, GGR99]. Nous renvoyons par exemple à [GR03, GvdV17] pour une exposition moderne et des variations autour de ces résultats. Toutefois, en raison de leur fécondité ultérieure, nous présentons maintenant les idées principales derrière la preuve de la consistance a posteriori de [Sch65].

Faisant donc ici l'hypothèse que les lois  $(\mathbf{P}^\theta)_{\theta \in \Theta}$  sont toutes absolument continues par rapport à une même mesure  $\mu$  sur  $\mathcal{X}$  (i.e. que le modèle est dominé), on note  $f_\theta$  la densité de  $\mathbf{P}^\theta$  par rapport à la mesure dominante  $\mu$ . La règle de Bayes s'applique alors et la loi a posteriori s'écrit pour tout entier  $n$  et tout ensemble mesurable  $A \subset \Theta$  comme

$$\Pi(A | X_{0:n}) = \frac{\int_A \prod_{i=0}^n f_\theta(X_i) \Pi(d\theta)}{\int_\Theta \prod_{i=0}^n f_\theta(X_i) \Pi(d\theta)}.$$

La technique développée par [Sch65] repose fondamentalement sur deux hypothèses :

- (i) D'une part, le paramètre  $\theta^*$  doit appartenir au support de Kullback-Leibler de la loi a priori, i.e. pour tout réel  $\delta > 0$

$$\Pi(K_\delta(\theta^*)) > 0 \quad \text{où} \quad K_\delta(\theta^*) = \left\{ \theta \in \Theta : \int_{\mathcal{X}} f_{\theta^*}(x) \log \left( \frac{f_{\theta^*}(x)}{f_\theta(x)} \right) \mu(dx) \leq \delta \right\}.$$

- (ii) D'autre part, pour n'importe quel voisinage  $U$  de  $\theta^*$ , il doit exister une suite d'applications  $(\phi_n)_{n \in \mathbb{N}^*}$  appelés *tests de l'hypothèse simple*  $H_0 : \theta = \theta^*$  contre l'hypothèse composite  $H_1 : \theta \in U^c$  telles que  $\phi_n : \mathbb{N}^n \rightarrow [0, 1]$  est mesurable et qui soit *uniformément exponentiellement consistante*, i.e. il existe  $C > 0$  et  $\beta > 0$  tels que pour  $n$  dans  $\mathbb{N}$

$$\mathbf{E}^{\theta^*} [\phi_n(X_{0:n})] \leq C e^{-\beta n} \quad \text{et} \quad \sup_{\theta \in U^c} \mathbf{E}^\theta [1 - \phi_n(X_{0:n})] \leq C e^{-\beta n}.$$

Considérant ici que le modèle vérifie les hypothèses (i) et (ii), nous munissons sans grande perte de généralité l'espace des paramètres  $\Theta$  d'une distance  $d$  (et de la topologie induite par cette distance). Définissant les ensembles  $A_\varepsilon = \{\theta \in \Theta : d(\theta^*, \theta) \geq \varepsilon\}$ , la consistance a posteriori pour  $\theta^*$  signifie alors que pour tout réel  $\varepsilon > 0$

$$\Pi(A_\varepsilon | X_{0:n}) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^{\theta^*}} 0.$$

Nous donnons ici très succinctement une idée de la preuve de convergence. L'ensemble  $A_\varepsilon$  étant un voisinage de  $\theta^*$  pour  $\varepsilon > 0$  fixé, il existe une famille de tests  $(\phi_n)_{n \in \mathbb{N}^*}$  satisfaisant



la condition (ii) et nous pouvons décomposer  $\Pi(A_\varepsilon | X_{0:n})$  de la façon suivante :

$$\Pi(A_\varepsilon | X_{0:n}) = \phi_n(X_{0:n})\Pi(A_\varepsilon | X_{0:n}) + (1 - \phi_n(X_{0:n}))\Pi(A_\varepsilon | X_{0:n}).$$

D'une part, la première inégalité de la condition (ii) assure

$$\mathbf{E}^{\theta^*} [\phi_n(X_{0:n})\Pi(A_\varepsilon | X_{0:n})] \leq \mathbf{E}^{\theta^*} [\phi_n(X_{0:n})] \leq C e^{-\beta n}.$$

D'autre part, le caractère i.i.d. des  $X_i$  combiné à la règle de Bayes permet de réécrire  $\Pi(A_\varepsilon | X_{0:n})$  comme le quotient de  $N_n(A_\varepsilon)$  par  $D_n$  où

$$N_n(A_\varepsilon) = \int_{A_\varepsilon} \prod_{i=0}^n \frac{f_\theta}{f_{\theta^*}}(X_i) \Pi(d\theta) \quad \text{et} \quad D_n = \int_{\Theta} \prod_{i=0}^n \frac{f_\theta}{f_{\theta^*}}(X_i) \Pi(d\theta).$$

La loi des grands nombres et la condition (i) combinée au théorème de Fubini permettent de montrer que pour tout  $\delta > 0$  et tout  $\delta' > \delta$  il existe  $n_0$  dans  $\mathbb{N}$  tel que pour tout entier  $n \geq n_0$

$$\mathbf{P}^{\theta^*} \left( D_n \geq \Pi(K_\delta(\theta^*)) e^{-\delta' n} \right) \xrightarrow[n \rightarrow +\infty]{} 1.$$

Le théorème de Fubini permet également de réécrire  $\mathbf{E}^{\theta^*} [(1 - \phi_n(X_{0:n}))N_n(A_\varepsilon)]$  comme

$$\mathbf{E}^{\theta^*} [(1 - \phi_n(X_{0:n}))N_n(A_\varepsilon)] = \int_{A_\varepsilon} \mathbf{E}^\theta [(1 - \phi_n(X_{0:n}))] \Pi(d\theta),$$

et grâce à la deuxième inégalité de la condition (ii), il s'ensuit

$$\mathbf{E}^{\theta^*} [(1 - \phi_n(X_{0:n}))N_n(A_\varepsilon)] \leq \Pi(A_\varepsilon) C e^{-n\beta} \leq C e^{-n\beta}.$$

Décomposant l'espace tout entier selon  $\{D_n < \Pi(K_\delta(\theta^*)) e^{-\delta' n}\}$  et son complémentaire

$$\mathbf{E}^{\theta^*} [(1 - \phi_n(X_{0:n}))\Pi(A_\varepsilon | X_{0:n})] \leq \frac{C}{\Pi(K_\delta(\theta^*))} e^{(\delta' - \beta)n} + \mathbf{P}^{\theta^*} \left( D_n < \Pi(K_\delta(\theta^*)) e^{-\delta' n} \right).$$

Le choix de valeurs  $\delta$  et  $\delta'$  telles que  $\delta < \delta' < \beta$  assure alors que la suite  $\mathbf{E}^{\theta^*} [\Pi(A_\varepsilon | X_{0:n})]$  converge vers 0 quand  $n$  tend vers  $+\infty$ . Pour tout  $\varepsilon > 0$ , l'inégalité de Markov donne donc

$$\Pi(A_\varepsilon | X_{0:n}) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^\nu} 0,$$

et la suite des distributions a posteriori est donc faiblement consistante en  $\theta^*$ .

Bien que ces techniques à base de tests proposent une solution aux problèmes soulevés par [Fre63, DF86], elles soulèvent la question de l'existence de familles de tests uniformément exponentiellement consistants. Cette question a déjà reçu diverses réponses dans le cadre i.i.d. . Nous renvoyons par exemple au travaux déjà cités [Sch65, Bar88, BSW99, GGR99] ou [GR03, GvdV17] pour voir des exemples de tests.

Par ailleurs, le cadre i.i.d. pour lequel la consistance est garantie, limite considérablement le cadre d'application de cette théorie. En effet, les liens entre les observations issues de données réelles sont souvent modélisés par des dépendances plus complexes. L'omniprésence de schéma de dépendances Markoviennes a ainsi amené [Ver15] et [DOR19] à s'intéresser aux problèmes de consistances respectivement pour des chaînes de Markov cachées et pour des modèles Markoviens partiellement observés.

Le Chapitre 5 de cette thèse met à profit la méthodologie bayésienne pour le problème d'estimation de loi du milieu de la MAMA i.i.d. sur  $\mathbb{Z}$  : nous y mettons en évidence la consistance a posteriori dans ce modèle qui n'est ni i.i.d., ni Markovien.

### 1.3.3 Consistance a posteriori de l'estimateur Bayésien : Chapitre 5

A l'inverse des travaux [CFL<sup>+</sup>14, DL18] ou de notre propre travail (exposé en Section 1.3.1), nous n'adoptons plus une approche fréquentiste mais bayésienne : la loi du milieu  $\nu$  est maintenant considérée comme aléatoire. Dans ce cadre, l'espace des paramètres  $\Theta$  est l'ensemble des mesures de probabilités sur  $[0, 1]$  absolument continues par rapport à la mesure de Lebesgue. Nous munissons cet ensemble d'une loi a priori  $\Pi$ . Pour tout  $\nu$  dans  $\Theta$ , nous notons  $f_\nu$  la densité de la loi  $\nu$  par rapport à la mesure de Lebesgue. Nous notons  $\text{supp}(\Pi) \subset \Theta$  le support de  $\Pi$  (en particulier  $\Pi(\text{supp}(\Pi)) = 1$ ). Nous considérons les hypothèses suivantes reliées soit aux propriétés de la MAMA exposées en Section 1.2.2 (hypothèses (i) et (ii)), soit aux techniques bayésiennes de preuve esquissées en Section 1.3.2 (hypothèses (iii), (iv), (v) et (vi)) :

- (i) il existe  $c_0 > 0$  tel que pour tout  $\nu$  dans  $\text{supp}(\Pi)$ ,  $\int_0^1 \log(\frac{1-t}{t})\nu(dt) < -c_0$ ,
- (ii) il existe  $\alpha$  dans  $(0, 1/2)$  tel que pour  $\nu$  dans  $\text{supp}(\Pi)$ ,  $\text{supp}(\nu) \subset [\alpha, 1 - \alpha]$ ,
- (iii) il existe  $m$  dans  $\mathbb{R}_+^*$  tel que pour  $\nu$  dans  $\text{supp}(\Pi)$  et  $t$  dans  $[\alpha, 1 - \alpha]$ ,  $f_\nu(t) \geq m$ ,
- (iv) il existe  $M$  dans  $\mathbb{R}_+^*$  tel que pour  $\nu$  dans  $\text{supp}(\Pi)$  et  $t$  dans  $[\alpha, 1 - \alpha]$ ,  $f_\nu(t) \leq M$ ,
- (v)  $\Pi$  attribue une probabilité non nulle à tout voisinage de  $\nu$  pour la norme  $\|\cdot\|_\infty$ ,
- (vi) pour tout  $\varepsilon > 0$ , il existe une famille  $(\Theta_n)_{n \in \mathbb{N}}$  de sous-ensembles mesurables de  $\text{supp}(\Pi)$  et une constante  $C_\varepsilon > 0$  telles que pour tout entier  $n$

$$\Pi(\Theta_n^c) \leq e^{-\frac{C_\varepsilon}{m}n} \quad \text{et} \quad \log(2N(\Theta_n, \|\cdot\|_\infty, \varepsilon)) \leq \frac{\varepsilon^2}{(M-m)^2}n,$$

où  $m$  et  $M$  sont donnés dans (iii) et (iv) et où  $N(\Theta_n, \|\cdot\|_\infty, \varepsilon)$  désigne le nombre minimal de  $\varepsilon$ -recouvrement de  $\Theta_n$  pour  $\|\cdot\|_\infty$  (voir l'Appendice B pour une définition).

D'autre part, reprenant le modèle d'observation adopté par [CFL<sup>+</sup>14, FGL14][FLM14, CFL16, DL18, HLM19], nous supposons disposer d'une trajectoire de la MAMA  $X_{0:T_n}$  jusqu'à ce qu'elle atteigne pour la première fois l'entier positif  $n$ .

L'apport majeur de ce chapitre est d'établir la consistance a posteriori de l'estimateur Bayésien : si la loi a priori  $\Pi$  vérifie les hypothèses (i) à (vi), alors, pour tout  $\varepsilon > 0$

$$\Pi\left(d^{\nu^*}(\nu^*, \nu) > \varepsilon \mid (L_k^n)_{0 \leq k \leq n}\right) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}_0^{\nu^*}} 0,$$

avec la quantité  $d^\nu(\nu', \nu'')$  définie pour tout  $\nu, \nu'$  et  $\nu''$  dans  $\Theta$  par

$$d^\nu(\nu', \nu'') = \mathbf{E}_{\mu^\nu} \left[ \log \left( \frac{\int_0^1 t^{Z_0+1} (1-t)^{Z_1} \nu'(dt)}{\int_0^1 t^{Z_0+1} (1-t)^{Z_1} \nu''(dt)} \right) \right],$$

où le processus canonique  $(Z_t)_{t \in \mathbb{N}}$  et  $\mathbf{Q}_{\mu^\nu}^\nu$  sont définis en Section 1.2.4.

Ce résultat découle en fait d'un résultat semblable pour les processus de branchement en milieu aléatoire grâce à l'égalité en loi entre les nombres de sauts à gauche  $(L_k^n)_{0 \leq k \leq n}$  de la MAMA et les  $(n+1)$  premières générations  $Z_{0:n}$  d'un processus de branchement avec

immigration en milieu aléatoire telle que nous l'avons exposée en Section 1.2.5. En effet, nous montrons au Chapitre 5 un résultat de consistance a posteriori pour les processus de branchement avec immigration en milieu aléatoire (définis en Section 1.2.4) : si le prior  $\Pi$  vérifie les hypothèses (i) à (vi), alors, pour tout  $\varepsilon > 0$

$$\Pi \left( d^{\nu^*}(\nu^*, \nu) > \varepsilon \mid (Z_k)_{0 \leq k \leq n} \right) \xrightarrow[n \rightarrow +\infty]{\mathbf{Q}_0^*} 0.$$

Sous cet aspect, la quantité  $d^{\nu^*}(\nu^*, \nu)$  prend ainsi tout son sens. Grâce à [CFL<sup>+</sup>14, Proposition 4.6], qui étudie le processus de branchement sous l'angle des chaînes de Markov, elle n'est en effet rien d'autre que la limite en probabilité de la différence des versions empiriques de la log-vraisemblance de  $Z_{0:n}$  en  $\nu^*$  et  $\nu$

$$\frac{1}{n} \sum_{i=0}^{n-1} \log(K^{\nu^*}(Z_i, Z_{i+1})) - \frac{1}{n} \sum_{i=0}^{n-1} \log(K^\nu(Z_i, Z_{i+1})) \xrightarrow[n \rightarrow +\infty]{\mathbf{Q}_0^*} d^{\nu^*}(\nu^*, \nu).$$

Les observations  $Z_{0:n}$  n'étant pas i.i.d., il nous est impossible d'appliquer directement la théorie de la consistance a posteriori de [Sch65, Bar88, BSW99, GGR99] pour la preuve d'un tel résultat mais nous nous inspirons des techniques de preuve exposées dans 1.3.2 requérant l'existence de tests uniformément consistants.

La preuve d'un tel résultat repose donc fondamentalement sur l'existence d'une famille  $\phi_n$  de tests de  $H_0 : \nu = \nu^*$  contre  $H_1 : \nu \in U^C$  (où  $U$  voisinage de  $\nu^*$ ) pour le modèle Markovien constitué par les  $(Z_k)_{k \in \mathbb{N}}$  telle qu'il existe  $C_1 > 0$  et  $C_2 > 0$  vérifiant pour tout  $n$  dans  $\mathbb{N}$

$$\mathbf{E}_0^{\nu^*} [\phi_n(Z_{0:n})] \leq C_1 e^{-C_2 n} \quad \text{et} \quad \sup_{\nu \in U^C} \mathbf{E}_0^\nu [1 - \phi_n(Z_{0:n})] \leq C_1 e^{-C_2 n}.$$

De telles garanties sur une famille spécifique de tests  $\phi_n$  sont directement reliées aux propriétés de concentration de la chaîne de Markov  $(Z_k)_{k \in \mathbb{N}}$  et nécessitent des résultats quantitatifs uniformément valables sur toute une classe de noyau Markovien : leur obtention est l'objet du Chapitre 4.

### 1.3.4 Inégalité de concentration pour transformation de chaînes de Markov avec la propriété de différences bornées : Chapitre 4

Les tests considérés au Chapitre 5 sont de la forme  $f(Z_{0:n-1}) > t$  où  $(Z_k)_{k \in \mathbb{N}}$  est un processus de branchement avec immigration en milieu aléatoire et  $t \in \mathbb{R}$  est un seuil. La fonction  $f : \mathbb{N}^n \rightarrow \mathbb{R}$  vérifie la propriété de différences bornées suivante : il existe  $\gamma = \gamma_{0:n-1}$  dans  $\mathbb{R}_+^n$  telle que, pour tous  $x_{0:n-1}$  et  $y_{0:n-1}$  dans  $\mathbb{N}^n$

$$|f(x_{0:n-1}) - f(y_{0:n-1})| \leq \sum_{i=0}^{n-1} \gamma_i \mathbb{1}_{\{x_i \neq y_i\}}.$$

La concentration de telles variables est connue dans le cas de données i.i.d. sous le nom d'inégalité de Mac Diarmid [McD89]. Différentes versions de cette inégalité sont également disponibles pour les chaînes de Markov, voir par exemple [Ada08, Pau15, DG15]. Reprenant la preuve de [DG15], [DMPS18, Theorem 23.3.1] établit, pour les chaînes de Markov  $V$ -géométriquement ergodiques, l'inégalité de concentration exponentielle suivante

$$(1.12) \quad \mathbb{P}_x (|f(X_{0:n-1}) - \mathbb{E}_x[f(X_{0:n-1})]| > t) \leq 2 \exp \left( - \frac{\beta t^2}{\sum_{l=0}^{n-1} \gamma_l^2} \right),$$

où la constante  $\beta$  n'est pas donnée explicitement.

Nous reprenons la preuve de [DMPS18, Theorem 23.3.1] de manière à expliciter la dépendance de  $\beta$  en certaines quantités caractéristiques de la chaîne de Markov. Nous proposons au Chapitre 4 une version quantitative de l'inégalité (1.12). Plus précisément, soit  $P$  un noyau Markovien irréductible et apériodique de distribution stationnaire  $\pi$ . Supposons qu'il existe un ensemble  $C$  et des constantes  $\delta > 1$ ,  $M > 0$ ,  $\zeta_0 > 0$  et  $r_0$  dans  $(0, 1)$  telles que pour tout  $x$  dans  $C$

$$(1.13) \quad \sup_{y \in C} (\mathbb{E}_y[\delta^{\sigma_C}]) \leq M \quad \text{et} \quad d_{TV}(\delta_x P^n, \pi) \leq \zeta_0 r_0^n.$$

Alors, l'inégalité (1.12) est valable pour tout  $x$  dans  $C$  avec une constante  $\beta$  dépendant explicitement uniquement des paramètres  $\delta$ ,  $M$ ,  $\zeta_0$  et  $r_0$ .

Considérons alors la famille de noyaux  $(K^\nu)_{\nu \in \Theta}$  où l'ensemble  $\Theta$  dépendant uniquement de paramètres  $c_0 > 0$  et  $\alpha$  dans  $(0, 1/2)$  est défini par

$$(1.14) \quad \Theta = \left\{ \nu \in \mathbb{M}^1([0, 1]) : \int_0^1 \log\left(\frac{1-t}{t}\right) \nu(dt) \leq -c_0 \quad \text{et} \quad \nu([\alpha, 1-\alpha]) = 1 \right\}.$$

Dans le Chapitre 4, nous obtenons alors une inégalité de type (1.12) qui soit uniforme sur la classe de chaînes de Markov définie par la famille de noyaux  $(K^\nu)_{\nu \in \Theta}$ . Il existe une constante  $\beta(c_0, \alpha)$  ne dépendant que de  $c_0$  et  $\alpha$  telle que, pour tout  $n$  dans  $\mathbb{N}^*$ , pour toute fonction  $f$  ayant la propriété de différences bornées de constante  $\gamma_{0:n-1}$  dans  $\mathbb{R}_+^{*n}$ , pour tout  $\nu$  dans  $\Theta$  et pour tout  $t > 0$

$$\mathbb{P}^\nu (|f(Z_{0:n-1}) - \mathbb{E}[f(Z_{0:n-1})]| > t) \leq 2 \exp\left(-\frac{\beta(c_0, \alpha) t^2}{\sum_{l=0}^{n-1} \gamma_l^2}\right).$$

Pour appliquer ce résultat, il faut alors montrer l'existence d'un ensemble  $C$  et de constantes  $\delta > 1$ ,  $M > 0$ ,  $\zeta_0 > 0$  et  $r_0$  dans  $(0, 1)$  pour lesquels les inégalités (1.13) sont valables pour la classe de chaînes de Markov définie par la famille de noyaux  $(K^\nu)_{\nu \in \Theta}$ . Nous montrons alors que ces hypothèses reliées aux propriétés d'ergodicité de cette famille de chaînes sont vérifiées du moment qu'il existe un moment exponentiel uniformément valable et borné pour cette famille, ce qui sera vérifié au Chapitre 3.

### 1.3.5 Contrôle uniforme de la queue de distribution du temps de retour en 0 du processus de branchement en milieu aléatoire : Chapitre 3

L'étude des queues de distribution du temps  $T$  de retour en 0 du processus de branchement en milieu aléatoire est motivée par l'inégalité de concentration uniforme discutée au Chapitre 4. Cette étude a été initiée par [KKS75]. En particulier, [KKS75, Lemma 1] garantit l'existence de constantes  $K_1 > 0$  et  $K_2 > 0$  vérifiant pour tout  $t > 0$

$$(1.15) \quad \mathbf{Q}^\nu (T > t) \leq K_1 e^{-K_2 t}.$$

On déduit simplement de ce résultat que

$$(1.16) \quad \mathbf{E}^\nu \left[ e^{\frac{K_2}{2} T} \right] \leq 1 + K_1.$$

L'inconvénient de ce résultat pour son application à notre cadre Bayésien est qu'il ne précise pas la dépendance des constantes  $K_1$  et  $K_2$  en fonction de la loi  $\nu$ . L'objet du Chapitre 3 est d'explicitier cette dépendance.

Nous reprenons la preuve de [KKS75, Lemma 1] et montrons l'existence de constantes  $K_1(c_0, \alpha) > 0$  et  $K_2(c_0, \alpha) > 0$  pour  $c_0 > 0$  et  $\alpha$  fixés telles que l'inégalité (1.15) soit vérifiée pour tout processus de branchement avec immigration en milieu aléatoire défini par  $\nu$ , probabilité sur  $[0, 1]$  telle que  $\int_0^1 \log(\frac{1-t}{t})\nu(dt) \leq -c_0$  et  $\nu([\alpha, 1 - \alpha])$ . Pour tout  $t > 0$ ,

$$(1.17) \quad \mathbf{Q}^\nu(T > t) \leq K_1(c_0, \alpha)e^{-K_2(c_0, \alpha)t}.$$

En particulier,

$$\mathbf{E}^\nu \left[ e^{\frac{K_2(c_0, \alpha)}{2}T} \right] \leq 1 + K_1(c_0, \alpha).$$

## 1.4 Conclusion et perspectives

La thèse s'intéresse au problème de l'estimation directe de la loi du milieu d'une marche aléatoire en milieu aléatoire sous divers angles. Ces résultats s'appuient en particulier sur des résultats provenant de divers sous-domaines des probabilités et statistiques.

Reprenant d'abord l'approche fréquentiste, nous fournissons un premier estimateur de la densité de la loi du milieu. Prenant en compte aussi bien la régularité de la densité cible que le régime de la marche, nous obtenons un contrôle du risque de cet estimateur. Ces vitesses de convergence sont sous optimales par rapport à la vitesse minimax d'estimation de la densité dans le cas i.i.d., voir par exemple [Tsy09] pour différents résultats dans ce sens et [BK10] pour une preuve utilisant des estimateurs de moment. Deux questions se posent naturellement. Nous pouvons d'abord nous demander si les contrôles des bornes de risque obtenus sont optimaux pour notre estimateur. Ensuite, étant donné la différence de comportement de l'estimateur entre régimes récurrent et transient, nous pouvons nous interroger sur la possibilité de trouver un unique estimateur qui atteigne non seulement notre vitesse de convergence en régime transient mais aussi retrouve la vitesse minimax du cas i.i.d. en régime récurrent.

Nous proposons également la première étude d'estimateurs bayésiens de la loi du milieu. Pour cela, nous avons établi une inégalité de concentration quantitative de type Mac Diarmid pour des chaînes de Markov ainsi qu'un contrôle quantitatif de la distribution du temps de retour en 0 du PBMA lié à la MAMA.

Un premier axe d'approfondissement serait la relaxation des hypothèses faites dans ce cadre Bayésien. Pour relâcher la condition d'ellipticité, on pourrait comme au Chapitre 2 tenter d'étendre les résultats au cas où la distribution de l'environnement assigne une probabilité faible aux extrémités du segment  $[0, 1]$ .

Un autre axe de recherche serait d'étendre au régime récurrent les résultats obtenus en régime transient. Les résultats obtenus jusqu'ici reposent principalement sur les propriétés d'ergodicité de la chaîne de Markov constituée par le processus des sauts à gauche. Cette chaîne n'étant pas récurrente lorsque la MAMA est récurrente, de nouvelles techniques sont alors à envisager.

Un dernier axe d'étude serait l'étude plus précise des propriétés asymptotiques de l'estimateur Bayésien. La consistance a posteriori étant désormais garantie, l'obtention de la vitesse de convergence en constitue la prochaine étape naturelle.

## Chapter 2

# Nonparametric density estimation of the RWRE

### 2.1 Introduction

Random walks in random environment (RWRE) provide simple models to describe various problems such as the propagation of heat, diffusion of matter through a physical medium, or DNA-unzipping experiments. In these situations, the medium is very irregular and irregularities can be modelled by a random environment. The definition of RWRE involves two ingredients: (i) the environment, which is an i.i.d. sample of some unknown distribution  $\nu$  and (ii) the random-walk whose transition probabilities are determined by the environment. This paper considers the problem of recovering the density  $f$  of the distribution  $\nu$  of the environment of a RWRE on  $\mathbb{Z}$  based on the observation of a single trajectory of the RWRE.

Since their introduction in [Che67], RWRE have been widely studied in the probability literature; see for example [Zei12] for a recent overview of probabilistic results on RWRE in  $\mathbb{Z}$  and more generally in  $\mathbb{Z}^d$ . On the other hand, statistical inference for RWRE has emerged only recently with the appearance of RWRE in several statistical models such as DNA-unzipping experiments or DNA-polymerase phenomenon in [AMJR12, HFR09, BBC<sup>+</sup>07, BBC<sup>+</sup>06, KSJW02]. In these applications, the main task is usually to recover the environment itself and, when it is the case, an estimator of  $\nu$  can be considered as a preliminary step in the construction of an empirical Bayes estimator.

The problem of recovering  $\nu$  was originally considered in [AE04]. They considered random walks in general state spaces and studied an estimator of the moments of the distribution but this estimator had poor statistical performance. In the last few years, [CFLL16, FLM14, FGL14] considered random walks on  $\mathbb{Z}$  and studied the asymptotic behaviour of the maximum likelihood estimator of  $\nu$  in a parametric setting. They proved its consistency and asymptotic normality in the ballistic and sub-ballistic regimes as well as its efficiency in the ballistic regime (see Section 2.2 for details regarding these notions). Consistency results have also been proved in the recurrent regime in [CFLL16] for a slightly different estimator and Markovian environments in [ALM15] have also been considered.

It is shown in [DL18] that the beta-moments of the distribution  $\nu$  can be estimated consistently from a single trajectory of the RWRE. These moment estimators are also used there to obtain an estimator of the cumulative distribution function (c.d.f.) of the random environment. In the present paper, we use these moment estimators to construct estimators of the probability density.

Recovering the probability density of a distribution from its moment is a classical problem in statistics called *moment reconstruction* (see [Akh65]). Density estimators of a distribution based on its beta-moments (also referred to as beta-kernel estimators) have already been studied when an i.i.d. sample  $(Z_1, \dots, Z_n)$  of the unknown distribution is observed (see [Che99, BR03, Mna08] and the references therein). In this direct observation setting, [Che99] estimated  $f$  at  $x \in [0, 1]$  by a single properly chosen  $\beta$ -moment

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{B}^{\frac{x}{h}+1, \frac{1-x}{h}+1}(Z_i),$$

where  $h$  is the smoothing bandwidth and for any  $a$  and  $b$  in  $\mathbb{R}_+^*$ ,  $\mathbb{B}^{a,b}$  is the beta function

$$(2.1) \quad \mathbb{B}^{a,b}(u) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} u^{a-1} (1-u)^{b-1} \mathbb{1}_{0 \leq u \leq 1}.$$

[Che99] showed in particular that this estimator does not have the boundary bias of standard kernel estimator. This beta-kernel density estimator has later been shown to be minimax under regularity assumptions on the density  $f$  in [BK10].

The first contribution of the paper is the construction of the first non-parametric density estimator of the random environment of a RWRE. The density  $f$  of  $\nu$  is approximated at every point  $x \in (0, 1)$  by a sequence of properly chosen beta kernels, following [Che99, Mna08]. These  $\beta$ -moments are estimated using the estimators introduced in [DL18]. The resulting density estimator depends on a regularization parameter. We propose to select automatically this parameter using the so-called *Goldenschluger-Lepski method* of [GL08]. This selection step allows to optimize the risk bounds of the preliminary estimators without prior knowledge of the regime of the walk or the regularity of the unknown density (see Section 2.4 for details). The Goldenschluger-Lepski method was used in [BK14] to build adaptive estimators in the i.i.d. setting from the minimax estimators of [BK10].

The second contribution is the derivation of the first non-parametric risk bounds for any density estimator of  $f$  based on the observation of a single trajectory of a RWRE. These are based on preliminary results obtained in [DL18] for the stochastic part of the risk and results on beta-kernel estimators presented in [BR03, Mna08, BK10] for the deterministic part of the risk. Our rates do not match minimax rates of density estimation in i.i.d. setting. Nevertheless, our results outperform those obtained in [CFLL16] in the recurrent regime where only consistency results for very particular densities were derived. The rates hold without prior knowledge on the regime of the walk contrary to those of [CFLL16].

We finally investigate the numerical behaviour of our estimators in a Monte Carlo experiment. Our estimator behaves reasonably in ballistic regimes where the chain has linear drift. This emphasizes the difference between population and individual parameters estimation. It is clear that recovering the environment itself is impossible in the ballistic regime, yet estimating the law of the random environment is still possible. Related results have already been reported in a parametric setting for example in [FLM14] but it comes perhaps as a surprise that the same behaviour is observed in the non-parametric setting. Both theoretical results and this first simulations consider the case where the chain is observed until it reaches a fixed site  $n$ . Depending on the regime, the number of observed steps of the random walk can therefore be very different, it is typically  $O(n)$  in the ballistic regime and  $e^{O(\sqrt{n})}$  in the recurrent regime. Our simulations also illustrate how the quality of estimation deteriorates when going from nearly recurrent regime to nearly ballistic regime. Finally, we implement

the Goldenshluger-Lepski algorithm and observe graphically how it performs according to variations in regime or regularity.

The paper is organised as follows. Section 2.2 gathers basic results on RWRE on  $\mathbb{Z}$  and on the likelihood. Section 2.3 presents the construction of basic estimators. Section 2.4 states the main results : convergence rates for basic estimators and the estimator selected by the Goldenshluger-Lepski method. Section 2.5 presents Monte Carlo experiment supporting our theoretical claims. Proofs of the main results are gathered in Section 2.6.

## 2.2 Random walks in random environment (RWRE)

Let  $E = (0, 1)^{\mathbb{Z}}$  denote the *set of environments*,  $E$  is endowed with the  $\sigma$ -field  $\mathcal{E} = \mathcal{B}((0, 1))^{\otimes \mathbb{Z}}$ , generated by the cylinders  $\prod_{k \in \mathbb{Z}} B_k$  where every  $B_k$  belongs to  $\mathcal{B}((0, 1))$  and  $B_k \neq (0, 1)$  only for a finite number of  $k$ . Let  $(X_t)_{t \in \mathbb{N}}$  denote the canonical process of the space  $\mathbb{Z}^{\mathbb{Z}^+}$  endowed with the  $\sigma$ -field generated by cylinders. For any  $\omega$  in  $E$ , the random walk in environment  $\omega$  is the time-homogeneous Markov chain with transition probabilities given for all  $x$  and  $y$  in  $\mathbb{Z}$  by

$$p_\omega(x, y) = \begin{cases} \omega_x & \text{if } y = x + 1 \\ 1 - \omega_x & \text{if } y = x - 1 \\ 0 & \text{otherwise} \end{cases} .$$

That is to say, given the environment  $\omega$ , the random walk currently at point  $x$  in  $\mathbb{Z}$ , will make a one-unit step to the right with probability  $\omega_x$  or to the left with probability  $1 - \omega_x$ . Hereafter, the Markov chain is started from 0. The distribution  $\mathbf{P}_\omega$  of the random walk in the environment  $\omega$  is called the *quenched distribution*.

All along the paper, the environment  $\omega = (\omega_x)_{x \in \mathbb{Z}}$  is a sequence of i.i.d. random variables with common distribution  $\nu$ . The *environment*  $\omega$  is distributed according to  $\mathbb{P}^\nu = \nu^{\otimes \mathbb{Z}}$ . Averaging the quenched probability with respect to the environment distribution, we obtain the *annealed distribution*  $\mathbf{P}^\nu$  on  $(2^{\mathbb{Z}})^{\otimes \mathbb{N}}$  defined by

$$\mathbf{P}^\nu(\cdot) = \int_E \mathbf{P}_\omega(\cdot) \mathbb{P}^\nu(d\omega) .$$

Expectation with respect to the annealed probability measure  $\mathbf{P}^\nu$  are denoted  $\mathbf{E}^\nu$ . The random walk  $(X_t)_{t \in \mathbb{N}}$  is a Markov chain only conditionally on the environment (i.e. with respect to the quenched distribution  $\mathbf{P}_\omega$ ) but the Markov property fails under the annealed probability measure  $\mathbf{P}^\nu$ . The past history provides information on the environment and can not be neglected. The random walk in i.i.d. random environment on  $\mathbb{Z}$  is the random sequence  $(X_t)_{t \in \mathbb{N}}$  considered under annealed distribution  $\mathbf{P}^\nu$ .

The asymptotic behaviour of the walk  $(X_t)_{t \in \mathbb{N}}$  depends on the distribution of the ratio

$$(2.2) \quad \rho_0 = \frac{1 - \omega_0}{\omega_0} .$$

If  $\mathbb{E}^\nu [|\log \rho_0|]$  is finite, [Sol75] proved the following classification:

- (i) if  $\mathbb{E}^\nu [\log \rho_0] \neq 0$ , then  $(X_t)_{t \in \mathbb{N}}$  is transient ( $\mathbf{P}^\nu$ -a.s.).  
Moreover if  $\mathbb{E}^\nu [\log \rho_0] < 0$ , then  $\lim_{t \rightarrow +\infty} X_t = +\infty$ ,  $\mathbf{P}^\nu$ -a.s., the process  $(X_t)_{t \in \mathbb{N}}$  is transient to the right,



(ii) if  $\mathbb{E}^\nu [\log \rho_0] = 0$ , then  $(X_t)_{t \in \mathbb{N}}$  is recurrent.

Moreover,  $\limsup_{t \rightarrow +\infty} X_t = +\infty$  and  $\liminf_{t \rightarrow +\infty} X_t = -\infty$ ,  $\mathbf{P}^\nu$ -a.s. .

In the transient case, the random walk  $X_t \rightarrow +\infty$ , with asymptotic rates proved by [Sol75, KKS75]. Let  $T_n$  denote the first hitting time of  $n$  in  $\mathbb{N}$ ,

$$(2.3) \quad T_n = \inf \{t \in \mathbb{N}, X_t = n\} .$$

Assuming that that the RWRE is transient  $\mathbb{E}^\nu [\log \rho_0] < 0$ , the following classification holds.

1. If  $\mathbb{E}^\nu [\rho_0] < 1$ , then  $\mathbf{P}^\nu$ -a.s.,

$$\frac{T_n}{n} \rightarrow \frac{1 + \mathbb{E}^\nu [\rho_0]}{1 - \mathbb{E}^\nu [\rho_0]} .$$

The RWRE is called *ballistic*.

2. If  $\mathbb{E}^\nu [\rho_0] \geq 1$ , then  $\mathbf{P}^\nu$ -a.s.,  $T_n/n \rightarrow +\infty$  and the RWRE is called sub-ballistic.

The fluctuations of  $T_n$  may be characterized more precisely. Suppose that the distribution of  $\log \rho_0$  is non arithmetic (that is the group generated by the support of  $\log \rho_0$  is dense in  $\mathbb{R}$ ) and that there exists  $\kappa$  in  $(0, \infty)$  such that

$$(2.4) \quad \mathbb{E}^\nu [\rho_0^\kappa] = 1 \quad \text{and} \quad \mathbb{E}^\nu [\rho_0^\kappa \log^+(\rho_0)] < \infty ,$$

where  $\log^+(x) = \log(x \vee 1)$ . A simple convexity argument shows that if  $\kappa$  exists then it is unique. The value of  $\kappa$  determines the asymptotic behaviour of  $(X_t)_{t \in \mathbb{N}}$ . It follows from [KKS75] that

1. if  $\kappa < 1$ ,  $T_n/n^{1/\kappa}$  and  $X_t/t^\kappa$  converge in  $\mathbf{P}^\nu$ -distribution to a non trivial distribution,
2. if  $\kappa = 1$ ,  $T_n/(n \log n)$  and  $(\log t/t)X_t$  converge in  $\mathbf{P}^\nu$ -probability to positive constants,
3. if  $\kappa > 1$ ,  $T_n/n$  and  $X_t/t$  converge in  $\mathbf{P}^\nu$ -probability to positive constants.

In cases (1) and (2), the random walk is sub-ballistic and in the last case,  $(X_t)_{t \in \mathbb{N}}$  is ballistic, both  $T_n$  and  $X_t$  grow linearly asymptotically.

When the RWRE is recurrent, the fluctuations of  $(X_t)_{t \in \mathbb{N}}$  have been evaluated by [Sin82]: suppose that  $\mathbb{E}^\nu [\log \rho_0] = 0$ ,  $\mathbb{E}^\nu [\log^2 \rho_0] > 0$  and that the support of  $\rho_0$  is included in  $(0, 1)$ , then  $X_t/(\log t)^2$  converges in  $\mathbf{P}^\nu$ -distribution to a non trivial limit (see also [Zei12] for some extensions under relaxed versions of this assumption).

Our results hold under the assumptions of [DL18], that is those of [KKS75], in the transient case, and a slightly weaker version of those of [Sin82] in the recurrent case. Consider the following assumptions:

**H1.**

$$\begin{cases} \mathbb{E}^\nu [\log \rho_0] = 0 \quad \text{and} \quad \mathbb{E}^\nu [\log^2 \rho_0] > 0 \\ \text{there exists } a > 0, \text{ such that } \mathbb{E}^\nu [\rho_0^a] + \mathbb{E}^\nu [\rho_0^{-a}] < \infty \end{cases} .$$

**H2.**

$$\begin{cases} \mathbb{E}^\nu [\log \rho_0] < 0 \\ \text{the distribution of } \log \rho_0 \text{ is non arithmetic} \\ \text{there exists } \kappa \in (0, \infty) \text{ such that } \mathbb{E}^\nu [\rho_0^\kappa] = 1 \text{ and } \mathbb{E}^\nu [\rho_0^\kappa \log^+(\rho_0)] < \infty \end{cases} .$$

Under **H1**,  $(X_t)_{t \in \mathbb{N}}$  is recurrent. Under **H2**,  $(X_t)_{t \in \mathbb{N}}$  is transient to the right. In both cases, the hitting time  $T_n$  is almost surely finite for any  $n$  in  $\mathbb{N}$ . Hereafter, the RWRE is observed until it hits the state  $n$ . The observations are given by  $X_0, \dots, X_{T_n}$ .

Following [CFL<sup>+</sup>14], for any positive integer  $t_0$  and  $y$  in  $\mathbb{Z}$ , define by  $L(t_0, y)$  and  $R(t_0, y)$  the number of steps to the left and to the right until time  $t_0$  and from site  $y$

$$\begin{aligned} L(t_0, y) &= \sum_{0 \leq t \leq t_0 - 1} \mathbb{1}_{\{X_t = y, X_{t+1} = y-1\}}, \\ R(t_0, y) &= \sum_{0 \leq t \leq t_0 - 1} \mathbb{1}_{\{X_t = y, X_{t+1} = y+1\}}. \end{aligned}$$

The likelihood  $\mathcal{L}_\nu(X_0, \dots, X_{T_n})$  of the observations (where  $T_n$  is defined in Equation (2.3)) can be expressed as follows (see [CFL<sup>+</sup>14] for details)

$$\int \left( \prod_{y \in \mathbb{Z}} \omega_y^{R(T_n, y)} (1 - \omega_y)^{L(T_n, y)} \right) \mathbb{P}^\nu(d\omega) = \prod_{y \in \mathbb{Z}} \int_0^1 a^{R(T_n, y)} (1 - a)^{L(T_n, y)} \nu(da).$$

Now, since  $T_n$  is the hitting time of the state  $n$ , for any  $y \geq n$ ,  $L(T_n, y) = 0$ , and

$$L(T_n, y + 1) = \begin{cases} R(T_n, y), & \text{for all } y < 0 \\ R(T_n, y) - 1, & \text{for all } y \in [0, n - 1] \end{cases}.$$

Hence the likelihood may be written

$$\mathcal{L}_\nu(X_0, \dots, X_{T_n}) = \prod_{y \leq n-1} \int_0^1 a^{L(T_n, y+1) + \mathbb{1}_{\{y \geq 0\}}} (1 - a)^{L(T_n, y)} \nu(da).$$

The collection  $(L(T_n, y))_{y \leq n}$  is a sufficient statistic that will be the building block of the estimation method. An important result of [KKS75], central in the analysis of the parametric case derived in [CFL<sup>+</sup>14] is that, under the annealed distribution  $\mathbf{P}^\nu$ , the process

$$(Z_y^n)_{0 \leq y \leq n} = (L(T_n, n - y))_{0 \leq y \leq n},$$

has the same distribution as  $(Z_y)_{0 \leq y \leq n}$  where  $(Z_y)_{y \in \mathbb{N}}$  is a branching process with immigration in random environment (BPIRE) defined for any  $y$  in  $\mathbb{N}^*$  by

$$(2.5) \quad Z_0 = 0 \quad \text{and} \quad Z_y = \sum_{i=0}^{Z_{y-1}} \xi_{y,i},$$

where  $(\xi_{y,i})_{(y,i) \in \mathbb{N}^* \times \mathbb{N}}$  is a family of geometrically distributed independent random variables, i.e. for any  $y$  in  $\mathbb{N}^*$ , any  $i$  in  $\mathbb{N}$  and any  $k$  in  $\mathbb{N}$ ,

$$\mathbf{P}_\omega(\xi_{y,i} = k) = \omega_y (1 - \omega_y)^k.$$

Under the annealed law  $\mathbf{P}^\nu$ , [CFL<sup>+</sup>14, Proposition 4.3] ensures that  $(Z_x)_{x \in \mathbb{N}}$  is an homogeneous Markov chain starting at 0 with transition kernel given for any  $i$  and  $j$  in  $\mathbb{N}$

$$(2.6) \quad K^\nu(i, j) = \binom{i+j}{i} \int_0^1 a^{i+1} (1 - a)^j \nu(da).$$

## 2.3 Estimator construction

Hereafter, the distribution  $\nu$  is assumed absolutely continuous with respect to the Lebesgue measure and  $f$  denotes its density. Let  $\beta > 0$ ,  $L > 0$  and  $m = \sup \{\ell \in \mathbb{N} : \ell < \beta\}$ . The set  $\Sigma(\beta, L)$  is the set of  $m$  times differentiable functions  $g : [0, 1] \rightarrow \mathbb{R}_+$  satisfying for any  $x$  and  $x'$  in  $[0, 1]^2$

$$(2.7) \quad \left| g^{(m)}(x) - g^{(m)}(x') \right| \leq L |x - x'|^{\beta-m} .$$

Equation (2.6) suggests to use the moments of  $\nu$  to build estimators. Recovering a density function from its moments is a classical instance of the Hausdorff moment problem (see [Akh65]). For any  $j$  in  $\mathbb{N}$ , let  $\mu_j[\nu]$  denote the  $j$ -th moment of the distribution  $\nu$ ,

$$\mu_j[\nu] = \int_0^1 t^j \nu(dt) = \int_0^1 t^j f(t) dt .$$

The moment problem is said to have unique solution if for two distributions  $\nu$  and  $\nu'$  over  $(0, 1)$ , the equations  $\mu_j[\nu] = \mu_j[\nu']$  for any positive integer  $j$  imply that  $\nu = \nu'$ . A distribution  $\nu$  satisfying this property is called *moment determinate*.  $\nu$  is moment determinate if it has a continuous density; see [MR03, Mna08].

When  $\nu$  is moment determinate, the operator mapping the c.d.f. of  $\nu$  into  $(\mu_j[\nu])_{j \in \mathbb{N}}$  is invertible. Given  $M$  in  $\mathbb{N}$ , this inverse is approximated by

$$(2.8) \quad \nu_M([0, x]) = \sum_{k=0}^{\lfloor Mx \rfloor} \sum_{j=k}^M \binom{M}{j} \binom{j}{k} (-1)^{j-k} \mu_j[\nu] .$$

[Mna08] showed that  $\lim_{M \rightarrow \infty} \nu_M([0, x]) = \nu([0, x])$  at any single point  $x$  in  $(0, 1)$  where the c.d.f. is continuous. If  $\nu$  has a density  $f$ , we consider then the following approximation, defined for any  $x$  in  $[0, 1]$  by

$$(2.9) \quad f^M(x) = (M+1) (\nu_M([0, x]) - \nu_M([0, x - M^{-1}])) .$$

[Mna08, Theorem 1] ensures that, when  $f$  is continuous,  $f^M$  converges uniformly to  $f$ . Lemma 8 shows moreover that, if  $f$  is in  $\Sigma(\beta, L)$  with  $\beta$  in  $(0, 2]$ , then, for any  $x$  in  $[0, 1]$

$$|f(x) - f^M(x)| \leq C_{\beta, L} M^{-\beta/2} .$$

In the following, we do not use the expression given in Equation (2.8) to construct our estimator. For any  $a$  and  $b$  in  $\mathbb{R}_+^*$ , denote by  $m^{a,b}$  the following beta-moments

$$(2.10) \quad m^{a,b} = \mathbb{E}^\nu \left[ \omega_0^a (1 - \omega_0)^b \right] = \int_0^1 u^a (1 - u)^b f(u) du .$$

**Proposition 1.** *For any  $M$  in  $\mathbb{N}$  and any  $x$  in  $[0, 1]$ , following identities hold*

$$(2.11) \quad f^M(x) = \int_0^1 \mathbb{B}^{\lfloor Mx \rfloor + 1, M - \lfloor Mx \rfloor + 1}(u) f(u) du$$

$$(2.12) \quad = (M+1) \binom{M}{\lfloor Mx \rfloor} m^{\lfloor Mx \rfloor, M - \lfloor Mx \rfloor} ,$$

where  $\mathbb{B}^{a,b}$  is the beta-kernel defined in Equation (2.1).

For sake of completeness, the proof of Proposition 1 which follows essentially [Mna08, Eq (12)] is given in Section 2.6.1. When  $a$  and  $b$  are integers,  $m^{a,b}$  can be estimated as in [DL18]. Using the convention  $0/0 = 0$ , denote, for all  $i$  and  $j$  in  $\mathbb{N}$

$$(2.13) \quad \phi^{a,b}(i, j) = \frac{\binom{i+j-(a+b)}{i-a}}{\binom{i+j}{i}} \mathbb{1}_{\{i \geq a, j \geq b\}} \quad \text{and} \quad N_n^a = \sum_{k=1}^n \mathbb{1}_{\{Z_{k-1}^n \geq a\}},$$

$$(2.14) \quad \widehat{m}_n^{a,b} = \frac{1}{N_n^a} \sum_{k=1}^n \phi^{a,b}(Z_{k-1}^n, Z_k^n),$$

where  $(Z_k^n)$  is defined in Equation (2.5). Therefore, an estimator of  $f(x)$  might be obtained by plugging an estimator of the beta moment in Equation (2.12) for any  $x$  in  $[0, 1]$

$$(2.15) \quad \widehat{f}_n^M(x) = (M+1) \binom{M}{\lfloor Mx \rfloor} \widehat{m}_n^{\lfloor Mx \rfloor, M - \lfloor Mx \rfloor}.$$

## 2.4 Main results

Let us first derive convergence rates of the basic estimators  $\widehat{f}_n^M$  for fixed values of the regularization parameter  $M$ .

**Theorem 2.** *Assume that the distribution  $\nu$  of the random environment has a density  $f$  in  $\Sigma(\beta, L)$  for some  $\beta$  in  $(0, 2]$  and  $L \geq 1$ . Let  $M \geq 1$  and  $\widehat{f}_n^M$  denote the estimator in Equation (2.15).*

(i) *Assume **H1**. For any  $M \geq 1$ , there exists a constant  $C_\nu$  such that*

$$(2.16) \quad \mathbf{E}^\nu \left[ \left\| f - \widehat{f}_n^M \right\|_\infty \right] \leq C_\nu \left( LM^{-\beta/2} + M \frac{\log n}{\sqrt{n}} \right),$$

(ii) *Assume **H2**. For any  $M \geq 1$*

$$(2.17) \quad \mathbf{E}^\nu \left[ \left\| f - \widehat{f}_n^M \right\|_\infty \right] \leq C_\nu \left( LM^{-\beta/2} + M^{1+\kappa} \sqrt{\frac{\log n}{n}} \right),$$

where  $\kappa$  is defined in **H2**.

Theorem 2 provides the first rates of convergence for non-parametric density estimation of the random environment of a RWRE. An important feature is that these rates are achieved by the same estimator in the recurrent and transient regimes. On the other hand, the value  $M_n$  optimizing these rates highly depends on both the regime of the walk and the regularity  $\beta$  of  $f$  that are both unknown in practice.

**Corollary 3.** (i) *Assume **H1**. Then, taking  $M_n = \lfloor (\sqrt{n}/\log(n))^{2/(\beta+2)} \rfloor$ ,*

$$\mathbf{E}^\nu \left[ \left\| f - \widehat{f}_n^{M_n} \right\|_\infty \right] \leq C_\nu \left( \frac{\log(n)}{\sqrt{n}} \right)^{\beta/(\beta+2)}.$$

(ii) *Assume **H2**. Then, taking  $M_n = \lfloor (n/\log(n))^{\frac{1}{\beta+2(1+\kappa)}} \rfloor$ ,*

$$\mathbf{E}^\nu \left[ \left\| f - \widehat{f}_n^{M_n} \right\|_\infty \right] \leq C_\nu \left( \frac{\log(n)}{n} \right)^{\frac{\beta}{2(\beta+2(1+\kappa))}}.$$

In the following, we propose a method to obtain an estimator simultaneously adaptive to the regime (recurrence / transience) and the regularity of the density of the random environment. We apply the Goldenshluger-Lepski strategy of [GL08] to obtain a data-driven choice of  $\widehat{M}_n$ . In the proof of Theorem 2, we build confidence regions for  $f$ . More precisely, Corollary 10 shows the existence of an event  $\Omega_0$  defined by

$$(2.18) \quad \Omega_0 = \bigcap_{M=1}^{n-1} \left\{ \left\| f^M - \widehat{f}_n^M \right\|_\infty \leq \frac{M+1}{N_n^M} \sqrt{\frac{n}{2} (\log(2n^3) + \log n)} \right\},$$

with probability larger than  $1 - n^{-2}$  on which for all  $M$  in  $\{1, \dots, n-1\}$ ,

$$(2.19) \quad \left\| f - \widehat{f}_n^M \right\|_\infty \leq 31LM^{-\beta/2} + \frac{M+1}{N_n^M} \sqrt{3n \log n}.$$

Define, for any  $M$  in  $\{1, \dots, n-1\}$ ,

$$B(M) = 31LM^{-\beta/2}, \quad V_n(M) = \frac{M+1}{N_n^M} \sqrt{3n \log n},$$

$$C_n(M) = \sup_{M' \in \{1, \dots, n-1\}} \left\{ \left\| \widehat{f}_n^M - \widehat{f}_n^{M \vee M'} \right\|_\infty - 2V_n(M') \right\}.$$

Goldenshluger-Lepski's choice of  $M$  is defined by

$$(2.20) \quad \widehat{M}_n \in \underset{M \in \{1, \dots, n-1\}}{\operatorname{argmin}} \{C_n(M) + 2V_n(M)\}.$$

The estimator  $\widehat{M}_n$  satisfies the following lemma.

**Lemma 4.** *On the event  $\Omega_0$ , for any  $M$  in  $\{1, \dots, n-1\}$*

$$\left\| f - \widehat{f}_n^{\widehat{M}_n} \right\|_\infty \leq 5B(M) + 5V_n(M).$$

Using Lemma 4 together with Equation (2.19) yields the following Theorem.

**Theorem 5.** *Let  $\beta$  in  $(0, 2]$  and  $L \geq 1$ . Assume that  $f$  lies in  $\Sigma(\beta, L)$ .*

(i) *Assume **H1**. Then, there exists a constant  $C_\nu$  such that*

$$(2.21) \quad \mathbf{P}^\nu \left( \left\| f - \widehat{f}_n^{\widehat{M}_n} \right\|_\infty \leq C_\nu \left( \frac{\log(n)}{\sqrt{n}} \right)^{\beta/(\beta+2)} \right) \geq 1 - C_\nu \frac{\log n}{\sqrt{n}}.$$

(ii) *Assume **H2**. Then, there exists a constant  $C_\nu$  such that*

$$\mathbf{E}^\nu \left[ \left\| f - \widehat{f}_n^{\widehat{M}_n} \right\|_\infty \right] \leq C_\nu \left( \frac{\log(n)}{n} \right)^{\frac{\beta}{2(\beta+2(1+\kappa))}}.$$

The estimator  $\widehat{f}_n^{\widehat{M}_n}$  achieves simultaneously in every regime the optimal rates detailed in Corollary 3. It is in this sense adaptive to the regime and the regularity of  $f$ . The rates do not match those one could reach if the environment  $\omega$  is observed (as shown by [BK10]). When the chain is recurrent, [DL18, Lemma 10] ensures that it visits typically more than

$n$  times almost all sites. Therefore, one can probably estimate the environment itself and plug these estimators in any standard density estimator. This might improve the rates in the recurrent regime but it would certainly provide a poor estimator in transient regimes. The problem of recovering the distribution of the environment is in general harder if one observes only one trajectory of the walk than if the environment is observed directly. It is not clear whether a single estimator can recover minimax rates of the i.i.d. setting in the recurrent regime and our rates in transient regimes.

The result holds in expectation in transient regimes and only with large probability in the recurrent regime. The reason is that the term  $N_n^M$  cannot be handled in the same way in transient regimes where the Markov chain  $(Z_k)$  admits a unique invariant probability and in recurrent regimes where it does not. However, a simple modification of these estimators can be proved to have the same rates of convergence in expectation in both regimes as shown by the following result.

**Proposition 6.** *Let  $\beta$  in  $(0, 2]$  and  $L \geq 1$ . Assume that  $f$  lies in  $\Sigma(\beta, L)$ . Let  $\tilde{f}_n = \min(\widehat{f}_n^{M_n}, \log n)$ .*

(i) *Assume **H1**. Then, there exists a constant  $C_\nu$  such that*

$$\mathbf{E}^\nu \left[ \left\| f - \tilde{f}_n \right\|_\infty \right] \leq C_\nu \left( \frac{\log(n)}{\sqrt{n}} \right)^{\beta/(\beta+2)}.$$

(ii) *Assume **H2**. Then, there exists a constant  $C_\nu$  such that*

$$\mathbf{E}^\nu \left[ \left\| f - \tilde{f}_n \right\|_\infty \right] \leq C_\nu \left( \frac{\log(n)}{n} \right)^{\frac{\beta}{2(\beta+2(1+\kappa))}}.$$

The proof of Proposition 6 is given in Section 2.6.5.

## 2.5 Simulation Study

In this section, we illustrate our theoretical results. We consider RWRE in different regimes (recurrent, transient ballistic and sub-ballistic). We also investigate the impact of the regularity of the density  $f$ . The number of steps of the random walk that are observed depends on the regime. It follows from our discussion in Section 2.2, that if the RWRE is transient to the right ( $\mathbb{E}^\nu[\log \rho_0] < 0$  where  $\rho_0$  is defined in Equation (2.2)) then either

1.  $T_n/n \rightarrow (1 + \mathbb{E}^\nu[\rho_0])/(1 - \mathbb{E}^\nu[\rho_0])$   $\mathbf{P}^\nu$ -a.s. in the ballistic regime  $\mathbb{E}^\nu[\rho_0] < 1$ , in which case the number of time steps grows linearly with  $n$ ,
2.  $T_n/n \rightarrow \infty$  converges  $\mathbf{P}^\nu$ -a.s. in the sub-ballistic regime  $\mathbb{E}^\nu[\rho_0] \geq 1$ , showing that the number of time steps grows superlinearly with  $n$ . In the sub-ballistic case, the fluctuations of  $T_n$  are controlled by the parameter  $\kappa$  defined in Equation (2.4).

We compare the behaviour of our estimators built with different values of the parameter  $M$  in three sub-ballistic regimes, with different  $\kappa$ . The estimator based on Goldenshluger-Lepski choice of  $M$ , analysed in Theorem 5 is also implemented.

### 2.5.1 Influence of the regularity

In this section, we consider 3 different distributions for the environment.

- The first example is the Beta distribution  $B(3,3)$  whose density  $\mathbb{B}^{3,3}$  is defined in Equation (2.1).
- The second example is the distribution with density

$$x \mapsto 4 \left( 1 - \frac{|2x - 0.6|}{0.15} \right) \mathbb{1}_{[0.225, 0.325]}(x) + \frac{14}{3} \left( 1 - \frac{|2x - 1.2|}{0.3} \right) \mathbb{1}_{[0.45, 0.75]}(x).$$

- The last example is the distribution with density

$$x \mapsto 0.6\mathbb{B}^{3,3}(2x)\mathbb{1}_{[0,0.5]}(x) + 1.4\mathbb{B}^{3,3}(2x-1)\mathbb{1}_{(0.5,1]}(x).$$

These densities are plotted in black in Figure 2.1. In the first example, the chain is recurrent, in the second and the third, it is sub-ballistic with respective values of  $\kappa$  0.17 and 0.57.

For each distribution of the environment, 100 trajectories of the RWRE have been simulated until the chain reaches 100.

In each case, our (piecewise constant) estimator  $\widehat{f}_n^M$ , defined in Equation (2.15), built for  $M$  in  $\{25, 50, 75\}$  is compared with the actual density  $f$ . More precisely, Figure 2.1 shows for each considered environment and for  $M$  in  $\{25, 50, 75\}$  a graphic representation of the true density (solid black line) and in each point  $x$  in  $(0, 1)$  the median of the estimator  $\widehat{f}_n^M(x)$  (dash-dotted blue line), its interquartile range (greyed area) and its lower and upper hinges (dotted blue line).

According to our theoretical bounds, the estimator  $\widehat{f}_n^M$  performs better when the true density is regular as in Examples 1 and 3 than in the non regular Example 2. The estimator  $\widehat{f}_n^M$ , based on a polynomial approximation, cannot properly approximate the two triangles shaped density of Example 2. More precisely, the regularization parameter  $M$  should probably be much larger for the bias of the estimator to be small.

### 2.5.2 Influence of the regime

Consider trajectories from RWRE that all are transient to the right and sub-ballistic. The environment is drawn from the distribution with density

$$(2.22) \quad x \mapsto 0.8\mathbb{B}^{3,3}(0.5 + 2(x - c_1)) + 1.2\mathbb{B}^{3,3}(0.5 + 2(x - c_2)),$$

and for three sets of the parameters  $c_1$  and  $c_2$  such that the parameter  $\kappa$  associated to the sub-ballistic RWRE takes values near the extreme values of this regime. In the following table, the chosen sets of parameters are provided with the corresponding (numerically approximated) values of  $\kappa$ .

$\kappa$	0.04	0.35	1
$c_1$	0.27	0.3	0.38
$c_2$	0.67	0.7	0.7

For each set of parameters for the environment distribution, 100 trajectories until the RWRE reaches  $n = 100$  have been generated.

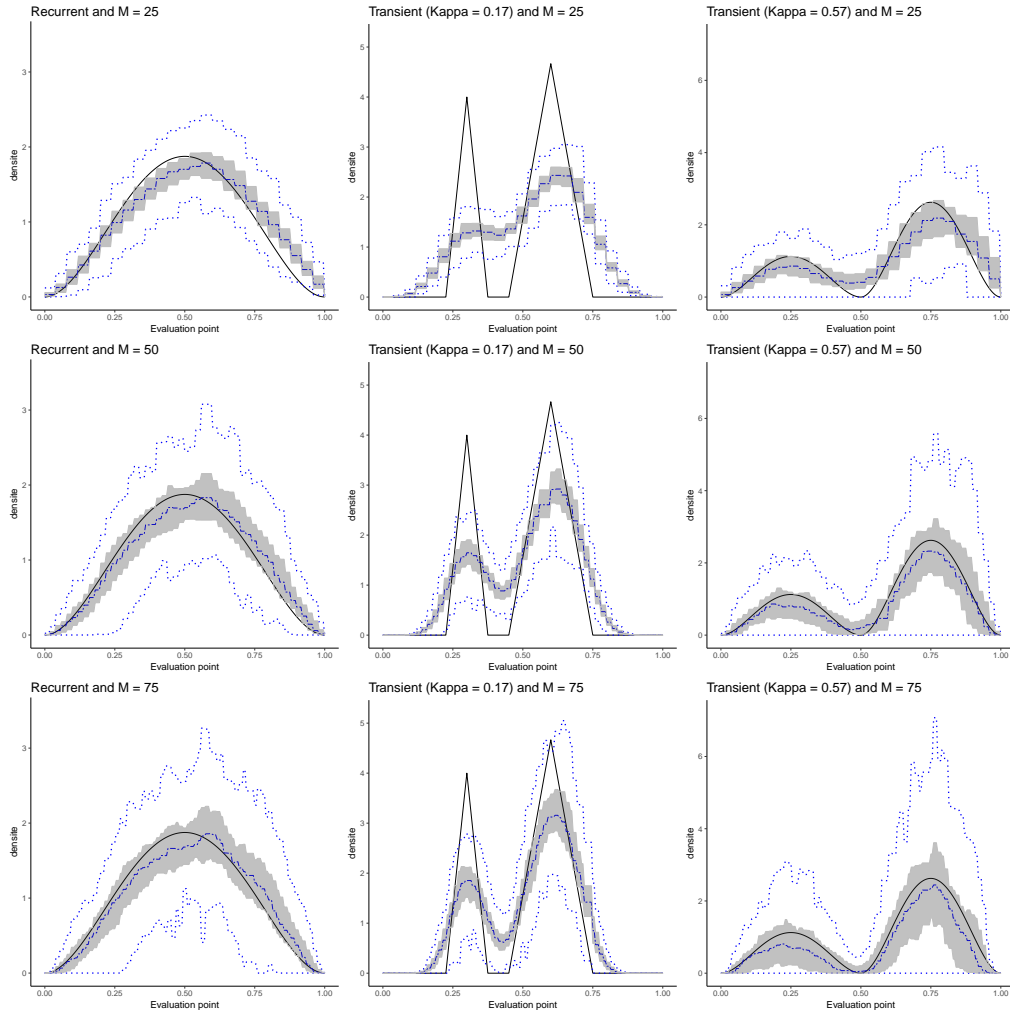


Figure 2.1: True density (solid black line) and in each point  $x$  in  $(0, 1)$  the median (dash-dotted black line), the interquartile range (greyed area) and the lower and upper hinges (dotted black line) of the estimator  $\hat{f}_{100}^M$  for the three considered environment (one environment per column) and for  $M$  in  $\{25, 50, 75\}$ .



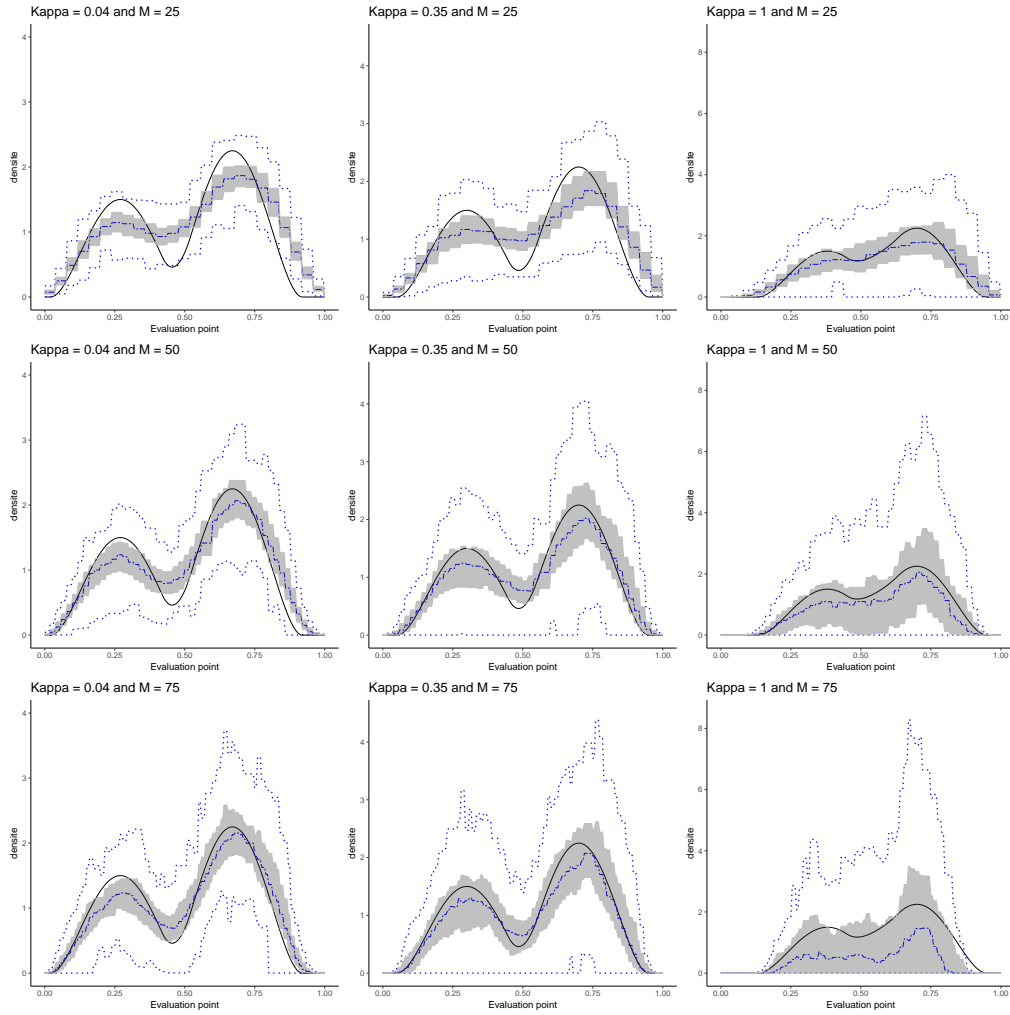


Figure 2.2: True density (solid black line) and in each point  $x$  in  $(0, 1)$  the median (dash-dotted black line), the interquartile range (greyed area) and the lower and upper hinges (dotted black line) of the estimator  $\hat{f}_{100}^M$  for three different values of parameter  $M$  (25, 50 and 75) and of  $\kappa$  (0.04, 0.35 and 1).

Figure 2.2 illustrates the variability of the estimator  $\hat{f}_{100}^M$ . For the densities defined in Equation (2.22) with  $\kappa$  in  $\{0.04, 0.35, 1\}$ , we plot the true density of the environment as well as the interquartile range and lower and upper hinges of  $\hat{f}_{100}^M$  computed over the 100 trajectories and for  $M$  in  $\{25, 50, 75\}$ .

As expected from our theoretical results, the variability of the estimator deteriorates with  $\kappa$ . When  $\kappa = 0.04$  or  $\kappa = 0.35$ , the estimator shows better bias and worse variability when  $M$  grows. The situation in the nearly ballistic case ( $\kappa = 1$ ) is different : the walk reaches faster  $n = 100$  and the trajectory is thus shorter, one single trajectory reaching  $n = 100$  does not seem sufficient to obtain a reliable estimator of  $f$ .

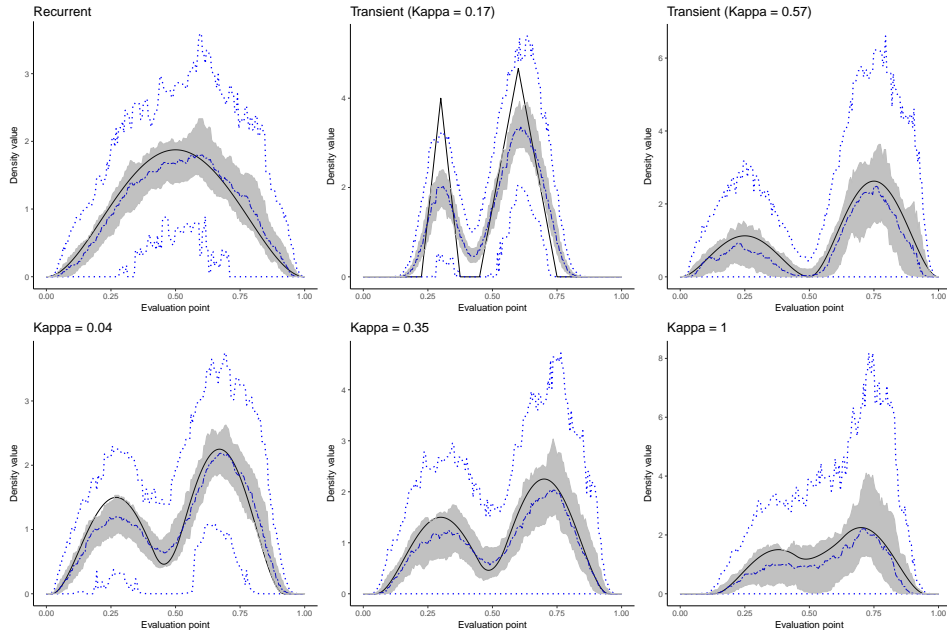


Figure 2.3: True density (solid black line) and in each point  $x$  in  $(0, 1)$  the median (dash-dotted black line), the interquartile range (greyed area) and the lower and upper hinges (dotted black line) of the estimator  $\widehat{f}_n^{\widehat{M}_n}$  for the three environment considered in 2.5.1 (first row) and for the three environment with different  $\kappa$  considered in 2.5.2 (second row) where  $\widehat{M}_n$  was selected by Goldenshluger-Lepski algorithm with  $V_n(M) = (M + 1)/N_n^M \sqrt{0.01 n \log n}$ .

### 2.5.3 Goldenshluger-Lepski estimator

We considered here the distributions of Subsections 2.5.1 and 2.5.2 and used anew the set of trajectories.

Figure 2.3 illustrates the statistical performance of the estimator  $\widehat{f}_n^{\widehat{M}_n}$  obtained by Goldenshluger-Lepski under different regularity and regime settings. For the densities defined in Subsection 2.5.1 and those defined in Equation (2.22), we plot the true density of the environment as well as the interquartile range and lower and upper hinges of  $\widehat{f}_n^{\widehat{M}_n}$  computed over the 100 trajectories.

For each trajectory, we first computed the estimator  $\widehat{f}_n^M$  for  $M$  ranging from 1 to 100, then selected  $\widehat{M}_n$  among those  $M$  through the Goldenshluger-Lepski algorithm as defined in Equation (2.20) and finally built the estimator  $\widehat{f}_n^{\widehat{M}_n}$ . The actual choice of  $V(M)$  provided in the theoretical analysis yields very conservative choice of  $\widehat{M}_n$ . In the simulations, we perform the Goldenshluger-Lepski step with

$$V_n(M) = (M + 1)\sqrt{0.01 n \log n}/N_n^M.$$

As expected, approximation of the true density and variability of the estimator deteriorates when regularity decreases or when the walk goes faster to  $n = 100$ . Nevertheless, graphical comparison of Figure 2.3 and Figure 2.2 shows that the estimator  $\widehat{f}_n^{\widehat{M}_n}$  has smaller bias and variability than the estimator  $\widehat{f}_n^M$  where  $M$  was chosen arbitrarily. Data-driven choice of the value of  $M$  thus performs the expected trade-off between bias and variance.

## 2.6 Proof

### 2.6.1 Proof of Proposition 1

By definition, for any  $M$  in  $\mathbb{N}^*$  and  $x$  in  $[0, 1]$ ,

$$\begin{aligned}
f^M(x) &= (M+1) \sum_{j=\lfloor Mx \rfloor}^M \binom{M}{j} \binom{j}{\lfloor Mx \rfloor} (-1)^{j-\lfloor Mx \rfloor} \mu_j[\nu] \\
&= (M+1) \sum_{j=0}^{M-\lfloor Mx \rfloor} \binom{M}{j+\lfloor Mx \rfloor} \binom{j+\lfloor Mx \rfloor}{\lfloor Mx \rfloor} (-1)^j \mu_{j+\lfloor Mx \rfloor}[\nu] \\
&= (M+1) \frac{M!}{\lfloor Mx \rfloor!} \sum_{j=0}^{M-\lfloor Mx \rfloor} \frac{(-1)^j}{(M-\lfloor Mx \rfloor-j)!j!} \mu_{j+\lfloor Mx \rfloor}[\nu] \\
&= \frac{(M+1)!}{\lfloor Mx \rfloor!(M-\lfloor Mx \rfloor)!} \int_0^1 t^{\lfloor Mx \rfloor} \sum_{j=0}^{M-\lfloor Mx \rfloor} \frac{(M-\lfloor Mx \rfloor)!}{(M-\lfloor Mx \rfloor-j)!j!} (-t)^j 1^{M-\lfloor Mx \rfloor-j} f(t) dt.
\end{aligned}$$

Thanks to binomial identity, we know that

$$\sum_{j=0}^{M-\lfloor Mx \rfloor} \frac{(M-\lfloor Mx \rfloor)!}{(M-\lfloor Mx \rfloor-j)!j!} (-t)^j 1^{M-\lfloor Mx \rfloor-j} = (1-t)^{M-\lfloor Mx \rfloor},$$

and so

$$\begin{aligned}
\int_0^1 t^{\lfloor Mx \rfloor} \sum_{j=0}^{M-\lfloor Mx \rfloor} \frac{(M-\lfloor Mx \rfloor)!}{(M-\lfloor Mx \rfloor-j)!j!} (-t)^j 1^{M-\lfloor Mx \rfloor-j} f(t) dt \\
= \int_0^1 t^{\lfloor Mx \rfloor} (1-t)^{M-\lfloor Mx \rfloor} f(t) dt.
\end{aligned}$$

We thus finally found following expressions for  $f^M$

$$\begin{aligned}
f^M(x) &= \int_0^1 \mathbb{B}^{\lfloor Mx \rfloor+1, M-\lfloor Mx \rfloor+1}(u) f(u) du \\
&= (M+1) \binom{M}{\lfloor Mx \rfloor} m^{\lfloor Mx \rfloor, M-\lfloor Mx \rfloor}.
\end{aligned}$$

### 2.6.2 Bounding $\widehat{f}_n^M$ and $f^M$ in sup-norm

**Proposition 7.** For any  $M$  in  $\mathbb{N}$  and any  $n$  in  $\mathbb{N}$

$$(2.23) \quad \|f^M\|_\infty \leq M+1 \quad \text{and} \quad \|\widehat{f}_n^M\|_\infty \leq M+1.$$

*Proof.* For any  $(c, d)$  in  $\mathbb{N}^2$  and any  $(i, j)$  in  $\mathbb{N}^2$  such that  $i \geq c$  and  $j \geq d$ , we have

$$\binom{i+j}{i} = \sum_{k=0}^i \binom{c+d}{k} \binom{i+j-c-d}{i-k} \geq \binom{c+d}{c} \binom{i-c+j-d}{i-c},$$

so that for any  $(i, j)$  in  $\mathbb{N}^2$ , we have

$$(2.24) \quad \phi_{c,d}(i, j) \leq \binom{c+d}{c}^{-1} \mathbb{1}_{i \geq c}.$$

Plugging this in Equation (2.14) and by definition of  $N_n^{\lfloor Mx \rfloor}$  in Equation (2.13), we thus find

$$(2.25) \quad 0 \leq \widehat{m}_n^{(M,x)} \leq \frac{1}{N_n^{\lfloor Mx \rfloor}} \sum_{k=1}^n \binom{M}{\lfloor Mx \rfloor}^{-1} \mathbb{1}_{\{Z_{k-1}^n \geq \lfloor Mx \rfloor\}} \leq \binom{M}{\lfloor Mx \rfloor}^{-1}.$$

Combining definition of  $K^\nu$  in Equation (2.6) and  $\phi^{c,d}$  in Equation (2.13), we have the identity

$$(2.26) \quad \begin{aligned} \mathbb{E}^\nu \left[ \phi^{c,d}(Z_{k-1}^n, Z_k^n) \middle| Z_{k-1}^n \right] &= \sum_{j=0}^{+\infty} \phi^{c,d}(Z_{k-1}^n, j) K^\nu(Z_{k-1}^n, j) \\ &= \mathbb{1}_{\{Z_{k-1}^n \geq c\}} \int_0^1 \sum_{j=d}^{+\infty} \binom{Z_{k-1}^n + j - (c+d)}{Z_{k-1}^n - c} a^{Z_{k-1}^n + 1} (1-a)^j \nu(da). \end{aligned}$$

By definition of the negative binomial distribution with probability of success  $1-a$  and number of trials  $Z_{k-1}^n - c + 1$ ,

$$a^c (1-a)^d \sum_{j=0}^{+\infty} \binom{Z_{k-1}^n - c + j}{j} a^{Z_{k-1}^n - c + 1} (1-a)^j = a^c (1-a)^d.$$

Plugging this result in Equation (2.26) gives, by definition of  $m^{c,d}$  in Equation (2.10),

$$\mathbb{E}^\nu \left[ \phi^{c,d}(Z_{k-1}^n, Z_k^n) \middle| Z_{k-1}^n \right] = \int_0^1 a^c (1-a)^d \nu(da) \mathbb{1}_{\{Z_{k-1}^n \geq c\}} = m^{c,d} \mathbb{1}_{\{Z_{k-1}^n \geq c\}}.$$

Hence, Equation (2.24) implies

$$\begin{aligned} N_n^{\lfloor Mx \rfloor} m^{\lfloor Mx \rfloor, M - \lfloor Mx \rfloor} &= \sum_{k=1}^n \mathbb{E}^\nu \left[ \phi^{\lfloor Mx \rfloor, M - \lfloor Mx \rfloor}(Z_{k-1}^n, Z_k^n) \middle| Z_{k-1}^n \right] \mathbb{1}_{\{Z_{k-1}^n \geq \lfloor Mx \rfloor\}} \\ &\leq \sum_{k=1}^n \binom{M}{\lfloor Mx \rfloor}^{-1} \mathbb{1}_{\{Z_{k-1}^n \geq \lfloor Mx \rfloor\}} = \binom{M}{\lfloor Mx \rfloor}^{-1} N_n^{\lfloor Mx \rfloor}. \end{aligned}$$

Thus

$$(2.27) \quad m^{\lfloor Mx \rfloor, M - \lfloor Mx \rfloor} \leq \binom{M}{\lfloor Mx \rfloor}^{-1}.$$

By definition of  $f^M$  and  $\widehat{f}_n^M$  in Equations (2.12) and (2.15), this yields

$$(2.28) \quad \|f^M\|_\infty \leq M + 1 \quad \text{and} \quad \|\widehat{f}_n^M\|_\infty \leq M + 1.$$

□

### 2.6.3 Proof of Theorem 2

By the triangular inequality

$$(2.29) \quad \left\| f - \widehat{f}_n^M \right\|_\infty \leq \|f^M - f\|_\infty + \left\| f^M - \widehat{f}_n^M \right\|_\infty.$$

The first term in the right-hand side is deterministic and called the bias, the second is stochastic and called the variance. Lemma 8 provides an upper bound for the bias. Random bounds on the variance, valid in any regime, are provided in Lemma 9. Lemma 11 uses these random bounds to derive deterministic bounds on the variance that, once plugged in Equation (2.29) together with the result of Lemma 8 allow to conclude the proof of Equation (2.17) and Equation (2.16). The choice of  $M_n$  optimizing the bounds is then immediate.

**Lemma 8.** *Let  $\beta$  in  $(0, 2]$  and  $L \geq 1$ . Assume that  $f$  belongs to  $\Sigma(\beta, L)$ . Then*

$$(2.30) \quad \left\| f - f^M \right\|_\infty \leq 31LM^{-\beta/2}.$$

*Proof.* Let  $x$  in  $[0, 1]$  and let  $B_{(M,x)}$  be a beta distributed random variable with parameters  $\lfloor Mx \rfloor + 1$  and  $M - \lfloor Mx \rfloor + 1$  on a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . By Equation (2.11)

$$f^M(x) - f(x) = \int_0^1 f(u) \mathbb{B}^{\lfloor Mx \rfloor + 1, M - \lfloor Mx \rfloor + 1}(u) du - f(x) = \mathbb{E} [f(B_{(M,x)}) - f(x)].$$

Consider first the case  $0 < \beta \leq 1$ . Since  $f$  is  $(\beta, L)$ -Hölder, by Jensen's inequality,

$$|f^M(x) - f(x)| \leq \mathbb{E} [|f(B_{(M,x)}) - f(x)|] \leq L \mathbb{E} [|B_{(M,x)} - x|^\beta] \leq L \mathbb{E} [|B_{(M,x)} - x|]^\beta.$$

For any  $\eta > 0$ , it holds

$$(2.31) \quad \mathbb{E} [|B_{(M,x)} - x|] \leq \eta + \mathbb{E} [|B_{(M,x)} - x| \mathbb{1}_{\{|B_{(M,x)} - x| > \eta\}}].$$

By Cauchy-Schwarz inequality

$$(2.32) \quad \mathbb{E} [|B_{(M,x)} - x| \mathbb{1}_{\{|B_{(M,x)} - x| > \eta\}}] \leq \sqrt{\mathbb{E} [(B_{(M,x)} - x)^2] \mathbb{P}(|B_{(M,x)} - x| > \eta)},$$

by Markov inequality

$$(2.33) \quad \mathbb{P}(|B_{(M,x)} - x| > \eta) = \mathbb{P}((B_{(M,x)} - x)^2 > \eta^2) \leq \frac{\mathbb{E} [(B_{(M,x)} - x)^2]}{\eta^2},$$

and by a bias-variance decomposition

$$(2.34) \quad \mathbb{E} [(B_{(M,x)} - x)^2] \leq \text{Var}(B_{(M,x)}) + (\mathbb{E}[B_{(M,x)}] - x)^2,$$

so that

$$(2.35) \quad \mathbb{E} [|B_{(M,x)} - x| \mathbb{1}_{\{|B_{(M,x)} - x| > \eta\}}] \leq \frac{\text{Var}(B_{(M,x)}) + (\mathbb{E}[B_{(M,x)}] - x)^2}{\eta}.$$

One has

$$\mathbb{E}[B_{(M,x)}] = \frac{\lfloor Mx \rfloor + 1}{M + 2} \quad \text{and} \quad \text{Var}(B_{(M,x)}) = \frac{(\lfloor Mx \rfloor + 1)(M - \lfloor Mx \rfloor + 1)}{(M + 2)^2(M + 3)}.$$

Therefore,

$$(2.36) \quad |\mathbb{E} [B_{(M,x)}] - x| \leq \frac{2}{M+2} \quad \text{and} \quad \text{Var} (B_{(M,x)}) \leq \frac{1}{4M}.$$

Plugging Equation (2.36) in Equation (2.35) yields

$$\mathbb{E} \left[ |B_{(M,x)} - x| \mathbb{1}_{\{|B_{(M,x)} - x| > \eta\}} \right] \leq \frac{1}{\eta} \left( \frac{1}{4M} + \frac{4}{M^2} \right) \frac{1}{4M\eta^2} \leq \frac{1}{\eta} \frac{17}{4M}.$$

Plugging this bound in Equation (2.31) and optimizing in  $\eta > 0$  yields

$$|f^M(x) - f(x)| \leq 17^{\beta/2} L M^{-\beta/2}.$$

Consider then the case  $1 < \beta \leq 2$ . Since  $f$  is differentiable, by the mean-value theorem, there exists a random variable  $\tilde{B}_{(M,x)}$  such that

$$\left| \tilde{B}_{(M,x)} - x \right| \leq |B_{(M,x)} - x| \quad \text{and} \quad f(B_{(M,x)}) - f(x) = f'(\tilde{B}_{(M,x)})(B_{(M,x)} - x).$$

Therefore

$$\begin{aligned} |f^M(x) - f(x)| &= \left| \mathbb{E} \left[ f'(x)(B_{(M,x)} - x) + \left( f'(\tilde{B}_{(M,x)}) - f'(x) \right) (B_{(M,x)} - x) \right] \right| \\ &\leq |f'(x)| |\mathbb{E} [B_{(M,x)}] - x| + \mathbb{E} \left[ \left| f'(\tilde{B}_{(M,x)}) - f'(x) \right| |B_{(M,x)} - x| \right]. \end{aligned}$$

Since  $f$  is  $(\beta, L)$ -Hölder, by Equation (2.36), one gets

$$|f^M(x) - f(x)| \leq \|f'\|_{\infty} \frac{2}{M} + L \mathbb{E} [|B_{(M,x)} - x|^{\beta}].$$

Furthermore,

$$\begin{aligned} \mathbb{E} [|B_{(M,x)} - x|^{\beta}] &\leq \mathbb{E} [|B_{(M,x)} - \mathbb{E} [B_{(M,x)}] + \mathbb{E} [B_{(M,x)}] - x|^{\beta}] \\ &\leq 2^{\beta-1} \left( \mathbb{E} [|B_{(M,x)} - \mathbb{E} [B_{(M,x)}]|^{\beta}] + |\mathbb{E} [B_{(M,x)}] - x|^{\beta} \right). \end{aligned}$$

By Equation (2.36)

$$\begin{aligned} \mathbb{E} [|B_{(M,x)} - x|^{\beta}] &\leq 2^{\beta-1} \left( \text{Var} (B_{(M,x)})^{\beta/2} + |\mathbb{E} [B_{(M,x)}] - x|^{\beta} \right) \\ &\leq 2^{\beta-1} \left( \frac{2^{\beta/2}}{M^{\beta/2}} + \frac{2^{\beta}}{M^{\beta}} \right) \leq \frac{4^{\beta}}{M^{\beta/2}}. \end{aligned}$$

By [Tsy09, Theorem 1.1]

$$\|f'\|_{\infty} \leq \frac{3}{\sqrt{4 - \sqrt{13}}} \frac{L + \beta + 1}{\beta + 1} \leq 7.2L.$$

It thus finally holds

$$|f^M(x) - f(x)| \leq 31LM^{-\beta/2}.$$

As this result holds for all  $x$  in  $[0, 1]$ , the proof is complete.  $\square$

Let us now turn to the random part of the risk. We start with a random upper bound.

**Lemma 9.** *Assuming  $\mathbf{H2}$  or  $\mathbf{H1}$ , then, for any  $z > 0$ ,*

$$\mathbf{P}^\nu \left( \left\| f^M - \widehat{f}_n^M \right\|_\infty > \frac{M+1}{N_n^M} \sqrt{\frac{n(z + \log M)}{2}} \right) \leq 2e^{-z}.$$

*Proof.* By definition, one has

$$f^M(x) - \widehat{f}_n^M(x) = (M+1) \binom{M}{\lfloor Mx \rfloor} \left( m^{\lfloor Mx \rfloor, M - \lfloor Mx \rfloor} - \widehat{m}_n^{\lfloor Mx \rfloor, M - \lfloor Mx \rfloor} \right),$$

and by [DL18, Theorem 4], for any  $z > 0$ ,

$$\mathbf{P}^\nu \left( \left| m^{\lfloor Mx \rfloor, M - \lfloor Mx \rfloor} - \widehat{m}_n^{\lfloor Mx \rfloor, M - \lfloor Mx \rfloor} \right| > \frac{n}{N_n^{\lfloor Mx \rfloor}} \binom{M}{\lfloor Mx \rfloor}^{-1} \sqrt{\frac{z}{2n}} \right) \leq 2e^{-z},$$

so that

$$\mathbf{P}^\nu \left( \left| f^M(x) - \widehat{f}_n^M(x) \right| > (M+1) \frac{n}{N_n^{\lfloor Mx \rfloor}} \sqrt{\frac{z}{2n}} \right) \leq 2e^{-z}.$$

Since  $N_n^{\lfloor Mx \rfloor} \geq N_n^M$ , we have a fortiori

$$\mathbf{P}^\nu \left( \left| f^M(x) - \widehat{f}_n^M(x) \right| > (M+1) \frac{n}{N_n^M} \sqrt{\frac{z}{2n}} \right) \leq 2e^{-z}.$$

As already stressed deterministic function  $f^M$  as well as random function  $\widehat{f}_n^M$  are stepwise constant (whatever  $n$  is) and intervals on which they are constant are the same (namely intervals  $[\frac{k}{M}, \frac{k+1}{M})$  for integers  $0 \leq k \leq M-1$ ). Therefore, the function  $f^M - \widehat{f}_n^M$  is stepwise constant and, for any integer  $0 \leq k \leq M-1$  and any  $x$  in  $[\frac{k}{M}, \frac{k+1}{M})$

$$\begin{aligned} & \left\{ \left| f^M(x) - \widehat{f}_n^M(x) \right| > (M+1) \frac{n}{N_n^M} \sqrt{\frac{z}{2n}} \right\} \\ &= \left\{ \sup_{x \in [\frac{k}{M}, \frac{k+1}{M})} \left| f^M(x) - \widehat{f}_n^M(x) \right| > (M+1) \frac{n}{N_n^M} \sqrt{\frac{z}{2n}} \right\}, \end{aligned}$$

and hence

$$\mathbf{P}^\nu \left( \sup_{x \in [\frac{k}{M}, \frac{k+1}{M})} \left| f^M(x) - \widehat{f}_n^M(x) \right| > (M+1) \frac{n}{N_n^M} \sqrt{\frac{z}{2n}} \right) \leq 2e^{-z}.$$

Since we have

$$\begin{aligned} & \left\{ \left\| f^M - \widehat{f}_n^M \right\| > (M+1) \frac{n}{N_n^M} \sqrt{\frac{z}{2n}} \right\} \\ &= \left\{ \sup_{x \in [0,1)} \left| f^M(x) - \widehat{f}_n^M(x) \right| > (M+1) \frac{n}{N_n^M} \sqrt{\frac{z}{2n}} \right\} \\ &= \left\{ \max_{0 \leq k \leq M-1} \left( \sup_{x \in [\frac{k}{M}, \frac{k+1}{M})} \left| f^M(x) - \widehat{f}_n^M(x) \right| \right) > (M+1) \frac{n}{N_n^M} \sqrt{\frac{z}{2n}} \right\} \\ &= \bigcup_{k=0}^{M-1} \left\{ \sup_{x \in [\frac{k}{M}, \frac{k+1}{M})} \left| f^M(x) - \widehat{f}_n^M(x) \right| > (M+1) \frac{n}{N_n^M} \sqrt{\frac{z}{2n}} \right\}, \end{aligned}$$

a union bound enables then to conclude the proof

$$\begin{aligned} \mathbf{P}^\nu \left( \left\| f^M - \widehat{f}_n^M \right\| > (M+1) \frac{n}{N_n^M} \sqrt{\frac{z}{2n}} \right) \\ \leq \sum_{k=0}^{M-1} \mathbf{P}^\nu \left( \sup_{x \in \left[ \frac{k}{M}, \frac{k+1}{M} \right)} \left| f^M(x) - \widehat{f}_n^M(x) \right| > (M+1) \frac{n}{N_n^M} \sqrt{\frac{z}{2n}} \right) \leq M 2 e^{-z}, \end{aligned}$$

by replacing here  $z$  by  $z + \log(M)$ .  $\square$

**Corollary 10.** *For any integer  $n \geq 2$  and any  $M$  in  $\{1, \dots, n-1\}$ , on the event  $\Omega_0$  defined in Equation (2.18), it holds*

$$(2.37) \quad \left\| f - \widehat{f}_n^M \right\|_\infty \leq 31LM^{-\beta/2} + \frac{M+1}{N_n^M} \sqrt{3n \log n}.$$

*Proof.* For any real positive  $z$  and any  $M$  in  $\{1, \dots, n-1\}$ , since  $\log M \leq \log n$ , we have

$$\begin{aligned} \mathbf{P}^\nu \left( \left\| f^M - \widehat{f}_n^M \right\|_\infty > \frac{M+1}{N_n^M} \sqrt{\frac{n(z + \log n)}{2}} \right) \\ \leq \mathbf{P}^\nu \left( \left\| f^M - \widehat{f}_n^M \right\|_\infty > \frac{M+1}{N_n^M} \sqrt{\frac{n(z + \log M)}{2}} \right). \end{aligned}$$

Therefore, using a union bound and Lemma 9 yields

$$\begin{aligned} \mathbf{P}^\nu(\Omega_0) &= 1 - \mathbf{P}^\nu(\Omega_0^C) \\ &= 1 - \mathbf{P}^\nu \left( \bigcup_{M=1}^{n-1} \left\{ \left\| f^M - \widehat{f}_n^M \right\|_\infty > \frac{M+1}{N_n^M} \sqrt{\frac{n}{2}(\log(2n^3) + \log n)} \right\} \right) \\ &\geq 1 - \sum_{M=1}^{n-1} \mathbf{P}^\nu \left( \left\| f^M - \widehat{f}_n^M \right\|_\infty > \frac{M+1}{N_n^M} \sqrt{\frac{n}{2}(\log(2n^3) + \log M)} \right) \\ &\geq 1 - \sum_{M=1}^{n-1} 2e^{-\log(2n^3)} = 1 - 2(n-1) \frac{1}{2n^3} \geq 1 - n^{-2}. \end{aligned}$$

Plugging Lemma 8 in Equation (2.29), we get, on  $\Omega_0$ , for any  $n \geq 2$  and  $1 \leq M \leq n-1$ ,

$$\begin{aligned} \left\| f - \widehat{f}_n^M \right\|_\infty &\leq 31LM^{-\beta/2} + \left\| f^M - \widehat{f}_n^M \right\|_\infty \leq 31LM^{-\beta/2} + \frac{M+1}{N_n^M} \sqrt{\frac{n}{2}(\log(2n^3) + \log n)} \\ &\leq 31LM^{-\beta/2} + \frac{M+1}{N_n^M} \sqrt{\frac{5}{2}n \log n} \leq 31LM^{-\beta/2} + \frac{M+1}{N_n^M} \sqrt{3n \log n}. \end{aligned}$$

$\square$

The following lemma provides deterministic rates of convergence that derive from Lemma 9.

**Lemma 11.** *Under  $H2$ ,  $X$  is transient and*

$$\mathbf{E}^\nu \left[ \left\| f^M - \widehat{f}_n^M \right\|_\infty \right] \leq C_\nu M^{1+\kappa} \sqrt{\frac{\log n}{n}}.$$



Under **H1**,  $X$  is recurrent and

$$\mathbf{E}^\nu \left[ \left\| f^M - \widehat{f}_n^M \right\|_\infty \right] \leq C_\nu M \frac{\log n}{\sqrt{n}}.$$

*Proof.* Fix some  $M \geq 1$ .

If the chain is transient, [DL18, Lemmas 8 and 9] ensures there exist positive constants  $C_{1,\nu}$  and  $C_{2,\nu}$  such that for any  $z > 0$

$$\mathbf{P}^\nu \left( \left| \frac{N_n^M}{n} - \Pi([M, +\infty)) \right| \geq C_{1,\nu} \sqrt{\frac{z}{n}} \right) \leq 2e^{-z} \quad \text{and} \quad \Pi([M, +\infty)) \geq \frac{C_{2,\nu}}{M^\kappa}.$$

Let  $\gamma = C_{2,\nu}/(2C_{1,\nu})$ . If the chain is transient, it follows that, for any  $z > 0$ ,

$$\mathbf{P}^\nu \left( \frac{N_n^M}{n} \geq C_{2,\nu} \left( \frac{1}{M^\kappa} - \frac{1}{2\gamma} \sqrt{\frac{z}{n}} \right) \right) \geq 1 - 2e^{-z}.$$

Therefore,

$$(2.38) \quad \mathbf{P}^\nu \left( \frac{N_n^M}{n} \geq \frac{C_{2,\nu}}{2} \frac{1}{M^\kappa} \right) \geq 1 - 2e^{-\gamma^2 n/M^{2\kappa}}.$$

Remark that

$$\begin{aligned} & \left\{ \frac{n}{N_n^M} \leq \frac{2M^\kappa}{C_{2,\nu}} \right\} \cap \left\{ \left\| f^M - \widehat{f}_n^M \right\|_\infty \leq \frac{M+1}{N_n^M} \sqrt{\frac{n(z+\log M)}{2}} \right\} \\ & \subset \left\{ \left\| f^M - \widehat{f}_n^M \right\|_\infty \leq (M+1)^{\kappa+1} \frac{\sqrt{2}}{C_{2,\nu}} \sqrt{\frac{z+\log M}{n}} \right\}. \end{aligned}$$

Let

$$R_n(M, z) = \frac{2^{\kappa+\frac{3}{2}}}{C_{2,\nu}} M^{(1+\kappa)} \sqrt{\frac{z+\log M}{n}} \geq (M+1)^{\kappa+1} \frac{\sqrt{2}}{C_{2,\nu}} \sqrt{\frac{z+\log M}{n}}.$$

Using a union bound with Equation (2.38) and Lemma 9, we get for any  $z$  in  $(0, \gamma^2 n/M^{2\kappa}]$

$$\mathbf{P}^\nu \left( \left\| f^M - \widehat{f}_n^M \right\|_\infty \leq R_n(M, z) \right) \geq 1 - 2e^{-z} - 2e^{-\gamma^2 n/M^{2\kappa}} \geq 1 - 4e^{-z}.$$

By Proposition 7, both  $f^M$  and  $\widehat{f}_n^M$  are uniformly upper-bounded by  $M+1$ . It follows that the result of Lemma 11 is trivial if  $M > (\gamma^2 n/\log n)^{1/2\kappa}$ .

Assume therefore that  $M \leq (\gamma^2 n/\log n)^{1/2\kappa}$  and apply the previous inequality with  $z = \log n$ . Denote by

$$R'(M, n) = \frac{2^{\kappa+\frac{3}{2}}}{C_{2,\nu}} M^{(1+\kappa)} \sqrt{\frac{2\log n}{n}} \geq R_n(M, \log n).$$

This yields

$$\begin{aligned} \mathbf{E}^\nu \left[ \left\| f^M - \widehat{f}_n^M \right\|_\infty \right] & \leq \mathbf{E}^\nu \left[ \left\| f^M - \widehat{f}_n^M \right\|_\infty \mathbb{1}_{\left\| f^M - \widehat{f}_n^M \right\|_\infty > R'(M, n)} \right] \\ & \quad + \mathbf{E}^\nu \left[ \left\| f^M - \widehat{f}_n^M \right\|_\infty \mathbb{1}_{\left\| f^M - \widehat{f}_n^M \right\|_\infty \leq R'(M, n)} \right] \\ & \leq \frac{4(M+1)}{n} + R'(M, n) \leq (1 + 4C_{2,\nu})R'(M, n). \end{aligned}$$

The last inequality holds since

$$R'(M, n) \geq \frac{2M}{C_{2,\nu}\sqrt{n}} \geq \frac{M+1}{C_{2,\nu}n}.$$

This concludes the proof in the transient case.

Assume now that the chain is recurrent, then [DL18, Lemma 10] states that there exists  $C_\nu$  such that

$$(2.39) \quad \mathbf{P}^\nu \left( N_n^n \leq \frac{n}{2} \right) \leq C_\nu \frac{\log n}{\sqrt{n}}.$$

Combining Equation (2.39) with that of Lemma 9 implies that, for any  $M$  in  $\{1, \dots, n\}$

$$\mathbf{P}^\nu \left( \left\| f^M - \widehat{f}_n^M \right\|_\infty > (M+1) \sqrt{\frac{2(z + \log M)}{n}} \right) \leq 2e^{-z} + C_\nu \frac{\log n}{\sqrt{n}}.$$

Choosing  $z = \log n$  and noticing that  $\|f^M\|_\infty \vee \|\widehat{f}_n^M\|_\infty \leq M+1$  concludes the proof.  $\square$

**Conclusion of the proof of Theorem 2** Plugging the conclusions of Lemmas 8 and 11 in Equation (2.29) yields:

1. if  $X$  is transient,

$$\mathbf{E}^\nu \left[ \left\| f - \widehat{f}_n^M \right\|_\infty \right] \leq 31LM^{-\beta/2} + C_\nu M^{1+\kappa} \sqrt{\frac{\log n}{n}},$$

2. if  $X$  is recurrent,

$$\mathbf{E}^\nu \left[ \left\| f - \widehat{f}_n^M \right\|_\infty \right] \leq 31LM^{-\beta/2} + C_\nu M \frac{\log n}{\sqrt{n}}.$$

The proof of Theorem 2 is complete.

## 2.6.4 Proof of Theorem 5

### 2.6.4.1 Proof of Lemma 4

For any bounded functions  $f$  and  $g$ , denote  $d(f, g) = \|f - g\|_\infty$ . By repeated applications of the triangular inequality, one obtains, for any integer  $n \geq 2$  and for any  $M$  in  $\{1, \dots, n-1\}$ , on  $\Omega_0$ ,

$$\begin{aligned} d(f, \widehat{f}_n^{\widehat{M}_n}) &\leq d(f, \widehat{f}_n^M) + d(\widehat{f}_n^M, \widehat{f}_n^{M \vee \widehat{M}_n}) + d(\widehat{f}_n^{M \vee \widehat{M}_n}, \widehat{f}_n^{\widehat{M}_n}) \\ &\leq B(M) + V_n(M) + C_n(M) + 2V_n(\widehat{M}_n) + C_n(\widehat{M}_n) + 2V_n(M) \\ &\leq B(M) + V_n(M) + C_n(M) + 2V_n(M) + C_n(M) + 2V_n(M) \\ &\leq B(M) + 5V_n(M) + 2C_n(M). \end{aligned}$$

Now, for any  $M$  in  $\{1, \dots, n-1\}$ ,  $B$  being non increasing and  $V_n$  non decreasing, on  $\Omega_0$

$$\begin{aligned} C_n(M) &\leq 0 \vee \sup_{M' > M} \left\{ d(f, \widehat{f}_n^M) + d(f, \widehat{f}_n^{M'}) - 2V_n(M') \right\} \\ &\leq 0 \vee \sup_{M' > M} \left\{ B(M) + B(M') + V_n(M) - V_n(M') \right\} \leq 2B(M). \end{aligned}$$

### 2.6.4.2 Conclusion of the proof

As  $B(M)$  and  $V(M)$  are respectively non-increasing and non-decreasing and both non-negative, Lemma 4 applies and one has therefore, on  $\Omega_0$

$$\left\| f - \widehat{f}_n^{\widehat{M}_n} \right\|_\infty \leq 5 \inf_{M \in \{1, \dots, n-1\}} \left\{ 31LM^{-\beta/2} + \frac{M+1}{N_n^M} \sqrt{3n \log n} \right\}.$$

Let  $M_{\max} \leq (\gamma^2 n / (3 \log n))^{1/2\kappa}$ . A union bound in Equation (2.38) gives that, if  $X$  is transient

$$\mathbf{P}^\nu \left( \forall M \leq M_{\max}, \frac{N_n^M}{n} \geq \frac{C_{2,\nu}}{2} \frac{1}{M^\kappa} \right) \geq 1 - 2 \sum_{M=1}^{M_{\max}} e^{-\gamma^2 n / M^{2\kappa}} \geq 1 - M_{\max} e^{-3 \log n}.$$

Therefore, on an event with probability larger than  $1 - (M_{\max} + 1)e^{-3 \log n} - n^{-2} \geq 1 - 2n^{-2}$

$$\begin{aligned} \left\| f - \widehat{f}_n^{\widehat{M}_n} \right\|_\infty &\leq \inf_{M \in \{1, \dots, M_{\max}\}} \left\{ 31LM^{-\beta/2} + C_\nu M^{\kappa+1} \sqrt{\frac{\log n}{n}} \right\} \\ (2.40) \quad &\leq C_\nu \left( \frac{\log n}{n} \right)^{\beta/(4+4\kappa+2\beta)}. \end{aligned}$$

The proof is concluded in the transient case since the fact that  $M_{\max} \leq n-1$  and Proposition 7 ensure that  $\left\| \widehat{f}_n^{\widehat{M}_n} \right\|_\infty \leq n+1$ . Since  $N_n^M \geq N_n^n$  for any  $M$  in  $\{1, \dots, n-1\}$ , in the recurrent case, one has therefore, by Equation (2.39), on an event with probability larger than  $1 - C_\nu \log n / \sqrt{n}$

$$\begin{aligned} \left\| f - \widehat{f}_n^{\widehat{M}_n} \right\|_\infty &\leq \inf_{M \in \{1, \dots, n-1\}} \left\{ 31LM^{-\beta/2} + C_\nu M \frac{\log n}{\sqrt{n}} \right\} \\ &\leq C_\nu \left( \frac{(\log n)^2}{n} \right)^{\beta/(2\beta+4)}. \end{aligned}$$

### 2.6.5 Proof of Proposition 6

First, as  $f$  is a density function on  $[0, 1]$  in  $\Sigma(\beta, L)$ , there exists a constant  $C_{\beta, L}$  such that  $\|f\|_\infty \leq C_{\beta, L}$ . Assuming now **H1**, by Theorem 5, on the event

$$\Omega = \left\{ \left\| f - \widehat{f}_n^{\widehat{M}_n} \right\|_\infty \leq C_\nu \left( \frac{\log(n)}{\sqrt{n}} \right)^{\beta/(\beta+2)} \right\},$$

it follows that

$$\|\widehat{f}_n^{\widehat{M}_n}\|_\infty \leq C_{\beta, L} + (2/e)C_\nu.$$

Therefore, for any  $n \geq n_0 = e^{C_{\beta, L} + (2/e)C_\nu}$ ,

$$\widetilde{f}_n = \widehat{f}_n^{\widehat{M}_n}.$$

As  $\widetilde{f}_n \leq \log n$ , for any  $n \geq n_0$ , it follows that

$$\begin{aligned} \mathbf{E}^\nu \left[ \left\| f - \widetilde{f}_n \right\|_\infty \right] &= \mathbf{E}^\nu \left[ \left\| f - \widetilde{f}_n \right\|_\infty \mathbf{1}_\Omega \right] + \mathbf{E}^\nu \left[ \left\| f - \widetilde{f}_n \right\|_\infty \mathbf{1}_{\Omega^c} \right] \\ &\leq \mathbf{E}^\nu \left[ \left\| f - \widehat{f}_n^{\widehat{M}_n} \right\|_\infty \mathbf{1}_\Omega \right] + (C_{\beta, L} + \log n) \mathbf{P}^\nu(\Omega^c) \\ &\leq C_\nu \left( \frac{\log(n)}{\sqrt{n}} \right)^{\beta/(\beta+2)} + (C_{\beta, L} + \log n) C_\nu \frac{\log n}{\sqrt{n}} \\ &\leq C'_\nu \left( \frac{\log(n)}{\sqrt{n}} \right)^{\beta/(\beta+2)}. \end{aligned}$$

The result holds for any  $n \geq n_0$ , therefore for any  $n$  provided that we enlarge the constant  $C'_\nu$  if necessary.

Assume now **H2**. Let

$$\Omega' = \left\{ \left\| f - \widehat{f}_n^{\widehat{M}_n} \right\|_\infty \leq C_\nu \left( \frac{\log n}{n} \right)^{\beta/(4+4\kappa+2\beta)} \right\}.$$

On  $\Omega'$ ,  $\|\widehat{f}_n^{\widehat{M}_n}\|_\infty \leq C_{\beta, L} + C_\nu$ . In particular, for any  $n \geq n_1 = e^{C_{\beta, L} + C_\nu}$ ,

$$\widetilde{f}_n = \widehat{f}_n^{\widehat{M}_n}.$$

As  $\widetilde{f}_n \leq \log n$ , for any  $n \geq n_1$ , it follows that

$$\begin{aligned} \mathbf{E}^\nu \left[ \left\| f - \widetilde{f}_n \right\|_\infty \right] &= \mathbf{E}^\nu \left[ \left\| f - \widetilde{f}_n \right\|_\infty \mathbf{1}_{\Omega'} \right] + \mathbf{E}^\nu \left[ \left\| f - \widetilde{f}_n \right\|_\infty \mathbf{1}_{(\Omega')^c} \right] \\ &\leq \mathbf{E}^\nu \left[ \left\| f - \widehat{f}_n^{\widehat{M}_n} \right\|_\infty \mathbf{1}_{\Omega'} \right] + (C_{\beta, L} + \log n) \mathbf{P}^\nu((\Omega')^c) \\ &\leq C_\nu \left( \frac{\log n}{n} \right)^{\beta/(4+4\kappa+2\beta)} + (C_{\beta, L} + \log n) \mathbf{P}^\nu((\Omega')^c). \end{aligned}$$

By Equation (2.40),  $\mathbf{P}^\nu((\Omega')^c) \leq 2n^{-2}$ , hence, there exists  $C'_\nu$  such that, for any  $n \geq n_1$ ,

$$\mathbf{E}^\nu \left[ \left\| f - \widetilde{f}_n \right\|_\infty \right] \leq C'_\nu \left( \frac{\log n}{n} \right)^{\beta/(4+4\kappa+2\beta)}.$$

The result holds for any  $n$  provided that we enlarge the constant  $C'_\nu$  if necessary.



## Chapter 3

# First return time of a branching process in random environment

The notion of *branching processes* appeared for the first time in 1875, when Galton and Watson investigate statistically the extinction of family names in their seminal paper [GW75]. Since family names pass from father to son, they study the evolution of male population along time. To modelize this problem, Galton and Watson assume that individuals of a generation  $n$  give birth to a random number of individuals at the generation  $n + 1$  independently from each other and according to the same probability distribution. Denoting  $Z_n$  the number of individuals in the population at generation  $n$ , the Galton-Watson process  $(Z_n)_{n \in \mathbb{N}}$  is defined, for every  $n$  in  $\mathbb{N}$ , by

$$Z_0 = 1 \quad \text{and} \quad Z_{n+1} = \sum_{k=1}^{Z_n} \xi_{n,k},$$

where  $\xi_{n,k}$  denotes the number of descendants that the individual  $k$  of generation  $n$  will have at generation  $n + 1$ . In this model, for any  $n$  in  $\mathbb{N}$ , the variables  $(\xi_{n,k})_{k \in \mathbb{N}}$  are i.i.d. . If  $Z_n = 0$  for some  $n$  in  $\mathbb{N}$ , then at any generation  $k \geq n$ ,  $Z_k = 0$  so that the family name disappears. Looking for the smallest  $n$  in  $\mathbb{N}$  when  $Z_n = 0$  is thus of main interest.

Despite Galton-Watson model reveals useful to model various problems, assuming that offspring distribution is the same at each generation doesn't reflect possible non-homogeneity along time. To model heterogeneity, we consider a more general branching process  $(Z_n)_{n \in \mathbb{N}}$  defined, for every  $n$  in  $\mathbb{N}$ , by

$$Z_0 = 1 \quad \text{and} \quad Z_{n+1} = \sum_{k=1}^{Z_n} \xi_{n,k},$$

where  $(\xi_{n,k})_{(n,k) \in \mathbb{N} \times \mathbb{N}^*}$  are still independent but where, for each  $n$  in  $\mathbb{N}$ , the random variables  $(\xi_{n,k})_{k \in \mathbb{N}^*}$  are distributed according to some probability  $\mu_n$  on  $\mathbb{N}$  (depending on  $n$ ). The first time  $n$  such that  $Z_n = 0$ , called *extinction time*, is still of central interest.

As expected, this model introduces heterogeneity through generations. To model some similarity within heterogeneity, we assume some statistical regularity between the generations. To do this, we consider that the offspring distributions  $(\mu_n)_{n \in \mathbb{N}}$  are realizations of independent and identically distributed random variables. This leads to a model named *branching process in random environment*.

**Definition 12.** Consider a sequence of random variables  $(\mu_l)_{l \in \mathbb{N}}$  defined on  $(\Omega, \mathbb{P})$  taking values in the space of probability distributions on  $\mathbb{N}$  that are independent and identically distributed. The sequence of random variables  $(Z_t)_{t \in \mathbb{N}}$  is a branching process in random environment if

$$(3.1) \quad Z_0 = 1 \quad \text{and} \quad Z_{t+1} = \sum_{j=1}^{Z_t} \xi_{t,j},$$

where the  $(\xi_{i,j})_{(i,j) \in \mathbb{N} \times \mathbb{N}}$  are independent random variables conditionally to the  $\sigma$ -field generated by  $(\mu_l)_{l \in \mathbb{N}}$  and such that for any  $i, j$  and  $n$  in  $\mathbb{N}$

$$\mathbb{P}(\xi_{i,j} = n | (\mu_l)_{l \in \mathbb{N}}) = \mu_i(n).$$

An other possible modification of the Galton-Watson process is to take into account the arrival in the population of individuals from outside. A common way to do this is to add *immigration* to the branching process. We make here the assumption that at generation  $n$ , an additional individual gives birth to a random number of individuals independently of and with the same offspring distribution as the  $Z_{n-1}$  individuals of the generation  $n-1$ . This additional feature is formalized in the model of branching process with immigration in random environment.

**Definition 13.** Consider a sequence of random variables  $(\mu_l)_{l \in \mathbb{N}}$  defined on  $(\Omega, \mathbb{P})$  taking values in the space of probability distributions on  $\mathbb{N}$  that are independent and identically distributed. The sequence of random variables  $(Z_t)_{t \in \mathbb{N}}$  is a branching process in random environment if

$$(3.2) \quad Z_0 = 1 \quad \text{and} \quad Z_{t+1} = \sum_{j=0}^{Z_t} \xi_{t,j},$$

where the  $(\xi_{i,j})_{(i,j) \in \mathbb{N} \times \mathbb{N}}$  are independent random variables conditionally to the  $\sigma$ -field generated by  $(\mu_l)_{l \in \mathbb{N}}$  and such that for any  $i, j$  and  $n$  in  $\mathbb{N}$

$$\mathbb{P}(\xi_{i,j} = n | (\mu_l)_{l \in \mathbb{N}}) = \mu_i(n).$$

The distribution of  $(Y_k)_{k \in \mathbb{N}}$  under  $\mathbb{P}$  is denoted  $\mathbf{P}^\nu$ .

At every time  $t$  such that  $Z_t = 0$ , the branching starts anew. This explains why the process  $\{Z_t = 0\}$  are of particular interest. The smallest time  $t$  in  $\mathbb{N}^*$  such that  $Z_t = 0$  is called the *return time to origin*. The purpose of this chapter is to study this random time  $T = \inf\{t \in \mathbb{N}^* : Z_t = 0\}$ .

Studying the long-term behaviour of the random walk in i.i.d. random environment on  $\mathbb{Z}$  (abbreviated *RWRE*), [KKS75] established an important parallel between left jump process of the RWRE and BPIRE. Assume we observe the trajectory of the RWRE starting from 0 with environment distribution  $\nu$  until it reaches for the first time the integer site  $n$ . Denote by  $L_k^n$  the number of steps from  $k$  to  $k+1$  made by the walk along the trajectory for any integer  $0 \leq k \leq n$ . Then, the sequence  $L_n^n, L_{n-1}^n, \dots, L_1^n, L_0^n$  has the same distribution as the first  $(n+1)$  generations of a BPIRE starting from 0 and with geometrically distributed offsprings whose parameters are i.i.d. with law  $\nu$  (denoted "*BPIREG*( $\nu, 0$ )").

In [KKS75], the control of the return time to origin  $T$  of the  $\text{BPIREG}(\nu, 0)$  is central. [KKS75, Lemma 1] states that, if the RWRE with environment distribution  $\nu$  is transient to the right and if  $\nu$  meets some technical assumptions, then, there exists positive constants  $K_1$  and  $K_2$  such that, for any positive  $t$

$$(3.3) \quad \mathbf{P}^\nu (T > t) \leq K_1 e^{-K_2 t} .$$

The control of the tail distribution of  $\text{BPIREG}(\nu, 0)$  given by [KKS75] is valid for fixed environment distribution  $\nu$  and it is not specified how constant  $K_1$  and  $K_2$  depend on this distribution. This information is necessary in some applications. For example, in the Bayesian approach of Chapter 5, we need a control of constants  $K_1$  and  $K_2$  uniform over a class of distributions.

The problem of direct estimation of the environment distribution for RWRE has been addressed in [CFL<sup>+</sup>14]. When the regime is transient,  $\text{BPIREG}(\nu, 0)$  is a positive recurrent Markov chain with transition kernel given in [CFL<sup>+</sup>14, Proposition 4.3]. In particular,  $\text{BPIREG}(\nu, 0)$  admits a unique invariant probability distribution [CFL<sup>+</sup>14, Proposition 4.5]. A sufficient condition for geometrical ergodicity of this chain is the existence of a set whose first return time admits a finite exponential moment. It is possible to obtain exponential concentration inequality for geometrically ergodic Markov chains. To make these concentration inequalities uniform over classes of the  $\text{BPIREG}(\nu, 0)$  with various environment distributions, we need that the first return time to 0 admits uniform exponential moment. To this end, we state and prove in this chapter a uniform version of the exponential control on the tail distribution of the return time to 0 of  $\text{BPIREG}(\nu, 0)$  initially obtained by [KKS75].

Section 3.1 is organized as follows. First, we set our framework defining assumptions as well as the particular BPIRE we look at. Then we state our main result : a uniform version of the exponential control on the tail distribution of the return time to 0 of  $\text{BPIREG}(\nu, 0)$  given by Inequality (3.3). Finally, we give an application of our main result showing the existence of exponential moment for the return time to 0 of  $\text{BPIREG}(\nu, 0)$ .

Section 3.2 is organized as follows. First we break the proof of our main result into two Theorems. Then, we give a sketch and a detailed version of their proof. Finally, we show how to combine them to prove our main result.

## 3.1 Setting and main result

### 3.1.1 Assumptions

In remaining of this chapter, we fix  $\nu$  a probability on  $[0, 1]$ . We consider a sequence  $(\omega_x)_{x \in \mathbb{Z}}$  of i.i.d. random variables with common distribution  $\nu$ . The random variable  $\omega = (\omega_x)_{x \in \mathbb{Z}}$  is thus distributed as  $\mathbb{P}^\nu = \nu^{\otimes \mathbb{Z}}$ . We define the sequence of random variables  $(\rho_i)_{i \in \mathbb{Z}}$  by

$$(3.4) \quad \rho_i = \frac{1 - \omega_i}{\omega_i} .$$

We denote

$$(3.5) \quad c_\nu = - \int_0^1 \log \left( \frac{1-t}{t} \right) \nu(dt) = -\mathbb{E}^\nu [\log(\rho_0)] .$$

Furthermore, we assume that  $\nu$  satisfies the following assumptions



**F1.** *There exists  $c_0$  in  $\mathbb{R}_+^*$  such that  $c_\nu \geq c_0$ .*

**F2.** *There exists  $\alpha$  in  $(0, 1/2)$  such that  $\text{supp}(\nu) \subset [\alpha, 1 - \alpha]$ .*

The assumptions **F** under which our main result is valid have to be discussed under the scope of random walk on  $\mathbb{Z}$  in i.i.d. random environment with common distribution  $\nu$ . In this framework, they indeed find a natural interpretation.

According to [Sol75], Assumption **F1** is the necessary and sufficient condition for the RWRE with environment distribution  $\nu$  to be transient to the left and thus concerns rather the long-time behaviour of the chain.

Assumption **F2** called *ellipticity condition* ensures that the value of the environment can not be too close to 0 or 1 with large probability. It is a technical condition commonly considered in the results on the RWRE. It prevents that the walk get trapped between two integers and thus concerns the local behaviour of the chain.

### 3.1.2 The case of RWRE

A general definition of the branching processes with or without immigration was respectively given in Definitions 13 and 12. We particularize here this definition to the specific BPIRE naturally present in the study of the RWRE.

Consider a sequence of i.i.d. random variables  $(\omega_x)_{x \in \mathbb{Z}}$  with common probability distribution  $\nu$  defined on a probability space  $(\Omega, \mathcal{A}, \mathbf{P})$  and taking values in  $[0, 1]$ . The random variable  $\omega = (\omega_x)_{x \in \mathbb{Z}}$  called *environment* is distributed as  $\mathbb{P}^\nu = \nu^{\otimes \mathbb{Z}}$  and takes value in the *set of environments*  $E = [0, 1]^{\mathbb{Z}}$  equipped with the  $\sigma$ -field generated by the cylinders  $\mathcal{E} = \mathcal{B}([0, 1]^{\otimes \mathbb{Z}})$ .

A *BPIREG*( $\nu, 0$ ) is a sequence of random variables  $(Z_y)_{y \in \mathbb{N}}$  defined on  $\Omega$  by

$$(3.6) \quad Z_0 = 0 \quad \text{and} \quad \forall y \in \mathbb{N}, Z_{y+1} = \sum_{i=0}^{Z_y} \xi_{y,i},$$

where the  $(\xi_{y,i})_{(y,i) \in \mathbb{N}^2}$  are independent conditionally to the  $\sigma$ -field generated by  $\omega$  and such that, for all  $y, i$  and  $k$  in  $\mathbb{N}$

$$\mathbf{P}(\xi_{y,i} = k | \omega) = \omega_y (1 - \omega_y)^k.$$

The distribution of  $(Z_y)_{y \in \mathbb{N}}$  under  $\mathbf{P}$  is denoted  $\mathbf{P}^\nu$ .

Given  $\omega$ , the sequence  $(Z_y)_{y \in \mathbb{N}}$  is a Galton-Watson branching process with immigration. It is thus an inhomogeneous Markov chain starting at 0 whose transition kernels  $(K_{\omega,k})_{k \in \mathbb{N}}$  are defined, for any  $k, i$  and  $j$  in  $\mathbb{N}$  and any  $\omega$  in  $E$ , by

$$(3.7) \quad K_{\omega,k}(i, j) = \binom{i+j}{i} \omega_k^{i+1} (1 - \omega_k)^j.$$

In the rest of the chapter, we denote  $\mathbf{P}_\omega = \mathbf{P}(\cdot | \omega)$  the probability distribution conditionally on  $\omega$  named *quenched distribution* and denote  $\mathbf{P}^\nu$  the probability distribution of  $(Z_y)_{y \in \mathbb{N}}$  under  $\mathbf{P}$  named *annealed distribution*.

### 3.1.3 Main result

The goal of this chapter is to establish the following quantitative version of the control on the tail distribution of the first return time to 0 uniform over a class of  $\text{BPIREG}(\nu, 0)$ .

**Theorem 14.** *Assume that Hypotheses **F1** and **F2** hold. Consider the 2 positive constants  $K_1(c_0, \alpha)$  and  $K_2(c_0, \alpha)$  (depending only on  $c_0$  and  $\alpha$ ) whose explicit expressions are given respectively in Equations (3.85) and (3.86). Let  $T$  be the first return time to  $\{0\}$  of the  $\text{BPIREG}(\nu, 0)$*

$$(3.8) \quad T = \min\{t \in \mathbb{N}^* / Z_t = 0\}.$$

Then, for any  $t > 0$

$$\mathbf{P}^\nu(T > t) \leq K_1(c_0, \alpha)e^{-K_2(c_0, \alpha)t}.$$

**Corollary 15.** *Consider a distribution  $\nu$  on  $[0, 1]$  such that  $\int_0^1 \log(\frac{1-t}{t})\nu(dt) \leq -c_0 < 0$  and such that  $\text{supp}(\nu) \subset [\alpha, 1 - \alpha]$  for some  $\alpha$  in  $[0, \frac{1}{2}]$ . Let  $T$  denote the first return time to  $\{0\}$  of the  $\text{BPIREG}(\nu, 0)$*

$$T = \min\{t \in \mathbb{N}^* / Z_t = 0\}.$$

Then, the positive constants  $K_1(c_0, \alpha)$  and  $K_2(c_0, \alpha)$  defined in Equations (3.85) and (3.86) are such that

$$\mathbf{E}^\nu \left[ e^{\frac{K_2(c_0, \alpha)}{2}T} \right] \leq 1 + K_1(c_0, \alpha) < +\infty.$$

## 3.2 Proofs

The proof of Theorem 14 involves two auxiliary objects, the (simple) random walk  $(S_n)_{n \in \mathbb{N}}$  and its successive downward ladder indices  $(N_n)_{n \in \mathbb{N}}$ . These sequences of random variables are defined, for any  $n$  in  $\mathbb{N}$ , by

$$(3.9) \quad S_0 = 0 \quad \text{and} \quad S_{n+1} = \sum_{i=0}^n \left( \log(\rho_i) + \frac{c_\nu}{2} \right)$$

$$(3.10) \quad N_0 = 0 \quad \text{and} \quad N_{n+1} = \inf \{n > N_n : S_n \leq S_{N_n}\},$$

where the  $\rho_i$  are defined in Equation (3.4).

The proof of Theorem 14 can be decomposed in two main theorems, concerning  $T$  and the  $N_k$  : Theorems 16 and 22.

First, we concentrate on Theorem 16 : Sections 3.2.1, 3.2.2 and 3.2.3 are respectively dedicated to its statement, to a sketch of its proof and to its detailed proof.

Then, in Section 3.2.4, we present a new interpretation of the  $\text{BPIREG}(\nu, 0)$  necessary to the exposition of Theorem 22.

Then, we concentrate on Theorem 22 : Sections 3.2.6, 3.2.6 and 3.2.7 are respectively dedicated to its statement, to a sketch of its proof and to its detailed proof.

Finally, in Section 3.2.8, we show why Theorems 16 and 22 yield Theorem 14.

### 3.2.1 Theorem 16

Theorem 16 controls the tail distribution of  $N_k$  under distribution  $\mathbb{P}^\nu = \nu^{\otimes \mathbb{Z}}$ .

**Theorem 16.** *Assume that Hypotheses **F1** and **F2** hold. Consider the positive constants  $b(c_0, \alpha)$  and  $K_3(c_0, \alpha)$  depending only on  $c_0$  and  $\alpha$  defined by*

$$(3.11) \quad b(c_0, \alpha) = \frac{c_0}{2 \log^2 \left( \frac{1-\alpha}{\alpha} \right)},$$

and

$$(3.12) \quad K_3(c_0, \alpha) = \frac{c_0 b(c_0, \alpha)}{16} \left( 1 - e^{-\frac{c_0 b(c_0, \alpha)}{8}} \right) e^{-\frac{c_0 b(c_0, \alpha)}{8}},$$

and some integer  $l_0 = l_0(c_0, \alpha)$  (also depending only on both  $c_0$  and  $\alpha$ ) such that

$$(3.13) \quad l_0 \geq 2 - \frac{4 \log \left( 1 - e^{-\frac{c_0 b(c_0, \alpha)}{8}} \right)}{c_0 b(c_0, \alpha)} \quad \text{and} \quad \sum_{k=l_0-1}^{+\infty} k e^{-\frac{c_0 b(c_0, \alpha)}{8}(k-2)} \leq \frac{1}{2}.$$

Then, for any  $k$  in  $\mathbb{N}$

$$(3.14) \quad \mathbb{P}^\nu(N_k > l_0(c_0, \alpha) \mathbb{E}^\nu[N_1] k) \leq e^{-K_3(c_0, \alpha) k}.$$

### 3.2.2 Sketch of proof of Theorem 16

From the definition of  $(S_n)_{n \in \mathbb{N}}$  and  $(N_k)_{k \in \mathbb{N}}$  in Equations (3.9) and (3.10), it follows that the couples  $(N_k - N_{k-1}, S_{N_k} - S_{N_{k-1}})_{k \in \mathbb{N}^*}$  are i.i.d. and distributed as  $(N_1, S_1)$  under  $\mathbb{P}^\nu$ . And since the following decomposition is valid for any  $k$  in  $\mathbb{N}^*$

$$(3.15) \quad N_k = \sum_{j=1}^k (N_j - N_{j-1}),$$

$N_k$  is the telescopic sum of  $k$  i.i.d. random variables distributed as  $N_1$  under  $\mathbb{P}^\nu$ . This suggests the following strategy for the proof:

- (i) First, we study the distribution of  $N_1$  and obtain a control on its tail.

**Corollary 17.** *Under Assumptions **F1** and **F2**, for all integer  $n$*

$$(3.16) \quad \mathbb{P}^\nu(N_1 > n) \leq e^{-\frac{c_0 b(c_0, \alpha)}{4} n},$$

where  $b(c_0, \alpha)$  is defined in Equation (3.11).

*Proof.* As  $\{N_1 > n\} \subset \{S_n > 0\}$  for any positive integer  $n$ , we have

$$\mathbb{P}^\nu(N_1 > n) \leq \mathbb{P}^\nu(S_n > 0).$$

For any positive real number  $\theta$ , Markov inequality then ensures

$$(3.17) \quad \mathbb{P}^\nu(S_n > 0) = \mathbb{P}^\nu(e^{\theta S_n} > 1) \leq \mathbb{E}^\nu[e^{\theta S_n}],$$

so that it follows from Equation (3.9)

$$\mathbb{P}^\nu(S_n > 0) \leq \left( e^{\frac{c_\nu}{2}\theta} \mathbb{E}^\nu[\rho_0^\theta] \right)^n.$$

In particular, for any positive integer  $n$

$$(3.18) \quad \mathbb{P}^\nu(N_1 > n) \leq \left( e^{\frac{c_\nu}{2}\theta} \mathbb{E}^\nu[\rho_0^\theta] \right)^n.$$

The proof is concluded by the application of following Lemma 18 used with  $\theta = b(c_0, \alpha)$  where  $b(c_0, \alpha)$  is defined in Equation (3.11).  $\square$

Lemma 18 investigates the convergence and the rate of convergence of  $\mathbb{E}^\nu[\rho_0^\theta]^{\frac{1}{\theta}}$  when  $\theta$  goes to 0. A detailed proof is given in Sections 3.2.3.2 and 3.2.3.1.

**Lemma 18.** *Consider that Assumptions **F1** and **F2** are satisfied. Then, the following inequality is satisfied for any  $\theta$  in  $\mathbb{R}_+^*$*

$$(3.19) \quad \left| \frac{1}{\theta} \log \mathbb{E}^\nu[\rho_0^\theta] + c_\nu \right| \leq \frac{1}{2} \log^2 \left( \frac{1-\alpha}{\alpha} \right) \theta,$$

where  $c_\nu$  was defined in Equation (3.5).

In particular, the following convergence holds

$$(3.20) \quad \frac{1}{\theta} \log \mathbb{E}^\nu[\rho_0^\theta] \xrightarrow{\theta \rightarrow 0^+} -c_\nu.$$

**Remark 19.** *Corollary 17 implies that  $N_1$  admits moments of any orders under  $\mathbb{P}^\nu$  and that Laplace-transform  $\lambda \mapsto \mathbb{E}^\nu[e^{\lambda N_1}]$  exists and is infinitely differentiable on the interval  $(-c_0 b(c_0, \alpha)/4, c_0 b(c_0, \alpha)/4)$ .*

- (ii) Then, we use the independence of the random variables distributed as  $N_1$  in decomposition (3.15) to derive the control of the tails of the  $N_k$ . This final step proving Theorem 16 is developed in Sections 3.2.3.4 and 3.2.3.5.

### 3.2.3 Detailed proof of Theorem 16

#### 3.2.3.1 A preliminary result for the proof of Lemma 18

**Lemma 20.** *Let us define for all  $\theta$  in  $\mathbb{R}_+$*

$$(3.21) \quad G(\theta) = \log(\mathbb{E}^\nu[\rho_0^\theta]) = \log \left( \int_0^1 e^{\theta \log(\frac{1-t}{t})} \nu(dt) \right).$$

Under **F2**, following properties hold:

(i) *the application  $G : \mathbb{R}_+ \rightarrow \mathbb{R}$  is well defined.*

(ii)  *$G$  is twice-differentiable over  $\mathbb{R}_+^*$  with first and second derivative given, for any  $\theta$  in  $\mathbb{R}_+^*$ , by*

$$(3.22) \quad G^{(1)}(\theta) = \mathbb{E}^\nu \left[ \log(\rho_0) \frac{\rho_0^\theta}{\mathbb{E}^\nu[\rho_0^\theta]} \right],$$

and

$$(3.23) \quad G^{(2)}(\theta) = \mathbb{E}^\nu \left[ \log^2(\rho_0) \frac{\rho_0^\theta}{\mathbb{E}^\nu[\rho_0^\theta]} \right] - \mathbb{E}^\nu \left[ \log(\rho_0) \frac{\rho_0^\theta}{\mathbb{E}^\nu[\rho_0^\theta]} \right]^2.$$

(iii)  $G^{(2)}$  is:

- (a) positive, so  $G$  is convex,
- (b) uniformly bounded over  $\mathbb{R}_+^*$  by  $\log^2\left(\frac{1-\alpha}{\alpha}\right)$ .

**Proof of Lemma 20.** (i) Assumption **F2** ensures that, for any  $\theta$  in  $\mathbb{R}_+$

$$(3.24) \quad \left(\frac{\alpha}{1-\alpha}\right)^\theta \leq \rho_0^\theta \leq \left(\frac{1-\alpha}{\alpha}\right)^\theta.$$

Hence, the function is well defined.

(ii) Since  $G = \log(g)$  where  $g : \theta \in \mathbb{R}_+ \mapsto \mathbb{E}^\nu[\rho_0^\theta] \in \mathbb{R}_+^*$ ,  $G$  is twice-differentiable if and only if  $g$  is twice-differentiable and in this case, it furthermore holds

$$(3.25) \quad G^{(1)}(\theta) = \frac{g^{(1)}(\theta)}{g(\theta)} \quad \text{and} \quad G^{(2)}(\theta) = \frac{g^{(2)}(\theta)g(\theta) - (g^{(1)}(\theta))^2}{g^2(\theta)}.$$

Fix some  $d$  in  $\mathbb{R}_+^*$  and let  $f : (t, \theta) \in [\alpha, 1-\alpha] \times [0, d] \mapsto \left(\frac{1-t}{t}\right)^\theta = e^{\theta \log\left(\frac{1-t}{t}\right)}$ . Then  $g$  is the integral  $g : \theta \in [0, d] \mapsto \int_\alpha^{1-\alpha} f(t, \theta) \nu(dt)$  and we want to prove that deriving twice under integral is permitted.

Fix  $t$  in  $[\alpha, 1-\alpha]$ , the first and second derivative of  $f$  according to  $\theta$  exist and, for all  $\theta$  in  $[0, d]$

$$\frac{\partial f}{\partial \theta}(t, \theta) = \left(\frac{1-t}{t}\right)^\theta \log\left(\frac{1-t}{t}\right) \quad \text{and} \quad \frac{\partial^2 f}{\partial^2 \theta}(t, \theta) = \left(\frac{1-t}{t}\right)^\theta \log^2\left(\frac{1-t}{t}\right).$$

Since  $\rho_0^\theta$  is  $\mathbb{P}^\nu$ -integrable for all  $\theta$  in  $[0, d]$  (thanks to Inequality (3.24)), the function  $t \mapsto \left(\frac{1-t}{t}\right)^\theta$  is thus  $\nu$ -integrable over  $[\alpha, 1-\alpha]$ . For fixed  $\theta$  in  $[0, d]$ , for any  $t$  in  $[\alpha, 1-\alpha]$  and for  $k$  in  $\{1, 2\}$

$$\left| \left(\frac{1-t}{t}\right)^\theta \log^k\left(\frac{1-t}{t}\right) \right| \leq \left(\frac{1-\alpha}{\alpha}\right)^d \log^k\left(\frac{1-\alpha}{\alpha}\right).$$

The first and second partial derivatives  $\frac{\partial f}{\partial \theta}(t, \theta)$  and  $\frac{\partial^2 f}{\partial^2 \theta}(t, \theta)$  are thus dominated by positive, measurable and  $\nu$ -integrable functions over  $[\alpha, 1-\alpha]$  not depending on  $\theta$  in  $[0, d]$ . Hence, deriving twice under integral is permitted. The function  $g$  is twice differentiable with first and second derivatives given, for all  $\theta$  in  $\mathbb{R}_+^*$  and  $k$  in  $\{1, 2\}$ , by

$$(3.26) \quad g^{(k)}(\theta) = \int_\alpha^{1-\alpha} \left(\frac{1-t}{t}\right)^\theta \log^k\left(\frac{1-t}{t}\right) \nu(dt) = \mathbb{E}^\nu \left[ \rho_0^\theta \log^k(\rho_0) \right].$$

As the result holds for any  $d$  in  $\mathbb{R}_+^*$ , twice-differentiability and Identity (3.26) hold over  $\mathbb{R}_+^*$  (not only over  $[0, d]$ ).

The function  $G$  is twice-differentiable over  $\mathbb{R}_+^*$  and by Identities (3.25), the first and second derivatives of  $G$  are defined, for any  $\theta$  in  $\mathbb{R}_+^*$ , by

$$G^{(1)}(\theta) = \frac{\mathbb{E}^\nu [\rho_0^\theta \log(\rho_0)]}{\mathbb{E}^\nu [\rho_0^\theta]} = \mathbb{E}^\nu \left[ \log(\rho_0) \frac{\rho_0^\theta}{\mathbb{E}^\nu [\rho_0^\theta]} \right],$$

and

$$\begin{aligned} G^{(2)}(\theta) &= \frac{\mathbb{E}^\nu [\rho_0^\theta \log^2(\rho_0)] \mathbb{E}^\nu [\rho_0^\theta] - \mathbb{E}^\nu [\rho_0^\theta \log(\rho_0)]^2}{\mathbb{E}^\nu [\rho_0^\theta]^2} \\ &= \mathbb{E}^\nu \left[ \log^2(\rho_0) \frac{\rho_0^\theta}{\mathbb{E}^\nu [\rho_0^\theta]} \right] - \mathbb{E}^\nu \left[ \log(\rho_0) \frac{\rho_0^\theta}{\mathbb{E}^\nu [\rho_0^\theta]} \right]^2. \end{aligned}$$

- (iii) From Equations (3.22) and (3.23), it appears that the first and second derivative of  $G$  are respectively the expectation and the variance of the real-valued random variable  $\log((1-Y)/Y)$  where  $Y$  is a random variable whose density with respect to the measure  $\nu$  is the application defined by

$$t \in (0, 1) \mapsto \left( \frac{1-t}{t} \right)^\theta \left( \int_0^1 \left( \frac{1-t}{t} \right)^\theta \nu(dt) \right)^{-1}.$$

For any probability distribution  $\mathbb{Q}$  and any real-valued random variable  $W$  on  $(F, \mathcal{F})$  with second moment, denoting  $\mathbb{E}_\mathbb{Q}$  and  $\text{Var}_\mathbb{Q}$  expectation and variance according to  $\mathbb{Q}$ , it is well known that for any  $a$  in  $\mathbb{R}$

$$(3.27) \quad 0 \leq \text{Var}_\mathbb{Q}(W) \leq \mathbb{E}_\mathbb{Q}[(W-a)^2].$$

Interpreting  $G^{(2)}$  as a variance, we derive from inequality (3.27) applied with  $F = \Omega$ ,  $\mathbb{Q} = (\rho_0^\theta / \mathbb{E}^\nu[\rho_0^\theta]) \nu$ ,  $W = \log(\rho_0)$  and  $a = 0$

- (a) For any  $\theta$  in  $\mathbb{R}_+^*$ ,  $G^{(2)}(\theta) \geq 0$  so  $G$  is convex.  
(b) It holds

$$-\log\left(\frac{1-\alpha}{\alpha}\right) \leq \log(\rho_0) \leq \log\left(\frac{1-\alpha}{\alpha}\right) \iff \log^2(\rho_0) \leq \log^2\left(\frac{1-\alpha}{\alpha}\right).$$

Hence, for any  $\theta$  in  $\mathbb{R}_+^*$

$$G^{(2)}(\theta) \leq \log^2\left(\frac{1-\alpha}{\alpha}\right).$$

□

### 3.2.3.2 Proof of Lemma 18

The function  $G$  defined in Equation (3.21) satisfies, for any  $\theta$  in  $\mathbb{R}_+^*$

$$\frac{1}{\theta} \log(\mathbb{E}^\nu[\rho_0^\theta]) = \frac{\log(\mathbb{E}^\nu[\rho_0^\theta]) - \log(1)}{\theta - 0} = \frac{G(\theta) - G(0)}{\theta - 0}.$$

Since  $G$  is differentiable and satisfies  $G^{(1)}(0) = -c_\nu = \mathbb{E}^\nu[\log(\rho_0^\theta)]$ , the convergence of  $\frac{1}{\theta} \log(\mathbb{E}^\nu[\rho_0^\theta])$  towards  $-c_\nu$  when  $\theta$  tends to  $0^+$  is granted.

Since  $G$  is twice differentiable, integral form of Taylor's theorem yields, for any  $\theta$  in  $\mathbb{R}_+$ ,

$$G(\theta) = G(0) + \theta G^{(1)}(0) + \theta^2 \int_0^1 (1-t) G^{(2)}(t\theta) dt,$$

Since Lemma 20 ensures that  $G^{(2)}$  is uniformly bounded by  $\log^2((1-\alpha)/\alpha)$  over  $\mathbb{R}_+^*$ , we get for any  $\theta$  in  $\mathbb{R}_+^*$

$$(3.28) \quad \left| \frac{G(\theta) - G(0)}{\theta - 0} - G^{(1)}(0) \right| \leq \theta \left( \int_0^1 (1-t) dt \right) \sup_{s \in [0, \theta]} \left( G^{(2)}(s) \right) \leq \frac{1}{2} \log^2 \left( \frac{1-\alpha}{\alpha} \right) \theta.$$

Hence the error made approximating  $G^{(1)}(0) = -c_\nu$  by  $\frac{G(\theta)}{\theta}$  is explicitly controlled. We have proved that for all  $\theta$  in  $\mathbb{R}_+^*$

$$\left| \frac{1}{\theta} \log(\mathbb{E}^\nu[\rho_0^\theta]) + c_\nu \right| \leq \frac{1}{2} \log^2 \left( \frac{1-\alpha}{\alpha} \right) \theta.$$

In particular,  $\frac{1}{\theta} \log(\mathbb{E}^\nu[\rho_0^\theta])$  converges to  $-c_\nu$  when  $\theta$  tends to  $0^+$ .

### 3.2.3.3 Proof of Corollary 17

For any  $\theta$  in  $\mathbb{R}_+^*$ , we have the following series of equivalence

$$\mathbb{E}^\nu[\rho_0^\theta] \leq e^{-\frac{3}{4}c_\nu\theta} \Leftrightarrow \frac{1}{\theta} \log(\mathbb{E}^\nu[\rho_0^\theta]) \leq -\frac{3}{4}c_\nu \Leftrightarrow \frac{1}{\theta} \log(\mathbb{E}^\nu[\rho_0^\theta]) + c_\nu \leq \frac{1}{4}c_\nu.$$

Thanks to Inequality (3.19), we also have for any  $\theta$  in  $\mathbb{R}_+^*$

$$\frac{1}{\theta} \log(\mathbb{E}^\nu[\rho_0^\theta]) + c_\nu \leq \left| \frac{1}{\theta} \log(\mathbb{E}^\nu[\rho_0^\theta]) + c_\nu \right| \leq \frac{1}{2} \log^2 \left( \frac{1-\alpha}{\alpha} \right) \theta.$$

Hence, for any  $\theta$  in  $\mathbb{R}_+^*$

$$\theta \leq \frac{c_\nu}{2 \log^2 \left( \frac{1-\alpha}{\alpha} \right)} \implies \mathbb{E}^\nu[\rho_0^\theta] \leq e^{-\frac{3}{4}c_\nu\theta}.$$

Defining  $b(c_0, \alpha)$  as in Equation (3.11), since  $c_\nu \geq c_0$ , Inequality (3.18) yields

$$\mathbb{P}^\nu(N_1 > n) \leq e^{-\frac{c_0 b(c_0, \alpha)}{4} n}.$$

### 3.2.3.4 An intermediate convexity result for the proof of Theorem 16

**Proposition 21.** *Define, for any integer  $l \geq 2$ , the function  $f_l$  by*

$$(3.29) \quad f_l : \lambda \in \mathbb{R} \mapsto \log \left( \mathbb{E}^\nu \left[ e^{\lambda(N_1 - \mathbb{E}^\nu[N_1])} \right] \right) - \lambda(l-1)\mathbb{E}^\nu[N_1].$$

*The function  $f_l$  is well defined, it is convex on the interval  $(-c_0 b(c_0, \alpha)/4, c_0 b(c_0, \alpha)/4)$  and, if  $l = l_0(c_0, \alpha)$  where  $l_0(c_0, \alpha)$  is defined in Equation (3.13), then*

$$(3.30) \quad f_{l_0(c_0, \alpha)}^{(1)} \left( \frac{c_0 b(c_0, \alpha)}{8} \right) \leq \frac{-1}{2} \left( 1 - e^{-\frac{c_0 b(c_0, \alpha)}{8}} \right) e^{-\frac{c_0 b(c_0, \alpha)}{8}},$$

and

$$(3.31) \quad f_{l_0(c_0, \alpha)} \left( \frac{c_0 b(c_0, \alpha)}{8} \right) \leq -\frac{c_0 b(c_0, \alpha)}{16} \left( 1 - e^{-\frac{c_0 b(c_0, \alpha)}{8}} \right) e^{-\frac{c_0 b(c_0, \alpha)}{8}} < 0.$$

*Proof.* By Remark 19, the function  $f_l$  defined in Equation (3.29) is well defined and infinitely differentiable on the interval  $(-c_0b(c_0, \alpha)/4, c_0b(c_0, \alpha)/4)$ .

For any  $\lambda$  in  $(-c_0b(c_0, \alpha)/4, c_0b(c_0, \alpha)/4)$ , usual differentiation rules give its first and second derivatives

$$(3.32) \quad f_l^{(1)}(\lambda) = \frac{\mathbb{E}^\nu \left[ (N_1 - l\mathbb{E}^\nu[N_1]) e^{\lambda N_1} \right]}{\mathbb{E}^\nu \left[ e^{\lambda N_1} \right]},$$

and

$$\begin{aligned} f_l^{(2)}(\lambda) &= \mathbb{E}^\nu \left[ (N_1 - \mathbb{E}^\nu[N_1])^2 \frac{e^{\lambda(N_1 - \mathbb{E}^\nu[N_1])}}{\mathbb{E}^\nu \left[ e^{\lambda(N_1 - \mathbb{E}^\nu[N_1])} \right]} \right] - \mathbb{E}^\nu \left[ (N_1 - \mathbb{E}^\nu[N_1]) \frac{e^{\lambda(N_1 - \mathbb{E}^\nu[N_1])}}{\mathbb{E}^\nu \left[ e^{\lambda(N_1 - \mathbb{E}^\nu[N_1])} \right]} \right]^2 \\ &= \mathbb{E}^\nu \left[ \left( N_1 - \mathbb{E}^\nu[N_1] \frac{e^{\lambda(N_1 - \mathbb{E}^\nu[N_1])}}{\mathbb{E}^\nu \left[ e^{\lambda(N_1 - \mathbb{E}^\nu[N_1])} \right]} \right)^2 \frac{e^{\lambda(N_1 - \mathbb{E}^\nu[N_1])}}{\mathbb{E}^\nu \left[ e^{\lambda(N_1 - \mathbb{E}^\nu[N_1])} \right]} \right] \geq 0. \end{aligned}$$

Proof of (3.30):

For any  $K$  in  $\mathbb{N}$ , any  $\lambda$  in  $(-c_0b(c_0, \alpha)/4, c_0b(c_0, \alpha)/4)$  and any integer  $l \geq 2$ , we have

$$\begin{aligned} \mathbb{E}^\nu \left[ (N_1 - l\mathbb{E}^\nu[N_1]) e^{\lambda N_1} \mathbb{1}_{N_1 \geq K} \right] &\leq \mathbb{E}^\nu \left[ N_1 e^{\lambda N_1} \mathbb{1}_{N_1 \geq K} \right] \\ &\leq \sum_{k=K}^{+\infty} k e^{\lambda k} \mathbb{P}^\nu(N_1 = k) \leq \sum_{k=K}^{+\infty} k e^{\lambda k} \mathbb{P}^\nu(N_1 > k - 1), \end{aligned}$$

Hence, by Corollary 17

$$(3.33) \quad \mathbb{E}^\nu \left[ (N_1 - l\mathbb{E}^\nu[N_1]) e^{\lambda N_1} \mathbb{1}_{N_1 \geq K} \right] \leq e^{\frac{c_0b(c_0, \alpha)}{4}} \sum_{k=K}^{+\infty} k e^{\left(\lambda - \frac{c_0b(c_0, \alpha)}{4}\right)k}.$$

The definition of  $N_1$  in Equation (3.10) implies that  $\mathbb{P}^\nu$  almost surely  $N_1 \geq 1$  so that  $\mathbb{E}^\nu[N_1] \geq 1$  and for any non negative  $\lambda$ ,  $\mathbb{P}^\nu$  almost surely  $e^{\lambda N_1} \geq e^\lambda$ . Thus, for any  $\lambda$  in  $(0, c_0b(c_0, \alpha)/4)$  and any integers  $l$  and  $K$  such that  $l \geq K$

$$\mathbb{E}^\nu \left[ (N_1 - l\mathbb{E}^\nu[N_1]) e^{\lambda N_1} \mathbb{1}_{N_1 < K} \right] \leq \mathbb{E}^\nu \left[ (K - l\mathbb{E}^\nu[N_1]) e^{\lambda N_1} \mathbb{1}_{N_1 < K} \right] \leq (K - l) e^\lambda \mathbb{E}^\nu \left[ \mathbb{1}_{N_1 < K} \right].$$

Corollary 17 ensures that

$$\mathbb{E}^\nu \left[ \mathbb{1}_{N_1 < K} \right] \geq (1 - \mathbb{P}^\nu(\mathbb{1}_{N_1 > K-1})) \geq \left(1 - e^{-\frac{c_0b(c_0, \alpha)}{4}(K-1)}\right).$$

As  $(K - l) \leq 0$

$$(3.34) \quad \begin{aligned} \mathbb{E}^\nu \left[ (N_1 - l\mathbb{E}^\nu[N_1]) e^{\lambda N_1} \mathbb{1}_{N_1 < K} \right] &\leq (K - l) e^\lambda (1 - \mathbb{P}^\nu(N_1 \geq K - 1)) \\ &\leq (K - l) e^\lambda \left(1 - e^{-\frac{c_0b(c_0, \alpha)}{4}(K-1)}\right). \end{aligned}$$

Hence, for any  $l \geq K$ , combining Equation (3.33) and Equation (3.34)

$$(3.35) \quad \begin{aligned} &\mathbb{E}^\nu \left[ (N_1 - l\mathbb{E}^\nu[N_1]) e^{\lambda N_1} \right] \\ &= \mathbb{E}^\nu \left[ (N_1 - l\mathbb{E}^\nu[N_1]) e^{\lambda N_1} (\mathbb{1}_{N_1 < K} + \mathbb{1}_{K \leq N_1}) \right] \\ &\leq (K - l) e^\lambda \left(1 - e^{-\frac{c_0b(c_0, \alpha)}{4}(K-1)}\right) + e^{\frac{c_0b(c_0, \alpha)}{4}} \sum_{k=K}^{+\infty} k e^{\left(\lambda - \frac{c_0b(c_0, \alpha)}{4}\right)k}. \end{aligned}$$



Since  $\exp(-c_0 b(c_0, \alpha)/8)$  is less than 1, the series with terms  $k \exp(-c_0 b(c_0, \alpha)k/8)$  is convergent and it is possible to choose  $\lambda_0 = \lambda_0(c_0, \alpha) = c_0 b(c_0, \alpha)/8$  such that the rest of the series with terms  $k \exp((\lambda_0 - c_0 b(c_0, \alpha)/4)k)$  tends to 0. Since  $\exp(-c_0 b(c_0, \alpha)/4)$  is less than 1, geometric progression with common ratio  $\exp(-c_0 b(c_0, \alpha)/4)$  also converges to 0, it is then possible to choose for  $K$  a value  $K_0 = K_0(c_0, \alpha)$  in  $\mathbb{N}$  so that

$$e^{\frac{c_0 b(c_0, \alpha)}{4}} \sum_{k=K_0}^{+\infty} k e^{\left(\lambda_0 - \frac{c_0 b(c_0, \alpha)}{4}\right)k} = \sum_{k=K_0}^{+\infty} k e^{-\frac{c_0 b(c_0, \alpha)}{8}(k-2)} \leq \frac{1}{2},$$

and simultaneously

$$e^{\lambda_0} \left(1 - e^{-\frac{c_0 b(c_0, \alpha)}{4}(K_0-1)}\right) = e^{\frac{c_0 b(c_0, \alpha)}{8}} \left(1 - e^{-\frac{c_0 b(c_0, \alpha)}{4}(K_0-1)}\right) \geq 1.$$

Choosing for  $l$  the value  $l_0 = l_0(c_0, \alpha)$  as defined in Equation (3.13),  $l_0 \geq K_0 + 1$  and thus Equation (3.35) becomes

$$(3.36) \quad \mathbb{E}^\nu \left[ (N_1 - l_0(c_0, \alpha)) \mathbb{E}^\nu [N_1] e^{\lambda_0(c_0, \alpha)N_1} \right] \leq -1 + \frac{1}{2} \leq -\frac{1}{2}.$$

As  $N_1 > 0$   $\mathbb{P}^\nu$ -a.s.,  $\mathbb{P}^\nu(N_1 = 0) = 0$ . Hence, Corollary 17 yields

$$(3.37) \quad \begin{aligned} \mathbb{E}^\nu \left[ e^{\lambda_0 N_1} \right] &= \sum_{n=1}^{+\infty} e^{\lambda_0 n} \mathbb{P}^\nu(N_1 = n) \leq \sum_{n=1}^{+\infty} e^{\lambda_0 n} \mathbb{P}^\nu(N_1 > n-1) \\ &\leq e^{\lambda_0} \left( \sum_{n=1}^{+\infty} e^{\lambda_0(n-1)} e^{-\frac{c_0 b(c_0, \alpha)}{4}(n-1)} \right) \leq \frac{e^{\lambda_0}}{1 - e^{\lambda_0 - \frac{c_0 b(c_0, \alpha)}{4}}}. \end{aligned}$$

Combining Equations (3.36) and (3.37), we derive that

$$\frac{\mathbb{E}^\nu \left[ (N_1 - l_0(c_0, \alpha)) \mathbb{E}^\nu [N_1] e^{\lambda_0(c_0, \alpha)N_1} \right]}{\mathbb{E}^\nu \left[ e^{\lambda_0(c_0, \alpha)N_1} \right]} \leq \frac{-1}{2} \left(1 - e^{-\lambda_0}\right) e^{-\lambda_0},$$

which is exactly Equation (3.30).

Proof of (3.31):

Since  $f_{l_0}^{(2)}$  is non negative,  $f_{l_0}^{(1)}$  is non decreasing and so for any  $\lambda$  in  $[0, \lambda_0(c_0, \alpha)]$

$$f_{l_0}^{(1)}(\lambda) \leq f_{l_0}^{(1)}(\lambda_0) \leq \frac{-1}{2} \left(1 - e^{-\lambda_0}\right) e^{-\lambda_0}.$$

Thanks to the fundamental theorem of calculus, we derive that, for any  $\lambda$  in  $[0, \lambda_0(c_0, \alpha)]$

$$f_{l_0}(\lambda) = \int_0^\lambda f_{l_0}^{(1)}(t) dt \leq \int_0^\lambda f_{l_0}^{(1)}(\lambda_0) dt \leq \frac{-\lambda}{2} \left(1 - e^{-\lambda_0}\right) e^{-\lambda_0}.$$

This shows Equation (3.31) for  $\lambda = \lambda_0$ . □

### 3.2.3.5 Final step of the proof of Theorem 16

When  $k = 0$ , Inequality (3.14) is immediate since the event is empty and the upper-bound takes value 1. Hence, we assume hereafter that  $k > 0$ .

Since the variables  $(N_j - N_{j-1})_{j \in \mathbb{N}^*}$  have the same distribution as  $N_1$  under  $\mathbb{P}^\nu$  (and so the same expectation), we can decompose  $N_k$  for any  $k$  in  $\mathbb{N}^*$  as

$$N_k = \sum_{j=1}^k (N_j - N_{j-1}) = k\mathbb{E}^\nu [N_1] + \sum_{j=1}^k (N_j - N_{j-1} - \mathbb{E}^\nu [N_j - N_{j-1}]) .$$

Hence,

$$N_k > l\mathbb{E}^\nu [N_1]k \quad \iff \quad \sum_{j=1}^k (N_j - N_{j-1} - \mathbb{E}^\nu [N_j - N_{j-1}]) > (l-1)\mathbb{E}^\nu [N_1]k .$$

For any integer  $l \geq 2$ , any  $k$  in  $\mathbb{N}^*$  and any  $\lambda > 0$ , Markov inequality thus yields

$$(3.38) \quad \mathbb{P}^\nu(N_k > l\mathbb{E}^\nu [N_1]k) = \mathbb{P}^\nu \left( e^{\lambda \sum_{j=1}^k (N_j - N_{j-1} - \mathbb{E}^\nu [N_j - N_{j-1}])} > e^{\lambda(l-1)\mathbb{E}^\nu [N_1]k} \right) \\ \leq e^{-k\lambda(l-1)\mathbb{E}^\nu [N_1]} \mathbb{E}^\nu \left[ \prod_{j=1}^k e^{\lambda(N_j - N_{j-1} - \mathbb{E}^\nu [N_j - N_{j-1}])} \right] .$$

Since the variables  $(N_k - N_{k-1})_{k \in \mathbb{N}^*}$  are i.i.d. and distributed as  $N_1$ , it follows that, for any  $k$  in  $\mathbb{N}^*$

$$(3.39) \quad \mathbb{E}^\nu \left[ \prod_{j=1}^k e^{\lambda(N_j - N_{j-1} - \mathbb{E}^\nu [N_j - N_{j-1}])} \right] = \prod_{j=1}^k \mathbb{E}^\nu \left[ e^{\lambda(N_j - N_{j-1} - \mathbb{E}^\nu [N_j - N_{j-1}])} \right] \\ = \left( \mathbb{E}^\nu \left[ e^{\lambda(N_1 - \mathbb{E}^\nu [N_1])} \right] \right)^k .$$

Plugging Identity (3.39) in inequality (3.38) thus yields

$$(3.40) \quad \mathbb{P}^\nu(N_k > l\mathbb{E}^\nu [N_1]k) \leq \left( e^{\log(\mathbb{E}^\nu [e^{\lambda(N_1 - \mathbb{E}^\nu [N_1])}]) - \lambda(l-1)\mathbb{E}^\nu [N_1]} \right)^k .$$

Taking  $l = l_0(c_0, \alpha)$  and defining the positive constant  $K_3(c_0, \alpha)$  as in Equation (3.12), Equation (3.31) of Proposition 21 ensures that there exists  $\lambda_0 = -(c_0 b(c_0, \alpha))/8$  such that

$$\log \left( \mathbb{E}^\nu \left[ e^{\lambda_0(N_1 - \mathbb{E}^\nu [N_1])} \right] \right) - \lambda_0(l_0(c_0, \alpha) - 1)\mathbb{E}^\nu [N_1] \leq -K_3(c_0, \alpha) .$$

Plugging this inequality in Equation (3.40), for any  $k$  in  $\mathbb{N}$

$$\mathbb{P}^\nu(N_k > l_0(c_0, \alpha)\mathbb{E}^\nu [N_1]k) \leq e^{-K_3(c_0, \alpha)k} .$$

Theorem 16 is thus proved.

### 3.2.4 A different interpretation of the BPIRE

The second part of the proof of Theorem 14 strongly relies on a different interpretation of the BPIRE that is presented in this section.

Since the BPIRE  $(Z_t)_{t \in \mathbb{N}}$  defined in (3.6) starts from 0 ( $Z_0 = 0$ ) and since, at each generation, an individual immigrates in the population, it is possible, at each generation,

to distinguish individuals within the population according to their immigrant ancestor. In other words,  $Z_t$  is the sum over  $s$  in  $\{0, \dots, t-1\}$  of the number  $Z_{s,t}$  of individuals born at time  $t$  which are descendants from the immigrant (only one per generation by definition) that entered at time  $s$ :

$$(3.41) \quad Z_t = \sum_{s=0}^{t-1} Z_{s,t}.$$

For any fixed integer  $s$ , the sequence  $(Z_{s,s+t})_{t \in \mathbb{N}}$  constitutes a branching process (without immigration) in random environment:  $Z_{s,s} = 0$  and for any  $t$  in  $\mathbb{N}$ ,  $Z_{s,s+t+1}$  can be decomposed as

$$(3.42) \quad Z_{s,s+t+1} = \sum_{j=1}^{Z_{s,s+t}} \xi_j^{s,s+t},$$

where the  $(\xi_j^{s,s+t})_{(s,t,j) \in \mathbb{N} \times \mathbb{N} \times \mathbb{N}^*}$  are independent conditionally to the  $\sigma$ -field generated by  $\omega$  and such that, for any  $s, t$  and  $k$  in  $\mathbb{N}$  and any  $j$  in  $\mathbb{N}^*$

$$(3.43) \quad \mathbf{P}(\xi_j^{s,s+t} = k | \omega) = \omega_{s+t} (1 - \omega_{s+t})^k.$$

### 3.2.5 Theorem 22

The links between the sequence  $N_k$  and the reinterpretation of the BPIRE given in Equation (3.41), enables to control the probability that  $T$  overpasses  $N_k$  under quenched probability.

**Theorem 22.** *Assume that Hypotheses **F1** and **F2** hold. Consider the positive constant  $K_6(c_0)$  defined by*

$$(3.44) \quad K_6(c_0) = \min \left( \frac{c_0}{8} - \frac{\log(K_4(c_0))}{2K_5(c_0)}, -\frac{\log(1 - z(c_0))}{2K_5(c_0)} \right),$$

where the constants  $K_4(c_0)$  and  $K_5(c_0)$  are defined in Equation (3.47). Then, for any  $\omega$  in  $E$  and any  $k$  in  $\mathbb{N}$ ,

$$(3.45) \quad \mathbf{P}_\omega(T > N_k) \leq 2 e^{-K_6(c_0)k}.$$

### 3.2.6 Sketch of proof of Theorem 22

**In this section, we consider a positive integer  $k$  and an environment  $\omega$  in  $E$  both fixed. All results are obtained under  $\mathbf{P}_\omega$ .**

The proof of Theorem 22 is based on the reinterpretation of the BPIRE presented in Section 3.2.4 and on the introduction of a specific sequence of integer indices related to this interpretation. First, we control the tail distribution of the decrements of this sequence of indices. Then, we relate these indices to the instants where the BPIREG( $\nu, 0$ ) returns to 0. Finally, we combine these phenomena to bound from above the probability of  $\{T > N_k\}$ .

### 3.2.6.1 Definition of the sequence $(k_i)_{i \in \mathbb{N}}$ and control of its decrements

Let us first introduce the specific sequence of integers  $(k_i)_{i \in \mathbb{N}}$  from the reinterpretation of the BPIRE given in Section 3.2.4.

**Definition 23.** Define on the canonical space the random sequence  $(k_i)_{i \in \mathbb{N}}$  by  $k_0 = k$  and for all  $i > 0$

1. if  $k_i > 0$ , we set

- (a) on  $\{Z_{N_{k_i}} > 0\}$ ,  $k_{i+1} = \max\left(l < k_i : \sum_{s=N_l}^{N_{k_i}-1} Z_{s, N_{k_i}} > 0 \text{ and } \sum_{s=N_{l+1}}^{N_{k_i}-1} Z_{s, N_{k_i}} = 0\right)$ .
- (b) on  $\{Z_{N_{k_i}} = 0\}$ ,  $k_{i+1} = k_i - 1$ .

2. if  $k_i = 0$ , we set  $k_{i+1} = 0$ .

**Remark 24.** The definition of the sequence  $(k_i)_{i \in \mathbb{N}}$  in Definition 23 is totally clear excepted in one situation: when  $Z_{N_{k_i}} > 0$  and  $k_i > 0$ , why would  $k_{i+1}$  be well-defined by (1a) ?

Let us explain why  $k_{i+1}$  is well defined by (1a) on the set  $\{Z_{N_{k_i}} > 0\} \cap \{k_i > 0\}$ .

When  $k_i > 0$ , then,  $N_{k_i} > 0$  so that  $Z_{N_{k_i}}$  is defined by

$$Z_{N_{k_i}} = \sum_{s=0}^{N_{k_i}-1} Z_{s, N_{k_i}}.$$

Since the random variables  $(Z_{s, N_{k_i}})_{0 \leq s \leq N_{k_i}}$  are non negative,  $Z_{N_{k_i}}$  is positive, if and only if, it exists some integer  $s$  in  $\{0, \dots, N_{k_i} - 1\}$  such that  $Z_{s, N_{k_i}}$  is positive.

Denote  $s_0$  the greatest integer in  $\{0, \dots, N_{k_i} - 1\}$  such that  $Z_{s_0, N_{k_i}} > 0$  : for any integer  $s$  in  $\{s_0 + 1, \dots, N_{k_i} - 1\}$ , we have  $Z_{s, N_{k_i}} = 0$ .

For any integer  $l$  in  $\{0, \dots, k_i - 1\}$  such that  $s_0 < N_l$ , it holds

$$\sum_{s=N_l}^{N_{k_i}-1} Z_{s, N_{k_i}} = 0 \quad \text{and a fortiori} \quad \sum_{s=N_{l+1}}^{N_{k_i}-1} Z_{s, N_{k_i}} = 0,$$

and for the only integer  $l$  in  $\{0, \dots, k_i - 1\}$  such that  $N_l \leq s_0 < N_{l+1}$  (necessary  $l \leq s_0$ ), it holds

$$\sum_{s=N_l}^{N_{k_i}-1} Z_{s, N_{k_i}} > 0 \quad \text{and} \quad \sum_{s=N_{l+1}}^{N_{k_i}-1} Z_{s, N_{k_i}} = 0.$$

In this context,  $k_{i+1}$  is thus well-defined :  $k_{i+1}$  is the integer  $l$  in  $\{0, \dots, k_i - 1\}$  such that  $N_l \leq s_0 < N_{l+1}$ .

The sequence  $(k_i)_{i \in \mathbb{N}}$  is strictly decreasing until it reaches 0. Then, it remains constantly equal to 0. Thanks to Lemmas 34 and 35, we can quantify its rate of convergence to 0. Let us state some results on the tail distribution of  $k_i - k_{i+1}$ .

**Lemma 25.** For any  $i$  in  $\mathbb{N}$  and any  $r$  in  $\mathbb{N}$

$$(3.46) \quad \mathbf{P}_\omega(k_i - k_{i+1} > r | k_0, \dots, k_i) \leq \frac{e^{-r \frac{c_V}{2}}}{1 - e^{-\frac{c_V}{2}}}.$$

Let us derive from this control the existence of an exponential moment for the  $k_i - k_{i+1}$  and a control on the existing exponential moment.

**Lemma 26.** *Let  $K_4(c_0)$  and  $K_5(c_0)$  denote the following quantities*

$$(3.47) \quad K_4(c_0) = 1 + \frac{e^{-\frac{c_0}{2}}}{\left(1 - e^{-\frac{c_0}{4}}\right)^2} \quad \text{and} \quad K_5(c_0) = 1 + \left\lfloor \frac{4 \log(K_4(c_0))}{c_0} \right\rfloor,$$

Let  $(k_i)_{i \in \mathbb{N}}$  denote the sequence introduced in Definition 23. Then, for any  $i$  in  $\mathbb{N}$  and any  $\lambda$  in  $(0, c_0/4]$

$$(3.48) \quad \mathbf{E}_\omega \left[ e^{\lambda(k_i - k_{i+1})} \mid k_0, \dots, k_i \right] \leq K_4(c_0) e^{\lambda K_5(c_0)}.$$

For any integer  $l$ , we can then control the probability that there exists  $i \leq l$  such that  $k_i = 0$ .

**Corollary 27.** *For any  $l$  in  $\mathbb{N}$  and any  $\lambda$  in  $(0, \frac{c_0}{4}]$*

$$(3.49) \quad \mathbf{P}_\omega \left( \bigcup_{i=0}^l \{k_i = 0\} \right) \leq e^{-\lambda k} \left( K_4(c_0) e^{\lambda K_5(c_0)} \right)^l,$$

where  $(k_i)_{i \in \mathbb{N}}$  is determined in Definition 23 and  $K_4(c_0)$  and  $K_5(c_0)$  are both defined in (3.47).

**Remark 28.** *Let  $l = l_1 = l_1(c_0, k)$  where*

$$(3.50) \quad l_1(c_0, k) = \left\lfloor \frac{k}{2K_5(c_0)} \right\rfloor.$$

Corollary 27 implies that, for any  $\lambda$  in  $(0, \frac{c}{4}]$

$$(3.51) \quad \begin{aligned} \mathbf{P}_\omega \left( \bigcup_{i=0}^{l_1} \{k_i = 0\} \right) &\leq e^{-\lambda k} \left( K_4(c_0) e^{\lambda K_5(c_0)} \right)^{l_1} \leq e^{-\lambda k} \left( K_4(c_0) e^{\lambda K_5(c_0)} \right)^{\left\lfloor \frac{k}{2K_5(c_0)} \right\rfloor} \\ &\leq e^{-\lambda k} \left( K_4(c_0) e^{\lambda K_5(c_0)} \right)^{\frac{k}{2K_5(c_0)}} \leq e^{-\lambda k} K_4(c_0)^{\frac{k}{2K_5(c_0)}} e^{\lambda \frac{k}{2}} \\ &\leq e^{-\frac{\lambda}{2} k} K_4(c_0)^{\frac{k}{2K_5(c_0)}}. \end{aligned}$$

### 3.2.6.2 Bounding the probability of return to 0 at $N_{k_i}$ of the BPIRE

Define the two functions  $y : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$  and  $z : \mathbb{R}_+^* \rightarrow (0, 1)$  that will be useful in the sequel defining for any  $c$  in  $\mathbb{R}_+^*$

$$y(c) = \frac{e^{-\frac{c}{2}}}{1 - e^{-\frac{c}{2}}} \quad \text{and} \quad z(c) = \frac{1}{4} e^{-4y(c)}.$$

**Lemma 29.** *For any  $i$  in  $\mathbb{N}^*$*

$$(3.52) \quad \mathbf{P}_\omega \left( Z_{N_{k_i}} = 0 \mid k_0, \dots, k_i, (Z_{s,t})_{N_{k_i} \leq s < t \leq k} \right) \geq z(c_\nu),$$

where  $(k_i)_{i \in \mathbb{N}}$  is defined in Definition 23,  $(Z_n)_{n \in \mathbb{N}}$  in Equation (3.6),  $(N_k)_{k \in \mathbb{N}}$  in Equation (3.10) and  $(Z_{s,t})_{(s,t) \in \mathbb{N}^* \times \mathbb{N}}$  in Equations (3.42) and (3.43).

This lower-bound enables to upper-bound the probability that the BPIRE does not return to 0 at any of the moments  $N_{k_0}, \dots, N_{k_l}$ .

**Lemma 30.** *For any  $l$  in  $\mathbb{N}^*$*

$$(3.53) \quad \mathbf{P}_\omega \left( \{k_l > 0\} \cap \left( \bigcap_{i=0}^l \{Z_{N_{k_i}} > 0\} \right) \right) \leq (1 - z(c_\nu))^{l+1},$$

where  $(k_i)_{i \in \mathbb{N}}$  is defined in (23) and  $(Z_n)_{n \in \mathbb{N}}$  and  $(N_k)_{k \in \mathbb{N}}$  are respectively defined in Equations (3.6) and (3.10).

The reinterpretation of the BPIREG( $\nu, 0$ ) given in Section 3.2.4 combined with Remark 28 and Lemma 30 finally enables to prove Theorem 22 in Section 3.2.7.7.

### 3.2.7 Detailed proof of Theorem 22

#### 3.2.7.1 Intermermediate results for the proof of Lemmas 25 and 29

To establish Lemmas 25 and 29, the following intermediate results are useful.

**Lemma 31.** *For any environment  $\omega$  in  $E$ , it holds*

$$(i) \text{ for any } k \text{ and } s \text{ in } \mathbb{N} \text{ such that } s \leq N_k - 1, \sum_{i=s}^{N_k-1} \log(\rho_i) \leq -(N_k - s) \frac{c_\nu}{2},$$

$$(ii) \text{ for any } t \text{ and } s \text{ in } \mathbb{N} \text{ such that } t > s, \mathbf{E}_\omega[Z_{s,t}] = \exp \left( \sum_{i=s}^{t-1} \log(\rho_i) \right),$$

$$(iii) \text{ for any } k \text{ in } \mathbb{N}, \mathbf{E}_\omega[Z_{N_k}] = \sum_{s=0}^{N_k-1} \exp \left( \sum_{i=s}^{N_k-1} \log(\rho_i) \right),$$

$$(iv) \text{ for any } k \text{ in } \mathbb{N}^*, \mathbf{E}_\omega[Z_{N_k}] > \rho_{N_k-1} \mathbf{E}_\omega[Z_{N_k-1}],$$

where for any  $i$  in  $\mathbb{Z}$ , the ratio  $\rho_i$  is defined in Equation (3.4).

**Proof of Lemma 31.** Consider a fixed environment  $\omega$  in  $E$ .

(i) Fix  $k$  in  $\mathbb{N}^*$  and  $s$  in  $\mathbb{N}$  such that  $s \leq N_k - 1$ .

By definition of  $(S_n)_{n \in \mathbb{N}}$  in Equation (3.9)

$$\sum_{i=s}^{N_k-1} \log(\rho_i) = (S_{N_k} - S_s) - (N_k - s) \frac{c_\nu}{2},$$

and by definition of  $(N_k)_{k \in \mathbb{N}}$  as downward-ladder indices in Equation (3.10), we necessarily have  $S_{N_k} - S_s \leq 0$  so that

$$\sum_{i=s}^{N_k-1} \log(\rho_i) \leq -(N_k - s) \frac{c_\nu}{2}.$$

(ii) Fix  $t$  in  $\mathbb{N}^*$  and  $s$  in  $\mathbb{N}$  such that  $t > s$ .

Thanks to the definition of  $(\xi_j^{s,s+t})_{(s,t,j) \in \mathbb{N} \times \mathbb{N}^* \times \mathbb{N}^*}$  in Section 3.2.4, for any  $i$  in  $\mathbb{N}$  such that  $i \geq s + 1$ , the random variables  $(\xi_j^{s,i})_{j \in \mathbb{N}^*}$  are i.i.d. with common distribution  $\mathcal{G}_{\mathbb{N}}(\omega_i)$  (whose mean is  $(1 - \omega_i)/\omega_i = \rho_i$ ). Hence Wald's identity ensures

$$(3.54) \quad \mathbf{E}_{\omega} \left[ \sum_{j=1}^{Z_{s,i}} \xi_j^{s,i} \right] = \mathbf{E}_{\omega} [Z_{s,i}] \mathbf{E}_{\omega} [\xi_1^{s,i}] = \mathbf{E}_{\omega} [Z_{s,i}] \frac{1 - \omega_i}{\omega_i} = \mathbf{E}_{\omega} [Z_{s,i}] \rho_i.$$

Definition of  $(Z_{s,s+t})_{(s,t) \in \mathbb{N} \times \mathbb{N}^*}$  by means of the  $(\xi_j^{s,s+t})_{(s,t,j) \in \mathbb{N} \times \mathbb{N}^* \times \mathbb{N}^*}$  in Section 3.2.4 and recursive use of Equation (3.54) thus lead to

$$(3.55) \quad \mathbf{E}_{\omega} [Z_{s,t}] = \mathbf{E}_{\omega} [Z_{s,s+1}] \prod_{i=s+1}^{t-1} \rho_i = \prod_{i=s}^{t-1} \rho_i = \exp \left( \sum_{i=s}^{t-1} \log(\rho_i) \right).$$

(iii) From interpretation of  $(Z_t)_{t \in \mathbb{N}}$  given in Equation (3.41), for any  $k$  in  $\mathbb{N}$

$$Z_{N_k} = \sum_{s=0}^{N_k-1} Z_{s,N_k}.$$

Since  $\omega$  is fixed,  $N_k$  is deterministic under  $\mathbf{P}_{\omega}$  so that linearity of the expectation yields

$$\mathbf{E}_{\omega} [Z_{N_k}] = \sum_{s=0}^{N_k-1} \mathbf{E}_{\omega} [Z_{s,N_k}].$$

By Equation (3.55), we then derive that for any  $k$  in  $\mathbb{N}$

$$(3.56) \quad \mathbf{E}_{\omega} [Z_{N_k}] = \sum_{s=0}^{N_k-1} \prod_{i=s}^{N_k-1} \rho_i = \sum_{s=0}^{N_k-1} \exp \left( \sum_{i=s}^{N_k-1} \log(\rho_i) \right).$$

(iv) For any  $k$  in  $\mathbb{N}$ , Equation (3.56) yields

$$\mathbf{E}_{\omega} [Z_{N_k}] = \sum_{s=0}^{N_k-1} \prod_{i=s}^{N_k-1} \rho_i = \rho_{N_k-1} \left( \sum_{s=0}^{N_k-1} \prod_{i=s}^{N_k-2} \rho_i \right) = \rho_{N_k-1} \left( 1 + \sum_{s=0}^{N_k-2} \prod_{i=s}^{N_k-2} \rho_i \right).$$

But Equation (3.56) also yields

$$\sum_{s=0}^{N_k-2} \prod_{i=s}^{N_k-2} \rho_i = \mathbf{E}_{\omega} [Z_{N_{k-1}}].$$

Hence,

$$\mathbf{E}_{\omega} [Z_{N_k}] = \rho_{N_k-1} (1 + \mathbf{E}_{\omega} [Z_{N_{k-1}}]) > \rho_{N_k-1} \mathbf{E}_{\omega} [Z_{N_{k-1}}].$$

□

**Remark 32.** *Plugging Inequality (i) in Identity (iii) of Lemma 31, for any  $k$  in  $\mathbb{N}$*

$$\mathbf{E}_\omega [Z_{N_k}] \leq \sum_{s=0}^{N_k-1} \exp\left(-s \frac{c_\nu}{2}\right) = \sum_{s=1}^{N_k} \exp\left(-s \frac{c_\nu}{2}\right) \leq e^{-\frac{c_\nu}{2}} \frac{1 - e^{-N_k \frac{c_\nu}{2}}}{1 - e^{-\frac{c_\nu}{2}}} \leq \frac{e^{-\frac{c_\nu}{2}}}{1 - e^{-\frac{c_\nu}{2}}} \leq y(c_\nu).$$

*Plugged in Inequality (iv) of Lemma 31, this gives, for any  $k$  in  $\mathbb{N}^*$*

$$(3.57) \quad \mathbf{E}_\omega [Z_{N_{k-1}}] \leq \frac{y(c_\nu)}{\rho_{N_{k-1}}}.$$

**Remark 33.** *By definition of the sequence  $(N_k)_{k \in \mathbb{N}}$  in Equation (3.10), for any  $k$  in  $\mathbb{N}^*$*

$$S_{N_k} \leq S_{N_{k-1}} \leq S_{N_{k-1}}.$$

*By definition of  $(S_n)_{n \in \mathbb{N}}$ , this yields*

$$S_{N_k} - S_{N_{k-1}} = \log(\rho_{N_{k-1}}) + \frac{c_\nu}{2} \leq 0.$$

*Since  $c_\nu > 0$ , this implies that  $\log(\rho_{N_{k-1}}) \leq 0$  which is equivalent to*

$$(3.58) \quad \omega_{N_{k-1}} > \frac{1}{2}.$$

**Lemma 34.** *For any  $k$  in  $\mathbb{N}^*$  and any  $\omega$  in  $E$*

$$(3.59) \quad \mathbf{P}_\omega (T \leq N_k) \geq \mathbf{P}_\omega (Z_{N_k} = 0) \geq z(c_\nu).$$

**Proof of Lemma 34.** Since  $T$  is defined as the smallest positive moment at which  $(Z_t)_{t \in \mathbb{N}}$  returns to 0 and since  $N_k$  is a positive integer, for any positive integer  $k$

$$(3.60) \quad \{Z_{N_k} = 0\} \subset \{T \leq N_k\}.$$

Hence,  $\mathbf{P}_\omega (T \leq N_k) \geq \mathbf{P}_\omega (Z_{N_k} = 0)$  and any lower-bound of  $\mathbf{P}_\omega (Z_{N_k} = 0)$  is thus also a lower-bound of  $\mathbf{P}_\omega (T \leq N_k)$ . Hereafter, we determine a lower-bound for  $\mathbf{P}_\omega (Z_{N_k} = 0)$ .

Since  $\omega$  is fixed, each of the indices  $(N_k)_{k \in \mathbb{N}}$  is deterministic under  $\mathbf{P}_\omega$ . The process  $(Z_t)_{t \in \mathbb{N}}$  being an inhomogeneous Markov chain under  $\mathbf{P}_\omega$  with kernels given by Equation (3.7), Markov property yields

$$(3.61) \quad \mathbf{P}_\omega (Z_{N_k} = 0 | Z_0, \dots, Z_{N_{k-1}}) = \mathbf{P}_\omega (Z_{N_k} = 0 | Z_{N_{k-1}}) = (\omega_{N_{k-1}})^{Z_{N_{k-1}}+1}.$$

Hence, for any  $L$  in  $\mathbb{R}_+^*$

$$(3.62) \quad \begin{aligned} \mathbf{P}_\omega (Z_{N_k} = 0) &= \sum_{l=0}^{+\infty} \mathbf{P}_\omega (Z_{N_k} = 0 | Z_{N_{k-1}} = l) \mathbf{P}_\omega (Z_{N_{k-1}} = l) \\ &= \sum_{l=0}^{+\infty} (\omega_{N_{k-1}})^{l+1} \mathbf{P}_\omega (Z_{N_{k-1}} = l) \geq \sum_{\substack{l \in \mathbb{N} \\ l \leq L}} (\omega_{N_{k-1}})^{l+1} \mathbf{P}_\omega (Z_{N_{k-1}} = l) \\ &\geq \sum_{\substack{l \in \mathbb{N} \\ l \leq L}} (\omega_{N_{k-1}})^{L+1} \mathbf{P}_\omega (Z_{N_{k-1}} = l) = \mathbf{P}_\omega (Z_{N_{k-1}} \leq L) (\omega_{N_{k-1}})^{L+1}. \end{aligned}$$



Choosing for  $L$  the value  $L_0 = 2y(c)/\rho_{N_k-1}$ , Equation (3.57) ensures that  $L \geq 2\mathbf{E}_\omega [Z_{N_k-1}]$  and since  $Z_{N_k-1} \geq 0$ , we have

$$\begin{aligned} \mathbf{E}_\omega [Z_{N_k-1}] &= \mathbf{E}_\omega [Z_{N_k-1} \mathbf{1}_{Z_{N_k-1} > L}] + \mathbf{E}_\omega [Z_{N_k-1} \mathbf{1}_{Z_{N_k-1} \leq L}] \\ &\geq \mathbf{E}_\omega [Z_{N_k-1} \mathbf{1}_{Z_{N_k-1} > L}] \geq L \mathbf{P}_\omega (Z_{N_k-1} > L) \\ &\geq 2\mathbf{E}_\omega [Z_{N_k-1}] \mathbf{P}_\omega (Z_{N_k-1} > L). \end{aligned}$$

Since  $\mathbf{E}_\omega [Z_{N_k-1}]$  is positive, it thus necessarily holds

$$(3.63) \quad \mathbf{P}_\omega (Z_{N_k-1} > L) \leq \frac{1}{2} \quad \text{or equivalently} \quad \mathbf{P}_\omega (Z_{N_k-1} \leq L) \geq \frac{1}{2}.$$

Plugging Equation (3.63) in Equation (3.62), we thus get

$$(3.64) \quad \mathbf{P}_\omega (Z_{N_k} = 0) \geq \frac{1}{2} (\omega_{N_k-1})^{L+1}.$$

The function  $g : x \in \mathbb{R} \mapsto 1 - x - e^{-2x}$  being increasing on  $(-\infty, \log(\sqrt{2})]$  and decreasing on  $[\log(\sqrt{2}), +\infty)$  with  $g(0) = 0$  and  $g(0.5) > 0.13 > 0$ ,  $g$  is non negative on  $[0, 0.5]$  and so for any  $x$  in  $[0, 0.5]$

$$(3.65) \quad 1 - x \geq e^{-2x}.$$

By the lower-bound (3.58),  $1 - \omega_{N_k-1}$  in  $[0, 1/2)$  so that Inequality (3.65) yields

$$\omega_{N_k-1} = 1 - (1 - \omega_{N_k-1}) \geq e^{-2(1-\omega_{N_k-1})},$$

and using again that  $\omega_{N_k-1}$  in  $(1/2, 1]$  ensured by (3.58), we get

$$\begin{aligned} (\omega_{N_k-1})^{L+1} &\geq \omega_{N_k-1} (\omega_{N_k-1})^L \geq \frac{1}{2} e^{-2(1-\omega_{N_k-1})L} \\ &\geq \frac{1}{2} e^{-2(1-\omega_{N_k-1})\frac{2y(c_\nu)}{\rho_{N_k-1}}} \geq \frac{1}{2} e^{-4y(c_\nu)\omega_{N_k-1}} \geq \frac{1}{2} e^{-4y(c_\nu)} = 2z(c_\nu). \end{aligned}$$

Plugging that lower-bound on  $(\omega_{N_k-1})^{L+1}$  in Inequality (3.64), it leads to

$$\mathbf{P}_\omega (Z_{N_k} = 0) \geq \mathbf{P}_\omega (Z_{N_k-1} \leq L) (\omega_{N_k-1})^{L+1} \geq \frac{1}{2} 2z(c_\nu) \geq z(c_\nu).$$

By Inclusion (3.60), we finally get the desired result

$$(3.66) \quad \mathbf{P}_\omega (T \leq N_k) \geq \mathbf{P}_\omega (Z_{N_k} = 0) \geq \frac{1}{4} e^{-4y(c_\nu)} = z(c_\nu).$$

□

**Lemma 35.** For any  $k$  in  $\mathbb{N}^*$ , any  $j$  in  $\mathbb{N}$  and any  $\omega$  in  $E$

$$(3.67) \quad \mathbf{P}_\omega \left( \sum_{s=0}^{N_j} Z_{s, N_{j+k}} > 0 \right) \leq \frac{e^{-k\frac{c_\nu}{2}}}{1 - e^{-\frac{c_\nu}{2}}}.$$

**Proof of Lemma 35.** Fix  $k$  in  $\mathbb{N}^*$ ,  $j$  in  $\mathbb{N}$  and  $\omega$  in  $E$ .

First, by Markov inequality

$$(3.68) \quad \mathbf{P}_\omega \left( \sum_{s=0}^{N_j} Z_{s, N_{j+k}} > 0 \right) = \mathbf{P}_\omega \left( \sum_{s=0}^{N_j} Z_{s, N_{j+k}} \geq 1 \right) \leq \mathbf{E}_\omega \left[ \sum_{s=0}^{N_j} Z_{s, N_{j+k}} \right].$$

Then by Item (ii) of Lemma 31

$$(3.69) \quad \mathbf{E}_\omega \left[ \sum_{s=0}^{N_j} Z_{s, N_{j+k}} \right] = \sum_{s=0}^{N_j} \mathbf{E}_\omega [Z_{s, N_{j+k}}] = \sum_{s=0}^{N_j} \exp \left( \sum_{i=s}^{N_{j+k}-1} \log(\rho_i) \right),$$

and by Item (i) of Lemma 31

$$(3.70) \quad \sum_{s=0}^{N_j} \exp \left( \sum_{i=s}^{N_{j+k}-1} \log(\rho_i) \right) \leq \sum_{s=0}^{N_j} \exp \left( -(N_{j+k} - s) \frac{c_\nu}{2} \right) = \sum_{l=N_{j+k}-N_j}^{N_{j+k}} \exp \left( -l \frac{c_\nu}{2} \right).$$

Finally, since for any  $m > n$ ,  $N_m - N_n \geq m - n$ , positivity of the exponential function ensures

$$(3.71) \quad \sum_{l=N_{j+k}-N_j}^{N_{j+k}} \exp \left( -l \frac{c_\nu}{2} \right) \leq \sum_{l=k}^{N_{j+k}} \exp \left( -l \frac{c_\nu}{2} \right) \leq \sum_{l=k}^{+\infty} \exp \left( -l \frac{c_\nu}{2} \right) = \frac{e^{-k \frac{c_\nu}{2}}}{1 - e^{-\frac{c_\nu}{2}}}.$$

Combining inequalities (3.68), (3.69), (3.70) and (3.71), leads to the expected result

$$\mathbf{P}_\omega \left( \sum_{s=0}^{N_j} Z_{s, N_{j+k}} > 0 \right) \leq \frac{e^{-k \frac{c_\nu}{2}}}{1 - e^{-\frac{c_\nu}{2}}}.$$

□

### 3.2.7.2 Proof of Lemma 25

Fix  $i$  in  $\mathbb{N}$ . For any  $l < k_i$ , we have

$$(3.72) \quad \begin{aligned} & \left\{ \sum_{s=N_l}^{N_{k_i}-1} Z_{s, N_{k_i}} > 0 \right\} \cap \left\{ \sum_{s=N_{l+1}}^{N_{k_i}-1} Z_{s, N_{k_i}} = 0 \right\} \\ &= \left( \bigcap_{s=N_{l+1}}^{N_{k_i}-1} \{Z_{s, N_{k_i}} = 0\} \right) \cap \left( \bigcup_{s=N_l}^{N_{l+1}-1} \{Z_{s, N_{k_i}} > 0\} \right). \end{aligned}$$

Hence, the event  $\{k_{i+1} = l\}$  only depends on the random variables  $(Z_{s,t})_{N_l \leq s < t \leq N_{k_i}}$ .

By definition of the sequence  $(k_j)_{j \in \mathbb{N}}$ , two situations can occur:

(a) If  $k_i > 0$ , by definition of the sequence  $(k_j)_{j \in \mathbb{N}}$ , two situations may occur:

(i) Suppose that  $k_i$  is such that  $Z_{N_{k_i}} > 0$ .

Let's first consider the case where  $r$  in  $\mathbb{N}^*$ .

When  $r \geq k_i$ , the set  $\{k_i - k_{i+1} > r\}$  is empty. Hence, if the set  $\{k_i - k_{i+1} > r\}$

is not empty, then,  $r \leq k_i - 1$ . As Inequality (3.46) is immediate when the set  $\{k_i - k_{i+1} > r\}$  is empty, let's assume  $r \leq k_i - 1$ . Using Equation (3.72), we derive the inclusions

$$\{k_i - k_{i+1} > r\} = \{k_{i+1} \leq k_i - r - 1\} \subset \left\{ \sum_{s=0}^{N_{k_i-r-1}} Z_{s, N_{k_i}} > 0 \right\},$$

and so,

$$(3.73) \quad \mathbf{P}_\omega(k_i - k_{i+1} > r | k_0, \dots, k_i) \leq \mathbf{P}_\omega \left( \sum_{s=0}^{N_{k_i-r-1}} Z_{s, N_{k_i}} > 0 \mid k_0, \dots, k_i \right).$$

Let's now upper-bound the left-hand side of (3.73) by the left-hand side of Inequality (3.46).

By the properties of conditional expectation with respect to discrete random variables, we derive that  $\mathbf{P}_\omega$ -p.s.

$$\begin{aligned} & \mathbf{P}_\omega \left( \sum_{s=0}^{N_{k_i-r-1}} Z_{s, N_{k_i}} > 0 \mid k_0, \dots, k_i \right) \\ &= \mathbf{E}_\omega \left[ \mathbb{1}_{\left\{ \sum_{s=0}^{N_{k_i-r-1}} Z_{s, N_{k_i}} > 0 \right\}} \mid k_0, \dots, k_i \right] \\ &= \sum_{j_{0:i} \in \mathbb{N}^{i+1} : \mathbf{P}_\omega(k_{0:i} = j_{0:i}) \neq 0} \mathbf{E}_\omega \left[ \mathbb{1}_{\left\{ \sum_{s=0}^{N_{k_i-r-1}} Z_{s, N_{k_i}} > 0 \right\}} \mid k_{0:i} = j_{0:i} \right] \mathbb{1}_{k_{0:i} = j_{0:i}} \\ &= \sum_{j_{0:i} \in \mathbb{N}^{i+1} : \mathbf{P}_\omega(k_{0:i} = j_{0:i}) \neq 0} \mathbf{E}_\omega \left[ \mathbb{1}_{\left\{ \sum_{s=0}^{N_{k_i-r-1}} Z_{s, N_{k_i}} > 0 \right\}} \mathbb{1}_{k_{0:i} = j_{0:i}} \right] \frac{\mathbb{1}_{k_{0:i} = j_{0:i}}}{\mathbf{P}_\omega(k_{0:i} = j_{0:i})}. \end{aligned}$$

By definition of  $(Z_{s,s+t})_{(s,t) \in \mathbb{N} \times \mathbb{N}^*}$  in Section 3.2.4 the groups of random variables  $(Z_{s_1, t_1})_{s_1 < t_1 \leq N_l}$  and  $(Z_{s_2, t_2})_{N_l \leq s_2 < t_2}$  are independent for any  $l$  in  $\mathbb{N}$ . Furthermore, the variables  $k_0, \dots, k_i$  only depend on the variables  $(Z_{s,t})_{N_{k_i} \leq s < t \leq N_k}$ .

Combining these two facts, we derive that, for any  $j_{0:i}$  in  $\mathbb{N}^{i+1}$ , the sets  $\{k_{0:i} = j_{0:i}\}$  and  $\{\sum_{s=0}^{N_{j_i-r-1}} Z_{s, N_{j_i}} > 0\}$  are independent and it holds  $\mathbf{P}_\omega$ -p.s.

$$\mathbf{E}_\omega \left[ \mathbb{1}_{\left\{ \sum_{s=0}^{N_{j_i-r-1}} Z_{s, N_{j_i}} > 0 \right\}} \mathbb{1}_{k_{0:i} = j_{0:i}} \right] = \mathbf{E}_\omega \left[ \mathbb{1}_{\left\{ \sum_{s=0}^{N_{j_i-r-1}} Z_{s, N_{j_i}} > 0 \right\}} \right] \mathbf{P}_\omega(k_{0:i} = j_{0:i}).$$

Hence, we have  $\mathbf{P}_\omega$ -p.s.

$$\begin{aligned} & \sum_{j_{0:i} \in \mathbb{N}^{i+1} : \mathbf{P}_\omega(k_{0:i} = j_{0:i}) \neq 0} \mathbf{E}_\omega \left[ \mathbb{1}_{\left\{ \sum_{s=0}^{N_{k_i-r-1}} Z_{s, N_{k_i}} > 0 \right\}} \mathbb{1}_{k_{0:i} = j_{0:i}} \right] \frac{\mathbb{1}_{k_{0:i} = j_{0:i}}}{\mathbf{P}_\omega(k_{0:i} = j_{0:i})} \\ & \leq \sum_{j_{0:i} \in \mathbb{N}^{i+1}} \mathbf{E}_\omega \left[ \mathbb{1}_{\left\{ \sum_{s=0}^{N_{j_i-r-1}} Z_{s, N_{j_i}} > 0 \right\}} \right] \mathbb{1}_{k_{0:i} = j_{0:i}} \\ & = \sum_{j_i \in \mathbb{N}} \mathbf{P}_\omega \left( \sum_{s=0}^{N_{j_i-r-1}} Z_{s, N_{j_i}} > 0 \right) \mathbb{1}_{k_i = j_i} \leq \sum_{j_i \in \mathbb{N}} \mathbf{P}_\omega \left( \sum_{s=0}^{N_{j_i-r}} Z_{s, N_{j_i}} > 0 \right) \mathbb{1}_{k_i = j_i}. \end{aligned}$$

Using  $r > 0$  and Inequality (3.67), we get, for any  $j_i$  in  $\mathbb{N}$ ,

$$\mathbf{P}_\omega \left( \sum_{s=0}^{N_{j_i-r}} Z_{s, N_{j_i}} > 0 \right) = \mathbf{P}_\omega \left( \sum_{s=0}^{N_{j_i-r}} Z_{s, N_{(j_i-r)+r}} > 0 \right) \leq \frac{e^{-r \frac{c\nu}{2}}}{1 - e^{-\frac{c\nu}{2}}}.$$

We thus finally derive Inequality (3.46)

$$\mathbf{P}_\omega (k_i - k_{i+1} > r | k_0, \dots, k_i) \leq \sum_{j_i \in \mathbb{N}} \frac{e^{-r \frac{c\nu}{2}}}{1 - e^{-\frac{c\nu}{2}}} \mathbb{1}_{k_i=j_i} \leq \frac{e^{-r \frac{c\nu}{2}}}{1 - e^{-\frac{c\nu}{2}}}.$$

Let's now deal with the case where  $r = 0$ .

Since the left-hand side of Inequality (3.46) is greater than 1 for  $r = 0$  and since  $\mathbf{P}_\omega$ -almost surely  $k_i - k_{i+1} > 0$ , Inequality (3.46) is trivially valid

$$\mathbf{P}_\omega (k_i - k_{i+1} > r | k_0, \dots, k_i) = 1 \leq \frac{1}{1 - e^{-\frac{c\nu}{2}}}.$$

(ii) If  $k_i$  is such that  $Z_{N_{k_i}} = 0$ , then,  $k_i - k_{i+1} = 1$ .

- For any  $r$  in  $\mathbb{N}^*$ ,  $\{k_i - k_{i+1} > r\} = \emptyset$  and thus

$$\mathbf{P}_\omega (k_{i+1} - k_i > r | k_0, \dots, k_i) = 0.$$

Inequality (3.46) is true for  $r$  in  $\mathbb{N}^*$ .

- For  $r = 0$ ,  $\{k_i - k_{i+1} > r\} = \Omega$  and thus

$$\mathbf{P}_\omega (k_{i+1} - k_i > r | k_0, \dots, k_i) = 1 \leq \frac{e^{-r \frac{c\nu}{2}}}{1 - e^{-\frac{c\nu}{2}}}.$$

Inequality (3.46) is also true for  $r = 0$ .

(b) If  $k_i = 0$ , then  $k_{i+1} = 0$  and for  $r$  in  $\mathbb{N}$ ,  $\{k_i - k_{i+1} > r\} = \emptyset$  so that

$$\mathbf{P}_\omega (k_i - k_{i+1} > r | k_0, \dots, k_i) = 0,$$

and Inequality (3.46) is true.

### 3.2.7.3 Proof of Lemma 26

Fix  $i$  in  $\mathbb{N}$  and  $\omega$  in  $E$ .

On the one hand, since for any positive real  $\lambda$

$$e^{\lambda(k_i - k_{i+1})} \mathbb{1}_{\{k_i - k_{i+1} \leq K\}} \leq e^{\lambda K},$$

we get

$$(3.74) \quad \mathbf{E}_\omega \left[ e^{\lambda(k_i - k_{i+1})} \mathbb{1}_{\{k_i - k_{i+1} \leq K\}} \middle| k_0, \dots, k_i \right] \leq e^{\lambda K}.$$

On the other hand, for any  $\lambda$  in  $(0, c_\nu/2)$

$$\begin{aligned} \mathbf{E}_\omega \left[ e^{\lambda(k_i - k_{i+1})} \mathbb{1}_{\{k_i - k_{i+1} > K\}} \middle| k_0, \dots, k_i \right] &= \sum_{j=K+1}^{+\infty} e^{\lambda j} \mathbf{P}_\omega(k_i - k_{i+1} = j | k_0, \dots, k_i) \\ &\leq \sum_{j=K}^{+\infty} e^{\lambda j} \mathbf{P}_\omega(k_i - k_{i+1} > j | k_0, \dots, k_i). \end{aligned}$$

By Lemma 25, for any positive integers  $j$  and  $K$  such that  $j \geq K$

$$\mathbf{P}_\omega(k_i - k_{i+1} > j | k_0, \dots, k_i) \leq \frac{e^{-j \frac{c_\nu}{2}}}{1 - e^{-\frac{c_\nu}{2}}}.$$

Hence, for any  $\lambda$  in  $(0, c_\nu/2)$

$$\begin{aligned} \sum_{j=K}^{+\infty} e^{\lambda j} \mathbf{P}_\omega(k_i - k_{i+1} > j | k_0, \dots, k_i) &\leq \sum_{j=K}^{+\infty} e^{\lambda j} \frac{e^{-j \frac{c_\nu}{2}}}{1 - e^{-\frac{c_\nu}{2}}} \leq e^{K(\lambda - \frac{c_\nu}{2})} \sum_{j=0}^{+\infty} \frac{e^{j(\lambda - \frac{c_\nu}{2})}}{1 - e^{-\frac{c_\nu}{2}}} \\ &\leq \frac{e^{K(\lambda - \frac{c_\nu}{2})}}{\left(1 - e^{-\frac{c_\nu}{2}}\right) \left(1 - e^{\lambda - \frac{c_\nu}{2}}\right)}. \end{aligned}$$

Hence, for any  $\lambda$  in  $(0, c_\nu/4]$

$$(3.75) \quad \mathbf{E}_\omega \left[ e^{\lambda(k_i - k_{i+1})} \mathbb{1}_{\{k_i - k_{i+1} > K\}} \middle| k_0, \dots, k_i \right] \leq \frac{e^{-K \frac{c_\nu}{2}}}{\left(1 - e^{-\frac{c_\nu}{2}}\right) \left(1 - e^{-\frac{c_\nu}{4}}\right)} e^{\lambda K}.$$

Decomposing  $e^{\lambda(k_i - k_{i+1})}$  as

$$e^{\lambda(k_i - k_{i+1})} = e^{\lambda(k_i - k_{i+1})} \left[ \mathbb{1}_{\{k_i - k_{i+1} \leq K\}} + \mathbb{1}_{\{k_i - k_{i+1} > K\}} \right],$$

Combining Equations (3.75) and (3.74), for any  $\lambda$  in  $(0, c_\nu/4]$

$$\begin{aligned} \mathbf{E}_\omega \left[ e^{\lambda(k_i - k_{i+1})} \middle| k_0, \dots, k_i \right] &\leq e^{\lambda K} + \frac{e^{-K \frac{c_\nu}{2}}}{\left(1 - e^{-\frac{c_\nu}{2}}\right) \left(1 - e^{-\frac{c_\nu}{4}}\right)} e^{\lambda K} \\ &\leq e^{\lambda K} \left( 1 + \frac{e^{-\frac{c_\nu}{2}}}{\left(1 - e^{-\frac{c_\nu}{2}}\right) \left(1 - e^{-\frac{c_\nu}{4}}\right)} \right) \leq e^{\lambda K} \left( 1 + \frac{e^{-\frac{c_0}{2}}}{\left(1 - e^{-\frac{c_0}{2}}\right) \left(1 - e^{-\frac{c_0}{4}}\right)} \right) \\ &\leq e^{\lambda K} \left( 1 + \frac{e^{-\frac{c_0}{2}}}{\left(1 - e^{-\frac{c_0}{4}}\right)^2} \right) \leq K_4(c_0) e^{\lambda K}. \end{aligned}$$

Hence, Inequality (3.48) holds with  $K_4(c_0)$  and  $K_5(c_0)$  defined in Equation (3.47).

#### 3.2.7.4 Proof of Corollary 27

Fix  $l$  in  $\mathbb{N}$  and  $\omega$  in  $E$ .

Let's first consider the case where  $l$  in  $\mathbb{N}^*$ .

By Definition 23 of the sequence  $(k_i)_{i \in \mathbb{N}}$ , we derive the following inclusion

$$(3.76) \quad \bigcup_{i=0}^l \{k_i = 0\} \subset \{k_0 - k_l \geq k\},$$

so

$$(3.77) \quad \mathbf{P}_\omega \left( \bigcup_{i=0}^l \{k_i = 0\} \right) \leq \mathbf{P}_\omega (k_0 - k_l \geq k).$$

Decompose  $k_0 - k_l$  as the telescopic sum of the  $(k_i - k_{i+1})_{0 \leq i \leq l-1}$ . Thus, by Markov inequality, for any  $\lambda$  in  $\mathbb{R}_+^*$

$$(3.78) \quad \mathbf{P}_\omega (k_0 - k_l \geq k) \leq \mathbf{P}_\omega \left( e^{\lambda(k_0 - k_l)} \geq e^{\lambda k} \right) \leq e^{-\lambda k} \mathbf{E}_\omega \left[ e^{\lambda(\sum_{i=0}^{l-1} (k_i - k_{i+1}))} \right].$$

Moreover, for any  $\lambda$  in  $\mathbb{R}_+^*$  and  $m$  in  $\mathbb{N}^*$

$$\begin{aligned} \mathbf{E}_\omega \left[ e^{\lambda(\sum_{i=0}^{m-1} (k_i - k_{i+1}))} \right] &= \mathbf{E}_\omega \left[ \mathbf{E}_\omega \left[ \prod_{i=0}^{m-1} e^{\lambda(k_i - k_{i+1})} \middle| k_0, \dots, k_{m-1} \right] \right] \\ &= \mathbf{E}_\omega \left[ \prod_{i=0}^{m-2} e^{\lambda(k_i - k_{i+1})} \mathbf{E}_\omega \left[ e^{\lambda(k_{m-1} - k_m)} \middle| k_0, \dots, k_{m-1} \right] \right]. \end{aligned}$$

By Lemma 26, for any  $\lambda$  in  $(0, c_0/4]$  and  $m$  in  $\mathbb{N}^*$

$$\begin{aligned} \mathbf{E}_\omega \left[ \prod_{i=0}^{m-2} e^{\lambda(k_i - k_{i+1})} \mathbf{E}_\omega \left[ e^{\lambda(k_{m-1} - k_m)} \middle| k_0, \dots, k_{m-1} \right] \right] &\leq \mathbf{E}_\omega \left[ \prod_{i=0}^{m-2} e^{\lambda(k_i - k_{i+1})} K_4(c_0) e^{\lambda K_5(c_0)} \right] \\ &\leq K_4(c_0) e^{\lambda K_5(c_0)} \mathbf{E}_\omega \left[ e^{\lambda(\sum_{i=0}^{m-2} (k_i - k_{i+1}))} \right]. \end{aligned}$$

Hence, for any  $\lambda$  in  $(0, c_0/4]$  and  $m$  in  $\mathbb{N}^*$

$$\mathbf{E}_\omega \left[ e^{\lambda(\sum_{i=0}^{m-1} (k_i - k_{i+1}))} \right] \leq K_4(c_0) e^{\lambda K_5(c_0)} \mathbf{E}_\omega \left[ e^{\lambda(\sum_{i=0}^{m-2} (k_i - k_{i+1}))} \right].$$

By recurrence, for any  $\lambda$  in  $(0, c_0/4]$  and  $l$  in  $\mathbb{N}^*$

$$(3.79) \quad \mathbf{E}_\omega \left[ e^{\lambda(\sum_{i=0}^{l-1} (k_i - k_{i+1}))} \right] \leq \left( K_4(c_0) e^{\lambda K_5(c_0)} \right)^l.$$

Combining Inequalities (3.77), (3.78) and (3.79) finally leads to Inequality (3.49).

Let's now consider the case where  $l = 0$ .

We have :

- either  $k = 0$  and  $\mathbf{P}_\omega (\{k_0 = 0\}) = 1 \leq 1 = e^{-\lambda k}$ ,
- or  $k \neq 0$  and  $\mathbf{P}_\omega (\{k_0 = 0\}) = 0 < e^{-\lambda k}$ .

so that Inequality (3.49) is always satisfied.

### 3.2.7.5 Proof of Lemma 29

Fix  $i$  in  $\mathbb{N}^*$  and  $\omega$  in  $E$ .

By definition of  $Z_{N_{k_i}}$ , we have

$$\begin{aligned}
& \mathbf{P}_\omega \left( Z_{N_{k_i}} = 0 \mid k_0, \dots, k_i, (Z_{s,t})_{N_{k_i} \leq s < t \leq k} \right) \\
&= \mathbf{E}_\omega \left[ \mathbb{1}_{\left\{ \sum_{s=0}^{N_{k_i}-1} Z_{s, N_{k_i}} = 0 \right\}} \mid k_0, \dots, k_i, (Z_{s,t})_{N_{k_i} \leq s < t \leq k} \right] \\
&= \mathbf{E}_\omega \left[ \sum_{j=0}^{k_i-1-1} \mathbb{1}_{\left\{ \sum_{s=0}^{N_j-1} Z_{s, N_j} = 0 \right\}} \mathbb{1}_{k_i=j} \mid k_0, \dots, k_i, (Z_{s,t})_{N_{k_i} \leq s < t \leq k} \right] \\
&= \sum_{j=0}^{k_i-1-1} \mathbf{E}_\omega \left[ \mathbb{1}_{\left\{ \sum_{s=0}^{N_j-1} Z_{s, N_j} = 0 \right\}} \mid k_0, \dots, k_i, (Z_{s,t})_{N_{k_i} \leq s < t \leq k} \right] \mathbb{1}_{k_i=j}.
\end{aligned}$$

Since the variables  $k_0, \dots, k_i$  only depend on the variables  $(Z_{s,t})_{N_{k_i} \leq s < t \leq N_k}$  and since the variables  $(Z_{s_1, t_1})_{s_1 < t_1 \leq N_{k_i}}$  and  $(Z_{s_2, t_2})_{N_{k_i} \leq s_2 < t_2}$  are independent, we can handle the same way as in Section 3.2.7.2 and thus omit the conditioning

$$\mathbf{E}_\omega \left[ \mathbb{1}_{\left\{ \sum_{s=0}^{N_j-1} Z_{s, N_j} = 0 \right\}} \mid k_0, \dots, k_i, (Z_{s,t})_{N_{k_i} \leq s < t \leq k} \right] \mathbb{1}_{k_i=j} = \mathbf{E}_\omega \left[ \mathbb{1}_{\left\{ \sum_{s=0}^{N_j-1} Z_{s, N_j} = 0 \right\}} \right] \mathbb{1}_{k_i=j}.$$

Hence,

$$\begin{aligned}
\mathbf{P}_\omega \left( Z_{N_{k_i}} = 0 \mid k_0, \dots, k_i, (Z_{s,t})_{N_{k_i} \leq s < t \leq k} \right) &= \sum_{j=0}^{k_i-1-1} \mathbf{E}_\omega \left[ \mathbb{1}_{\left\{ \sum_{s=0}^{N_j-1} Z_{s, N_j} = 0 \right\}} \right] \mathbb{1}_{k_i=j} \\
&= \sum_{j=0}^{k_i-1-1} \mathbf{P}_\omega \left( \{ Z_{N_j} = 0 \} \right) \mathbb{1}_{k_i=j}.
\end{aligned}$$

Using Equation (3.59), we finally derive that

$$\mathbf{P}_\omega \left( Z_{N_{k_i}} = 0 \mid k_0, \dots, k_i, (Z_{s,t})_{N_{k_i} \leq s < t \leq k} \right) \geq z(c_\nu) \left( \sum_{j=0}^{k_i-1-1} \mathbb{1}_{k_i=j} \right) \geq z(c_\nu).$$

### 3.2.7.6 Proof of Lemma 30

Fix  $l$  in  $\mathbb{N}$  and  $\omega$  in  $E$ .

Assume that  $l = 0$ .

In this case, Lemma 30 is a direct consequence of Lemma 34.

Assume that  $l$  belongs to  $\mathbb{N}^*$ .

By definition,  $N_0 = 0$  and  $Z_0 = 0$  so that for any  $i$  in  $\mathbb{N}^*$ , we have the following inclusions

$$\{k_{i-1} = 0\} \subset \{N_{k_{i-1}} = 0\} \subset \{Z_{N_{k_{i-1}}} = 0\},$$

or equivalently

$$\{Z_{N_{k_{i-1}}} > 0\} \subset \{N_{k_{i-1}} > 0\} \subset \{k_{i-1} > 0\}.$$

Since, for any  $i$  in  $\mathbb{N}^*$ ,  $k_{i-1} > 0$  on  $\{Z_{N_{k_{i-1}}} > 0\}$ ,  $k_i$  is defined on  $\{Z_{N_{k_{i-1}}} > 0\}$  as

$$k_i = \max \left( \left\{ m < k_{i-1} : \sum_{s=N_m}^{N_{k_{i-1}}} Z_{s, N_{k_{i-1}}} > 0 \quad \text{and} \quad \sum_{s=N_{m+1}}^{N_{k_{i-1}-1}} Z_{s, N_{k_{i-1}}} = 0 \right\} \right),$$

and thus for any  $r$  in  $\mathbb{N}$ , we get

$$\{k_i = r\} \cap \{Z_{N_{k_{i-1}}} > 0\} = \left\{ \sum_{s=N_r}^{N_{k_{i-1}-1}} Z_{s, N_{k_{i-1}}} > 0 \right\} \cap \left\{ \sum_{s=N_{r+1}}^{N_{k_{i-1}-1}} Z_{s, N_{k_{i-1}}} = 0 \right\}.$$

Hence, for any decreasing sequence of real numbers  $(r_i)_{1 \leq i \leq l}$  such that  $1 \leq r_l$  and  $r_1 \leq k-1$  and any  $t$  in  $\{1, \dots, l\}$ , the following identity holds

$$\bigcap_{i=1}^t (\{k_i = r_i\} \cap \{Z_{N_{k_{i-1}}} > 0\}) = \bigcap_{i=1}^t \left( \left\{ \sum_{s=N_{r_i}}^{N_{k_{i-1}-1}} Z_{s, N_{k_{i-1}}} > 0 \right\} \cap \left\{ \sum_{s=N_{r_i+1}}^{N_{k_{i-1}-1}} Z_{s, N_{k_{i-1}}} = 0 \right\} \right).$$

The event  $\bigcap_{i=1}^t (\{k_i = r_i\} \cap \{Z_{N_{k_{i-1}}} > 0\})$  thus only depends on the variables  $(Z_{s,t})_{N_{k_t} \leq s < t \leq N_k}$ . Hence

$$\begin{aligned} & \mathbf{E}_\omega \left[ \mathbb{1}_{Z_{N_{k_t}} > 0} \left| \bigcap_{i=1}^t (\{k_i = r_i\} \cap \{Z_{N_{k_{i-1}}} > 0\}) \right. \right] \\ &= \mathbf{E}_\omega \left[ \mathbf{E}_\omega \left[ \mathbb{1}_{Z_{N_{k_t}} > 0} \left| \bigcap_{i=1}^t (\{k_i = r_i\} \cap \{Z_{N_{k_{i-1}}} > 0\}) \right. \right] \middle| k_0, \dots, k_t, (Z_{s,t})_{N_{k_t} \leq s < t \leq N_k} \right] \\ &= \mathbf{E}_\omega \left[ \mathbf{E}_\omega \left[ \mathbb{1}_{Z_{N_{k_t}} > 0} \middle| k_0, \dots, k_t, (Z_{s,t})_{N_{k_t} \leq s < t \leq N_k} \right] \left| \bigcap_{i=1}^t (\{k_i = r_i\} \cap \{Z_{N_{k_{i-1}}} > 0\}) \right. \right]. \end{aligned}$$

By Lemma 29

$$\mathbf{P}_\omega \left( Z_{N_{k_t}} > 0 \middle| k_0, \dots, k_t, (Z_{s,t})_{N_{k_t} \leq s < t \leq N_k} \right) \leq 1 - z(c_\nu).$$

Hence, it holds, for any decreasing sequence of real numbers  $(r_i)_{1 \leq i \leq l}$  with  $1 \leq r_l$  and  $r_1 \leq k-1$  and any  $t$  in  $\{1, \dots, l\}$

$$(3.80) \quad \mathbf{P}_\omega \left( Z_{N_{k_t}} > 0 \left| \bigcap_{i=1}^t (\{k_i = r_i\} \cap \{Z_{N_{k_{i-1}}} > 0\}) \right. \right) \leq 1 - z(c_\nu).$$

Consider a decreasing sequence of real numbers  $(r_i)_{0 \leq i \leq l}$  with  $k = k_0 = r_0$ ,  $1 \leq r_l$  and  $r_1 \leq k-1$ .

For any  $t$  in  $\{0, \dots, l-1\}$ , on the set  $Z_{N_{k_{t+1}}} > 0$ , we necessarily have  $k_{t+1} \geq 1$ . Hence, given  $k_t$ ,  $k_{t+1}$  lies in  $\{1, \dots, k_t-1\}$  on the set  $Z_{N_{k_{t+1}}} > 0$ .



Decomposing according to the value of  $k_{t+1}$  and conditioning

$$\begin{aligned}
& \mathbf{P}_\omega \left( \{k_l > 0\} \cap \left( \bigcap_{i=t}^l \{Z_{N_{k_i}} > 0\} \right) \middle| \bigcap_{i=1}^t (\{k_i = r_i\} \cap \{Z_{N_{k_{i-1}}} > 0\}) \right) \\
&= \sum_{r_{t+1}=1}^{r_t-1} \mathbf{P}_\omega \left( \{k_{t+1} = r_{t+1}\} \cap \{Z_{N_{k_t}} > 0\} \middle| \bigcap_{i=1}^t (\{k_i = r_i\} \cap \{Z_{N_{k_{i-1}}} > 0\}) \right) \\
&\times \mathbf{P}_\omega \left( \{k_l > 0\} \cap \left( \bigcap_{i=t+1}^l \{Z_{N_{k_i}} > 0\} \right) \middle| \bigcap_{i=1}^{t+1} (\{k_i = r_i\} \cap \{Z_{N_{k_{i-1}}} > 0\}) \right) \\
&\leq \mathbf{P}_\omega \left( Z_{N_{k_t}} > 0 \middle| \bigcap_{i=1}^{t+1} (\{k_i = r_i\} \cap \{Z_{N_{k_{i-1}}} > 0\}) \right) \\
&\times \max_{1 \leq r_{t+1} \leq r_t-1} \mathbf{P}_\omega \left( \{k_l > 0\} \cap \left( \bigcap_{i=t+1}^l \{Z_{N_{k_i}} > 0\} \right) \middle| \bigcap_{i=1}^{t+1} (\{k_i = r_i\} \cap \{Z_{N_{k_{i-1}}} > 0\}) \right).
\end{aligned}$$

By Equation (3.80)

$$\mathbf{P}_\omega \left( Z_{N_{k_t}} > 0 \middle| \bigcap_{i=1}^{t+1} (\{k_i = r_i\} \cap \{Z_{N_{k_{i-1}}} > 0\}) \right) \leq 1 - z(c_\nu).$$

Hence,

$$\begin{aligned}
& \mathbf{P}_\omega \left( \{k_l > 0\} \cap \left( \bigcap_{i=t}^l \{Z_{N_{k_i}} > 0\} \right) \middle| \bigcap_{i=1}^t (\{k_i = r_i\} \cap \{Z_{N_{k_{i-1}}} > 0\}) \right) \\
&\leq (1 - z(c)) \max_{1 \leq r_{t+1} \leq r_t-1} \mathbf{P}_\omega \left( \{k_l > 0\} \cap \left( \bigcap_{i=t+1}^l \{Z_{N_{k_i}} > 0\} \right) \middle| \bigcap_{i=1}^{t+1} (\{k_i = r_i\} \cap \{Z_{N_{k_{i-1}}} > 0\}) \right).
\end{aligned}$$

By recurrence, we get

$$\begin{aligned}
& \mathbf{P}_\omega \left( \{k_l > 0\} \cap \left( \bigcap_{i=0}^l \{Z_{N_{k_i}} > 0\} \right) \right) \\
&\leq (1 - z(c))^l \times \max_{1 \leq r_{l-1} < \dots < r_1 < k} \mathbf{P}_\omega \left( \{k_l > 0\} \cap \{Z_{N_{k_l}} > 0\} \middle| \bigcap_{i=1}^l (\{k_i = r_i\} \cap \{Z_{N_{k_{i-1}}} > 0\}) \right).
\end{aligned}$$

Since  $\{k_l > 0\} \cap \{Z_{N_{k_l}} > 0\} = \{Z_{N_{k_l}} > 0\}$ , using again Equation (3.80), we get

$$\mathbf{P}_\omega \left( \{k_l > 0\} \cap \left( \bigcap_{i=0}^l \{Z_{N_{k_i}} > 0\} \right) \right) \leq (1 - z(c_\nu))^{l+1}.$$

### 3.2.7.7 Final step of the proof of Theorem 22

For any  $i$  in  $\mathbb{N}$ ,  $k_i > 0$  if and only if  $N_{k_i} > 0$ .

By definition of  $T$ , we thus derive the following inclusion

$$\{k_i > 0\} \cap \{Z_{N_{k_i}} = 0\} \subset \{k_i > 0\} \cap \{T \leq N_{k_i}\}.$$

Using successively the fact that  $(N_n)_{n \in \mathbb{N}}$  is increasing and the definition of the  $(k_j)_{j \in \mathbb{N}}$

$$\{k_i > 0\} \subset \{k_i > 0\} \cap \{k_i < k\} \subset \{k_i > 0\} \cap \{N_{k_i} < N_k\}.$$

Hence, the following inclusion holds

$$\{k_i > 0\} \cap \{Z_{N_{k_i}} = 0\} \subset \{k_i > 0\} \cap \{T \leq N_k\}.$$

Taking cosets, it leads, for any  $i$  in  $\mathbb{N}$ , to the inclusion

$$\{k_i > 0\} \cap \{T > N_k\} \subset \{k_i > 0\} \cap \{Z_{N_{k_i}} > 0\}.$$

For any  $l$  in  $\mathbb{N}$ , this allow to derive the following inclusions

$$\begin{aligned} \{T > N_k\} \cap \left( \bigcup_{i=0}^l \{k_i = 0\} \right)^c &= \{T > N_k\} \cap \bigcap_{i=0}^l \{k_i > 0\} \\ &= \bigcap_{i=0}^l (\{T > N_k\} \cap \{k_i > 0\}) \subset \bigcap_{i=0}^l (\{k_i > 0\} \cap \{Z_{N_{k_i}} > 0\}) \\ &\subset \left( \bigcap_{i=0}^l \{k_i > 0\} \right) \cap \left( \bigcap_{i=0}^l \{Z_{N_{k_i}} > 0\} \right) \subset \{k_l > 0\} \cap \left( \bigcap_{i=0}^l \{Z_{N_{k_i}} > 0\} \right). \end{aligned}$$

And we also have the following inclusions

$$\{T > N_k\} \cap \bigcup_{i=0}^l \{k_i = 0\} \subset \bigcup_{i=0}^l \{k_i = 0\}.$$

Hence,

$$\begin{aligned} \{T > N_k\} &= \{T > N_k\} \cap \left( \bigcup_{i=0}^l \{k_i = 0\} \sqcup \left( \bigcup_{i=1}^l \{k_i = 0\} \right)^c \right) \\ &= \left( \{T > N_k\} \cap \bigcup_{i=0}^l \{k_i = 0\} \right) \sqcup \left( \{T > N_k\} \cap \left( \bigcup_{i=0}^l \{k_i = 0\} \right)^c \right) \\ (3.81) \quad &\subset \left( \bigcup_{i=0}^l \{k_i = 0\} \right) \sqcup \left( \{k_l > 0\} \cap \left( \bigcap_{i=0}^l \{Z_{N_{k_i}} > 0\} \right) \right). \end{aligned}$$

Using Equation (3.81) yields directly for any  $l$  in  $\mathbb{N}$

$$\mathbf{P}_\omega(T > N_k) \leq \mathbf{P}_\omega \left( \bigcup_{i=0}^{l_1} \{k_i = 0\} \right) + \mathbf{P}_\omega \left( \{k_0 > k_1 > \dots > k_{l_1}\} \cap \bigcap_{i=0}^{l_1} \{Z_{N_{k_i}} > 0\} \right).$$

Take  $l = l_1(c_0, k)$  (defined in Equation (3.50)) such that  $l_1 + 1 \geq l_1 \geq k/(2K_5(c_0))$ . Thanks to Equations (27) and (30) and since  $1 - z(c_\nu) < 1$ , we get, for any  $\lambda$  in  $(0, c_\nu/4]$

$$\begin{aligned} \mathbf{P}_\omega(T > N_k) &\leq e^{-\frac{\lambda}{2}k} K_4(c_0)^{\frac{k}{2K_5(c_0)}} + (1 - z(c_\nu))^{l_1+1} \\ (3.82) \quad &\leq e^{\left(-\frac{\lambda}{2} + \frac{\log(K_4(c_0))}{2K_5(c_0)}\right)k} + e^{\frac{\log(1-z(c_\nu))}{2K_5(c_0)}k} \leq 2e^{-\min\left(\frac{\lambda}{2} - \frac{\log(K_4(c_0))}{2K_5(c_0)}, -\frac{\log(1-z(c_\nu))}{2K_5(c_0)}\right)k}. \end{aligned}$$

Since  $c \mapsto z(c)$  is increasing

$$\min \left( \frac{c_0}{8} - \frac{\log(K_4(c_0))}{2K_5(c_0)}, -\frac{\log(1-z(c_0))}{2K_5(c_0)} \right) \leq \min \left( \frac{c}{8} - \frac{\log(K_4(c))}{2K_5(c)}, -\frac{\log(1-z(c))}{2K_5(c)} \right).$$

Taking  $\lambda = c_0/4$  in Equation (3.82), we find

$$\mathbf{P}_\omega(T > N_k) \leq 2e^{-\min \left( \frac{c_0}{8} - \frac{\log(K_4(c_0))}{2K_5(c_0)}, -\frac{\log(1-z(c_0))}{2K_5(c_0)} \right) k} \leq 2e^{-\min \left( \frac{c_0}{8} - \frac{\log(K_4(c_0))}{2K_5(c_0)}, -\frac{\log(1-z(c_0))}{2K_5(c_0)} \right) k}.$$

Setting

$$K_6(c_0) = \min \left( \frac{c_0}{8} - \frac{\log(K_4(c_0))}{2K_5(c_0)}, -\frac{\log(1-z(c_0))}{2K_5(c_0)} \right),$$

which only depends on  $c_0$  and is positive (hence the choice of  $K_5(c_0)$  according to  $K_4(c_0)$  mentioned earlier in Equation (3.47)), the result claimed by Theorem 22 finally holds for any  $\omega$  in  $E$  and any  $k$  in  $\mathbb{N}$

$$\mathbf{P}_\omega(T > N_k) \leq 2e^{-K_6(c_0)k}.$$

### 3.2.8 Proof of Theorem 14 through Theorems 16 and 22

First, in Theorem 22, the environment  $\omega$  in  $E$  and integer  $k$  are considered fixed but the upper-bound does not depend on  $\omega$ . We can thus integrate Inequality 3.45 with respect to  $\omega$  in  $E$  under  $\mathbb{P}^\nu = \nu^{\otimes \mathbb{Z}}$ . We get, for any  $k$  in  $\mathbb{N}$

$$(3.83) \quad \mathbf{P}^\nu(T > N_k) = \int_E \mathbf{P}_\omega(T > N_k) \mathbb{P}^\nu(d\omega) \leq 2e^{-K_6(c_0, \alpha)k}.$$

Then, the integer  $l_0(c_0, \alpha) \geq 2$  being fixed, for any  $t > 0$ , two situations can occur.

- If  $t \geq l_0(c_0, \alpha) \mathbb{E}^\nu[N_1]$ , there exists  $k$  in  $\mathbb{N}^*$  such that

$$l_0(c_0, \alpha)k \mathbb{E}^\nu[N_1] \leq t < l_0(c_0, \alpha)(k+1) \mathbb{E}^\nu[N_1],$$

so that

$$\{T > t\} \cap \{N_k < l_0(c_0, \alpha)k \mathbb{E}^\nu[N_1]\} \subset \{T > N_k\}.$$

Thus

$$\begin{aligned} \mathbf{P}^\nu(T > t) &= \mathbf{P}^\nu(\{T > t\} \cap \{N_k \geq l_0(c_0, \alpha)k \mathbb{E}^\nu[N_1]\}) + \mathbf{P}^\nu(\{T > t\} \cap \{N_k < l_0(c_0, \alpha)k \mathbb{E}^\nu[N_1]\}) \\ &\leq \mathbf{P}^\nu(N_k \geq l_0(c_0, \alpha)k \mathbb{E}^\nu[N_1]) + \mathbf{P}^\nu(T > N_k) \\ &= \mathbb{P}^\nu(N_k \geq l_0(c_0, \alpha)k \mathbb{E}^\nu[N_1]) + \mathbf{P}^\nu(T > N_k). \end{aligned}$$

Inequalities (3.14) and (3.45) then yield

$$\mathbf{P}^\nu(T > t) \leq e^{-K_3(c_0, \alpha)k} + 2e^{-K_6(c_0, \alpha)k} \leq (1+2)e^{-\min(K_3(c_0, \alpha), K_6(c_0, \alpha))k}.$$

Since  $k > (t/l_0(c_0, \alpha) \mathbb{E}^\nu[N_1]) - 1$  and thus

$$\mathbf{P}^\nu(T > t) \leq 3e^{\min(K_3(c_0, \alpha), K_6(c_0, \alpha))} e^{-\frac{\min(K_3(c_0, \alpha), K_6(c_0, \alpha))t}{l_0(c_0, \alpha) \mathbb{E}^\nu[N_1]}}.$$

Using the control on the tail of  $N_1$  given in Corollary 17, we get

$$(3.84) \quad \mathbb{E}^\nu[N_1] = \sum_{k=0}^{+\infty} \mathbb{P}^\nu(N_1 > k) \leq \sum_{k=0}^{+\infty} \left( e^{-\frac{c_0 b(c_0, \alpha)}{4}} \right)^k \leq \frac{1}{1 - e^{-\frac{c_0 b(c_0, \alpha)}{4}}},$$

so that it finally holds

$$\mathbf{P}^\nu(T > t) \leq \left( 3e^{\min(K_3(c_0, \alpha), K_6(c_0, \alpha))} \right) e^{-\frac{\min(K_3(c_0, \alpha), K_6(c_0, \alpha))}{l_0(c_0, \alpha)} \left( 1 - e^{-\frac{c_0 b(c_0, \alpha)}{4}} \right) t}.$$

- If  $t < l(c_0, \alpha)\mathbb{E}^\nu[N_1]$ , by to Inequality (3.84), we get

$$-\frac{\min(K_3(c_0, \alpha), K_6(c_0, \alpha))}{l_0(c_0, \alpha)} \left( 1 - e^{-\frac{c_0 b(c_0, \alpha)}{4}} \right) t \geq -\min(K_3(c_0, \alpha), K_6(c_0, \alpha)),$$

so that it also holds

$$\mathbf{P}^\nu(T > t) \leq 3 \leq \left( 3e^{\min(K_3(c_0, \alpha), K_6(c_0, \alpha))} \right) e^{-\frac{\min(K_3(c_0, \alpha), K_6(c_0, \alpha))}{l_0(c_0, \alpha)} \left( 1 - e^{-\frac{c_0 b(c_0, \alpha)}{4}} \right) t}.$$

Setting

$$(3.85) \quad K_1(c_0, \alpha) = 3e^{\min(K_3(c_0, \alpha), K_6(c_0, \alpha))},$$

and

$$(3.86) \quad K_2(c_0, \alpha) = \frac{\min(K_3(c_0, \alpha), K_6(c_0, \alpha))}{l_0(c_0, \alpha)} \left( 1 - e^{-\frac{c_0 b(c_0, \alpha)}{4}} \right),$$

it holds, for any positive  $t$

$$\mathbf{P}^\nu(T > t) \leq K_1(c_0, \alpha)e^{-K_2(c_0, \alpha)t}.$$

Hence, Theorem 14 is proved.

### 3.2.9 Proof of the existence of exponential moments for BPIREG( $\nu, 0$ )

Remember that for any probability distribution  $\mathbb{Q}$  on  $(F, \mathcal{F})$ , for any random variable  $W$  from  $(F, \mathcal{F})$  to  $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$  and for any differentiable function  $h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  such that  $h(0) = 0$  and  $h' > 0$ , the following integration by part formula holds

$$\mathbb{Q}[h(W)] = \int_0^{+\infty} h'(w)\mathbb{Q}(W > w) dw.$$

Applying this identity with the annealed probability  $\mathbf{P}^\nu$ , the positive random variable  $T$  defined in Equation (3.8) and the function  $h : t \in \mathbb{R}^+ \mapsto \kappa_0^t - 1 \in \mathbb{R}^+$  where  $\kappa_0 > 0$  yields

$$\mathbf{E}^\nu[\kappa_0^T] = 1 + \mathbf{E}^\nu[e^{\log(\kappa_0)T} - 1] = 1 + \int_0^{+\infty} \log(\kappa_0)e^{\log(\kappa_0)t} \mathbf{P}^\nu(T > t) dt.$$

Combined with Theorem 14, we find that for any  $\nu$  meeting assumptions **F1** and **F2**

$$(3.87) \quad \mathbf{E}^\nu[\kappa_0^T] \leq 1 + \log(\kappa_0)K_1(c_0, \alpha) \int_0^{+\infty} e^{[\log(\kappa_0) - K_2(c_0, \alpha)]t} dt.$$

The condition  $\log(\kappa_0) - K_2(c_0, \alpha) < 0$  being sufficient for the integral in the right hand side of Inequality (3.87) to be finite, we set

$$\kappa_0 = \kappa_0(c_0, \alpha) = e^{\frac{K_2(c_0, \alpha)}{2}} > 1,$$

so that

$$\log(\kappa_0(c_0, \alpha))K_1(c_0, \alpha) \int_0^{+\infty} e^{[\log(\kappa_0(c_0, \alpha)) - K_2(c_0, \alpha)]t} dt = K_1(c_0, \alpha) < +\infty.$$

We derive the existence of an exponential moment for the first return time to 0 of the BPIREG( $\nu, 0$ ) with the following uniform control

$$\mathbf{E}^\nu [\kappa_0(c_0, \alpha)^T] = \mathbf{E}^\nu \left[ e^{\frac{K_2(c_0, \alpha)}{2}T} \right] \leq 1 + K_1(c_0, \alpha) < +\infty.$$

The Corollary 15 sums up these findings.

## Chapter 4

# Concentration inequality for geometrically ergodic Markov chains

In probability theory, the laws of large numbers states under mild conditions that the empirical mean of independent random variables is close to their expectation with large probability. Given this property, the next natural question is to assess the fluctuations of the empirical mean around its expectation. This is a special instance of *concentration inequalities*: such inequalities quantify the probability that a function  $f$  of random variables  $X_1, \dots, X_n$  (not necessarily independent) differs from its mean by more than a certain amount

$$(4.1) \quad \mathbb{P}(f(X_1, \dots, X_n) - \mathbb{E}[f(X_1, \dots, X_n)] \geq t) \leq g(t),$$

where  $g$  is some real valued function defined on positive reals.

Concentration properties for sums of independent random variables were first intensively studied and rather well-understood in probability theory. Whereas those properties were only available for sums, the *martingale methods*, developed in the 1970's for example in [Yur76, Mau79, MS86, McD89], made possible the handling of concentration for more general functions of independent random variables. Talagrand in the mid-1990's ([Tal95, Tal96a, Tal96b]) offered a new insight onto the problem suggesting new techniques depending on the *smoothness* of the transformation applied to the independent random variables. For a complete exposition of this methods we refer to [BLM13]. The applications of the concentration inequalities to various fields such as discrete mathematics (via the probabilistic method [AS11]), computer science (running times of randomized algorithms [DP09], machine learning) and statistics (statistical learning, model selection [MP07], density estimation) give them a always bigger importance.

Concentration inequalities where the function  $g$  in Inequality (4.1) is of the form  $2 \exp(-t^2/C)$  are said *sub-gaussian* and are of special interest because they are directly related to moment upper-bounds for the variables. Similarly, situations where the function  $f$  is in fact a *transformation with bounded differences property* were also widely studied because of their occurrence in other situations.

**Definition 36.** A function  $f : \mathcal{X}^n \rightarrow \mathbb{R}$  is said to have the bounded differences property if there exist nonnegative constants  $\gamma_{0:n-1} = (\gamma_0, \dots, \gamma_{n-1})$  in  $\mathbb{R}_+^n$  such that for all  $x_{0:n-1}$  in

$\mathcal{X}^n$  and all  $y_{0:n-1}$  in  $\mathcal{X}^n$

$$(4.2) \quad |f(x_{0:n-1}) - f(y_{0:n-1})| \leq \sum_{i=0}^{n-1} \gamma_i \mathbb{1}_{\{x_i \neq y_i\}}.$$

The class of all functions  $f : \mathcal{X}^n \rightarrow \mathbb{R}$  satisfying Inequality (4.2) is denoted  $\mathbb{B}\mathbb{D}(\mathcal{X}^n, \gamma_{0:n-1})$ .

In particular, the Mac Diarmid's inequality derived from the martingale method states an exponential concentration inequality for transformation with bounded differences property of i.i.d. sequence of random variables.

**Theorem 37.** [McD89][Mac Diarmid's inequality] Let  $(X, \mathcal{X})$  be a measurable space and  $(\Omega, \mathcal{F}, \mathbb{P})$  a probability space on which is defined a  $n$ -tuple of independent random variables  $X_{0:n-1} = (X_0, \dots, X_{n-1})$ . If  $f$  is a function on  $\mathcal{X}^n$  with bounded differences property of constant  $\gamma_{0:n-1}$  with  $n$  in  $\mathbb{N}^*$  and  $\gamma_{0:n-1}$  in  $\mathbb{R}_+^n$ , then, for any  $t$  in  $\mathbb{R}_+^*$

$$(4.3) \quad \mathbb{P}(|f(X_{0:n-1}) - \mathbb{E}[f(X_{0:n-1})]| > t) \leq 2 \exp\left(-\frac{2t^2}{\sum_{l=0}^{n-1} \gamma_l^2}\right),$$

In this chapter, we are concerned with a generalization of Mac Diarmid's concentration inequality that holds for Markov chains (instead of i.i.d. random variables) and is explicit with respect to the characteristics of the chains. As illustration, we apply this inequality to the Markov kernel  $K^\nu$  defined for any  $\nu$  in  $\mathbb{M}^1([0, 1])$  and any  $i$  and  $j$  in  $\mathbb{N}$  by

$$(4.4) \quad K^\nu(i, j) = \binom{i+j}{i} \int_0^1 t^{i+1} (1-t)^j \nu(dt).$$

In [DMPS18], the authors give an exhaustive introduction to Markov chain theory and introduce the readers to concentration inequality for Markov chains. In particular, they provide with a sub-gaussian concentration inequality for transformations of *V-geometrically ergodic* Markov chains by functions having the *bounded differences property* which is analogous to Mac Diarmid's concentration inequality (4.3):

**Theorem 38.** [DMPS18, Theorem 23.3.1] If  $P$  is an irreducible and aperiodic Markov kernel on  $X \times \mathcal{X}$  with invariant probability  $\pi$ , then, for every geometrically recurrent small set  $C$ , there exists a positive constant  $\beta$  such that for all  $n$  in  $\mathbb{N}^*$ , all  $\gamma_{0:n-1}$  in  $\mathbb{R}_+^n$ , all function  $f$  on  $\mathcal{X}^n$  with bounded differences property of constant  $\gamma_{0:n-1}$  and all  $x$  in  $C$ , it holds for any  $t$  in  $\mathbb{R}_+^*$

$$\mathbb{P}_x(|f(X_{0:n-1}) - \mathbb{E}_x[f(X_{0:n-1})]| > t) \leq 2 \exp\left(-\frac{\beta t^2}{\sum_{l=0}^{n-1} \gamma_l^2}\right),$$

where  $(X_n)_{n \in \mathbb{N}}$  denotes the canonical Markov chain on  $X^{\mathbb{N}}$ .

In the concentration inequality given by Theorem 38, the upper-bound is function of some positive constant  $\beta$  that possibly depends on the transition kernel  $P$  and on the small set  $C$ . Nevertheless, there is no clue how  $\beta$  depends on those three quantities.

In [Pau15, Corollary 2.10], a similar concentration inequality is given for Markov chains with values in a Polish state space and functions  $f$  with the bounded differences property of constant  $\gamma = \gamma_{0:n-1}$  except that the function  $g$  now writes  $g(t) = 2 \exp(-t^2 / (\|\gamma\|_2^2 \tau_{\min}))$  where the quantity  $\tau_{\min}$  strongly depends on the mixing time of the Markov chain. Even if it

gives in some sense the explicit dependency on the kernel  $P$  of the constant  $\beta$  of [DMPS18], it seems hard to give an expression of the mixing time explicitly depending on the characteristic of the Markov chain without very restrictive assumptions on the chain (reversibility, finite state space, ...).

In [DMPS18, Theorem 23.2.2] Mac-Diarmid's concentration inequality is also extended to uniformly ergodic Markov kernel. In this setting, there exists  $\beta > 0$  such that the function  $g$  writes  $g(t) = 2 \exp(-t^2/(D_n(\beta)))$  and denoting  $\Delta(M)$  the Dobrushin coefficient of Markov kernel  $M$ ,  $D_n : \mathbb{R}_+^* \rightarrow \mathbb{R}^+$  is defined for any  $\alpha$  in  $\mathbb{R}_+^*$  by

$$D_n(\beta) = \sum_{l=0}^{n-1} \left( \gamma_l + \beta \sum_{m=l+1}^{n-1} \gamma_m \Delta(P^{m-l}) \right)^2.$$

The major drawback of such an inequality is the computation of the Dobrushin coefficient. For example, there exists no uniform minorization condition on the Markov kernel  $K^\nu$ , what makes the computation of the Dobrushin coefficient impossible.

In this chapter, revisiting the proof of [DMPS18, Theorem 23.3.1], we track the dependency of the constant  $\beta$  into the specific parameters of the Markov chain in order to derive a quantitative version of the Mac Diarmid's type concentration inequality ending into Theorem 39.

As an illustration, Theorem 40 and Corollary 41 ensure that this quantitative concentration inequality is valid for the Markov chain starting from 0 and with transition kernel  $K^\nu$  for  $\nu$  in  $\mathbb{M}^1([0, 1])$  provided there exists  $\alpha$  in  $(0, 1/2)$  and  $c_0 > 0$  such that

$$(4.5) \quad \text{supp}(\nu) \subset [\alpha, 1 - \alpha] \quad \text{and} \quad \int_0^1 \log\left(\frac{1-t}{t}\right) \nu(dt) \leq -c_0.$$

In Section 4.1, we introduce the Markovian framework of the Chapter. Furthermore, we define also some assumptions concerning long-time behaviour of the considered Markov chains.

In Section 4.2, we state the main results of this Chapter : a quantitative concentration inequality for Markov chains in Theorem 39 and its application to Markov chain with kernel  $K^\nu$  in Theorem 40 and Corollary 41.

In Section 4.3, we concentrate on the proof of Theorem 39 : we decompose it in a logical sequence of elementary intermediate steps and prove them precisely.

In Section 4.4, we prove Theorem 40 : we justify that assumptions for application of Theorem 39 to Markov kernels  $K^\nu$  are met when  $\nu$  satisfies conditions (4.5).

## 4.1 Framework

Before going further, we set the framework, remind some useful notations that are common in Markov chain theory and introduce some assumptions. For basic definitions and elementary results on Markov chains theory, we provide with some reminders in Appendix A and refer to the book [DMPS18] for more comprehensive exposition.

### 4.1.1 Markovian setting

Let us first define the Markovian framework and some notations used below



- $(X^{\mathbb{N}}, \mathcal{X})$  is a measurable space and  $P$  is a Markov kernel on  $(X^{\mathbb{N}}, \mathcal{X})$ ,
- $(X_n)_{n \in \mathbb{N}}$  is the *canonical process* on the *canonical filtered space*  $(X^{\mathbb{N}}, \mathcal{X}^{\otimes \mathbb{N}}, (\mathcal{F}_k^X)_{k \in \mathbb{N}})$  defined for any  $n$  in  $\mathbb{N}$  and any  $x = (x_k)_{k \in \mathbb{N}}$  in  $X^{\mathbb{N}}$  by

$$X_n(x) = x_n \quad \text{and} \quad \mathcal{F}_n^X = \sigma(X_0, \dots, X_n),$$

- $\theta$  denotes the *shift operator* on the canonical space defined by

$$\begin{aligned} \theta : \quad X^{\mathbb{N}} &\longrightarrow X^{\mathbb{N}} \\ x = (x_k)_{k \in \mathbb{N}} &\mapsto \theta(x) = (\theta(x)_k)_{k \in \mathbb{N}}, \end{aligned}$$

where for any  $k$  in  $\mathbb{N}$ ,  $\theta(x)_k = x_{k+1}$ .

- For any probability measure  $\xi$  on  $\mathcal{X}$ , the unique probability measure on  $(X^{\mathbb{N}}, \mathcal{X}^{\otimes \mathbb{N}})$  such that  $(X_n)_{n \in \mathbb{N}}$  is a Markov chain with kernel  $P$  and initial distribution  $\xi$  is denoted  $\mathbb{P}_\xi$  (hereafter, for any  $x$  in  $\mathcal{X}$ , we write  $\mathbb{P}_x$  instead of  $\mathbb{P}_{\delta_x}$ ).

Furthermore, define for any set  $B$  in  $\mathcal{X}$  the stopping times  $\sigma_B$  and  $(\tau_B^i)_{i \in \mathbb{N}}$  by

$$\sigma_B = \inf\{n \geq 1 : X_n \in B\} \quad \text{and} \quad \tau_B^i = \inf\{n \geq i : X_n \in B\}.$$

The stopping time  $\sigma_B$  is the return time of set  $B$  while  $\tau_B = \tau_B^0$  is the first hitting time of set  $B$  and  $\tau_B^i$  is the first hitting time of set  $B$  after time  $i$  (with the convention that  $\inf(\emptyset) = +\infty$ ).

### 4.1.2 Assumptions in Markovian framework

Let us also define some assumptions concerning the long-time behaviour of those chains.

**M1.**  $P$  is an irreducible and aperiodic Markov kernel on  $X \times \mathcal{X}$  admitting the invariant probability  $\pi$ .

**M2.** There exist a set  $C$  in  $\mathcal{X}$  and real numbers  $\delta > 1$  and  $M > 0$  such that

$$(4.6) \quad \sup_{x \in C} \mathbb{E}_x[\delta^{\sigma_C}] \leq M.$$

**M3.** There exist  $r_0$  in  $(0, 1)$  and  $\zeta_0 > 0$  such that for any  $x$  in  $C$  and any  $n$  in  $\mathbb{N}$

$$(4.7) \quad d_{\text{TV}}(\delta_x P^n, \pi) \leq \zeta_0 r_0^n.$$

In particular, under assumption **M3**, the Markov chain with transition kernel  $P$  is geometrically ergodic provided its initial distribution gives mass 1 to the set  $C$ .

## 4.2 Main results : Theorems 39 and 40

The main contribution of this chapter is to give a quantitative sub-gaussian concentration inequality for transformations of Markov chain with the bounded differences property.

**Theorem 39.** *Assume that **M1**, **M2** and **M3** hold. If  $f$  is a function with bounded differences property in  $\mathbb{B}\mathbb{D}(\mathcal{X}^n, \gamma_{0:n-1})$  where  $n$  is some positive integer and  $\gamma_{0:n-1}$  is some  $n$ -tuple of positive reals, then, for any  $x$  in  $C$  and any  $t$  in  $\mathbb{R}_+^*$*

$$(4.8) \quad \mathbb{P}_x (|f(X_{0:n-1}) - \mathbb{E}_x[f(X_{0:n-1})]| > t) \leq 2 \exp \left( -\frac{\beta(\zeta_0, r_0, \delta, M) t^2}{\sum_{l=0}^{n-1} \gamma_l^2} \right),$$

where  $\beta(\zeta_0, r_0, \delta, M)$  (depending only on the parameters  $\zeta_0, r_0, \delta$  and  $M$  given by the assumptions) is defined by

$$(4.9) \quad \beta(\zeta_0, r_0, \delta, M) = \frac{1}{4} \left( \frac{4 \left(1 + \frac{4\zeta_0}{1-r_0}\right)}{\log(\delta)} + \frac{16 \max(2\zeta_0, 1)^2}{1 - \max\left(r_0, \delta^{-\frac{1}{4}}\right)^2} M \right)^{-1}.$$

We now prove that under Assumptions **F1** and **F2** of Chapter 3, the kernel  $K^\nu$  meets requirements **M1**, **M2** and **M3**.

**Theorem 40.** *Fix a probability distribution  $\nu$  on  $[0, 1]$  and consider the Markov kernel  $K^\nu$  defined for any integers  $i$  and  $j$  by*

$$(4.10) \quad K^\nu(i, j) = \binom{i+j}{i} \int_0^1 t^{i+1} (1-t)^j \nu(dt).$$

If the distribution  $\nu$  meets Assumption **F1** and **F2** of Chapter 3, i.e. there exists  $c_0$  in  $\mathbb{R}_+^*$  and  $\alpha$  in  $(0, 1/2)$  such that

$$\int_0^1 \log \left( \frac{1-t}{t} \right) \nu(dt) \leq -c_0 \quad \text{and} \quad \nu([\alpha, 1-\alpha]) = 1,$$

then, the Markov kernel  $K^\nu$  meet Assumptions **M1**, **M2** and **M3** where  $C = \{0\}$  and the constants  $r_0 = r_0(c_0, \alpha)$ ,  $\zeta_0 = \zeta_0(c_0, \alpha)$ ,  $\delta = \delta(c_0, \alpha)$  and  $M = M(c_0, \alpha)$ , respectively defined in Equations (4.95), (4.96), (4.64) and (4.65), depend explicitly on  $c_0$  and  $\alpha$ .

In framework of Theorem 40, Theorem 39 applies immediately to kernel  $K^\nu$  : we thus obtain in Corollary 41 a concentration inequality for transformations with bounded differences property of the Markov chain starting from 0 with transition kernel  $K^\nu$ .

**Corollary 41.** *Fix a probability distribution  $\nu$  on  $[0, 1]$ . If the distribution  $\nu$  meets Assumption **F1** and **F2** of Chapter 3, i.e. there exists  $c_0$  in  $\mathbb{R}_+^*$  and  $\alpha$  in  $(0, 1/2)$  such that*

$$\int_0^1 \log \left( \frac{1-t}{t} \right) \nu(dt) \leq -c_0 \quad \text{and} \quad \nu([\alpha, 1-\alpha]) = 1,$$

then, the concentration inequality (4.8) is valid for the Markov chain with transition kernel  $K^\nu$  defined in Equation (4.10), the set  $C = \{0\}$  and the constant  $\beta(\zeta_0, r_0, \delta, M)$  defined in Equation (4.9) previously plugged with the constants  $r_0 = r_0(c_0, \alpha)$ ,  $\zeta_0 = \zeta_0(c_0, \alpha)$ ,  $\delta = \delta(c_0, \alpha)$  and  $M = M(c_0, \alpha)$ , respectively defined in Equations (4.95), (4.96), (4.64) and (4.65).

When considering a RWRE with environment distribution  $\nu$  in  $\mathbb{M}^1([0, 1])$ , the existence of positive real numbers  $c_0$  and  $\alpha$  in  $(0, 1/2)$  such that  $\int_0^1 \log \left( \frac{1-t}{t} \right) \nu(dt) \leq -c_0$  and  $\nu([\alpha, 1-\alpha]) = 1$  concern respectively long-time and local behaviours of the RWRE. First condition is equivalent to its transience to the right whereas the second, called *ellipticity condition*, prevents the walk to go systematically in one direction or to get trapped between two specific integers and is nearly systematic in probabilistic statement about RWRE.

Sections 4.3 and 4.4 are dedicated to the proofs of Theorems 39 and 40 respectively.

### 4.3 Proof of Theorem 39

We first give a bright oversight of the proof of Theorem 39

#### 4.3.1 Sketch of proof of Theorem 39

The first step of the proof of Theorem 39 consists in showing Inequality (4.11).

**Lemma 42** ([DMPS18, Theorem 23.3.1]). *For any  $\xi$  and  $\xi'$  in  $\mathbb{M}^1(\mathcal{X})$ , any  $n$  in  $\mathbb{N}$ , any  $\gamma_{0:n-1}$  in  $\mathbb{R}_+^n$  and any  $h$  in  $\mathbb{BD}(\mathcal{X}^n, \gamma_{0:n-1})$*

$$(4.11) \quad |\mathbb{E}_\xi[h(X_{0:n-1})] - \mathbb{E}_{\xi'}[h(X_{0:n-1})]| \leq 2 \sum_{i=0}^{n-1} \gamma_i d_{\text{TV}}(\xi P^i, \xi' P^i).$$

For the sake of completeness, its proof is given in Section 4.3.2.1.

As, for any  $0 \leq i \leq n-1$ , the stopping time  $\tau_C^i$  is well defined by **M2**, we can introduce the following sequence of random variables  $(G_i)_{0 \leq i \leq n-1}$  defined by

$$G_i = \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{\tau_C^i}^X \right].$$

One important property is that any two successive terms of this sequence,  $G_i$  and  $G_{i-1}$ , are equal excepted on the set  $\{\tau_C^{i-1} = i-1\}$  as Lemma 43 guarantees. Its proof is postponed to Section 4.3.2.2.

**Lemma 43.** *Considering the sequence  $(G_i)_{0 \leq i \leq n-1}$  defined for any integer  $1 \leq i \leq n-1$  by*

$$(4.12) \quad G_i = \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{\tau_C^i}^X \right],$$

*it holds for any integer  $1 \leq i \leq n-1$*

$$(4.13) \quad G_i - G_{i-1} = (G_i - G_{i-1}) \mathbb{1}_{\{\tau_C^{i-1} = i-1\}}.$$

We can control the distances between  $G_i$  and  $G_{i-1}$  thanks to the existence of a geometrically recurrent small set granted by **M2** and the geometric ergodicity of the chain granted by **M3**. This result is shown in the following Lemma 44, whose proof is postponed to Section 4.3.2.3.

**Lemma 44.** *Under **M1**, **M2** and **M3**, for all  $r$  in  $[r_0, 1)$ , for any  $n$  in  $\mathbb{N}$ , any integer  $1 \leq i \leq n-1$ , any  $\gamma_{0:n-1}$  in  $\mathbb{R}_+^n$ , any  $f$  in  $\mathbb{BD}(\mathcal{X}^n, \gamma_{0:n-1})$  and any  $x$  in  $C$ , it holds*

$$(4.14) \quad |G_i - G_{i-1}| \leq \zeta_1(\zeta_0, r_0) \max_{0 \leq j \leq n-1} (\gamma_j) \mathbb{1}_{\{\tau_C^{i-1} = i-1\}} \sigma_C \circ \theta^{i-1},$$

and

$$(4.15) \quad |G_i - G_{i-1}|^2 \leq \zeta_2(\zeta_0, r) \mathbb{1}_{\{\tau_C^{i-1} = i-1\}} r^{-2\sigma_C \circ \theta^{i-1}} \sum_{k=i}^{n-1} \gamma_k^2 r^{k-i},$$

where  $G_i$  is defined in Equation (4.12) and where

$$(4.16) \quad \zeta_1(\zeta_0, r_0) = 1 + \frac{4\zeta_0}{1-r_0},$$

and

$$(4.17) \quad \zeta_2(\zeta_0, r) = \frac{4 \max(2\zeta_0, 1)^2}{1-r}.$$

Relying on the controls (4.14) and (4.15) of the distances between  $G_i$  and  $G_{i-1}$ , a decomposition of the difference between  $f(X_{0:n-1})$  and  $\mathbb{E}_x[f(X_{0:n-1})]$  by mean of the  $G_i$  leads to Lemma 45. Its proof is postponed to Section 4.3.2.4.

**Lemma 45.** *Under M1, M2 and M3, for all  $\gamma_{0:n-1}$  in  $\mathbb{R}_+^n$ ,  $f$  in  $\mathbb{BD}(\mathcal{X}^n, \gamma_{0:n-1})$  and  $x$  in  $C$ , it holds*

$$(4.18) \quad \mathbb{E}_x \left[ e^{f(X_{0:n-1}) - \mathbb{E}_x[f(X_{0:n-1})]} \right] \leq \exp \left( \kappa(\zeta_0, r_0, \delta, M) \sum_{i=0}^{n-1} \gamma_i^2 \right),$$

with

$$(4.19) \quad \kappa(\zeta_0, r_0, \delta, M) = \frac{4\zeta_1(\zeta_0, r_0)}{\log(\delta)} + \frac{2\zeta_2 \left( \zeta_0, \frac{1 + \max(r_0, \delta^{-\frac{1}{4}})}{2} \right)}{1 - \max(r_0, \delta^{-\frac{1}{4}})} M,$$

where  $\zeta_1(\zeta_0, r_0)$  and  $\zeta_2(\zeta_0, r)$  are defined in Equations (4.16) and (4.17) of Lemma 44.

Using a "Chernov bound", Lemma 45 finally enables to completely prove Theorem 39.

**Proof of Theorem 39.** For any  $s > 0$ ,  $t > 0$  and  $g$  in  $\mathbb{BD}(\mathcal{X}^n, \gamma_{0:n-1})$ , it holds by Markov inequality and Lemma 45

$$(4.20) \quad \begin{aligned} \mathbb{P}_x (g(X_{0:n-1}) - \mathbb{E}_x[g(X_{0:n-1})] > t) &= \mathbb{P}_x \left( e^{st(g(X_{0:n-1}) - \mathbb{E}_x[g(X_{0:n-1})])} > e^{st} \right) \\ &\leq e^{-st} \mathbb{E}_x \left[ e^{s(g(X_{0:n-1}) - \mathbb{E}_x[g(X_{0:n-1})])} \right] \\ &\leq e^{-st + s^2 \kappa(\zeta_0, r_0, \delta, M) \sum_{i=0}^{n-1} s^2 \gamma_i^2}. \end{aligned}$$

Choosing

$$s = \frac{t}{2 \kappa(\zeta_0, r_0, \delta, M) \sum_{i=0}^{n-1} s^2 \gamma_i^2},$$

in Inequality (4.20), we get

$$(4.21) \quad \mathbb{P}_x (g(X_{0:n-1}) - \mathbb{E}_x[g(X_{0:n-1})] > t) \leq e^{-\frac{t^2}{4 \kappa(\zeta_0, r_0, \delta, M) \sum_{i=0}^{n-1} \gamma_i^2}}.$$

Since  $-g$  also belongs to  $\mathbb{BD}(\mathcal{X}^n, \gamma_{0:n-1})$ , Equation (4.21) also holds for  $-g$ . Hence, we finally proved that for all  $t > 0$

$$\mathbb{P}_x (|f(X_{0:n-1}) - \mathbb{E}_x[f(X_{0:n-1})]| > t) \leq 2 e^{-\frac{t^2}{4 \kappa(\zeta_0, r_0, \delta, M) \sum_{i=0}^{n-1} \gamma_i^2}}.$$

□

### 4.3.2 Intermediate results for proof of Theorem 39

In this section, we prove all intermediate results that have been introduced in Section 4.3.1. More or less standard technical properties of stopping times on filtered probability spaces are used and we refer to Appendix C for detailed proof of such results.

Recall that the total variation norm and the total variation distance can be respectively defined as follows for any  $\xi$  and  $\xi'$  in  $\mathbb{M}^1(\mathcal{X})$

$$\|\xi\|_{\text{TV}} = \sup (\{\xi(f) : \|f\|_\infty \leq 1\}) \quad \text{and} \quad d_{\text{TV}}(\xi, \xi') = \frac{1}{2} \|\xi - \xi'\|_{\text{TV}}.$$

### 4.3.2.1 Proof of Lemma 42

Fix  $\bar{x}$  in  $X$  and for all integer  $1 \leq i \leq n-1$ , define the function  $h_i$  by

$$h_i(x_{i:n-1}) = h(\bar{x}, \dots, \bar{x}, x_{i:n-1}).$$

Set  $h_0 = h$  and  $h_n = h(\bar{x}, \dots, \bar{x})$ . The function  $h$  can thus be decomposed as

$$(4.22) \quad h(x_{0:n-1}) = h_n + \sum_{i=0}^{n-1} (h_i(x_{i:n-1}) - h_{i+1}(x_{i+1:n-1})),$$

where we write abusively  $h_n(x_{n:n-1}) = h_n$  for all  $x_{0:n-1}$  in  $\mathcal{X}^n$ .

For any integer  $0 \leq i \leq n-1$  and  $x_0^i$  in  $X^{i+1}$ , let's define  $w_i(x_i)$  by

$$\begin{aligned} w_i(x_i) &= \int (h_i(x_{i:n-1}) - h_{i+1}(x_{i+1:n-1})) \prod_{l=i+1}^{n-1} P(x_{l-1}, dx_l) \\ &= \int (h(\bar{x}, \dots, \bar{x}, x_{i:n-1}) - h(\bar{x}, \dots, \bar{x}, x_{i+1:n-1})) \prod_{l=i+1}^{n-1} P(x_{l-1}, dx_l). \end{aligned}$$

Since  $(X_n)$  is a Markov chain, for any  $\xi$  in  $\mathbb{M}^1(\mathcal{X})$  and for any integer  $0 \leq i \leq n-1$

$$w_i(X_i) = \mathbb{E}_\xi [h_i(X_{i:n-1}) - h_{i+1}(X_{i+1:n-1}) | \mathcal{F}_i^X] \quad \mathbb{P}_\xi\text{-a.s.}$$

Thanks to the decomposition (4.22) of  $h$ , we obtain

$$\begin{aligned} \mathbb{E}_\xi [h(X_{0:n-1})] &= \mathbb{E}_\xi \left[ h_n + \sum_{i=0}^{n-1} (h_i(X_i^{n-1}) - h_{i+1}(X_{i+1}^{n-1})) \right] \\ &= h_n + \sum_{i=0}^{n-1} \mathbb{E}_\xi [\mathbb{E}_\xi [h_i(X_i^{n-1}) - h_{i+1}(X_{i+1}^{n-1}) | \mathcal{F}_i^X]] \\ &= h_n + \sum_{i=0}^{n-1} \mathbb{E}_\xi [w_i(X_i)] = h_n + \sum_{i=0}^{n-1} \xi P^i w_i, \end{aligned}$$

so that

$$(4.23) \quad \begin{aligned} \mathbb{E}_\xi [h(X_{0:n-1})] - \mathbb{E}_{\xi'} [h(X_{0:n-1})] &= h_n + \sum_{i=0}^{n-1} \xi P^i w_i - \left\{ h_n + \sum_{i=0}^{n-1} \xi' P^i w_i \right\} \\ &= \sum_{i=0}^{n-1} (\xi P^i - \xi' P^i) w_i. \end{aligned}$$

For any integer  $0 \leq i \leq n-1$ , we derive from the definition of the total variation norm that

$$(4.24) \quad |(\xi P^i - \xi' P^i) w_i| \leq \|w_i\|_\infty \|\xi P^i - \xi' P^i\|_{\text{TV}}.$$

Since  $h$  belongs to  $\mathbb{BD}(\mathcal{X}^n, \gamma_{0:n-1})$  and since the arguments of function  $h$  only differ from  $(i+1)^{\text{th}}$  component in the definitions of  $h_i$  and  $h_{i+1}$ , Equation (4.2) ensures that for any integer  $0 \leq i \leq n-1$

$$|h_i(x_{i:n-1}) - h_{i+1}(x_{i+1:n-1})| \leq \gamma_i,$$

and thus by definition of  $w_i$  and Jensen inequality, we thus get for any  $x_{0:n-1}$  in  $\mathcal{X}^n$

$$\begin{aligned} \sup_{x_i \in X} |w_i(x_i)| &= \sup_{x_i \in X} \left( \left| \int (h_i(x_{i:n-1}) - h_{i+1}(x_{i+1:n-1})) \prod_{l=i+1}^{n-1} P(x_{l-1}, dx_l) \right| \right) \\ &\leq \sup_{x_i \in X} \left( \int |h_i(x_{i:n-1}) - h_{i+1}(x_{i+1:n-1})| \prod_{l=i+1}^{n-1} P(x_{l-1}, dx_l) \right) \\ &\leq \int \gamma_i \prod_{l=i+1}^{n-1} P(x_{l-1}, dx_l) = \gamma_i, \end{aligned}$$

so that

$$(4.25) \quad \|w_i\|_\infty \leq \gamma_i.$$

Combining Inequalities (4.24) and (4.25), Equation (4.23) finally leads to

$$\begin{aligned} |\mathbb{E}_\xi [h(X_{0:n-1})] - \mathbb{E}_{\xi'} [h(X_{0:n-1})]| &= \left| \sum_{i=0}^{n-1} (\xi P^i - \xi' P^i) w_i \right| \leq \sum_{i=0}^{n-1} |(\xi P^i - \xi' P^i) w_i| \\ &\leq 2 \sum_{i=0}^{n-1} \gamma_i \text{dTV}(\xi P^i, \xi' P^i). \end{aligned}$$

#### 4.3.2.2 Proof of Lemma 43

For any integer  $1 \leq i \leq n-1$ , by definition of  $\tau_C^{i-1}$  and  $\tau_C^i$ , we get

$$\tau_C^{i-1} \geq i-1 \quad \text{and} \quad \tau_C^{i-1} > i-1 \Leftrightarrow \tau_C^{i-1} = \tau_C^i.$$

Hence,

$$X^\mathbb{N} = \{\tau_C^{i-1} = i-1\} \sqcup \{\tau_C^{i-1} = \tau_C^i, \tau_C^{i-1} \geq n-1\} \sqcup \{\tau_C^{i-1} = \tau_C^i, i-1 < \tau_C^{i-1} < n-1\}.$$

This leads to the following decomposition of  $G_i - G_{i-1}$

$$(4.26) \quad G_i - G_{i-1} = (G_i - G_{i-1}) \left( \mathbb{1}_{\{\tau_C^{i-1} = i-1\}} + \mathbb{1}_{\{\tau_C^{i-1} = \tau_C^i, i-1 < \tau_C^{i-1} < n-1\}} + \mathbb{1}_{\{\tau_C^{i-1} = \tau_C^i, \tau_C^{i-1} \geq n-1\}} \right).$$

For any integer  $0 \leq i \leq n-1$ , using Lemma 91 of Appendix C

$$\begin{aligned} \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{\tau_C^i}^X \right] \mathbb{1}_{\{\tau_C^i < n-1\}} &= \sum_{j=i}^{n-2} \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{\tau_C^i}^X \right] \mathbb{1}_{\{\tau_C^i = j\}} \\ &= \sum_{j=i}^{n-2} \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_j^X \right] \mathbb{1}_{\{\tau_C^i = j\}}. \end{aligned}$$

Hence, for any integer  $1 \leq i \leq n-1$ , the second term in the right-hand side of Equation (4.26) becomes

$$\begin{aligned}
(G_i - G_{i-1}) \mathbb{1}_{\{\tau_C^{i-1} = \tau_C^i, i-1 < \tau_C^{i-1} < n-1\}} &= \mathbb{E}_x \left[ f(X_{0:n-1}) \Big| \mathcal{F}_{\tau_C^i}^X \right] \mathbb{1}_{\{i-1 < \tau_C^i < n-1\}} \mathbb{1}_{\{\tau_C^{i-1} = \tau_C^i\}} \\
&\quad - \mathbb{E}_x \left[ f(X_{0:n-1}) \Big| \mathcal{F}_{\tau_C^{i-1}}^X \right] \mathbb{1}_{\{i-1 < \tau_C^{i-1} < n-1\}} \mathbb{1}_{\{\tau_C^{i-1} = \tau_C^i\}} \\
&= \sum_{j=i}^{n-2} \mathbb{E}_x \left[ f(X_{0:n-1}) \Big| \mathcal{F}_j^X \right] \mathbb{1}_{\{\tau_C^i = j\}} \mathbb{1}_{\{\tau_C^{i-1} = \tau_C^i\}} \\
&\quad - \sum_{j=i}^{n-2} \mathbb{E}_x \left[ f(X_{0:n-1}) \Big| \mathcal{F}_j^X \right] \mathbb{1}_{\{\tau_C^{i-1} = j\}} \mathbb{1}_{\{\tau_C^{i-1} = \tau_C^i\}} = 0.
\end{aligned}$$

Moreover, by Lemma 91 of Appendix C,

$$\begin{aligned}
G_i \mathbb{1}_{\{\tau_C^{i-1} = \tau_C^i, \tau_C^{i-1} \geq n-1\}} &= \mathbb{E}_x \left[ f(X_{0:n-1}) \Big| \mathcal{F}_{\tau_C^i}^X \right] \mathbb{1}_{\{\tau_C^i \geq n-1\}} \mathbb{1}_{\{\tau_C^{i-1} = \tau_C^i\}} \\
&= \mathbb{E}_x \left[ f(X_{0:n-1}) \Big| \mathcal{F}_{n-1}^X \right] \mathbb{1}_{\{\tau_C^i \geq n-1\}} \mathbb{1}_{\{\tau_C^{i-1} = \tau_C^i\}} \\
&= f(X_{0:n-1}) \mathbb{1}_{\{\tau_C^{i-1} \geq n-1\}} \mathbb{1}_{\{\tau_C^{i-1} = \tau_C^i\}}.
\end{aligned}$$

Likewise,

$$\begin{aligned}
G_{i-1} \mathbb{1}_{\{\tau_C^{i-1} = \tau_C^i, \tau_C^{i-1} \geq n-1\}} &= \mathbb{E}_x \left[ f(X_{0:n-1}) \Big| \mathcal{F}_{\tau_C^{i-1}}^X \right] \mathbb{1}_{\{\tau_C^{i-1} \geq n-1\}} \mathbb{1}_{\{\tau_C^{i-1} = \tau_C^i\}} \\
&= \mathbb{E}_x \left[ f(X_{0:n-1}) \Big| \mathcal{F}_{n-1}^X \right] \mathbb{1}_{\{\tau_C^{i-1} \geq n-1\}} \mathbb{1}_{\{\tau_C^{i-1} = \tau_C^i\}} \\
&= f(X_{0:n-1}) \mathbb{1}_{\{\tau_C^{i-1} \geq n-1\}} \mathbb{1}_{\{\tau_C^{i-1} = \tau_C^i\}}.
\end{aligned}$$

Hence, the third term in the right-hand side of Equation (4.26) becomes

$$(G_i - G_{i-1}) \mathbb{1}_{\{\tau_C^{i-1} = \tau_C^i, \tau_C^{i-1} \geq n-1\}} = (f(X_{0:n-1}) - f(X_{0:n-1})) \mathbb{1}_{\{\tau_C^{i-1} \geq n-1\}} \mathbb{1}_{\{\tau_C^{i-1} = \tau_C^i\}} = 0.$$

This concludes the proof of Lemma 43.

### 4.3.2.3 Proof of Lemma 44

First, Lemma 43 enables to decompose the difference  $G_i - G_{i-1}$  in two terms. Each of the two resulting terms are again decomposed in Equations (4.30) and (4.32) and then controlled thanks to the bounded differences property of  $f$ . A preliminary bound on  $|G_i - G_{i-1}|$  is thus derived in Inequality (4.38). Finally, we obtain Inequalities (4.14) and (4.15) that give respectively expected bounds on  $|G_i - G_{i-1}|$  and  $|G_i - G_{i-1}|^2$ .

Fix  $x$  in  $C$ . For any integer  $1 \leq i \leq n$ , define  $G_{i,1}$  and  $G_{i,2}$  by

$$G_{i,1} = \mathbb{E}_x \left[ f(X_{0:n-1}) \Big| \mathcal{F}_{\tau_C^{i-1}}^X \right] \mathbb{1}_{\{\tau_C^{i-1} = i-1\}} \quad \text{and} \quad G_{i,2} = \mathbb{E}_x \left[ f(X_{0:n-1}) \Big| \mathcal{F}_{\tau_C^i}^X \right] \mathbb{1}_{\{\tau_C^{i-1} = i-1\}}.$$

Using Equation (4.13) of Lemma 43, we get

$$(4.27) \quad G_i - G_{i-1} = (G_i - G_{i-1}) \mathbb{1}_{\{\tau_C^{i-1} = i-1\}} = G_{i,2} - G_{i,1}.$$

Define the functions  $(g_i)_{0 \leq i \leq n-1}$  by  $g_{n-1} = f$  and for any integer  $0 \leq i \leq n-2$

$$g_i(x_{0:i}) = \mathbb{E}_{x_i} [f(x_{0:i}, X_{1:n-1-i})],$$

Define similarly the functions  $(g_{i,\pi})_{0 \leq i \leq n-1}$  by  $g_{n-1,\pi} = f$  and for any integer  $0 \leq i \leq n-2$

$$g_{i,\pi}(x_{0:i}) = \mathbb{E}_\pi [f(x_{0:i}, X_{1:n-1-i})] = \mathbb{E}_\pi [f(x_{0:i}, X_{i+1:n-1})].$$

Thanks to Markov property, for any integer  $0 \leq i \leq n-1$  and all  $x$  in  $X$

$$\mathbb{E}_x [f(X_{0:n-1}) | \mathcal{F}_i^X] = g_i(X_{0:i}) \quad \mathbb{P}_x\text{-a.s.}$$

Choose  $i$  in  $\{0, \dots, n-1\}$  and  $x_{0:i}$  in  $\mathcal{X}^{i+1}$  fixed. It holds

$$|g_i(x_{0:i}) - g_{i,\pi}(x_{0:i})| = |\mathbb{E}_{x_i} [f(x_{0:i}, X_{1:n-1-i})] - \mathbb{E}_\pi [f(x_{0:i}, X_{1:n-1-i})]|.$$

Since  $f$  is in  $\mathbb{BD}(\mathcal{X}^n, \gamma_{0:n-1})$ , the function  $\tilde{f}_i : y_{0:n-2-i} \in \mathcal{X}^{n-1-i} \mapsto f(x_{0:i}, y_{0:n-2-i}) \in \mathbb{R}$  satisfies for any  $y_{0:n-2-i}$  in  $\mathcal{X}^{n-1-i}$  and any  $\tilde{y}_{0:n-2-i}$  in  $\mathcal{X}^{n-1-i}$

$$\begin{aligned} \left| \tilde{f}_i(y_{0:n-2-i}) - \tilde{f}_i(\tilde{y}_{0:n-2-i}) \right| &= \left| f(x_{0:i}^i, y_{0:n-2-i}) - f(x_{0:i}^i, \tilde{y}_{0:n-2-i}) \right| \\ &\leq \sum_{k=0}^i \gamma_k \mathbb{1}_{\{x_k \neq \tilde{x}_k\}} + \sum_{k=0}^{n-2-i} \gamma_{i+k+1} \mathbb{1}_{\{y_k \neq \tilde{y}_k\}} \leq \sum_{k=0}^{n-2-i} \gamma_{i+k+1} \mathbb{1}_{\{y_k \neq \tilde{y}_k\}}, \end{aligned}$$

what exactly means that  $\tilde{f}_i$  has bounded differences property and even that  $\tilde{f}_i$  belongs to  $\mathbb{BD}(\mathcal{X}^{n-1-i}, \gamma_{i+1:n-1})$ . Lemma 42 applied to  $\tilde{f}_i$  and the invariance of  $\pi$  for  $P$  lead to

$$\begin{aligned} (4.28) \quad |g_i(x_{0:i}) - g_{i,\pi}(x_{0:i})| &= |\mathbb{E}_{x_i} [f(x_{0:i}, X_{1:n-1-i})] - \mathbb{E}_\pi [f(x_{0:i}, X_{1:n-1-i})]| \\ &= \left| \mathbb{E}_{\delta_{x_i} P} [\tilde{f}_i(X_{0:n-2-i})] - \mathbb{E}_\pi [\tilde{f}_i(X_{0:n-2-i})] \right| \\ &\leq 2 \sum_{k=i+1}^{n-1} \gamma_k d_{\text{TV}}(\delta_{x_i} P^{k-i}, \pi). \end{aligned}$$

Assumption **M3** ensures that there exist  $\zeta_0$  in  $\mathbb{R}_+^*$  and  $r_0$  in  $(0, 1)$  such that for any  $x$  in  $C$  and any  $n$  in  $\mathbb{N}$ ,  $d_{\text{TV}}(\delta_x P^n, \pi) \leq \zeta_0 r_0^n$ . Hence Inequality (4.28) implies for any  $x_{0:i} = (x_0, \dots, x_i)$  in  $\mathcal{X}^i \times C$

$$(4.29) \quad |g_i(x_{0:i}) - g_{i,\pi}(x_{0:i})| \leq 2\zeta_0 \sum_{k=i+1}^{n-1} \gamma_k r_0^{k-i}.$$

For any integer  $1 \leq i \leq n$ , applying Lemma 91 of Appendix C with  $T = \tau_C^{i-1}$

$$\begin{aligned} (4.30) \quad G_{i,1} &= \mathbb{E}_x \left[ f(X_{0:n-1}) \Big| \mathcal{F}_{\tau_C^{i-1}}^X \right] \mathbb{1}_{\{\tau_C^{i-1} = i-1\}} \\ &= \mathbb{E}_x \left[ f(X_{0:n-1}) \Big| \mathcal{F}_{i-1}^X \right] \mathbb{1}_{\{\tau_C^{i-1} = i-1\}} \\ &= g_{i-1}(X_{0:i-1}) \mathbb{1}_{\{\tau_C^{i-1} = i-1\}} = R_{i,1} + g_{i-1,\pi}(X_{0:i-1}) \mathbb{1}_{\{\tau_C^{i-1} = i-1\}}, \end{aligned}$$

where  $R_{i,1} = g_{i-1}(X_{0:i-1}) \mathbb{1}_{\{\tau_C^{i-1} = i-1\}} - g_{i-1,\pi}(X_{0:i-1}) \mathbb{1}_{\{\tau_C^{i-1} = i-1\}}$ .

Using Inequality (4.29), we get

$$(4.31) \quad |R_{i,1}| \leq |g_{i-1}(X_{0:i-1}) - g_{i-1,\pi}(X_{0:i-1})| \mathbb{1}_{\{\tau_C^{i-1} = i-1\}} \leq 2\zeta_0 \left( \sum_{k=i}^{n-1} \gamma_k r_0^{k-i+1} \right) \mathbb{1}_{\{\tau_C^{i-1} = i-1\}}.$$



On the other hand, since  $\tau_C^i \geq i$ , we have

$$\mathcal{X}^{\mathbb{N}} = \{\tau_C^i \geq n-1\} \sqcup \{i \leq \tau_C^i \leq n-2\} = \{\tau_C^i \geq n-1\} \sqcup \left( \bigsqcup_{j=i}^{n-2} \{\tau_C^i = j\} \right),$$

so that  $G_{i,2}$  decomposes as

$$\begin{aligned} G_{i,2} &= \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{\tau_C^i}^X \right] \mathbb{1}_{\{\tau_C^{i-1}=i-1\}} \\ &= \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{\tau_C^i}^X \right] \left( \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i \geq n-1\}} + \mathbb{1}_{\{\tau_C^{i-1}=i-1, i \leq \tau_C^i \leq n-2\}} \right) \\ &= \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{\tau_C^i}^X \right] \left( \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i \geq n-1\}} + \sum_{j=i}^{n-2} \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i=j\}} \right). \end{aligned}$$

Applying Lemma 91 of Appendix C with  $T = \tau_C^i$ , for any integers  $i \leq j \leq n-2$

$$\begin{aligned} \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{\tau_C^i}^X \right] \sum_{j=i}^{n-2} \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i=j\}} &= \sum_{j=i}^{n-2} \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{\tau_C^i}^X \right] \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i=j\}} \\ &= \sum_{j=i}^{n-2} \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_j^X \right] \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i=j\}} \\ &= \sum_{j=i}^{n-2} g_j(X_{0:j}) \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i=j\}}, \end{aligned}$$

and similarly

$$\begin{aligned} \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{\tau_C^i}^X \right] \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i \geq n-1\}} &= \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{\tau_C^i}^X \right] \mathbb{1}_{\{\tau_C^i \geq n-1\}} \mathbb{1}_{\{\tau_C^{i-1}=i-1\}} \\ &= \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{n-1}^X \right] \mathbb{1}_{\{\tau_C^i \geq n-1\}} \mathbb{1}_{\{\tau_C^{i-1}=i-1\}} \\ &= f(X_{0:n-1}) \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i \geq n-1\}}, \end{aligned}$$

so that  $G_{i,2}$  finally decomposes as

$$(4.32) \quad G_{i,2} = f(X_{0:n-1}) \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i \geq n-1\}} + \sum_{j=i}^{n-2} g_j(X_{0:j}) \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i=j\}}.$$

Let us define

$$(4.33) \quad R_{i,2} = \sum_{j=i}^{n-2} (g_j(X_{0:j}) - g_{j,\pi}(X_{0:j})) \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i=j\}}.$$

Noting that for each  $i \leq j \leq n-2$ ,  $\{\tau_C^i = j\} \subset \{X_j \in C\}$ , Inequality (4.29) ensures

$$\begin{aligned} |g_j(X_{0:j}) - g_{j,\pi}(X_{0:j})| \mathbb{1}_{\{\tau_C^i=j\}} &= |g_j(X_{0:j}) - g_{j,\pi}(X_{0:j})| \mathbb{1}_{\{\tau_C^i=j\}} \\ &\leq 2\zeta_0 \sum_{k=j+1}^{n-1} \gamma_k r_0^{k-j} \mathbb{1}_{\{\tau_C^i=j\}} \leq 2\zeta_0 \sum_{k=\tau_C^i+1}^{n-1} \gamma_k r_0^{k-\tau_C^i} \mathbb{1}_{\{\tau_C^i=j\}} \end{aligned}$$

and hence

$$\begin{aligned}
(4.34) \quad |R_{i,2}| &= \left| \sum_{j=i}^{n-2} (g_j(X_{0:j}) - g_{j,\pi}(X_{0:j})) \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i=j\}} \right| \\
&\leq \sum_{j=i}^{n-2} |g_j(X_{0:j}) - g_{j,\pi}(X_{0:j})| \mathbb{1}_{\{\tau_C^i=j\}} \mathbb{1}_{\{\tau_C^{i-1}=i-1\}} \\
&\leq 2\zeta_0 \left( \sum_{k=1+\tau_C^i}^{n-1} \gamma_k r_0^{k-\tau_C^i} \right) \mathbb{1}_{\tau_C^{i-1}=i-1}.
\end{aligned}$$

Using Equation (4.30), we derive

$$G_{i,1} = R_{i,1} + g_{i-1,\pi}(X_{0:i-1}) \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i \geq n-1\}} + \sum_{j=i}^{n-2} g_{i-1,\pi}(X_{0:i-1}) \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i=j\}},$$

and using Equation (4.32) combined with definition of  $R_{i,2}$  in Equation (4.33), we derive

$$G_{i,2} = f(X_{0:n-1}) \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i \geq n-1\}} + R_{i,2} + \sum_{j=i}^{n-2} g_{j,\pi}(X_{0:j}) \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i=j\}},$$

Combining those decompositions of  $G_{i,1}$  and  $G_{i,2}$  with Equation (4.27), we get

$$\begin{aligned}
(4.35) \quad G_i - G_{i-1} &= G_{i,2} - G_{i,1} \\
&= R_{i,2} - R_{i,1} + (f(X_{0:n-1}) - g_{i-1,\pi}(X_{0:i-1})) \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i \geq n-1\}} \\
&\quad + \sum_{j=i}^{n-2} (g_{j,\pi}(X_{0:j}) - g_{i-1,\pi}(X_{0:i-1})) \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i=j\}}.
\end{aligned}$$

By definition of  $g_{i-1,\pi}$  and Jensen inequality

$$\begin{aligned}
|f(X_{0:n-1}) - g_{i-1,\pi}(x_{0:i-1})| &= |f(X_{0:n-1}) - \mathbb{E}_\pi[f(x_{0:i-1}, X_{1:n-i})]| \\
&= |\mathbb{E}_\pi[f(X_{0:n-1})] - \mathbb{E}_\pi[f(x_{0:i-1}, X_{1:n-i})]| \\
&= |\mathbb{E}_\pi[f(X_{0:n-1}) - f(x_{0:i-1}, X_{1:n-i})]| \\
&\leq \mathbb{E}_\pi[|f(X_{0:n-1}) - f(x_{0:i-1}, X_{1:n-i})|].
\end{aligned}$$

Since  $(X_{0:n-1})$  and  $(x_{0:i-1}, X_{1:n-i})$  differ at most by their  $n-i$  last components, bounded differences property of  $f$  in  $\mathbb{BD}(\mathcal{X}^n, \gamma_{0:n-1})$  ensures

$$|f(X_{0:n-1}) - f(x_{0:i-1}, X_{1:n-i})| \leq \sum_{k=i}^{n-1} \gamma_k,$$

so that

$$(4.36) \quad |f(X_{0:n-1}) - g_{i-1,\pi}(x_{0:i-1})| \leq \bar{\mathbb{E}}_\pi[|f(X_{0:n-1}) - f(x_{0:i-1}, X_{1:n-i})|] \leq \sum_{k=i}^{n-1} \gamma_k.$$

For any integer  $i \leq j \leq n-2$ , by definition of  $g_{j,\pi}$  and  $g_{i-1,\pi}$  and by Jensen inequality

$$\begin{aligned} |g_{j,\pi}(x_{0:j}) - g_{i-1,\pi}(x_{0:i-1})| &= |\mathbb{E}_\pi[f(x_{0:j}, X_{j+1:n-1})] - \mathbb{E}_\pi[f(x_{0:i-1}, X_{i:n-1})]| \\ &= |\mathbb{E}_\pi[f(x_{0:j}, X_{j+1:n-1}) - f(x_{0:i-1}, X_{i:n-1})]| \\ &\leq \mathbb{E}_\pi[|f(x_{0:j}, X_{j+1:n-1}) - f(x_{0:i-1}, X_{i:n-1})|] \end{aligned}$$

Since  $(x_{0:j}, X_{j+1:n-1})$  and  $(x_{0:i-1}, X_{i:n-1})$  differ at worst from the  $i$ -th to  $j$ -th components, bounded differences property of  $f$  in  $\mathbb{BD}(\mathcal{X}^n, \gamma_{0:n-1})$  ensures

$$|f(x_{0:j}, X_{j+1:n-1}) - f(x_{0:i-1}, X_{i:n-1})| \leq \sum_{k=i}^j \gamma_k,$$

so that for  $i \leq j \leq n-2$

$$|g_{j,\pi}(x_{0:j}) - g_{i-1,\pi}(x_{0:i-1})| \leq \mathbb{E}_\pi[|f(x_{0:j}, X_{j+1:n-1}) - f(x_{0:i-1}, X_{i:n-1})|] \leq \sum_{k=i}^j \gamma_k,$$

and it leads to

$$(4.37) \quad \sum_{j=i}^{n-2} |g_{j,\pi}(X_{0:j}) - g_{i-1,\pi}(X_{0:i-1})| \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i=j\}} \leq \sum_{j=i}^{n-2} \sum_{k=i}^j \gamma_k \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i=j\}} \\ \leq \sum_{k=i}^{\tau_C^i} \gamma_k \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i \leq n-2\}}.$$

Plugging Inequalities (4.34), (4.31), (4.36) and (4.37) in Inequality (4.35)

$$\begin{aligned} |G_i - G_{i-1}| &\leq |R_{i,2}| + |R_{i,1}| + \left| (f(X_{0:n-1}) - g_{i-1,\pi}(X_{0:i-1})) \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i \geq n-1\}} \right| \\ &\quad + \sum_{j=i}^{n-2} |g_{j,\pi}(X_{0:j}) - g_{i-1,\pi}(X_{0:i-1})| \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i=j\}} \\ &\leq 2\zeta_0 \left( \sum_{k=1+\tau_C^i}^{n-1} \gamma_k r_0^{k-\tau_C^i} \right) \mathbb{1}_{\tau_C^{i-1}=i-1} + 2\zeta_0 \left( \sum_{k=i}^{n-1} \gamma_k r_0^{k-i+1} \right) \mathbb{1}_{\tau_C^{i-1}=i-1} \\ &\quad + \sum_{k=i}^{n-1} \gamma_k \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i \geq n-1\}} + \sum_{k=i}^{\tau_C^i} \gamma_k \mathbb{1}_{\{\tau_C^{i-1}=i-1, \tau_C^i \leq n-2\}}, \end{aligned}$$

so that finally

$$(4.38) \quad |G_i - G_{i-1}| \leq \left( 2\zeta_0 \sum_{k=i}^{n-1} \gamma_k r_0^{k-i+1} + 2\zeta_0 \sum_{k=\tau_C^i+1}^{n-1} \gamma_k r_0^{k-\tau_C^i} + \sum_{k=i}^{\min(\tau_C^i, n-1)} \gamma_k \right) \mathbb{1}_{\tau_C^{i-1}=i-1}.$$

Definitions of  $\tau_C^i$ ,  $\tau_C$  and  $\sigma_C$  ensure

$$(4.39) \quad \tau_C^i = i + \tau_C \circ \theta^i \geq i \quad \text{and} \quad \sigma_C = 1 + \tau_C \circ \theta,$$

so that

$$(4.40) \quad \sigma_C \circ \theta^{i-1} = 1 + \tau_C \circ \theta^i = 1 + \tau_C^i - i \geq 1 > r_0.$$

Using Inequalities (4.40) and (4.38), we are now able to prove Inequalities (4.14) and (4.15).

(I) First, we concentrate on the upper-bound on  $|G_i - G_{i-1}|$  claimed in Inequality (4.14) denoting  $\bar{\gamma} = \max\{\gamma_0, \dots, \gamma_{n-1}\}$ .

Since  $r_0 < 1$ , we have on the one hand

$$(4.41) \quad \sum_{k=i}^{n-1} \gamma_k r_0^{k-i+1} \leq \bar{\gamma} r_0 \sum_{k=i}^{n-1} r_0^{k-i} = \bar{\gamma} r_0 \frac{1 - r_0^{n-i}}{1 - r_0} \leq \bar{\gamma} r_0 \frac{1}{1 - r_0},$$

and on the other hand

$$(4.42) \quad \sum_{k=\tau_C^i+1}^{n-1} \gamma_k r_0^{k-\tau_C^i} \leq \bar{\gamma} r_0 \sum_{k=0}^{n-\tau_C^i-2} r_0^k = \bar{\gamma} r_0 \frac{1 - r_0^{n-1-\tau_C^i}}{1 - r_0} \leq \bar{\gamma} r_0 \frac{1}{1 - r_0},$$

and since  $\tau_C^i - i + 1 = \sigma_C \circ \theta^{i-1}$  by Equation (4.40), we have

$$(4.43) \quad \sum_{k=i}^{\min(\tau_C^i, n-1)} \gamma_k \leq \bar{\gamma} (\min(\tau_C^i, n-1) - i + 1) \leq \bar{\gamma} (\tau_C^i - i + 1) \leq \bar{\gamma} \sigma_C \circ \theta^{i-1}.$$

Hence, plugging Inequalities (4.41), (4.42) and (4.43) in Inequality (4.38), with the fact that  $r_0 \leq \sigma_C \circ \theta^{i-1}$  granted by Inequality (4.40), we finally derive

$$\begin{aligned} |G_i - G_{i-1}| &\leq \left( \bar{\gamma} \frac{4\zeta_0}{1 - r_0} r_0 + \bar{\gamma} \sigma_C \circ \theta^{i-1} \right) \mathbb{1}_{\tau_C^{i-1}=i-1} \\ &\leq \bar{\gamma} \left( \frac{4\zeta_0}{1 - r_0} + 1 \right) \sigma_C \circ \theta^{i-1} \mathbb{1}_{\tau_C^{i-1}=i-1} \leq \zeta_1(\zeta_0, r_0) \bar{\gamma} \sigma_C \circ \theta^{i-1} \mathbb{1}_{\tau_C^{i-1}=i-1}. \end{aligned}$$

(II) Let us now get the upper-bound on  $|G_i - G_{i-1}|^2$  claimed in Inequality (4.15).

For any  $r$  in  $(r_0, 1)$ , since  $k - i + 1 \geq 0$  for any  $i \leq k \leq n - 1$ , we have  $r_0^{k-i+1} \leq r^{k-i+1}$  so that

$$\sum_{k=i}^{n-1} \gamma_k r_0^{k-i+1} \leq \sum_{k=i}^{n-1} \gamma_k r^{k-i+1} = r \sum_{k=i}^{n-1} \gamma_k r^{k-i} \leq \sum_{k=i}^{n-1} \gamma_k r^{k-i},$$

and similarly since  $k - \tau_C^i \geq 0$  for any  $\tau_C^i + 1 \leq k \leq n - 1$ , we have  $r_0^{k-\tau_C^i} \leq r^{k-\tau_C^i}$  so that using Equation (4.40)

$$\begin{aligned} \sum_{k=\tau_C^i+1}^{n-1} \gamma_k r_0^{k-\tau_C^i} &\leq \sum_{k=\tau_C^i+1}^{n-1} \gamma_k r^{k-\tau_C^i} \leq r^{-\tau_C^i+i} \sum_{k=\tau_C^i+1}^{n-1} \gamma_k r^{k-i} \\ &\leq r^{1-\sigma_C \circ \theta^{i-1}} \sum_{k=\tau_C^i+1}^{n-1} \gamma_k r^{k-i} \leq r^{-\sigma_C \circ \theta^{i-1}} \left( \sum_{k=\tau_C^i+1}^{n-1} \gamma_k r^{k-i} \right). \end{aligned}$$

Similarly, for any  $r$  in  $(0, 1)$ , since  $k - \tau_C^i \leq 0$  for any  $i \leq k \leq \min(\tau_C^i, n - 1)$ , we have  $1 \leq r^{k-\tau_C^i}$  so that using Equation (4.40)

$$\begin{aligned} \sum_{k=i}^{\min(\tau_C^i, n-1)} \gamma_k &\leq \sum_{k=i}^{\min(\tau_C^i, n-1)} \gamma_k r^{k-\tau_C^i} \leq r^{-\tau_C^i+i} \sum_{k=i}^{\min(\tau_C^i, n-1)} \gamma_k r^{k-i} \\ &\leq r^{1-\sigma_C \circ \theta^{i-1}} \sum_{k=i}^{\min(\tau_C^i, n-1)} \gamma_k r^{k-i} \leq r^{-\sigma_C \circ \theta^{i-1}} \left( \sum_{k=i}^{\min(\tau_C^i, n-1)} \gamma_k r^{k-i} \right). \end{aligned}$$

Hence, for any  $r$  in  $(r_0, 1)$ , from Inequality (4.38), we derive on  $\{\tau_C^{i-1} = i-1\}$

$$\begin{aligned}
& |G_i - G_{i-1}| \\
& \leq \left[ 2\zeta_0 \sum_{k=i}^{n-1} \gamma_k r^{k-i} + 2\zeta_0 r^{-\sigma_C \circ \theta^{i-1}} \left( \sum_{k=\tau_C^i+1}^{n-1} \gamma_k r^{k-i} \right) + r^{-\sigma_C \circ \theta^{i-1}} \left( \sum_{k=i}^{\min(\tau_C^i, n-1)} \gamma_k r^{k-i} \right) \right] \\
& \leq \left( 2\zeta_0 \sum_{k=i}^{n-1} \gamma_k r^{k-i} + \max(2\zeta_0, 1) r^{-\sigma_C \circ \theta^{i-1}} \left( \sum_{k=i}^{n-1} \gamma_k r^{k-i} \right) \right) \\
& \leq 2 \max(2\zeta_0, 1) \left( \sum_{k=i}^{n-1} \gamma_k r^{k-i} \right) r^{-\sigma_C \circ \theta^{i-1}}.
\end{aligned}$$

Cauchy-Schwarz inequality yields

$$\begin{aligned}
\left( \sum_{k=i}^{n-1} \gamma_k r^{k-i} \right)^2 &= \left( \sum_{k=i}^{n-1} (\gamma_k \sqrt{r^{k-i}}) \sqrt{r^{k-i}} \right)^2 \\
&\leq \left( \sum_{k=i}^{n-1} \gamma_k^2 r^{k-i} \right) \left( \sum_{k=i}^{n-1} r^{k-i} \right) \leq \left( \sum_{k=i}^{n-1} \gamma_k^2 r^{k-i} \right) \frac{1}{1-r},
\end{aligned}$$

so that finally

$$\begin{aligned}
|G_i - G_{i-1}|^2 &\leq 4 \max(2\zeta_0, 1)^2 \left( \sum_{k=i}^{n-1} \gamma_k r^{k-i} \right)^2 r^{-2\sigma_C \circ \theta^{i-1}} \mathbb{1}_{\tau_C^{i-1}=i-1} \\
&\leq \zeta_2(\zeta_0, r) r^{-2\sigma_C \circ \theta^{i-1}} \left( \sum_{k=i}^{n-1} \gamma_k^2 r^{k-i} \right) \mathbb{1}_{\tau_C^{i-1}=i-1}.
\end{aligned}$$

#### 4.3.2.4 Proof of Lemma 45

Fix  $x$  in  $C$ .

Since  $\tau_C = \tau_C^0 = 0$  under  $\mathbb{P}_x$  for  $x$  in  $C$ , we have for all  $x$  in  $C$  and all  $n$  in  $\mathbb{N}^*$

$$\mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{\tau_C^0}^X \right] = \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_0^X \right] = \mathbb{E}_x \left[ f(X_{0:n-1}) \right] \quad \mathbb{P}_x\text{-a.s.},$$

so that  $\mathbb{P}_x$ -almost surely  $G_0 = \mathbb{E}_x \left[ f(X_{0:n-1}) \right]$ .

By definition,  $\tau_C^{n-1} \geq n-1$  so that  $\mathbb{1}_{\tau_C^{n-1} \geq n-1} = 1$ . Hence, Lemma 91 of Appendix C ensures that  $\mathbb{P}_x$ -almost surely for all  $n$  in  $\mathbb{N}^*$

$$\begin{aligned}
\mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{\tau_C^{n-1}}^X \right] &= \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{\tau_C^{n-1}}^X \right] \mathbb{1}_{\tau_C^{n-1} \geq n-1} = \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{n-1}^X \right] \mathbb{1}_{\tau_C^{n-1} \geq n-1} \\
&= \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{n-1}^X \right] = f(X_{0:n-1}),
\end{aligned}$$

so that  $\mathbb{P}_x$ -almost surely  $G_{n-1} = f(X_{0:n-1})$ .

Hence, we get the following decomposition of  $f(X_{0:n-1}) - \mathbb{E}_x \left[ f(X_{0:n-1}) \right]$

$$(4.44) \quad f(X_{0:n-1}) - \mathbb{E}_x \left[ f(X_{0:n-1}) \right] = G_{n-1} - G_0 = \sum_{i=1}^{n-1} (G_i - G_{i-1}).$$

For all integer  $1 \leq i \leq n-1$ , using Lemma 43

$$G_i - G_{i-1} = (G_i - G_{i-1}) \mathbb{1}_{\{\tau_C^{i-1} = i-1\}}.$$

By Lemma 44, for all  $r$  in  $[r_0, 1)$ , there exist constants  $\zeta_1(\zeta_0, r_0)$  and  $\zeta_2(\zeta_0, r)$  such that for any  $n$  in  $\mathbb{N}$ , any integer  $1 \leq i \leq n-1$ , any  $\gamma_{0:n-1}$  in  $\mathbb{R}_+^n$ , any  $f$  in  $\mathbb{BD}(\mathcal{X}^n, \gamma_{0:n-1})$  and any  $x$  in  $C$

$$(4.45) \quad |G_i - G_{i-1}| \leq \zeta_1(\zeta_0, r_0) \max_{0 \leq j \leq n-1} (\gamma_j) \mathbb{1}_{\{\tau_C^{i-1} = i-1\}} \sigma_C \circ \theta^{i-1},$$

and

$$(4.46) \quad |G_i - G_{i-1}|^2 \leq \zeta_2(\zeta_0, r) \mathbb{1}_{\{\tau_C^{i-1} = i-1\}} r^{-2\sigma_C \circ \theta^{i-1}} \sum_{k=i}^{n-1} \gamma_k^2 r^{k-i}.$$

Using that  $e^t \leq 1 + t + t^2 e^{|t|}$  for every real  $t$ , yields

$$\mathbb{E}_x \left[ e^{G_i - G_{i-1}} \middle| \mathcal{F}_{\tau_C^{i-1}}^X \right] \leq \mathbb{E}_x \left[ 1 + (G_i - G_{i-1}) + (G_i - G_{i-1})^2 e^{|G_i - G_{i-1}|} \middle| \mathcal{F}_{\tau_C^{i-1}}^X \right].$$

For any  $1 \leq i \leq n-1$ , since  $\tau_C^{i-1} \leq \tau_C^i$ ,  $\mathcal{F}_{\tau_C^{i-1}}^X \subset \mathcal{F}_{\tau_C^i}^X$  and tower property of conditional expectation in Lemma 92-Item (iii) of Appendix C thus ensures

$$\mathbb{E}_x \left[ \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{\tau_C^i}^X \right] \middle| \mathcal{F}_{\tau_C^{i-1}}^X \right] = \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{\tau_C^{i-1}}^X \right],$$

so that

$$\begin{aligned} & \mathbb{E}_x \left[ G_i - G_{i-1} \middle| \mathcal{F}_{\tau_C^{i-1}}^X \right] = \mathbb{E}_x \left[ G_i \middle| \mathcal{F}_{\tau_C^{i-1}}^X \right] - \mathbb{E}_x \left[ G_{i-1} \middle| \mathcal{F}_{\tau_C^{i-1}}^X \right] \\ &= \mathbb{E}_x \left[ \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{\tau_C^i}^X \right] \middle| \mathcal{F}_{\tau_C^{i-1}}^X \right] - \mathbb{E}_x \left[ \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{\tau_C^{i-1}}^X \right] \middle| \mathcal{F}_{\tau_C^{i-1}}^X \right] \\ &= \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{\tau_C^i}^X \right] - \mathbb{E}_x \left[ f(X_{0:n-1}) \middle| \mathcal{F}_{\tau_C^{i-1}}^X \right] = 0, \end{aligned}$$

and thus

$$\begin{aligned} & \mathbb{E}_x \left[ 1 + (G_i - G_{i-1}) + (G_i - G_{i-1})^2 e^{|G_i - G_{i-1}|} \middle| \mathcal{F}_{\tau_C^{i-1}}^X \right] \\ & \leq 1 + \mathbb{E}_x \left[ (G_i - G_{i-1})^2 e^{|G_i - G_{i-1}|} \middle| \mathcal{F}_{\tau_C^{i-1}}^X \right]. \end{aligned}$$

Since  $G_i - G_{i-1} = 0$  outside  $\{\tau_C^{i-1} = i-1\}$  by Lemma 43, Lemma 91 of Appendix C yields

$$\begin{aligned} \mathbb{E}_x \left[ (G_i - G_{i-1})^2 e^{|G_i - G_{i-1}|} \middle| \mathcal{F}_{\tau_C^{i-1}}^X \right] &= \mathbb{E}_x \left[ (G_i - G_{i-1})^2 \mathbb{1}_{\{\tau_C^{i-1} = i-1\}} e^{|G_i - G_{i-1}|} \middle| \mathcal{F}_{\tau_C^{i-1}}^X \right] \\ &\leq \mathbb{E}_x \left[ (G_i - G_{i-1})^2 e^{|G_i - G_{i-1}|} \middle| \mathcal{F}_{\tau_C^{i-1}}^X \right] \mathbb{1}_{\{\tau_C^{i-1} = i-1\}}. \end{aligned}$$

We set

$$\varepsilon = \frac{\log(\delta)}{2\zeta_1(\zeta_0, r_0)} \quad \text{and} \quad r = \frac{1 + \max\left(r_0, \delta^{-\frac{1}{4}}\right)}{2},$$

where  $\delta$  is defined in **M2** so that  $\varepsilon$  lies in  $(0, \zeta_1(\zeta_0, r_0)^{-1} \log(\delta))$  and  $r$  in  $[r_0, 1)$  and thus

$$(4.47) \quad -2 \log(r) + \zeta_1(\zeta_0, r_0) \varepsilon \leq \log(\delta).$$

With Inequalities (4.45) and (4.46), we get

$$\begin{aligned}
& \mathbb{E}_x \left[ (G_i - G_{i-1})^2 e^{|G_i - G_{i-1}|} \middle| \mathcal{F}_{i-1}^X \right] \\
& \leq \mathbb{E}_x \left[ \zeta_2(\zeta_0, r) r^{-2\sigma_C \circ \theta^{i-1}} \sum_{k=i}^{n-1} \gamma_k^2 r^{k-i} e^{\zeta_1(\zeta_0, r_0) \max_{0 \leq i \leq n-1} (\gamma_i) \sigma_C \circ \theta^{i-1}} \middle| \mathcal{F}_{i-1}^X \right] \\
& \leq \zeta_2(\zeta_0, r) \left( \sum_{k=i}^{n-1} \gamma_k^2 r^{k-i} \right) \mathbb{E}_x \left[ e^{\left( -2 \log(r) + \zeta_1(\zeta_0, r_0) \max_{0 \leq i \leq n-1} (\gamma_i) \right) \sigma_C \circ \theta^{i-1}} \middle| \mathcal{F}_{i-1}^X \right],
\end{aligned}$$

but Markov property ensures

$$\mathbb{E}_x \left[ e^{\left( -2 \log(r) + \zeta_1(\zeta_0, r_0) \max_{0 \leq i \leq n-1} (\gamma_i) \right) \sigma_C \circ \theta^{i-1}} \middle| \mathcal{F}_{i-1}^X \right] = \mathbb{E}_{X_{i-1}} \left[ e^{\left( -2 \log(r) + \zeta_1(\zeta_0, r_0) \max_{0 \leq i \leq n-1} (\gamma_i) \right) \sigma_C} \right],$$

so that

$$\begin{aligned}
(4.48) \quad & \mathbb{E}_x \left[ (G_i - G_{i-1})^2 e^{|G_i - G_{i-1}|} \middle| \mathcal{F}_{i-1}^X \right] \\
& \leq \zeta_2(\zeta_0, r) \left( \sum_{k=i}^{n-1} \gamma_k^2 r^{k-i} \right) \mathbb{E}_{X_{i-1}} \left[ e^{\left( -2 \log(r) + \zeta_1(\zeta_0, r_0) \max_{0 \leq i \leq n-1} (\gamma_i) \right) \sigma_C} \right],
\end{aligned}$$

(I) We first deal with the case where  $\max(\gamma_0, \dots, \gamma_{n-1}) \leq \varepsilon$ .

From definition of  $\tau_C^{i-1}$ ,  $X_{i-1}$  lies in  $C$  on  $\{\tau_C^{i-1} = i-1\}$  so that

$$\begin{aligned}
(4.49) \quad & \mathbb{E}_{X_{i-1}} \left[ e^{\left( -2 \log(r) + \zeta_1(\zeta_0, r_0) \max_{0 \leq i \leq n-1} (\gamma_i) \right) \sigma_C} \right] \mathbb{1}_{\{\tau_C^{i-1} = i-1\}} \\
& \leq \sup_{x \in C} \left( \mathbb{E}_x \left[ e^{\left( -2 \log(r) + \zeta_1(\zeta_0, r_0) \max_{0 \leq i \leq n-1} (\gamma_i) \right) \sigma_C} \right] \right) \mathbb{1}_{\{\tau_C^{i-1} = i-1\}}
\end{aligned}$$

Using Inequality (4.47)

$$\begin{aligned}
(4.50) \quad & \sup_{x \in C} \left( \mathbb{E}_x \left[ e^{\left( -2 \log(r) + \zeta_1(\zeta_0, r_0) \max_{0 \leq i \leq n-1} (\gamma_i) \right) \sigma_C} \right] \right) \\
& \leq \sup_{x \in C} \left( \mathbb{E}_x \left[ e^{(-2 \log(r) + \zeta_1(\zeta_0, r_0) \varepsilon) \sigma_C} \right] \right) \\
& \leq \sup_{x \in C} \left( \mathbb{E}_x \left[ e^{\log(\delta) \sigma_C} \right] \right) \leq \sup_{x \in C} (\mathbb{E}_x [\delta^{\sigma_C}]) \leq M.
\end{aligned}$$

Using the convexity inequality  $1 + t \leq e^t$  valid for any real  $t$ , we get

$$(4.51) \quad 1 + \zeta_2(\zeta_0, r) \left( \sum_{k=i}^{n-1} \gamma_k^2 r^{k-i} \right) M \leq \exp \left( \zeta_2(\zeta_0, r) \left( \sum_{k=i}^{n-1} \gamma_k^2 r^{k-i} \right) M \right).$$

Plugging successively Inequalities (4.49), (4.50) and (4.51) in Inequality (4.48) leads for any integer  $1 \leq i \leq n-1$  to

$$(4.52) \quad \mathbb{E}_x \left[ e^{G_i - G_{i-1}} \middle| \mathcal{F}_{\tau_C^{i-1}}^X \right] \leq \exp \left( \zeta_2(\zeta_0, r) \left( \sum_{k=i}^{n-1} \gamma_k^2 r^{k-i} \right) M \right).$$

For integer  $1 \leq j \leq n-1$ , we have  $\tau_C^{n-j-1} \geq n-j-1$  so that Item (ii) in Definition 89 guarantees the inclusion  $\mathcal{F}_{\tau_C^{n-j-1}}^X \subset \mathcal{F}_{\tau_C^{n-j-1}}^X$ . The random variable  $e^{G_i-G_{i-1}}$  is thus  $\mathcal{F}_{\tau_C^{n-j-1}}$ -measurable for any integer  $1 \leq i \leq n-j-1$  so that

$$\mathbb{E}_x \left[ \prod_{i=1}^{n-j} e^{G_i-G_{i-1}} \middle| \mathcal{F}_{\tau_C^{n-j-1}}^X \right] = \prod_{i=1}^{n-j-1} e^{G_i-G_{i-1}} \mathbb{E}_x \left[ e^{G_{n-j}-G_{n-j-1}} \middle| \mathcal{F}_{\tau_C^{n-j-1}}^X \right].$$

and since  $1 \leq n-j \leq n-1$ , Inequality (4.51) yields

$$\begin{aligned} (4.53) \quad & \mathbb{E}_x \left[ \prod_{i=1}^{n-j} e^{G_i-G_{i-1}} \right] \\ &= \mathbb{E}_x \left[ \mathbb{E}_x \left[ \prod_{i=1}^{n-j} e^{G_i-G_{i-1}} \middle| \mathcal{F}_{\tau_C^{n-j-1}}^X \right] \right] \\ &= \mathbb{E}_x \left[ \prod_{i=1}^{n-j-1} e^{G_i-G_{i-1}} \mathbb{E}_x \left[ e^{G_{n-j}-G_{n-j-1}} \middle| \mathcal{F}_{\tau_C^{n-j-1}}^X \right] \right] \\ &\leq \mathbb{E}_x \left[ \prod_{i=1}^{n-j-1} e^{G_i-G_{i-1}} \right] \exp \left( \zeta_2(\zeta_0, r) \left( \sum_{k=n-j}^{n-1} \gamma_k^2 r^{k-(n-j)} \right) M \right). \end{aligned}$$

Hence, decomposing  $f(X_{0:n-1}) - \mathbb{E}_x[f(X_{0:n-1})]$  according to Equation (4.44) and using then recursively Inequality (4.53)

$$\begin{aligned} \mathbb{E}_x \left[ e^{f(X_{0:n-1}) - \mathbb{E}_x[f(X_{0:n-1})]} \right] &= \mathbb{E}_x \left[ e^{\sum_{i=1}^{n-1} (G_i - G_{i-1})} \right] = \mathbb{E}_x \left[ \prod_{i=1}^{n-1} e^{G_i - G_{i-1}} \right] \\ &\leq \prod_{j=1}^{n-1} \exp \left( \zeta_2(\zeta_0, r) \left( \sum_{k=n-j}^{n-1} \gamma_k^2 r^{k-(n-j)} \right) M \right) = \prod_{j=1}^{n-1} \exp \left( \zeta_2(\zeta_0, r) \left( \sum_{k=j}^{n-1} \gamma_k^2 r^{k-j} \right) M \right) \\ &\leq \exp \left( \zeta_2(\zeta_0, r) M \left( \sum_{j=1}^{n-1} \sum_{k=j}^{n-1} \gamma_k^2 r^{k-j} \right) \right). \end{aligned}$$

Since

$$\sum_{i=1}^{n-1} \sum_{k=i}^{n-1} \gamma_k^2 r^{k-i} = \sum_{k=1}^{n-1} \gamma_k^2 \sum_{i=1}^k r^{k-i} = \sum_{k=1}^{n-1} \gamma_k^2 \sum_{j=0}^{k-1} r^j = \sum_{k=1}^{n-1} \gamma_k^2 \frac{1-r^k}{1-r} \leq \frac{1}{1-r} \sum_{k=0}^{n-1} \gamma_k^2,$$

we finally have

$$(4.54) \quad \mathbb{E}_x \left[ e^{f(X_{0:n-1}) - \mathbb{E}_x[f(X_{0:n-1})]} \right] \leq \exp \left( \frac{\zeta_2(\zeta_0, r)}{1-r} M \sum_{k=0}^{n-1} \gamma_k^2 \right).$$

(II) We now deal with the second case :  $\max(\gamma_0, \dots, \gamma_{n-1}) > \varepsilon$ .

Fix  $\tilde{x}$  in  $\mathcal{X}$  and define an auxiliary function  $\tilde{f} : \mathcal{X}^n \rightarrow \mathbb{R}$  from  $f$  by denoting for any  $x_{0:n-1}$  in  $\mathcal{X}^n$

$$\tilde{f}(X_{0:n-1}) = f(x_0 \mathbb{1}_{\{\gamma_0 \leq \varepsilon\}} + \tilde{x} \mathbb{1}_{\{\gamma_0 > \varepsilon\}}, \dots, x_{n-1} \mathbb{1}_{\{\gamma_{n-1} \leq \varepsilon\}} + \tilde{x} \mathbb{1}_{\{\gamma_{n-1} > \varepsilon\}}).$$



Function  $f$  belonging to  $\mathbb{BD}(\mathcal{X}^n, \gamma_{0:n-1})$ , auxiliary function  $\tilde{f}$  automatically also belongs to  $\mathbb{BD}(\mathcal{X}^n, \tilde{\gamma}_{0:n-1})$  where

$$\tilde{\gamma}_0^{n-1} = (\gamma_0 \mathbb{1}_{\{\gamma_0 \leq \varepsilon\}}, \dots, \gamma_{n-1} \mathbb{1}_{\{\gamma_{n-1} \leq \varepsilon\}}).$$

Since  $\max(\tilde{\gamma}_0, \dots, \tilde{\gamma}_{n-1}) < \varepsilon$ ,  $\tilde{f}$  satisfies Inequality (4.54)

$$(4.55) \quad \begin{aligned} \mathbb{E}_x \left[ e^{\tilde{f}(X_{0:n-1}) - \mathbb{E}_x[\tilde{f}(X_{0:n-1})]} \right] &\leq \exp \left( \frac{\zeta_2(\zeta_0, r)}{1-r} M \sum_{k=0}^{n-1} \tilde{\gamma}_k^2 \right) \\ &\leq \exp \left( \frac{\zeta_2(\zeta_0, r)}{1-r} M \sum_{k=0}^{n-1} \gamma_k^2 \mathbb{1}_{\{\gamma_k \leq \varepsilon\}} \right) \\ &\leq \exp \left( \frac{\zeta_2(\zeta_0, r)}{1-r} M \sum_{k=0}^{n-1} \gamma_k^2 \right). \end{aligned}$$

Furthermore, by definition of  $\tilde{f}$  and since  $f$  is in  $\mathbb{BD}(\mathcal{X}^n, \gamma_{0:n-1})$ , for any  $x_{0:n-1}$  in  $\mathcal{X}^n$

$$(4.56) \quad |f(x_{0:n-1}) - \tilde{f}(x_{0:n-1})| = \sum_{i=0}^{n-1} \gamma_i \mathbb{1}_{\{\gamma_i > \varepsilon\}} \leq \sum_{i=0}^{n-1} \gamma_i \frac{\gamma_i}{\varepsilon} \leq \frac{1}{\varepsilon} \sum_{i=0}^{n-1} \gamma_i^2.$$

Since Inequality (4.56) yields on the one hand

$$f(X_{0:n-1}) - \tilde{f}(X_{0:n-1}) \leq \left| f(X_{0:n-1}) - \tilde{f}(X_{0:n-1}) \right| \leq \frac{1}{\varepsilon} \sum_{i=0}^{n-1} \gamma_i^2,$$

and on the other hand combined with Jensen Inequality

$$\begin{aligned} &\mathbb{E}_x \left[ \tilde{f}(X_{0:n-1}) \right] - \mathbb{E}_x [f(X_{0:n-1})] = \mathbb{E}_x \left[ \tilde{f}(X_{0:n-1}) - f(X_{0:n-1}) \right] \\ &\leq \left| \mathbb{E}_x \left[ \tilde{f}(X_{0:n-1}) - f(X_{0:n-1}) \right] \right| \leq \mathbb{E}_x \left[ \left| \tilde{f}(X_{0:n-1}) - f(X_{0:n-1}) \right| \right] \\ &\leq \mathbb{E}_x \left[ \frac{1}{\varepsilon} \sum_{i=0}^{n-1} \gamma_i^2 \right] \leq \frac{1}{\varepsilon} \sum_{i=0}^{n-1} \gamma_i^2, \end{aligned}$$

decomposing  $f(X_{0:n-1}) - \mathbb{E}_x[f(X_{0:n-1})]$  as the sum

$$\begin{aligned} &f(X_{0:n-1}) - \tilde{f}(X_{0:n-1}) + \tilde{f}(X_{0:n-1}) \\ &- \mathbb{E}_x \left[ \tilde{f}(X_{0:n-1}) \right] + \mathbb{E}_x \left[ \tilde{f}(X_{0:n-1}) \right] - \mathbb{E}_x [f(X_{0:n-1})], \end{aligned}$$

enables to write

$$\begin{aligned} &e^{f(X_{0:n-1}) - \mathbb{E}_x[f(X_{0:n-1})]} \\ &= e^{f(X_{0:n-1}) - \tilde{f}(X_{0:n-1})} e^{\tilde{f}(X_{0:n-1}) - \mathbb{E}_x[\tilde{f}(X_{0:n-1})]} e^{\mathbb{E}_x[\tilde{f}(X_{0:n-1})] - \mathbb{E}_x[f(X_{0:n-1})]} \\ &\leq e^{\frac{1}{\varepsilon} \sum_{i=0}^{n-1} \gamma_i^2} e^{\tilde{f}(X_{0:n-1}) - \mathbb{E}_x[\tilde{f}(X_{0:n-1})]} e^{\mathbb{E}_x[\frac{1}{\varepsilon} \sum_{i=0}^{n-1} \gamma_i^2]} \\ &\leq e^{\frac{2}{\varepsilon} \sum_{i=0}^{n-1} \gamma_i^2} e^{\tilde{f}(X_{0:n-1}) - \mathbb{E}_x[\tilde{f}(X_{0:n-1})]}. \end{aligned}$$

And hence

$$\begin{aligned} \mathbb{E}_x \left[ e^{f(X_{0:n-1}) - \mathbb{E}_x[f(X_{0:n-1})]} \right] &\leq \mathbb{E}_x \left[ e^{\frac{2}{\varepsilon} \sum_{i=0}^{n-1} \gamma_i^2} e^{\tilde{f}(X_{0:n-1}) - \mathbb{E}_x[\tilde{f}(X_{0:n-1})]} \right] \\ &\leq e^{\frac{2}{\varepsilon} \sum_{i=0}^{n-1} \gamma_i^2} \mathbb{E}_x \left[ e^{\tilde{f}(X_{0:n-1}) - \mathbb{E}_x[\tilde{f}(X_{0:n-1})]} \right], \end{aligned}$$

so that, Inequality (4.55) yields

$$\begin{aligned} \mathbb{E}_x \left[ e^{f(X_{0:n-1}) - \mathbb{E}_x[f(X_{0:n-1})]} \right] &\leq e^{\frac{2}{\varepsilon} \sum_{i=0}^{n-1} \gamma_i^2} \exp \left( \frac{\zeta_2(\zeta_0, r)}{1-r} M \sum_{k=0}^{n-1} \gamma_k^2 \mathbb{1}_{\{\gamma_k \leq \varepsilon\}} \right) \\ &\leq \exp \left( \left( \frac{2}{\varepsilon} + \frac{\zeta_2(\zeta_0, r)}{1-r} M \right) \sum_{k=0}^{n-1} \gamma_k^2 \right). \end{aligned}$$

Whatever the value of  $\max(\gamma_0, \dots, \gamma_{n-1})$ , we just proved that for all  $\gamma_{0:n-1}$  in  $\mathbb{R}_+^*$ ,  $f$  in  $\mathbb{BD}(\mathcal{X}^n, \gamma_{0:n-1})$  and  $x$  in  $C$

$$\mathbb{E}_x \left[ e^{f(X_{0:n-1}) - \mathbb{E}_x[f(X_{0:n-1})]} \right] \leq \exp \left( \kappa \sum_{i=0}^{n-1} \gamma_i^2 \right),$$

where explicit expression for  $\kappa(\zeta_0, r_0, \delta, M)$  is given in Equation (4.19).

## 4.4 Proof of Theorem 40

In this section, we want to prove that assumptions necessary for application of Theorem 39 to the specific transition kernel  $K^\nu$  given by Equation (4.10) are satisfied provided  $\nu$  meets some requirements.

In Section 4.4.1, we provides some theoretical prerequisites. In Section 4.4.2, we check that the Assumptions **M1**, **M2** and **M3** are indeed entailed by Assumptions **F1** and **F2** of Chapter 3.

### 4.4.1 $V$ -geometric ergodicity

Recall first the notion of Döblin set and the definition of the  $V$ -norm.

**Definition 46.** *If  $P$  is a Markov kernel on  $X \times \mathcal{X}$ , a set  $C$  in  $\mathcal{X}$  is a  $(m, \varepsilon)$ -Döblin set for  $P$  where  $m$  in  $\mathbb{N}^*$  and  $\varepsilon > 0$ , if and only if, for any  $x$  and  $x'$  in  $C$*

$$d_{\text{TV}}(P^m(x, \cdot), P^m(x', \cdot)) \leq 1 - \varepsilon.$$

**Remark 47.** *When the state space  $X$  is discrete, for every couple of probability measures  $(\mu_1, \mu_2)$  on  $\mathcal{X}$ , it holds*

$$d_{\text{TV}}(\mu_1, \mu_2) = 1 - \sum_{z \in X} \min(\mu_1(z), \mu_2(z)),$$

so that a set  $B$  in  $\mathcal{X}$  is a  $(m, \varepsilon)$ -Döblin set for Markov kernel  $P$  on  $X \times \mathcal{X}$ , if and only if, it holds

$$(4.57) \quad \inf_{(x, x') \in B \times B} \left( \sum_{z \in X} \min(P^m(x, z), P^m(x', z)) \right) \geq \varepsilon.$$

Therefore, if there exists  $z$  in  $X$  such that  $\inf_{x \in B} P^m(x, z) \geq \varepsilon$ , then  $B$  is a  $(m, \varepsilon)$ -Döblin set for  $P$ .

**Definition 48.** If  $V$  is a function from  $X$  into  $[1, +\infty]$ , the  $V$ -norm of any finite signed measure  $\xi$  on  $X$  such that  $|\xi|(V) < +\infty$  (where  $|\xi| = \xi^+ + \xi^-$  with  $(\xi^+, \xi^-)$  the Jordan decomposition of  $\xi$ ) is the positive real number denoted  $\|\xi\|_V$  and defined as

$$\|\xi\|_V = |\xi|(V).$$

**Remark 49.** Since  $\mathbb{1}_X \leq V$  and  $|\xi|$  is a (positive) measure for any signed measure  $\xi$ , it holds

$$(4.58) \quad \|\xi\|_{\text{TV}} = |\xi|(X) = |\xi|(\mathbb{1}_X) \leq |\xi|(V) = \|\xi\|_V.$$

And hence, if the  $V$ -norm of a signed measure is upper-bounded so is also its TV-norm and the upper-bound is the same.

The condition of uniform geometric ergodicity of the chain in total variation distance specified in **M3** asserts that the distribution of some Markov chain converges geometrically to its invariant distribution and even at which rate. This condition of geometric ergodicity with explicit rates can be granted by an adaptation of [DMPS18, Theorem 19.4.1] stated for completeness in Theorem 50.

**Theorem 50.** [DMPS18, Theorem 19.4.1] Let  $P$  be a Markov kernel on  $X \times \mathcal{X}$ . Assume that  $P$  satisfies the following conditions:

1. there exist  $V : X \rightarrow [1, +\infty)$ ,  $\lambda$  in  $[0, 1)$  and  $b$  in  $\mathbb{R}_+$  such that

$$PV \leq \lambda V + b,$$

2. there exist  $m$  in  $\mathbb{N}^*$ ,  $\varepsilon$  in  $(0, 1]$  and  $d > 0$  such that the level set  $\{V \leq d\}$  is a  $(m, \varepsilon)$ -Döblin set and  $\lambda + 2b/(1 + d) < 1$ .

Then,  $P$  admits a unique invariant probability measure  $\pi$ .

Furthermore,  $\pi(V) < +\infty$  and for all  $\xi$  in  $\mathcal{M}^1(\mathcal{X})$  and all  $n$  in  $\mathbb{N}^*$ , it holds

$$(4.59) \quad d_V(\xi P^n, \pi) \leq c_m (\pi(V) + \xi(V)) \rho^{\lfloor \frac{n}{m} \rfloor},$$

where

$$\begin{aligned} \log(\rho) &= \frac{\log(1 - \varepsilon) \log(\bar{\lambda}_m)}{\log(1 - \varepsilon) + \log(\bar{\lambda}_m) - \log(\bar{b}_m)}, \\ \bar{\lambda}_m &= \lambda^m + \frac{2b_m}{1 + d} \quad \text{and} \quad \bar{b}_m = \lambda^m b_m + d \quad \text{and} \quad b_m = b \frac{1 - \lambda^m}{1 - \lambda}, \\ c_m &= \frac{\lambda^m + \frac{1 - \lambda^m}{1 - \lambda}}{1 + \frac{b_m}{(1 - \varepsilon)(1 - \lambda_m)}}. \end{aligned}$$

**Remark 51.** The combination of Theorem 50-Item 1 and of the conclusions of Theorem 50 has an other interesting consequence. Since  $\pi$  is the invariant distribution for  $P$ , it holds  $\pi P = \pi$  and  $\pi(V) < +\infty$ . Then, we have

$$\begin{aligned} PV \leq \lambda V + b &\Rightarrow (\pi P)(V) \leq \lambda \pi(V) + b \\ &\Rightarrow \pi(V) \leq \lambda \pi(V) + b \\ &\Rightarrow \pi(V) \leq \frac{b}{1 - \lambda}. \end{aligned}$$

The term  $\pi(V)$  of Inequality (4.59) is thus not only finite but also precisely upper-bounded thanks to the drift condition Theorem 50-Item 1.

Theorem 50-Item 1, called *drift condition*, is a common condition in Markov chain theory closely related to the asymptotic behaviour of the chain. Indeed, [DMPS18, Proposition 4.3.3] (stated for completeness in Proposition 52) shows that the existence of a drift condition of the form  $PV \leq \lambda V + b\mathbb{1}_C$  is equivalent to the existence of a finite exponential moment for the return time to the set  $C$ .

**Proposition 52.** [DMPS18, Proposition 4.3.3] *Let  $P$  be a Markov kernel on  $X \times \mathcal{X}$  and  $C$  in  $\mathcal{X}$ .*

(i) *If there exists  $\beta$  in  $[1, +\infty)$  such that  $b = \sup_{x \in C} (\mathbb{E}_x[\beta^{\sigma_C}]) < +\infty$ , then, the function  $V : X \rightarrow (1, +\infty)$  defined by  $V(x) = \mathbb{E}_x[\beta^{\tau_C}]$  is such that*

$$(4.60) \quad PV \leq \beta^{-1}V + b\mathbb{1}_C.$$

(ii) *If there exist a function  $V : X \rightarrow [1, +\infty)$  and constants  $\lambda$  in  $[0, 1)$  and  $b < +\infty$  such that  $PV \leq \lambda V + b\mathbb{1}_C$ , then, for all  $x$  in  $X$*

$$(4.61) \quad \mathbb{E}_x[\lambda^{-\sigma_C}] \leq V(x) + b\lambda^{-1}.$$

**Remark 53.** *In the framework of Section 4.1, if Assumption M2 is satisfied there exist  $C$  in  $\mathcal{X}$ ,  $\delta$  in  $(1, +\infty)$  and  $M$  in  $(0, +\infty)$  such that  $\sup_{x \in C} (\mathbb{E}_x[\delta^{\sigma_C}]) \leq M$ . Defining for any  $x$  in  $\mathcal{X}$ , the drift function  $V$  by*

$$V(x) = \mathbb{E}_x[\delta^{\tau_C}],$$

[DMPS18, Proposition 4.3.3] ensures that

$$(4.62) \quad PV \leq \delta^{-1}V + M\mathbb{1}_C \leq \delta^{-1}V + M,$$

so that Theorem 50-Item 1 is satisfied.

But the function  $V$  has also an other interesting property : since  $\tau_C = 0$   $\mathbb{P}_x$ -a.s. for any  $x$  in  $C$ , we have  $\delta_x(V) = V(x) = \mathbb{E}_x[\delta^{\tau_C}] = \mathbb{E}_x[\delta^0] = 1$  and it holds

$$(4.63) \quad \sup_{x \in C} (\delta_x(V)) = \delta_0(V) = V(0) = 1.$$

This is in particular interesting for the control of term  $\xi(V)$  of the Inequality (4.59) when  $\text{supp}(\xi) \subset C$ .

#### 4.4.2 Satisfaction of Assumptions M1, M2 and M3

For fixed  $c_0 > 0$  and fixed  $\alpha$  in  $(0, 1/2)$ , consider  $\nu$  in  $\mathbb{M}^1([0, 1])$  such that  $\nu([\alpha, 1 - \alpha]) = 1$  and  $-\infty < \int_0^1 \log\left(\frac{1-t}{t}\right) \nu(dt) \leq -c_0$ . Hereafter, we set  $(\mathcal{X}, \mathcal{X}) = (\mathbb{N}, \mathcal{P}(\mathbb{N}))$ ,  $P = K^\nu$  and  $C = \{0\}$ .

For any  $(i, j) \in \mathbb{N}$ ,  $K^\nu(i, j)$  is strictly positive so that  $K^\nu$  is aperiodic. The set  $C$  is an atom and is therefore small. Furthermore, since  $K^\nu(i, 0) > 0$  for any  $i \in \mathbb{N}$ ,  $C$  is also an accessible set. It follows that, for any  $\nu$  in  $\Theta$ ,  $K^\nu$  is an irreducible and aperiodic Markov

kernel on  $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ .  $K^\nu$  admits thus an unique invariant probability denoted  $\pi_\nu$ . The Assumption **M1** is thus satisfied.

Consider the positive constants  $K_1(c_0, \alpha)$  and  $K_2(c, \alpha)$  given in Equations (3.85) and (3.86) of Chapter 3 and set

$$(4.64) \quad \delta = \delta(c_0, \alpha) e^{\frac{K_2(c_0, \alpha)}{2}} > 1,$$

and

$$(4.65) \quad M = M(c_0, \alpha) = 1 + K_1(c_0, \alpha) > 0.$$

By Corollary 15 of Chapter 3, we get

$$(4.66) \quad \sup_{x \in C} \mathbb{E}_x[\delta^{\sigma_C}] \leq M,$$

so that the Assumption **M2** is satisfied.

Now, we finally check that **M3** is satisfied by application of Theorem 50. Since Assumption **M2** is satisfied, Remark 53 ensures that the function  $V_\nu$  defined for all  $x$  in  $\mathbb{N}$  by  $V_\nu(x) = \mathbb{E}_x[\delta^{\tau_C}]$  satisfies

$$(4.67) \quad K^\nu V_\nu \leq \delta^{-1} V_\nu + M \mathbb{1}_C \leq \delta^{-1} V_\nu + M.$$

Hence the drift condition of Theorem 50-Item 1 is satisfied.

To check Theorem 50-Item 2, we have to show that there exist  $d > -1 + 2M/(1 - \delta^{-1})$ , a real number  $\varepsilon > 0$  such that  $\{V_\nu \leq d\}$  is an  $(1, \varepsilon)$ -Döbblin set for  $K^\nu$ . Thanks to Remark 47, it suffices to show that there exist  $d > -1 + 2M/(1 - \delta^{-1})$  and  $\varepsilon > 0$  such that

$$(4.68) \quad \inf_{x \in \{V_\nu \leq d\}} K^\nu(x, 0) \geq \varepsilon.$$

From definition of  $K^\nu$  and the fact that  $\text{supp}(\nu) \subset [\alpha, 1 - \alpha]$ , we obtain for any  $x$  in  $\{V_\nu \leq d\}$

$$\begin{aligned} K^\nu(x, 0) &= \binom{x+0}{x} \int_0^1 t^{x+1} (1-t)^0 \nu(dt) = \int_0^1 t^{x+1} \nu(dt) \\ &= \int_\alpha^{1-\alpha} t^{x+1} \nu(dt) \geq \int_\alpha^{1-\alpha} \alpha^{x+1} \nu(dt) \geq \alpha^{x+1}. \end{aligned}$$

To reach Inequality (4.68), it remains to prove that there exist  $d$  in  $\mathbb{R}_+^*$  and  $B_d$  in  $\mathbb{N}$  such that

$$(4.69) \quad [B_d, +\infty) \cap \mathbb{N} \subset \{V_\nu > d\}.$$

Indeed,  $[B_d, +\infty) \cap \mathbb{N} \subset \{V_\nu > d\}$  being equivalent to  $\{V_\nu \leq d\} \subset [0, B_d] \cap \mathbb{N}$ , this would imply that

$$(4.70) \quad \inf_{x \in \{V_\nu \leq d\}} (K^\nu(x, 0)) \geq \inf_{x \in [0, B_d] \cap \mathbb{N}} (K^\nu(x, 0)) \geq \inf_{x \in [0, B_d] \cap \mathbb{N}} (\alpha^{x+1}) \geq \alpha^{B_d+1},$$

and the set  $\{V_\nu \leq d\}$  would then be a  $(1, \alpha^{B_d+1})$ -Döbblin set.

Since  $\delta > 1$ , we first notice that for any  $N$  in  $\mathbb{N}^*$  and any  $x$  in  $\mathbb{N}$

$$(4.71) \quad \begin{aligned} V_\nu(x) &= \mathbb{E}_x[\delta^{\tau_C}] = \mathbb{E}_x[\delta^{\tau_C} \mathbb{1}_{\tau_C < N}] + \mathbb{E}_x[\delta^{\tau_C} \mathbb{1}_{\tau_C \geq N}] \\ &\geq 0 + \mathbb{E}_x[\delta^N \mathbb{1}_{\tau_C \geq N}] = \delta^N \mathbb{P}_x(\tau_C \geq N). \end{aligned}$$

Inequality (4.71) suggest to lower-bound the probability of the event  $\{\tau_0 \geq N\}$  under  $\mathbb{P}_x$  as a function of  $x$  in order to find  $B_d$  as in Inclusion (4.69). The event  $\{\tau_0 \geq N\}$  corresponds to the fact that the chain does not reach  $\{0\}$  before time step  $N$ . This can in particular happen when the chain  $(X_n)_{n \in \mathbb{N}}$  starting from big enough integer  $i$  doesn't decrease more than a specific constant proportion, denoted  $f(\alpha, p)$  (defined in Equation (4.73)), of the precedent position at each time step. This event is denoted  $A_N(\alpha, p)$  and defined formally in Equation (4.74). Indeed, if we choose the integer  $i$  and the proportion accordingly, after  $N - 1$  steps, the chain has not reached  $\{0\}$  yet.

First, we prove in Lemma 59 that the probability that the walk decreases more than a constant proportion  $f(\alpha, p)$  of the precedent position in one single time step, is upper-bounded by some  $p$  in  $(0, 1)$ . Then in Lemma 61, we proceed by recursion to lower-bound the probability  $\mathbb{P}_i$  of the particular event  $A_N(\alpha, p)$  included in  $\{\tau_0 \geq N\}$  for a specific value  $i_N$  of  $i$  defined in Equation (4.75).

Let us introduce some notations:

- For any  $p$  in  $(0, 1)$ ,  $i_0(\alpha, p)$  is the integer defined by

$$(4.72) \quad i_0(\alpha, p) = \left\lfloor \frac{(1 - \alpha)^3}{p\alpha^4} \right\rfloor.$$

- For any  $p$  in  $(0, 1)$ ,  $f(\alpha, p)$  is defined by

$$(4.73) \quad f(\alpha, p) = \frac{\alpha}{1 - \alpha} - \frac{1}{\sqrt{p}} \frac{\sqrt{1 - \alpha}}{\alpha} \frac{1}{\sqrt{i_0(\alpha, p) + 1}}.$$

- For any  $p$  in  $(0, 1)$  and any  $N$  in  $\mathbb{N}$ , the set  $A_N(\alpha, p)$  is defined by

$$(4.74) \quad A_N(\alpha, p) = \bigcap_{j=1}^N \{X_j > \lfloor f(\alpha, p) X_{j-1} \rfloor\} \in \mathcal{F}_N^X.$$

- For any  $p$  in  $(0, 1)$  and any  $N$  in  $\mathbb{N}^*$ , the integer  $i_N$  is defined by

$$(4.75) \quad i_N = i_N(\alpha, p) = 1 + \left\lfloor \frac{i_0(\alpha, p)}{f(\alpha, p)^N} \right\rfloor.$$

**Remark 54.** For any  $p$  in  $(0, 1)$ , since

$$1 - \alpha \geq \frac{1}{2} \quad \text{and} \quad \frac{1}{\alpha} > 2 \quad \text{and} \quad \frac{1}{p} > 1,$$

it follows

$$i_0(\alpha, p) = \left\lfloor \frac{(1 - \alpha)^3}{p\alpha^4} \right\rfloor \geq 2.$$

**Remark 55.** For any  $p$  in  $(0, 1)$  and any  $N$  in  $\mathbb{N}$ , Equation (4.74) ensures

$$(4.76) \quad A_{N+1}(\alpha, p) = A_N(\alpha, p) \cap \{X_{N+1} > \lfloor f(\alpha, p)X_N \rfloor\},$$

so that the sequence of sets  $(A_N(\alpha, p))_{N \in \mathbb{N}}$  is decreasing

$$A_0(\alpha, p) \supset A_1(\alpha, p) \supset \cdots \supset A_{N-1}(\alpha, p) \supset A_N(\alpha, p) \supset A_{N+1}(\alpha, p) \supset \cdots.$$

**Remark 56.** For any  $p$  in  $(0, 1)$  and for any  $i$  in  $\mathbb{N}$

$$\begin{aligned} & \frac{\alpha}{1-\alpha} - \frac{1}{\sqrt{p}} \frac{\sqrt{1-\alpha}}{\alpha} \frac{1}{\sqrt{i+1}} > 0 \\ \Leftrightarrow & \frac{\alpha}{1-\alpha} > \frac{1}{\sqrt{p}} \frac{\sqrt{1-\alpha}}{\alpha} \frac{1}{\sqrt{i+1}} \Leftrightarrow \frac{p\alpha^4}{(1-\alpha)^3} > \frac{1}{i+1} \\ \Leftrightarrow & \frac{p\alpha^4}{(1-\alpha)^3} > \frac{1}{i+1} \Leftrightarrow i > \frac{(1-\alpha)^3}{p\alpha^4} - 1 \Leftrightarrow i \geq \left\lfloor \frac{(1-\alpha)^3}{p\alpha^4} \right\rfloor, \end{aligned}$$

so that the value chosen for the integer  $i_0(\alpha, p)$  ensures that the function  $f$  takes indeed value in  $(0, 1)$ .

**Remark 57.** If  $N$  is a non negative integer, by definition of  $A_N(\alpha, p)$  in Equation (4.74), on the set  $A_N(\alpha, p)$ , we have  $X_1 > \lfloor f(\alpha, p)X_0 \rfloor$ .

Assume now that for some integer  $N \geq 2$  and some non negative integer  $k \leq N-1$ , the inequality  $X_k > f(\alpha, p)^k X_0$  holds on the set  $A_N(\alpha, p)$ .  $X_k$  being an integer  $X_k > \lfloor f(\alpha, p)^k X_0 \rfloor$  is equivalent to  $X_k > f(\alpha, p)X_0$  and similarly  $X_{k+1}$  being an integer  $X_{k+1} > \lfloor f(\alpha, p)X_k \rfloor$  is equivalent to  $X_{k+1} > f(\alpha, p)X_k$  and thus on the set  $A_N(\alpha, p)$

$$X_{k+1} > f(\alpha, p)X_k > f(\alpha, p)f(\alpha, p)^k X_0 > f(\alpha, p)^{k+1}X_0 \geq \lfloor f(\alpha, p)^{k+1}X_0 \rfloor.$$

By induction, for any  $p$  in  $(0, 1)$  and any non negative integer  $N$ , on the set  $A_N(\alpha, p)$ , we get for any integer  $1 \leq k \leq N$

$$X_k > \lfloor f(\alpha, p)^k X_0 \rfloor,$$

and thus

$$(4.77) \quad A_N(\alpha, p) \subset \bigcap_{k=1}^N \{X_k > \lfloor f(\alpha, p)^k X_0 \rfloor\}.$$

**Remark 58.** For any  $p$  in  $(0, 1)$  and any non negative integers  $N$  and  $k$  such that  $k \leq N$ , since  $f(\alpha, p)$  lies in  $(0, 1)$  and  $N-k$  is a positive integer,  $f(\alpha, p)^{N-k}$  lies also in  $(0, 1)$  and thus for any integer  $i \geq i_N$

$$\begin{aligned} \lfloor f(\alpha, p)^k i \rfloor & \geq \lfloor f(\alpha, p)^k i_N \rfloor \geq \left\lfloor f(\alpha, p)^k \left(1 + \left\lfloor \frac{i_0(\alpha, p)}{f(\alpha, p)^N} \right\rfloor\right) \right\rfloor \\ & \geq \left\lfloor f(\alpha, p)^k \frac{i_0(\alpha, p)}{f(\alpha, p)^N} \right\rfloor \geq \left\lfloor \frac{i_0(\alpha, p)}{f(\alpha, p)^{N-k}} \right\rfloor \geq \lfloor i_0(\alpha, p) \rfloor \\ & \geq i_0(\alpha, p). \end{aligned}$$

Thanks to inclusion (4.77) of Remark 57 and Remark 58, we derive for any  $p$  in  $(0, 1)$ , for any  $N$  in  $\mathbb{N}$ , for any  $i \geq i_N$  and for any  $1 \leq n \leq N - 1$ , the following inclusions

$$(4.78) \quad \begin{aligned} A_n(\alpha, p) \cap \{X_0 = i\} &\subset \bigcap_{k=1}^n \left\{ X_k > \lfloor f(\alpha, p)^k X_0 \rfloor \right\} \cap \{X_0 = i\} \\ &\subset \bigcap_{k=1}^n \left\{ X_k > \lfloor f(\alpha, p)^k i \rfloor \right\} \subset \bigcap_{k=1}^n \{X_k > i_0(\alpha, p)\} . \end{aligned}$$

Remark 54 ensures that  $i_0(\alpha, p) \geq 1$  so that  $\bigcap_{k=1}^{n-1} \{X_k > i_0(\alpha, p)\} \subset \{\tau_0 \geq n\}$  and then

$$(4.79) \quad A_{n-1}(\alpha, p) \cap \{X_0 = i\} \subset \{\tau_0 \geq n\} .$$

Since Equation (4.79) implies that for any integer  $i \geq i_N$

$$(4.80) \quad \mathbb{P}_i(A_{N-1}(\alpha, p)) \leq \mathbb{P}_i(\tau_0 \geq N) ,$$

it is sufficient to find a lower-bound on  $\mathbb{P}_i(A_{N-1}(\alpha, p))$  to obtain a lower-bound on  $\mathbb{P}_i(\tau_0 \geq N)$ ,

**Lemma 59.** *For any  $p$  in  $(0, 1)$  and any  $i$  in  $\mathbb{N}$  such that  $i \geq i_0(\alpha, p)$*

$$(4.81) \quad K^\nu(i, \{j : j \leq \lfloor f(\alpha, p)i_0(\alpha, p) \rfloor\}) \leq K^\nu(i, \{j : j \leq \lfloor f(\alpha, p)i \rfloor\}) \leq p .$$

*Proof.* Fix  $p$  in  $(0, 1)$  and an integer  $i \geq i_0(\alpha, p)$ .

Let's introduce a family of probability distributions  $(\mathbb{Q}_{t,j})_{(t,j) \in [0,1] \times \mathbb{N}}$  (as well as  $\mathbb{E}_{\mathbb{Q}_{t,j}}$  and  $\text{Var}_{\mathbb{Q}_{t,j}}$  the associated expectation and variance symbols) and a random variable  $Y$  whose distribution under  $\mathbb{Q}_{t,j}$  is a negative binomial with parameters  $j$  and  $t$ , i.e. for any  $k$  in  $\mathbb{N}$

$$\mathbb{Q}_{t,j}(Y = k) = \binom{j+k}{k} t^{j+1} (1-t)^k .$$

Recall that

$$(4.82) \quad \mathbb{E}_{\mathbb{Q}_{t,j}}[Y] = (j+1) \frac{1-t}{t} \quad \text{and} \quad \text{Var}_{\mathbb{Q}_{t,j}}(Y) = (j+1) \frac{1-t}{t^2} .$$

For any integer  $j$ , the functions  $t \in [\alpha, 1-\alpha] \mapsto \mathbb{E}_{\mathbb{Q}_{t,j}}[Y]$  and  $t \in [\alpha, 1-\alpha] \mapsto \text{Var}_{\mathbb{Q}_{t,j}}(Y)$  are hence both decreasing so that for any  $j$  in  $\mathbb{N}$  and any  $t$  in  $[\alpha, 1-\alpha]$

$$(4.83) \quad \mathbb{E}_{\mathbb{Q}_{\alpha,j}}[Y] \geq \mathbb{E}_{\mathbb{Q}_{t,j}}[Y] \geq \mathbb{E}_{\mathbb{Q}_{1-\alpha,j}}[Y] \quad \text{and} \quad \text{Var}_{\mathbb{Q}_{\alpha,j}}(Y) \geq \text{Var}_{\mathbb{Q}_{t,j}}(Y) \geq \text{Var}_{\mathbb{Q}_{1-\alpha,j}}(Y) .$$

Since we have the inclusions

$$\{j : j \leq \lfloor f(\alpha, p)i_0(\alpha, p) \rfloor\} \subset \{j : j \leq \lfloor f(\alpha, p)i \rfloor\} \subset \{j : j \leq \lfloor f(\alpha, p)(i+1) \rfloor\} ,$$

it holds

$$\begin{aligned} K^\nu(i, \{j : j \leq \lfloor f(\alpha, p)i_0(\alpha, p) \rfloor\}) &\leq K^\nu(i, \{j : j \leq \lfloor f(\alpha, p)i \rfloor\}) \\ &\leq K^\nu(i, \{j : j \leq \lfloor f(\alpha, p)(i+1) \rfloor\}) . \end{aligned}$$



Combining the definition of  $K^\nu$  and the fact that  $\text{supp}(\nu) \subset [\alpha, 1 - \alpha]$ , we derive

$$\begin{aligned}
K^\nu(i, \{j : j \leq \lfloor f(\alpha, p)(i+1) \rfloor\}) &= \sum_{j=0}^{\lfloor f(\alpha, p)(i+1) \rfloor} K^\nu(i, j) \\
&= \sum_{j=0}^{\lfloor f(\alpha, p)(i+1) \rfloor} \binom{i+j}{j} \int_0^1 t^{i+1}(1-t)^j \nu(dt) = \int_0^1 \left( \sum_{j=0}^{\lfloor f(\alpha, p)(i+1) \rfloor} \binom{i+j}{j} t^{i+1}(1-t)^j \right) \nu(dt) \\
&= \int_0^1 \mathbb{Q}_{t,i}(Y \leq \lfloor f(\alpha, p)(i+1) \rfloor) \nu(dt) \leq \int_0^1 \mathbb{Q}_{t,i}(Y \leq f(\alpha, p)(i+1)) \nu(dt),
\end{aligned}$$

so that it holds

$$\begin{aligned}
(4.84) \quad K^\nu(i, \{j \in \mathbb{N} : j \leq \lfloor f(\alpha, p)i_0(\alpha, p) \rfloor\}) &\leq K^\nu(i, \{j \in \mathbb{N} : j \leq \lfloor f(\alpha, p)i \rfloor\}) \\
&\leq \int_0^1 \mathbb{Q}_{t,i}(Y \leq f(\alpha, p)(i+1)) \nu(dt).
\end{aligned}$$

By the definition of  $f(\alpha, p)$  and Equation (4.82), for any  $j$  in  $\mathbb{N}$

$$\begin{aligned}
f(\alpha, p)(j+1) &= \left( \frac{\alpha}{1-\alpha} - \frac{1}{\sqrt{p}} \frac{\sqrt{1-\alpha}}{\alpha} \frac{1}{\sqrt{i_0(\alpha, p)+1}} \right) (j+1) \\
&= \frac{\alpha}{1-\alpha} (j+1) - \sqrt{\frac{(j+1) \frac{1-\alpha}{\alpha^2}}{p}} \sqrt{\frac{j+1}{i_0(\alpha, p)+1}} \\
&= \mathbb{E}_{\mathbb{Q}_{1-\alpha, j}}[Y] \leq -\sqrt{\frac{\text{Var}_{\mathbb{Q}_{\alpha, j}}(Y)}{p}} \sqrt{\frac{j+1}{i_0(\alpha, p)+1}}.
\end{aligned}$$

Combining latest inequality with Inequalities (4.83) and the fact that  $i \geq i_0(\alpha, p)$ , we thus derive the following sequence of inclusions

$$\begin{aligned}
\{Y \leq f(\alpha, p)(i+1)\} &\subset \left\{ Y - \mathbb{E}_{\mathbb{Q}_{1-\alpha, i}}[Y] \leq -\sqrt{\frac{\text{Var}_{\mathbb{Q}_{\alpha, i}}(Y)}{p}} \sqrt{\frac{i+1}{i_0(\alpha, p)+1}} \right\} \\
&\subset \left\{ Y - \mathbb{E}_{\mathbb{Q}_{t, i}}[Y] \leq -\sqrt{\frac{\text{Var}_{\mathbb{Q}_{t, i}}(Y)}{p}} \sqrt{\frac{i+1}{i_0(\alpha, p)+1}} \right\} \\
&\subset \left\{ Y - \mathbb{E}_{\mathbb{Q}_{t, i}}[Y] \leq -\sqrt{\frac{\text{Var}_{\mathbb{Q}_{t, i}}(Y)}{p}} \right\} \subset \left\{ |Y - \mathbb{E}_{\mathbb{Q}_{t, i}}[Y]| \geq \sqrt{\frac{\text{Var}_{\mathbb{Q}_{t, i}}(Y)}{p}} \right\}.
\end{aligned}$$

By Tchebychev inequality, we have, for any  $j$  in  $\mathbb{N}$  and any  $t$  in  $(0, 1)$

$$\mathbb{Q}_{t, j} \left( |Y - \mathbb{E}_{\mathbb{Q}_{t, j}}[Y]| \geq \sqrt{\frac{\text{Var}_{\mathbb{Q}_{t, j}}(Y)}{p}} \right) \leq \frac{\text{Var}_{\mathbb{Q}_{t, j}}(Y)}{\left( \sqrt{\frac{\text{Var}_{\mathbb{Q}_{t, j}}(Y)}{p}} \right)^2} \leq p.$$

Hence, previous inclusions yield for any  $t$  in  $(0, 1)$

$$(4.85) \quad \int_0^1 \mathbb{Q}_{t, i}(Y \leq f(\alpha, p)(i+1)) \nu(dt) \leq p.$$

Combining Equations (4.84) and (4.85), we finally proved that

$$K^\nu(i, \{j \in \mathbb{N} : j \leq \lfloor f(\alpha, p)i_0(\alpha, p) \rfloor\}) \leq K^\nu(i, \{j \in \mathbb{N} : j \leq \lfloor f(\alpha, p)i \rfloor\}) \leq p.$$

□

**Remark 60.** Expressions of  $i_0(\alpha, p)$  and  $f(\alpha, p)$  in Equations (4.72) and (4.73) have been designed explicitly to grant Lemma 59.

Proof of Lemma 61 uses a common property of Markov chains on discrete state space given in Lemma 85 of Appendix A.

**Lemma 61.** For any positive integer  $N$ , any integer  $1 \leq k \leq N - 1$ , any  $p$  in  $(0, 1)$  and any integer  $i \geq i_N$

$$(4.86) \quad \mathbb{P}_i(A_k(\alpha, p)) \geq (1 - p)^k,$$

where  $A_k(\alpha, p)$  is defined in Equation (4.74).

*Proof.* Fix an integer  $i \geq i_N$ . We will proceed by induction on  $k$ . Since  $\mathbb{P}_i$ -a.s.  $X_0 = i$ , we have for any integer  $k$  in  $\mathbb{N}$

$$\mathbb{P}_i(A_k(\alpha, p)) = \mathbb{P}_i(A_k(\alpha, p) \cap \{X_0 = i\}).$$

**Base case :** By Remark 58,  $\lfloor f(\alpha, p)i_N \rfloor \geq i_0(\alpha, p)$  so that  $\lfloor f(\alpha, p)i \rfloor \geq \lfloor f(\alpha, p)i_N \rfloor \geq i_0(\alpha, p)$  and Inequality (4.81) ensures

$$\begin{aligned} \mathbb{P}_i(A_k(\alpha, p)) &= \mathbb{P}_i(\{X_0 = i\} \cap \{X_1 > \lfloor f(\alpha, p)X_0 \rfloor\}) \\ &= \mathbb{P}_i(\{X_0 = i\} \cap \{X_1 > \lfloor f(\alpha, p)i \rfloor\}) \\ &= K^\nu(i, \{j \in \mathbb{N} : j > \lfloor f(\alpha, p)i \rfloor\}) \\ &= 1 - K^\nu(i, \{j \in \mathbb{N} : j \leq \lfloor f(\alpha, p)i \rfloor\}) \geq 1 - p, \end{aligned}$$

and induction is thus initialized.

**Induction step :** Let us now assume that Inequality (4.86) holds for some integer  $1 \leq k \leq N - 2$ , that is

$$\mathbb{P}_i(A_k(\alpha, p)) \geq (1 - p)^k.$$

By the Inclusion (4.78), we know that

$$A_{k+1}(\alpha, p) \cap \{X_0 = i\} \subset \bigcap_{j=1}^{k+1} \{X_j > \lfloor f(\alpha, p)^j i \rfloor\},$$

and by Equation (4.76) of Remark 55 that

$$\begin{aligned} A_{k+1}(\alpha, p) &= \{X_{k+1} > \lfloor f(\alpha, p)X_k \rfloor\} \cap A_k(\alpha, p) \\ &= \bigsqcup_{j_k \in \mathbb{N}} \{X_{k+1} > \lfloor f(\alpha, p)j_k \rfloor\} \cap \{X_k = j_k\} \cap A_k(\alpha, p), \end{aligned}$$

so that following identity ensues

$$\begin{aligned}
& A_{k+1}(\alpha, p) \cap \{X_0 = i\} \\
&= A_{k+1}(\alpha, p) \cap \{X_0 = i\} \cap \{X_k > \lfloor f(\alpha, p)^k i \rfloor\} \\
&= \bigsqcup_{j_k > \lfloor f(\alpha, p)^k i \rfloor} \{X_{k+1} > \lfloor f(\alpha, p) j_k \rfloor\} \cap \{X_k = j_k\} \cap A_k(\alpha, p) \cap \{X_0 = i\} .
\end{aligned}$$

Hence,

$$\begin{aligned}
(4.87) \quad & \mathbb{P}_i(A_{k+1}(\alpha, p) \cap \{X_0 = i\}) \\
&= \mathbb{P}_i \left( \bigsqcup_{j_k > \lfloor f(\alpha, p)^k i \rfloor} \{X_{k+1} > \lfloor f(\alpha, p) j_k \rfloor\} \cap \{X_k = j_k\} \cap A_k(\alpha, p) \cap \{X_0 = i\} \right) \\
&= \sum_{j_k > \lfloor f(\alpha, p)^k i \rfloor} \mathbb{P}_i(\{X_{k+1} > \lfloor f(\alpha, p) j_k \rfloor\} \cap \{X_k = j_k\} \cap A_k(\alpha, p) \cap \{X_0 = i\}) .
\end{aligned}$$

Equation (4.74) ensures that  $\{X_k = j_k\} \cap A_k(\alpha, p) \cap \{X_0 = i\}$  belongs to the  $\sigma$ -field  $\mathcal{F}_k^X$  so that for any  $j_k > \lfloor f(\alpha, p)^k i \rfloor$  such that  $\mathbb{P}_i(\{X_k = j_k\} \cap A_k(\alpha, p) \cap \{X_0 = i\}) \neq 0$ , Remark 86 yields

$$\begin{aligned}
& \mathbb{P}_i(\{X_{k+1} > \lfloor f(\alpha, p) j_k \rfloor\} \mid \{X_k = j_k\} \cap A_k(\alpha, p) \cap \{X_0 = i\}) \\
&= K^\nu(j_k, \{j \in \mathbb{N} : j > \lfloor f(\alpha, p) j_k \rfloor\}) ,
\end{aligned}$$

and thus

$$\begin{aligned}
(4.88) \quad & \mathbb{P}_i(\{X_{k+1} > \lfloor f(\alpha, p) j_k \rfloor\} \cap \{X_k = j_k\} \cap A_k(\alpha, p) \cap \{X_0 = i\}) \\
&= K^\nu(j_k, \{j \in \mathbb{N} : j > \lfloor f(\alpha, p) j_k \rfloor\}) \mathbb{P}_i(\{X_k = j_k\} \cap A_k(\alpha, p) \cap \{X_0 = i\}) .
\end{aligned}$$

Since  $k \leq k+1 \leq N$ , by Remark 58 we have  $\lfloor f(\alpha, p)^k i \rfloor \geq i_0(\alpha, p)$ , and Inequality (4.81) thus yields that for any integer  $j_k > \lfloor f(\alpha, p)^k i \rfloor$

$$\begin{aligned}
(4.89) \quad & K^\nu(j_k, \{j \in \mathbb{N} : j > \lfloor f(\alpha, p) j_k \rfloor\}) = 1 - K^\nu(j_k, \{j \in \mathbb{N} : j \leq \lfloor f(\alpha, p) j_k \rfloor\}) \geq 1 - p .
\end{aligned}$$

Combination of Inequalities (4.88) and (4.89) plugged in Equation (4.87) yields

$$\begin{aligned}
\mathbb{P}_i(A_{k+1}(\alpha, p)) &\geq \sum_{j_k \geq \lfloor f(\alpha, p)^k i \rfloor} (1 - p) \mathbb{P}_i(\{X_k = j_k\} \cap A_k(\alpha, p) \cap \{X_0 = i\}) \\
&= (1 - p) \mathbb{P}_i(A_k(\alpha, p) \cap \{X_0 = i\}) = (1 - p) \mathbb{P}_i(A_k(\alpha, p)) ,
\end{aligned}$$

what leads, using induction hypothesis, to

$$\mathbb{P}_i(A_{k+1}(\alpha, p)) \geq (1 - p)(1 - p)^k \geq (1 - p)^{k+1} .$$

Hence, Inequality (4.86) holds for  $k+1$ .

Finally, we indeed prove by induction that Inequality (4.86) holds for  $1 \leq k \leq N-1$ .  $\square$

Lemma 61 in particular holds for  $k = N - 1$  and combined with Inequality (4.80), this leads for any integer  $i \geq i_N$  to

$$\mathbb{P}_i(\tau_0 \geq N) \geq \mathbb{P}_i(A_{N-1}(\alpha, p)) \geq (1 - p)^{N-1}.$$

Plugging this in Inequality (4.71), we get for all  $N$  in  $\mathbb{N}^*$ , any  $p$  in  $(0, 1)$  and any integer  $i \geq i_N = 1 + \lfloor i_0(\alpha, p)/f(\alpha, p)^N \rfloor$

$$V_\nu(i) \geq \delta^N \mathbb{P}_i(\tau_0 \geq N) \geq \delta^N (1 - p)^{N-1} = \delta^N (1 - p)^{N-1}.$$

From Inequality (4.70), we deduce that for all  $N$  in  $\mathbb{N}^*$

$$\inf_{x \in \{V_\nu \leq \delta^N (1-p)^{N-1}\}} (K^\nu(x, 0)) \geq \alpha^{2 + \lfloor \frac{i_0(\alpha, p)}{f(\alpha, p)^N} \rfloor}.$$

As this holds for any  $p$  in  $(0, 1)$  and for any  $N$  in  $\mathbb{N}^*$ , one can choose  $p$  so that  $\delta(1 - p) > 1$  and then  $N$  so that  $\delta^N (1 - p)^{N-1} > -1 + 2M/(1 - \delta^{-1})$ : we set thus

$$(4.90) \quad p = p(c_0, \alpha) = \frac{\delta(c_0, \alpha) - 1}{2\delta(c_0, \alpha)},$$

and

$$(4.91) \quad N = N(c_0, \alpha) = 1 + \left\lfloor \frac{\log \left( \frac{\delta(c_0, \alpha) + 1}{2\delta(c_0, \alpha)} \left( \frac{(2M(c_0, \alpha) + 1)\delta(c_0, \alpha) - 1}{\delta(c_0, \alpha) - 1} \right) \right)}{\log \left( \frac{\delta(c_0, \alpha) + 1}{2} \right)} \right\rfloor.$$

Choosing

$$(4.92) \quad d = d(c_0, \alpha) = \delta(c_0, \alpha)^{N(c_0, \alpha)} (1 - p(c_0, \alpha))^{N(c_0, \alpha) - 1},$$

and

$$(4.93) \quad \varepsilon = \varepsilon(c_0, \alpha) = \alpha^{2 + \left\lfloor \frac{i_0(\alpha, p(c_0, \alpha))}{f(\alpha, p(c_0, \alpha))^{N(c_0, \alpha)}} \right\rfloor},$$

we finally proved that  $\{V_\nu \leq d\}$  is an  $(1, \varepsilon)$ -Döblin set for  $K^\nu$ . Theorem 50-Item 2 is thus satisfied.

Application of Theorem 50 and control of the  $d_{TV}$ -distance by the  $d_{V_\nu}$ -distance yields

$$d_{V_\nu}(\delta_0(K^\nu)^n, \pi_\nu) \leq d_{TV}(\delta_0(K^\nu)^n, \pi_\nu) \leq c_1 (\pi_\nu(V_\nu) + \delta_0(V_\nu)) \rho^{\lfloor \frac{n}{1} \rfloor} \leq c_1 (\pi_\nu(V_\nu) + V_\nu(0)) \rho^n,$$

where

$$\log(\rho) = \frac{\log(1 - \varepsilon) \log \left( \frac{1}{\delta} + \frac{2M}{1+d} \right)}{\log(1 - \varepsilon) + \log \left( \frac{1}{\delta} + \frac{2M}{1+d} \right) - \log \left( \frac{M}{\delta} + d \right)},$$

and

$$c_1 = \frac{1 + \delta}{\delta + \frac{M + d\delta}{(1 - \varepsilon)(1 - \frac{M}{\delta} - d)}}.$$

By Remark 51 and Inequality (4.67), we have

$$\pi_\nu(V_\nu) \leq \frac{M}{1 - \delta^{-1}} = \frac{M\delta}{\delta - 1},$$

and by Equation (4.63) in Remark 53, we have

$$\sup_{x \in C}(\delta_x(V_\nu)) = \sup_{x \in \{0\}}(\delta_x(V_\nu)) = \delta_0(V_\nu) = V_\nu(0) = 1.$$

Hence, for any  $\nu$  in  $\text{supp}(\pi)$  and any  $n$  in  $\mathbb{N}$

$$(4.94) \quad d_{\text{TV}}(\delta_0 P_\nu^n, \pi_\nu) \leq \zeta_0(c_0, \alpha) r_0(c_0, \alpha)^n,$$

where

$$(4.95) \quad \log(r_0(c_0, \alpha)) = \frac{\log(1 - \varepsilon) \log\left(\frac{1}{\delta} + \frac{2M}{1+d}\right)}{\log(1 - \varepsilon) + \log\left(\frac{1}{\delta} + \frac{2M}{1+d}\right) - \log\left(\frac{M}{\delta} + d\right)} < 0,$$

and

$$(4.96) \quad \zeta_0(c_0, \alpha) = \frac{1 + \delta}{\delta + \frac{M+d\delta}{(1-\varepsilon)\left(1-\frac{M}{\delta}-d\right)}},$$

with  $\delta = \delta(c_0, \alpha)$ ,  $M = M(c_0, \alpha)$ ,  $d = d(c_0, \alpha)$  and  $\varepsilon = \varepsilon(c_0, \alpha)$  respectively defined in Equations (4.64), (4.65), (4.92) and (4.93).

Assumption **M3** is satisfied and proof of Theorem 40 is over.

## Chapter 5

# Posterior consistency of Bayes estimator of the environment

The estimation of the environment distribution from the observation of a single trajectory of a random walk in random environment on  $\mathbb{Z}$  was first investigated in a parametric framework in [CFLL16, CFL<sup>+</sup>14, FLM14, FGL14]. In these works, the environment is i.i.d. with common distribution  $\nu^*$ . The distribution  $\nu^*$  belongs to a family of probabilities  $(\nu_\theta)_{\theta \in \Theta}$  on  $[0, 1]$ , where  $\Theta \subset \mathbb{R}^d$ . They build and study different  $M$ -estimators of the distribution parameter  $\theta^*$  from a trajectory of the walk stopped when it reaches for the first time a fixed  $n$  in  $\mathbb{N}$ . They prove consistency of this estimator in all regimes and its asymptotic normality in the transient regimes when  $n$  tends to infinity.

A non-parametric approach was then developed in [DL18]. This work considered the estimation of the cumulative distribution function (*c.d.f.*) of the environment distribution  $\nu^*$ . They built an estimator of the moments of the true distribution and, using a deterministic approximation of the true c.d.f. by these moments, they deduce an estimator of the true c.d.f.. They prove concentration properties, consistency and asymptotic normality of the estimator of the moments. The final estimator of the c.d.f. does not depend on the regime of the walk and it is adaptive to the regularity of the c.d.f.. Using similar ideas, we built in Chapter 2 estimators of the density of the environment. All these works adopt a frequentist point of view. In this chapter, we develop the first Bayesian estimator for this problem.

Given a statistical model  $(P_\theta)_{\theta \in \Theta}$  with observations  $X$  (and  $\Theta$  some given set), the Bayesian paradigm assigns a probability distribution to the couple  $(\theta, X)$ . To this end, it considers that the parameter  $\theta$  is a random variable distributed according to a *prior* (or *a priori*) probability  $\Pi$  on  $\Theta$  and looks at the distribution of the observation under  $P_\theta$  as the conditional distribution of the observation  $X$  given  $\theta$ . The *Bayesian estimator* of  $\theta$  is the conditional distribution of  $\theta$  given  $X$ . It is called *posterior* or *a posteriori distribution*. In words, the Bayesian strategy proceeds as follows: first, we consider the parameter as random with distribution  $\Pi$  and then, we update this distribution using the observations. In this approach, a crucial point is the choice of the prior. This raises an important question: how does this prior influence the long-term behaviour of the posterior? We address this question by making the (frequentist) assumption that the observations were sampled under  $P_{\theta^*}$  for a specific  $\theta^*$  in  $\Theta$ . Then we wonder if the posterior distribution concentrates around  $\theta^*$  when the number of observations tends to infinity. The Bayes estimator is said to be consistent at true parameter  $\theta^*$  if the posterior assigns to any neighbourhood of  $\theta^*$  a mass converging to 1 as  $n$  grows to infinity.

This approach has a long history that dates back to the pioneering works of Doob [Doo49], Freedman [Fre63] or Diaconis [DF86]. In i.i.d. frameworks, [Doo49] showed that if observations take values in an Euclidean space, if the model is identifiable and if the parameter space is a complete separable metric space equipped with prior  $\Pi$ , then the sequence of posterior distributions is consistent in  $\Pi$ -almost all  $\theta^*$ . Posterior consistency in parametric framework holds therefore under weak assumptions. More precise results can also be proven such as Bernstein-von Mises theorem giving conditions for convergence of posterior distributions toward normal distribution. On the other hand, in non-parametric settings, [Fre63, Fre65] pointed out that consistency is not automatic even if the true parameter belongs to the support of the prior and [DF86] even showed that this situation is standard.

When the model is i.i.d. and dominated (i.e. there exists a probability  $\mu$  on  $\mathcal{X}$  such that, for all  $\theta$  in  $\Theta$ ,  $\mathbf{P}^\theta$  is absolutely continuous with respect to  $\mu$ ), a theory of posterior consistency was nevertheless developed in [Sch65, Bar88, BSW99, GGR99]. When  $\Theta$  is equipped with a distance  $d$ , those results establish under some conditions that, for  $\varepsilon > 0$ ,

$$\Pi(\theta \in \Theta : d(\theta^*, \theta) \geq \varepsilon | X_{0:n}) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^{\theta^*}} 0.$$

Hence, the sequence of posterior distributions is weakly consistent in  $\theta^*$ . We refer to [GR03, GvdV17] for recent overviews.

In the last few years, [Ver15] and [DOR19] among others started to investigate the problem of posterior consistency in non-i.i.d. settings by considering respectively hidden Markov models and partially observed Markov models. This chapter also analyses Bayesian estimators when the observations are dependent. As in [HLM19, DL18, CFL16, CFL<sup>+</sup>14, FLM14, FGL14], the observations are actually the successive positions of a random walk in i.i.d. random environment with distribution  $\nu^*$  on  $\mathbb{Z}$  (abbreviated *RWRE*) until it reaches  $n$  in  $\mathbb{N}^*$  for the first time. The main result of this chapter (Theorem 63) shows posterior consistency of the Bayesian estimator of the environment under technical restrictions on the prior distribution. This result is the consequence of a more general property showing posterior consistency of the Bayesian estimator of the environment of specific branching processes with immigration in random environment (abbreviated *BPIRE*). The remaining of the chapter is structured as follows.

Section 5.1 presents the mathematical framework: we define random walks in i.i.d. random environment on  $\mathbb{Z}$  and introduce their left jump processes, we recall that these left jump processes are branching processes with immigration in random environment and we introduce the notion of posterior consistency useful for the statement of our main result.

In Section 5.2, after defining some assumptions on potential prior, we set a Bayesian framework as well for RWRE as for BPIRE and we state our main results: posterior consistency of the Bayesian estimator of the environment distribution for RWRE as well as for BPIRE.

Section 5.3 gathers the proof of our consistency results.

## 5.1 RWRE, BPIRE and the Bayesian setting

We study Bayesian estimators of the environment distribution. We define the Bayesian framework after remembering briefly the definitions of a random walk in random environment and of a branching process in random environment as well as the link between them.

### 5.1.1 RWRE

Denote by  $E = [0, 1]^{\mathbb{Z}}$  the *set of environments* endowed with  $\mathcal{E} = \mathcal{B}([0, 1]^{\mathbb{Z}})$ , the  $\sigma$ -field generated by the cylinders. Denote by  $X = (X_t)_{t \in \mathbb{N}}$  the canonical process of the space  $\mathbb{Z}^{\mathbb{N}}$  endowed with the  $\sigma$ -field  $\mathcal{S}$  generated by the cylinders, i.e. for any  $t$  in  $\mathbb{N}$  and any  $x = (x_t)_{t \in \mathbb{Z}}$

$$(5.1) \quad X_t(x) = x_t.$$

For any environment  $\omega$  in  $E$  and any  $z$  in  $\mathbb{Z}$ , there exists a unique probability measure  $\mathbf{Q}_{z, \omega}$  on  $\mathcal{S}$  under which  $(X_t)_{t \in \mathbb{N}}$  is a time-homogeneous Markov chain starting from atom  $z$  with transition probabilities given, for all  $x$  and  $y$  in  $\mathbb{Z}$ , by

$$p_{\omega}(x, y) = \begin{cases} \omega_x & \text{if } y = x + 1 \\ 1 - \omega_x & \text{if } y = x - 1 \\ 0 & \text{otherwise} \end{cases}.$$

Consider a fixed probability  $\nu$  on  $[0, 1]$ . The probability distribution  $\mathbf{Q}_z^{\nu}$  on  $\mathcal{S}$  is then defined as the expectation of  $\mathbf{Q}_{z, \omega}$  with respect to the distribution  $\mathbb{P}^{\nu} = \nu^{\otimes \mathbb{Z}}$ : for any  $B$  in  $\mathcal{S}$ ,

$$(5.2) \quad \mathbf{Q}_z^{\nu}(B) = \int_E \mathbf{Q}_{z, \omega}(B) \mathbb{P}^{\nu}(d\omega).$$

For  $n$  in  $\mathbb{N}$ , we consider the function  $T_n : \mathbb{Z}^{\mathbb{N}} \rightarrow \mathbb{N} \cup \{+\infty\}$  defined for any  $x = (x_t)_{t \in \mathbb{Z}}$  by

$$T_n(x) = \inf\{t \in \mathbb{N} : x_t = n\},$$

and we also note  $T_n = T_n(X)$  where  $X$  is defined in Equation (5.1). When the environment distribution  $\nu$  is such that  $\int_0^1 \log(\frac{1-t}{t}) \nu(dt) < 0$ , [Sol75] proves that the RWRE is  $\mathbf{Q}_z^{\nu}$ -almost surely transient to the right (whatever  $z$  in  $\mathbb{Z}$ ) that is

$$(5.3) \quad \mathbf{Q}_z^{\nu}\text{-a.s.} \quad X_n \xrightarrow[n \rightarrow +\infty]{} +\infty,$$

so that any integer  $n$  is reached  $\mathbf{Q}_0^{\nu}$ -a.s. by the RWRE :  $T_n$  is well-defined and finite  $\mathbf{Q}_0^{\nu}$ -a.s.

For any  $k$  in  $\mathbb{Z}$ , any  $n$  in  $\mathbb{N}$  and for any sequence  $x = (x_t)_{t \in \mathbb{N}}$  in  $\mathbb{Z}^{\mathbb{N}}$  such that  $T_n(x) < +\infty$ , we define the number  $L(k, x, n)$  of left jumps from  $k$  of the sequence  $x$  before reaching  $n$  by

$$(5.4) \quad L(k, x, n) = \sum_{t=0}^{T_n(x)-1} \mathbb{1}_{\{x_t=k, x_{t+1}=k-1\}},$$

and derive the following fundamental equation, for any  $x = (x_t)_{t \in \mathbb{N}}$  in  $\{x \in \mathcal{S}_0 : T_n(x) < +\infty\}$

$$(5.5) \quad \begin{aligned} & \mathbf{Q}_0^{\nu}(X_{0:T_n} = x_{0:T_n}(x)) \\ &= \int_E \prod_{t=0}^{T_n(x)-1} (\omega_{x_t} \mathbb{1}_{x_{t+1}=x_t+1} + (1 - \omega_{x_t}) \mathbb{1}_{x_{t+1}=x_t-1}) \nu^{\otimes \mathbb{Z}}(d\omega) \\ &= \prod_{k=-\infty}^{-1} \int_0^1 t^{L(k+1, x, n)} (1-t)^{L(k, x, n)} \nu(dt) \times \prod_{k=0}^{n-1} \int_0^1 t^{L(k+1, x, n)+1} (1-t)^{L(k, x, n)} \nu(dt). \end{aligned}$$



Defining  $L_k^n$ , for any  $n$  in  $\mathbb{N}$  and any  $k$  in  $\mathbb{N}$ , on the set  $\{x \in \mathcal{S}_0 : T_n(x) < +\infty\}$  by

$$(5.6) \quad L_k^n = L(k, X, n) = \sum_{t=0}^{T_n-1} \mathbb{1}_{\{X_t=k, X_{t+1}=k-1\}},$$

Equation (5.5) imply that the family  $(L_k^n)_{k \leq n}$  is a sufficient statistic.

As first noted by [KKS75] and then successfully used by [CFL16, CFL<sup>+</sup>14, FLM14, FGL14], [KKS75], under  $\mathbf{Q}_0^\nu$ , the process

$$(Z_y^n)_{0 \leq y \leq n} = (L_{n-y}^n)_{0 \leq y \leq n},$$

has the same distribution as the first  $n+1$  generations of a *branching process with immigration in random environment starting from 0 with offspring distributions geometric distributions whose parameters are i.i.d. with common law  $\nu$* .

This link is at the basis of our main result. For this reason, following section is dedicated to the definition of branching process with immigration in random environment (denoted hereafter *BPIRE*) and to some properties of BPIRE.

### 5.1.2 BPIRE

**Definition 62.** Consider two probability measures  $\nu$  and  $\mu$  respectively on  $[0, 1]$  and  $\mathbb{N}$ . Consider a sequence of i.i.d. random variables  $\omega = (\omega_k)_{k \in \mathbb{Z}}$  in  $E = [0, 1]^{\mathbb{Z}}$  defined on  $(\Omega, \mathcal{T}, \mathbb{P})$  and with common probability distribution  $\nu$  in  $\mathbb{M}^1([0, 1])$ , so that  $\omega$  admits distribution  $\mathbb{P}^\nu = \nu^{\otimes \mathbb{Z}}$ . A branching process in random environment with immigration, starting from  $\mu$  in  $\mathbb{M}^1(\mathbb{N})$  and with offspring distributions geometric distributions whose parameters are i.i.d. with common law  $\nu$  (abbreviated *BPIREG*( $\nu, \mu$ ) hereafter) is a sequence of random variables  $(Z_k)_{k \in \mathbb{N}}$  such that  $Z_0$  has distribution  $\mu$  and is independent of  $\omega$  and for any  $k$  in  $\mathbb{N}$

$$Z_{k+1} = \sum_{i=0}^{Z_k} \xi_{k,i},$$

where the  $(\xi_{k,i})_{(k,i) \in \mathbb{N}^2}$  are random variables independent conditionally to the  $\sigma$ -field generated by  $\omega$  and geometrically distributed such that for all  $k, i$  and  $j$  in  $\mathbb{N}$

$$\mathbb{P}(\xi_{k,i} = j | \omega) = \omega_k (1 - \omega_k)^j.$$

The distribution of  $(Z_k)_{k \in \mathbb{N}}$  under  $\mathbb{P}$  defines a probability on  $\mathbb{M}^1(\mathbb{N}^{\mathbb{N}})$  denoted  $\mathbf{P}_\mu^\nu$ .

When addressing the problem of direct estimation of the distribution of a RWRE from a single realization of the trajectory  $X_{0:T_n}$  and because of the link between between RWRE and *BPIREG*( $\nu, 0$ ), [CFL<sup>+</sup>14] undertakes the study of *BPIREG*( $\nu, \mu$ ). We cite hereafter some of its results.

First, [CFL<sup>+</sup>14, Proposition 4.3] establishes that, for any random environment distribution  $\nu$  in  $\mathbb{M}([0, 1])$  and any starting probability distribution  $\mu$  in  $\mathbb{M}^1(\mathbb{N})$ , *BPIREG*( $\nu, \mu$ ) is an homogeneous Markov chain with initial distribution  $\mu$  and transition kernel  $K^\nu$  defined, for any integers  $i$  and  $j$ , by

$$(5.7) \quad K^\nu(i, j) = \binom{i+j}{i} \int_0^1 t^{i+1} (1-t)^j \nu(dt).$$

In the remaining of this Chapter,  $\text{BPIREG}(\nu, \mu)$  is considered under the scope of Markov chain theory. To this end, let  $(Z_x)_{x \in \mathbb{N}}$  denote the canonical process of the space  $\mathbb{N}^{\mathbb{N}}$  endowed with the  $\sigma$ -field generated by the cylinders and, for every  $\nu$  in  $\mathbb{M}^1([0, 1])$  and  $\mu$  in  $\mathbb{M}^1(\mathbb{N})$ , let  $\mathbf{P}_\mu^\nu$  denote the unique probability distribution on  $\mathbb{N}^{\mathbb{N}}$  under which  $(Z_x)_{x \in \mathbb{N}}$  is a Markov chain with initial distribution  $\mu$  and transition kernel  $K^\nu$  given in Equation (5.7). In this framework, if  $l_n$  is defined, for  $n$  in  $\mathbb{N}$ ,  $z_{0:n}$  in  $\mathbb{N}^{n+1}$  and  $\nu$  in  $\mathbb{M}^1([0, 1])$ , by

$$(5.8) \quad l_n(\nu, z_{0:n}) = \log(\delta_0(z_0)) + \log \left( \prod_{i=0}^{n-1} K^\nu(z_i, z_{i+1}) \right),$$

the link between left jump process of a RWRE and  $\text{BPIREG}(\nu, 0)$  means that for  $z_{0:n}$  in  $\mathbb{N}^{n+1}$

$$\log \mathbf{P}_0^\nu(Z_{0:n} = z_{0:n}) = l_n(\nu, z_{0:n}) = \log \mathbf{Q}_0^\nu((L_{n-k}^n)_{0 \leq k \leq n} = z_{0:n}).$$

Secondly, if the random environment distribution  $\nu$  is such that  $\int_0^1 \log(\frac{1-t}{t})\nu(dt) < 0$ , [CFL<sup>+</sup>14, Proposition 4.5] states that the Markov chain  $(Z_x)_{x \in \mathbb{N}}$  is positive recurrent and admits a unique invariant probability measure  $\mu^\nu$ . Using then ergodic theorem for Markov chains, [CFL<sup>+</sup>14, Proposition 4.6] derives that, if the random environment distribution  $\nu$  is such that  $\int_0^1 \log(\frac{1-t}{t})\nu(dt) < 0$ , then for any  $\nu'$  in  $\mathbb{M}^1([0, 1])$

$$(5.9) \quad \frac{1}{n} \sum_{i=0}^{n-1} \log \left( \frac{K^\nu(Z_i, Z_{i+1})}{\binom{Z_i + Z_{i+1}}{Z_i}} \right) = \frac{1}{n} \sum_{i=0}^{n-1} \log \left( \int_0^1 t^{Z_i+1} (1-t)^{Z_{i+1}} \nu'(dt) \right) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}_0^\nu} m^\nu(\nu'),$$

where for any  $\nu$  and  $\nu'$  in  $\mathbb{M}^1([0, 1])$

$$(5.10) \quad m^\nu(\nu') = \mathbf{E}_{\mu^\nu}^\nu \left[ \log \left( \int_0^1 t^{Z_0+1} (1-t)^{Z_1} \nu'(dt) \right) \right] \in [-\infty, 0].$$

### 5.1.3 Towards posterior consistency

If  $Z_{0:n}$  and  $\Pi$  denote respectively the sequence of observations and the prior and if  $d$  is some fixed distance on (metrizable) parameter space  $\Theta$ , posterior consistency in  $\nu^*$  usually means that for any  $\varepsilon > 0$

$$(5.11) \quad \Pi(d(\nu^*, \nu) \geq \varepsilon | Z_{0:n}) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^{\nu^*}} 0.$$

The main results of this Chapter concern posterior consistency in Bayesian framework but in the following sense : as in [LLV18], instead of a distance  $d$  to define vicinity of  $\nu^*$ , we choose a specific function  $d^{\nu^*}$  (depending on  $\nu^*$ ) defined on  $\Theta^2$  as a limit such that  $d^{\nu^*}(\nu^*, \nu) > 0$ . Let us explicit somehow the construction of  $d^{\nu^*}$  and understand the motivation behind this choice.

The convergence result (5.9) of [CFL<sup>+</sup>14, Proposition 4.5] ensures that, if the distribution  $\nu$  is such that  $\int_0^1 \log(\frac{1-t}{t})\nu(dt) < 0$ , then, for any  $\nu'$  and  $\nu''$  in  $\mathbb{M}^1([0, 1])$

$$\frac{1}{n} (l_n(\nu', Z_{0:n}) - l_n(\nu'', Z_{0:n})) = \frac{1}{n} \sum_{i=0}^{n-1} \log \left( \frac{K^{\nu'}(Z_i, Z_{i+1})}{K^{\nu''}(Z_i, Z_{i+1})} \right) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}_0^\nu} m^\nu(\nu') - m^\nu(\nu'').$$

Define for any  $\nu, \nu'$  and  $\nu''$  in  $\Theta$

$$(5.12) \quad d^\nu(\nu', \nu'') = m^\nu(\nu') - m^\nu(\nu'') = \mathbf{E}_{\mu^\nu}^\nu \left[ \log \left( \frac{K^{\nu'}(Z_0, Z_1)}{K^{\nu''}(Z_0, Z_1)} \right) \right].$$

Proposition 75 grants that  $d^{\nu^*}(\nu^*, \nu) \geq 0$  for any  $\nu$  in  $\mathbb{M}^1([0, 1])$  and we have moreover  $d^{\nu^*}(\nu^*, \nu^*) = 0$ . The function  $d^{\nu^*}$  on  $\Theta^2$  thus defines a way to evaluate how far environment distributions are from the true one in a manner suited to BPIREG.

## 5.2 Bayesian setting and main results

In order to state our main results, we need to specify the Bayesian framework : the a priori distribution  $\Pi$  (i.e. a probability distribution on the set of parameters) and the statistical model (i.e. a collection of probability distribution parametrized by the elements of the set of parameters).

### 5.2.1 Assumptions on the prior

The parameter of interest is of course the environment distribution  $\nu$ . We assume that  $\nu$  belongs to the set  $\Theta$  of probability measures on  $[0, 1]$  that are absolutely continuous with respect to Lebesgue measure on  $[0, 1]$  and for any  $\nu$  in  $\Theta$ , we denote  $f_\nu = \frac{d\nu}{dt}$  the density function of  $\nu$  with respect to Lebesgue measure on  $[0, 1]$ .

We fix some prior distribution  $\Pi$  on  $\Theta$  and some probability distribution  $\nu^*$  in  $\Theta$ . In our work, we ensure posterior consistency of the Bayesian estimator under several restrictions on this prior distribution and on this parameter. Denoting by  $\text{supp}(\Pi)$  the support of  $\Pi$  for the remaining of this chapter, we define the following assumptions:

- B1.** *There exists  $c_0$  in  $\mathbb{R}_+^*$  such that for any  $\nu$  in  $\text{supp}(\Pi)$ ,  $\int_0^1 \log(\frac{1-t}{t})\nu(dt) < -c_0$ .*
- B2.** *There exists  $\alpha$  in  $(0, 1/2)$  such that for any  $\nu$  in  $\text{supp}(\Pi)$ ,  $\text{supp}(\nu) \subset [\alpha, 1 - \alpha]$ .*
- B3.** *There exists  $m$  in  $\mathbb{R}_+^*$  such that for any  $\nu$  in  $\text{supp}(\Pi)$  and any  $t$  in  $[\alpha, 1 - \alpha]$ ,  $f_\nu(t) \geq m$ .*
- B4.** *There exists  $M$  in  $\mathbb{R}_+^*$  such that for any  $\nu$  in  $\text{supp}(\Pi)$  and any  $t$  in  $[\alpha, 1 - \alpha]$ ,  $f_\nu(t) \leq M$ .*
- B5.** *The prior  $\Pi$  assigns a positive mass to every  $\|\cdot\|_\infty$ -neighbourhood of  $\nu^*$ .*
- B6.** *For any  $\varepsilon > 0$ , there exist a family  $(\Theta_n)_{n \in \mathbb{N}}$  of measurable subsets of  $\text{supp}(\Pi)$  and a positive constant  $C_\varepsilon$  such that for any integer  $n$*

$$(5.13) \quad \Pi(\Theta_n^c) \leq e^{-\frac{C_\varepsilon}{m}n} \quad \text{and} \quad \log(2N(\Theta_n, \|\cdot\|_\infty, \varepsilon)) \leq \frac{\varepsilon^2}{(M-m)^2}n,$$

where  $m$  and  $M$  are given in assumptions **B3** and **B4** and  $N(\Theta_n, \|\cdot\|_\infty, \varepsilon)$  is the  $\varepsilon$ -covering number of  $\Theta_n$  for  $\|\cdot\|_\infty$  as defined in Appendix B.

Assumption **B1** ensures by [Sol75] that the RWRE is  $\mathbf{P}_0^\nu$ -almost surely transient to the right (that is, tends towards infinity when  $n$  grows to infinity). In particular, hitting time of site  $n$  in  $\mathbb{N}$  by the walk is  $\mathbf{P}_0^\nu$ -almost surely defined and thanks to the parallel made by [KKS75] between RWRE and BPIRE, this allows to move the problem of estimation of the environment from RWRE to BPIRE. Nevertheless, all this would have been also granted if the RWRE was recurrent. In fact, we restrict here to transient regime and even to very specific transient regime because we were able to obtain a uniform control of test errors only under this assumption. Indeed, even if a control of the tail distribution of  $T_n$  was obtained by [Sol75] under the regime assumption of left-transience, the more specific assumption **B1** enabled us

in Chapter 3 to make the control of [Sol75] uniform with respect to the problem parameters  $\alpha$  and  $c_0$  and this uniformity lead in Chapter 4 to an uniform exponential concentration inequality for transformations of  $(Z_{0:n})$  with bounded differences property.

Assumption **B2** called *ellipticity condition* is considered in the probabilistic literature as a sufficient condition for most of the existing results on the RWRE. This assumption prevents in some sense that the walk goes systematically forward or systematically backward by ensuring that the  $\omega_k$  can not be close to 0 or 1.

It is relevant to stress that Assumptions **B1** and **B2** in our setting are only Bayesian version of Assumptions **F1** and **F2** of Chapter 3 : under Assumption **B**, Assumption **F** holds uniformly for any distribution  $\nu$  lying in the full-mass set  $\text{supp}(\Pi)$ . Hence, under some or all of the assumptions **B**, results valid in Chapter 3 under the corresponding assumptions **F** extend uniformly to any  $\nu$  in  $\text{supp}(\Pi)$ .

Assumptions **B3** and **B4** are restrictive assumptions : combined to **B2**, they imply that the densities have a graph included in the rectangle whose frontier is given by the 2 vertical lines  $x = \alpha$  and  $x = 1 - \alpha$  and the 2 horizontal lines  $y = m$  and  $y = M$ . Assumption **B3** is sufficient to obtain a control on the denominator of the posterior probability (while a weaker assumption is usually needed for that control that holds exclusively for  $\nu^*$  instead of every  $\nu$  in full mass set) and to obtain a control of the test errors (uniformity of the lower-bound of densities in  $\text{supp}(\Pi)$  is indeed crucial to grant the uniformity of the concentration inequality of Chapter 4). Assumption **B4** enables here to get an upper-bound on the diameter of the sets  $\Theta_n$  introduced in Assumption **B6**.

Kullback-Leibler type condition such as **B5** is introduced in [Sch65] and is common in the study of the posterior consistency in non-parametric framework : we assume that the prior assigns positive mass to every neighbourhood of the parameter  $\nu^*$ .

Assumption **B6** ensures the existence of sets of overwhelming prior mass whose topological complexity is not too big with respect to  $\varepsilon$  and  $n$ . It is also common in Bayesian non-parametric results to ensure posterior consistency (see [GGR99, Bar88]) or to obtain precise contraction rate (see [GvdV07a, GvdV07b]).

As in previous works, our main result for RWRE is a consequence of the fundamental parallel between the left jumps of the RWRE from positive sites and the BPIRE. For this reason, the full trajectory  $X_{0:T_n}$  of the RWRE until it reaches  $n$  for the first time is not our observation.

## 5.2.2 Bayesian framework for RWRE

In the framework of Section 5.1.1, the observation consists in the numbers  $(L_k^n)_{0 \leq k \leq n}$  of jumps from  $k$  to  $k - 1$  of the sequence  $X$  until it reaches  $n$  as defined in Equation (5.6). The family of distributions characterizing the statistical model is thus given by the log-likelihood of the left jumps  $(L_k^n)_{0 \leq k \leq n}$  of the RWRE under  $\mathbf{Q}'_0$  defined in Equation (5.8).

In this setting, Bayes rule ensures thus that the posterior writes, for any measurable  $A \subset \Theta$

$$(5.14) \quad \Pi(A | L_{0:n}^n) = \frac{\int_A e^{l_n(\nu, (L_{n-k}^n)_{0 \leq k \leq n})} \Pi(d\nu)}{\int_{\Theta} e^{l_n(\nu, (L_{n-k}^n)_{0 \leq k \leq n})} \Pi(d\nu)},$$

where, for any parameter  $\nu$  in  $\Theta$ ,  $l_n(\nu, \cdot)$  is defined in Equation (5.8).

### 5.2.3 Bayesian framework for BPIRE

When considering the BPIRE, observations are given by the sequence  $(Z_{0:n})_{n \in \mathbb{N}}$  of Section 5.1.2 and the family of distribution  $(\mathbf{P}_0^\nu)_{\nu \in \Theta}$  defines the statistical model.

Denoting  $l_n(\nu) = l_n(\nu, Z_{0:n})$  for any parameter  $\nu$  in  $\Theta$  (where  $l_n(\nu, z_{0:n})$  is defined in Equation (5.8)), Bayes rule ensures that the posterior writes for any measurable set  $A \subset \Theta$

$$(5.15) \quad \Pi(A | Z_{0:n}) = \frac{\int_A e^{l_n(\nu)} \Pi(d\nu)}{\int_{\Theta} e^{l_n(\nu)} \Pi(d\nu)}.$$

### 5.2.4 Main results : posterior consistency for RWRE and BPIREG( $\nu, 0$ )

First main result of this chapter is the posterior consistency of the Bayesian estimator of the environment distribution of a RWRE.

**Theorem 63.** *Consider a prior  $\Pi$  on the set  $\Theta$  of probability measure absolutely continuous with respect to the Lebesgue measure on  $[0, 1]$  and a parameter  $\nu^*$  in  $\text{supp}(\Pi)$  satisfying simultaneously Assumptions **B1**, **B2**, **B3**, **B4**, **B5** and **B6**.*

*In the Bayesian framework of Section 5.2.2, for any  $\varepsilon > 0$*

$$\Pi \left( d^{\nu^*}(\nu^*, \nu) > \varepsilon \mid (L_k^n)_{0 \leq k \leq n} \right) \xrightarrow[n \rightarrow +\infty]{\mathbf{Q}_0^{\nu^*}} 0,$$

where  $d^\nu(\nu', \nu'')$  is defined for any  $\nu, \nu'$  and  $\nu''$  in  $\Theta$  by Equation (5.12) and where for any  $n$  in  $\mathbb{N}^*$  and any  $0 \leq k \leq n$ ,  $L_k^n$  is defined from the trajectory  $X_{0:T_n}$  of the RWRE by

$$L_k^n = \sum_{i=0}^{T_n-1} \mathbb{1}_{\{X_i=k, X_{i+1}=k-1\}}.$$

While we know from [DL18] that  $(L_k^n)_{k \leq n}$  is a sufficient statistic, the conditioning in Theorem 63 is only made to respect to the sub-family  $(L_k^n)_{0 \leq k \leq n}$ . As done in all previous successful studies (see [CFL<sup>+</sup>14, DL18, HLM19]), only one part of the statistical information contained in the trajectory  $X_{0:T_n}$  is thus taken into account to update the prior  $\Pi$ .

In fact, this restriction to  $(L_k^n)_{0 \leq k \leq n}$  is motivated by the fruitful link between left jump process from non negative sites of RWRE and BPIREG( $\nu, 0$ ). In this chapter, Theorem 63 of posterior consistency for RWRE indeed derive from the following result concerning posterior consistency of the Bayesian estimator of the parameter  $\nu$  of a BPIREG( $\nu, 0$ ).

**Theorem 64.** *Consider a prior  $\Pi$  on the set  $\Theta$  of probability measure absolutely continuous with respect to the Lebesgue measure on  $[0, 1]$  and a parameter  $\nu^*$  in  $\text{supp}(\Pi)$  satisfying simultaneously Assumptions **B1**, **B2**, **B3**, **B4**, **B5** and **B6**.*

*In the Bayesian framework of Section 5.2.3, for any  $\varepsilon > 0$*

$$\Pi \left( d^{\nu^*}(\nu^*, \nu) > \varepsilon \mid Z_{0:n} \right) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}_0^{\nu^*}} 0,$$

where  $d^\nu(\nu', \nu'')$  is defined for any  $\nu, \nu'$  and  $\nu''$  in  $\Theta$  by Equation (5.12).

### 5.3 Proof of the main results

In Section 5.3.11, we derive Theorem 63 as a corollary of Theorem 64. Hence, it is sufficient to prove Theorem 64 and Theorem 63 will follow as a straightforward consequence. For this reason, the Bayesian framework of Section 5.2.3 is adopted in following sections (unless explicitly mentioned otherwise) and we denote  $l_n(\nu) = l_n(\nu, Z_{0:n})$  for any  $\nu$  in  $\Theta$  (where  $l_n(\nu, \cdot)$  is defined in Equation (5.8)).

The strategy of proof for our posterior consistency result on BPIRE is based on an adaptation of classical Bayesian non-parametric techniques with frequentist approach in i.i.d. case as it was developed in [Sch65, Bar88, BSW99, GGR99]. To enlighten those techniques as well as the need for Assumptions **B** of Section 5.2.1, we analyse our problem to derive the main ideas underlying the proof of Theorem 64.

#### 5.3.1 Analysis : a look into Bayesian techniques

Considering the function  $d^{\nu^*} : \mathbb{M}^1([0, 1])^2 \rightarrow \mathbb{R}$  given in Equation (5.12), we define the  $\varepsilon$ -vicinity of  $\nu^*$  for  $d^{\nu^*}$  for any  $\varepsilon > 0$  by

$$(5.16) \quad A_\varepsilon = \left\{ d^{\nu^*}(\nu^*, \nu) > \varepsilon \right\},$$

so that posterior consistency of Theorem 64 means that for any  $\delta > 0$

$$(5.17) \quad \mathbf{P}_0^{\nu^*} (\Pi(A_\varepsilon | Z_{0:n}) \geq \delta) \xrightarrow[n \rightarrow +\infty]{} 0.$$

Since Markov inequality ensures that for any  $\delta > 0$  and any  $n$  in  $\mathbb{N}$

$$\mathbf{P}_0^{\nu^*} (\Pi(A_\varepsilon | Z_{0:n}) \geq \delta) \leq \frac{\mathbf{E}_0^{\nu^*} [\Pi(A_\varepsilon | Z_{0:n})]}{\delta},$$

the convergence (5.17) follows if for any  $\varepsilon > 0$

$$(5.18) \quad \mathbf{E}_0^{\nu^*} [\Pi(A_\varepsilon | Z_{0:n})] \xrightarrow[n \rightarrow +\infty]{} 0.$$

In order to prove convergence (5.18), we have to focus on the control of  $\mathbf{E}^{\nu^*} [\Pi(A | Z_{0:n})]$  for any measurable  $A \subset \Theta$  (keeping in mind that  $A$  typically denotes the complementary of some vicinity of  $\nu^*$ ).

Defining for any measurable  $A \subset \Theta$

$$(5.19) \quad N_n(A) = \int_A e^{l_n(\nu) - l_n(\nu^*)} \Pi(d\nu) \quad \text{and} \quad D_n = \int_\Theta e^{l_n(\nu) - l_n(\nu^*)} \Pi(d\nu),$$

we derive from Equation (5.15) that

$$\Pi(A | Z_{0:n}) = \frac{\int_A e^{l_n(\nu)} \Pi(d\nu)}{\int_\Theta e^{l_n(\nu)} \Pi(d\nu)} = \frac{\int_A e^{l_n(\nu) - l_n(\nu^*)} \Pi(d\nu)}{\int_\Theta e^{l_n(\nu) - l_n(\nu^*)} \Pi(d\nu)} = \frac{N_n(A)}{D_n}.$$

For any measurable function  $\phi_n$  on  $\mathbb{N}^{n+1}$  and any measurable  $A \subset \Theta$ , we thus derive the following decomposition

$$(5.20) \quad \frac{N_n(A)}{D_n} = \frac{N_n(A)}{D_n} \phi_n(Z_{0:n}) + \frac{N_n(A)}{D_n} (1 - \phi_n(Z_{0:n})).$$

On one hand, if  $\phi_n \geq 0$ , since the ratio  $N_n(A)/D_n$  lies in  $[0, 1]$ , we have

$$\frac{N_n(A)}{D_n} \phi_n(Z_{0:n}) \leq \phi_n(Z_{0:n}) .$$

On the other hand, if  $\phi_n \leq 1$  so that  $1 - \phi_n \geq 0$  and if there exists, for any  $n$  in  $\mathbb{N}$ , a deterministic threshold  $b_n$  in  $\mathbb{R}_+^*$  such that

$$(5.21) \quad D_n \geq b_n ,$$

then

$$\frac{N_n(A)}{D_n} (1 - \phi_n(Z_{0:n})) \leq \frac{N_n(A)}{b_n} (1 - \phi_n(Z_{0:n})) .$$

Hence if  $0 \leq \phi_n \leq 1$  and if a sequence  $(b_n)_{n \in \mathbb{N}}$  meeting condition (5.21) exists, then

$$\frac{N_n(A)}{D_n} \leq \phi_n(Z_{0:n}) + \frac{N_n(A)}{b_n} (1 - \phi_n(Z_{0:n})) .$$

Taking expectation with respect to  $\mathbf{P}_0^{\nu^*}$ , we get

$$(5.22) \quad \mathbf{E}_0^{\nu^*} [\Pi(A | Z_{0:n})] \leq \mathbf{E}_0^{\nu^*} [\phi_n(Z_{0:n})] + \frac{1}{b_n} \mathbf{E}_0^{\nu^*} [N_n(A) (1 - \phi_n(Z_{0:n}))] .$$

As a measurable function on  $\mathbb{N}^{n+1}$  taking values in  $[0, 1]$ ,  $\phi_n$  is a *test function*.

If we choose for  $\phi_n$  a test function of  $H_0 : \nu = \nu^*$  against  $H_1 : \nu \neq \nu^*$ , the first term of the right hand side in Equation (5.22)  $\mathbf{E}_0^{\nu^*} [\phi_n(Z_{0:n})]$  is actually the first kind risk of  $\phi_n$ . Nevertheless, second kind risk of some test does not appear directly in the right hand side of Equation (5.22) but further decomposition of  $\mathbf{E}_0^{\nu^*} [N_n(A) (1 - \phi_n(Z_{0:n}))]$  put it into light.

For any measurable  $B \subset \Theta$ , we have

$$\mathbf{E}_0^{\nu^*} [N_n(B)(1 - \phi_n(Z_{0:n}))] = \sum_{z_{0:n} \in \mathbb{N}^{n+1}} \left( \int_B e^{l_n(\nu, z_{0:n}) - l_n(\nu^*, z_{0:n})} \Pi(d\nu) \right) (1 - \phi_n(z_{0:n})) e^{l_n(\nu^*, z_{0:n})} .$$

By Fubini's theorem,

$$\begin{aligned} & \sum_{z_{0:n} \in \mathbb{N}^{n+1}} \left( \int_B e^{l_n(\nu, z_{0:n}) - l_n(\nu^*, z_{0:n})} \Pi(d\nu) \right) (1 - \phi_n(z_{0:n})) e^{l_n(\nu^*, z_{0:n})} \\ &= \int_B \left( \sum_{z_{0:n} \in \mathbb{N}^{n+1}} (1 - \phi_n(z_{0:n})) e^{l_n(\nu^*, z_{0:n})} e^{l_n(\nu, z_{0:n}) - l_n(\nu^*, z_{0:n})} \right) \Pi(d\nu) \\ &= \int_B \left( \sum_{z_{0:n} \in \mathbb{N}^{n+1}} (1 - \phi_n(z_{0:n})) e^{l_n(\nu, z_{0:n})} \right) \Pi(d\nu) . \end{aligned}$$

Thus,

$$\begin{aligned} \mathbf{E}_0^{\nu^*} [N_n(B)(1 - \phi_n(Z_{0:n}, \nu^*))] &= \int_B \left( \sum_{z_{0:n} \in \mathbb{N}^{n+1}} (1 - \phi_n(z_{0:n})) e^{l_n(\nu, z_{0:n})} \right) \Pi(d\nu) \\ &= \int_B \mathbf{E}_0^{\nu} [1 - \phi_n(Z_{0:n})] \Pi(d\nu) \\ &\leq \Pi(B) \sup_{\nu \in B} (\mathbf{E}_0^{\nu} [1 - \phi_n(Z_{0:n})]) . \end{aligned}$$

For any measurable  $A \subset \Theta$  and  $C \subset \Theta$ , the decomposition  $N_n(A) = N_n(A \cap C) + N_n(A \cap C^c)$  is valid so that, considering a family  $(\Theta_n)_{n \in \mathbb{N}}$  of measurable subsets of  $\Theta$

$$\begin{aligned} & \mathbf{E}_0^{\nu^*} [N_n(A) (1 - \phi_n(Z_{0:n}))] \\ &= \mathbf{E}_0^{\nu^*} [N_n(A \cap \Theta_n) (1 - \phi_n(Z_{0:n}))] + \mathbf{E}_0^{\nu^*} [N_n(A \cap \Theta_n^c) (1 - \phi_n(Z_{0:n}))] \\ &\leq \Pi(A \cap \Theta_n) \sup_{\nu \in A \cap \Theta_n} \mathbf{E}_0^\nu [1 - \phi_n(Z_{0:n})] + \Pi(A \cap \Theta_n^c) \sup_{\nu \in A \cap \Theta_n^c} \mathbf{E}_0^\nu [1 - \phi_n(Z_{0:n})]. \end{aligned}$$

Since  $\Pi(A \cap \Theta_n)$  and  $\sup_{\nu \in A \cap \Theta_n^c} (\mathbf{E}_0^\nu [1 - \phi_n(Z_{0:n})])$  are both upper-bounded by 1 (because  $\Pi$  is a probability and  $1 - \phi_n(Z_{0:n})$  lies in  $[0, 1]$ ), we derive

$$\mathbf{E}_0^{\nu^*} [N_n(A) (1 - \phi_n(Z_{0:n}))] \leq \sup_{\nu \in A \cap \Theta_n} (\mathbf{E}_0^\nu [1 - \phi_n(Z_{0:n})]) + \Pi(\Theta_n^c).$$

Plugged in Inequality (5.22), this yields

$$(5.23) \quad \mathbf{E}_0^{\nu^*} [\Pi(A | Z_{0:n})] \leq \mathbf{E}_0^{\nu^*} [\phi_n(Z_{0:n})] + \frac{1}{b_n} \sup_{\nu \in A \cap \Theta_n} \mathbf{E}_0^\nu [1 - \phi_n(Z_{0:n})] + \frac{\Pi(\Theta_n^c)}{b_n}.$$

Replacing  $A$  with the set  $A_\varepsilon$  defined in Equation (5.16) and choosing for  $\phi_n$  a test function of hypothesis  $H_0 : \nu = \nu^*$  against hypothesis  $H_1 : \nu \in A_\varepsilon \cap \Theta_n$ , the terms  $\mathbf{E}_0^{\nu^*} [\phi_n(Z_{0:n})]$  and  $\sup \{ \mathbf{E}_0^\nu [1 - \phi_n(Z_{0:n})] : \nu \in A \cap \Theta_n \}$  in the right hand side of Inequality (5.23) are now first and second kind risks respectively.

Finally, for any  $\varepsilon > 0$ , we derive conditions granting the convergence of  $\mathbf{E}_0^{\nu^*} [\Pi(A_\varepsilon | Z_{0:n})]$  to 0 as  $n$  grows to infinity as required in condition (5.18). It is sufficient that for any  $\varepsilon > 0$  there exist a family of positive real numbers  $(b_n)_{n \in \mathbb{N}}$  and a family of test functions  $\phi_n : \mathbb{N}^{n+1} \rightarrow [0, 1]$  and a family  $(\Theta_n)_{n \in \mathbb{N}}$  of measurable subsets of  $\Theta$  such that

- (i)  $\mathbf{E}_0^{\nu^*} [\phi_n(Z_{0:n})]$  tends to 0 when  $n$  grows to infinity,
- (ii)  $\frac{1}{b_n} \sup_{\nu \in A_\varepsilon \cap \Theta_n} \mathbf{E}_0^\nu [1 - \phi_n(Z_{0:n})]$  tends to 0 when  $n$  grows to infinity,
- (iii)  $\frac{\Pi(\Theta_n^c)}{b_n}$  tends to 0 when  $n$  grows to infinity.
- (iv)  $D_n \geq b_n$  for any  $n$  in  $\mathbb{N}$ .

### 5.3.2 Sketch of proof of Theorem 64

Hereafter, we consider a fixed  $\varepsilon > 0$ .

In Section 5.3.1, we reduced the proof of Theorem 64 to the verification of sufficient conditions (i), (ii), (iii) and (iv). After specifying the family  $\Theta_n$  and defining precisely the families  $\phi_n$  and  $b_n$ , we expose here all intermediate results leading to these conditions.

First of all, for any  $n$  in  $\mathbb{N}^*$ , any  $z_{0:n}$  in  $\mathbb{N}^{n+1}$  and any  $\nu, \nu'$  and  $\nu''$  in  $\Theta$ :

- $m_n(\nu, z_{0:n})$  is defined by

$$(5.24) \quad m_n(\nu, z_{0:n}) = \log \left( \prod_{i=0}^{n-1} K^\nu(z_i, z_{i+1}) \right) = \sum_{i=0}^{n-1} \log (K^\nu(z_i, z_{i+1})),$$



- $d_n(\nu', \nu'', z_{0:n})$  is defined by

$$(5.25) \quad d_n(\nu', \nu'', z_{0:n}) = \frac{m_n(\nu', z_{0:n}) - m_n(\nu'', z_{0:n})}{n},$$

- $F_n(\nu', \nu'', z_{0:n}, \nu)$  is defined by

$$(5.26) \quad F_n(\nu', \nu'', z_{0:n}, \nu) = d_n(\nu', \nu'', z_{0:n}) - d^\nu(\nu', \nu'').$$

Then, by Assumption **B6**, there exists a family  $(\Theta_n)_{n \in \mathbb{N}}$  of measurable subsets of  $\Theta$  and a positive constant  $C_\varepsilon$  such that requirements (5.13) are satisfied. Thus, we define for any  $\nu^*$  in  $\mathbb{M}^1([0, 1])$  and any  $n$  in  $\mathbb{N}^*$  the test function  $\phi_n$  by

$$(5.27) \quad \phi_n(z_{0:n}) = \mathbb{1} \left\{ \sup_{\nu' \in \Theta_n} F_n(\nu', \nu^*, z_{0:n}, \nu^*) > \frac{\varepsilon}{2} \right\},$$

Finally, for any positive  $\delta$ , we consider the family  $(b_n^\delta)_{n \in \mathbb{N}}$  defined, for any  $n$  in  $\mathbb{N}$ , by

$$(5.28) \quad b_n^\delta = s_\delta e^{-\frac{\delta}{m} n},$$

where

$$(5.29) \quad s_\delta = \Pi(\nu \in \Theta : \|f_{\nu^*} - f_\nu\|_\infty \leq \delta),$$

is positive thanks to Assumption **B5**.

**Remark 65.** We point out that, for any probability on natural integers  $\mu$  and any  $\nu$  in  $\Theta$ , the log-likelihood of the BPIRE under  $\mathbf{P}_\mu^\nu$  can be written, for any  $z_{0:n}$  in  $\mathbb{N}^{n+1}$

$$\log(\mu(z_0)) + m_n(\nu, z_{0:n}).$$

In particular, for  $\mu = \delta_0$ , we have  $l_n(\nu, z_{0:n}) = \log(\delta_0(z_0)) + m_n(\nu, z_{0:n})$ .

Thanks to the convergence (5.9),  $d_n(\nu', \nu'', Z_{0:n})$  tends in  $\mathbf{P}_0^\nu$ -probability to  $d^\nu(\nu', \nu'')$  as  $n$  grows to infinity. We thus consider  $d_n(\nu', \nu'', Z_{0:n})$  as a random approximation of  $d^\nu(\nu', \nu'')$ .

Before evaluating the quality of this approximation, it is important to notice that  $d^\nu(\nu', \nu'')$  is in fact the expectation of  $d_n(\nu', \nu'', Z_{0:n})$  under  $\mathbf{P}_{\mu^\nu}^\nu$ .

**Proposition 66.** For any  $\nu, \nu'$  and  $\nu''$  in  $\Theta$

$$(5.30) \quad d^\nu(\nu', \nu'') = \mathbf{E}_{\mu^\nu}^\nu [d_n(\nu', \nu'', Z_{0:n})].$$

Decomposing  $F_n(\nu', \nu'', Z_{0:n}, \nu) = d_n(\nu', \nu'', Z_{0:n}) - d^\nu(\nu', \nu'')$  in a random and a deterministic part by inserting the central term  $\mathbf{E}_0^\nu [d_n(\nu', \nu'', Z_{0:n})]$  appears now meaningful

$$\begin{aligned} & F_n(\nu', \nu'', Z_{0:n}, \nu) \\ &= d_n(\nu', \nu'', Z_{0:n}) - \mathbf{E}_0^\nu [d_n(\nu', \nu'', Z_{0:n})] + \mathbf{E}_0^\nu [d_n(\nu', \nu'', Z_{0:n})] - \mathbf{E}_{\mu^\nu}^\nu [d_n(\nu', \nu'', Z_{0:n})], \end{aligned}$$

and the triangular inequality yields

$$(5.31) \quad |F_n(\nu', \nu'', Z_{0:n}, \nu)| \leq |d_n(\nu', \nu'', Z_{0:n}) - \mathbf{E}_0^\nu [d_n(\nu', \nu'', Z_{0:n})]| + |\mathbf{E}_0^\nu [d_n(\nu', \nu'', Z_{0:n})] - \mathbf{E}_{\mu^\nu}^\nu [d_n(\nu', \nu'', Z_{0:n})]|.$$

Control of  $|F_n(\nu', \nu'', Z_{0:n}, \nu)|$  will thus follow from the controls of:

- (a) the random term  $|\mathbf{d}_n(\nu', \nu'', Z_{0:n}) - \mathbf{E}'_0[\mathbf{d}_n(\nu', \nu'', Z_{0:n})]|$ ,
- (b) the deterministic term  $|\mathbf{E}'_0[\mathbf{d}_n(\nu', \nu'', Z_{0:n})] - \mathbf{E}'_{\mu\nu}[\mathbf{d}_n(\nu', \nu'', Z_{0:n})]|$ .

The following proposition provides a control of the random term (a).

**Proposition 67.** *Assume that **B1**, **B2** and **B3** hold and denote the constants introduced in those assumptions by  $\alpha$  in  $(0, 1/2)$  and by  $c_0$  and  $m$  in  $\mathbb{R}_+^*$ .*

*Then, for any  $n$  in  $\mathbb{N}^*$ , any  $t$  in  $\mathbb{R}_+$ , any  $\nu, \nu'$  and  $\nu''$  in  $\text{supp}(\Pi)$*

$$(5.32) \quad \mathbf{P}'_0 \left( |\mathbf{d}_n(\nu', \nu'', Z_{0:n}) - \mathbf{E}'_0[\mathbf{d}_n(\nu', \nu'', Z_{0:n})]| > t \right) \leq 2 \exp \left( -\frac{\beta(c_0, \alpha) m^2}{32 \|f_{\nu''} - f_{\nu'}\|_\infty^2} n t^2 \right),$$

where  $\beta(c_0, \alpha)$  is defined in Equation (4.9) of Chapter 4.

The following proposition provides a control of the deterministic term (b).

**Proposition 68.** *Assume that **B1**, **B2** and **B3** hold and denote the constants introduced in those assumptions by  $\alpha$  in  $(0, 1/2)$  and by  $c_0$  and  $m$  in  $\mathbb{R}_+^*$ .*

*Then, for any  $n$  in  $\mathbb{N}^*$ , any  $t$  in  $\mathbb{R}_+$ , any  $\nu, \nu'$  and  $\nu''$  in  $\text{supp}(\Pi)$*

$$(5.33) \quad |\mathbf{E}'_0[\mathbf{d}_n(\nu', \nu'', Z_{0:n})] - \mathbf{E}'_{\mu\nu}[\mathbf{d}_n(\nu', \nu'', Z_{0:n})]| \leq 2 \frac{\|f_{\nu'} - f_{\nu''}\|_\infty}{n m} \frac{\zeta_0(c_0, \alpha)}{1 - r_0(c_0, \alpha)},$$

where  $r_0(c_0, \alpha)$  and  $\zeta_0(c_0, \alpha)$  are respectively defined in Equations (4.95) and (4.96) of Chapter 4.

The controls provided by Propositions 67 and 68 plugged in Inequality (5.31) finally enable to get a probabilistic control of  $|F_n(\nu', \nu'', Z_{0:n}, \nu)|$  uniform for  $\nu'$  and  $\nu''$  in some measurable compact subset of  $\Theta$ .

Hereafter,  $\Delta(T)$  denotes the *diameter* of  $T$  and is defined for any  $T \subset \Theta$  by

$$(5.34) \quad \Delta(T) = \sup_{(\nu_1, \nu_2) \in T^2} \|f_{\nu_1} - f_{\nu_2}\|_\infty.$$

**Proposition 69.** *Assume that **B1**, **B2** and **B3** hold and denote the constants introduced in those assumptions by  $\alpha$  in  $(0, 1/2)$  and by  $c_0$  and  $m$  in  $\mathbb{R}_+^*$ . If  $\tilde{\Theta}$  is a subset of  $\Theta$  compact for the topology induced by  $\|\cdot\|_\infty$ , then, for any positive  $\tilde{\varepsilon}$ , any  $\nu$  in  $\tilde{\Theta}$ , any  $n$  in  $\mathbb{N}^*$  and any  $u$  in  $\mathbb{R}_+^*$ , with probability under  $\mathbf{P}'_0$  less than  $e^{-u^2}$*

$$\sup_{(\nu', \nu'') \in \tilde{\Theta}^2} |F_n(\nu', \nu'', Z_{0:n}, \nu)| \geq C(c_0, \alpha, m) \left( \tilde{\varepsilon} + \Delta(\tilde{\Theta}) \frac{u + \sqrt{\log \left( 2N(\tilde{\Theta}, \|\cdot\|_\infty, \tilde{\varepsilon}) \right)}}{\sqrt{n}} \right),$$

where  $N(\tilde{\Theta}, \|\cdot\|_\infty, \tilde{\varepsilon})$  are the covering numbers of set  $\tilde{\Theta}$  for  $\|\cdot\|_\infty$  (defined in Appendix B) and  $C(c_0, \alpha, m)$  is defined by

$$(5.35) \quad C(c_0, \alpha, m) = \frac{2\sqrt{2}}{m} \max \left( \sqrt{2}, \frac{1}{\sqrt{\beta(c_0, \alpha)}} + \frac{\zeta_0(c_0, \alpha)}{(1 - r_0(c_0, \alpha)) \sqrt{2 \log(2)}} \right),$$

with  $r_0(c_0, \alpha)$ ,  $\zeta_0(c_0, \alpha)$  and  $\beta(c_0, \alpha)$  defined in Equations (4.95), (4.96) and (4.9) of Chapter 4.

Thanks to Proposition 69, the first and second kind risks of the test  $\phi_n$  appearing in conditions (i) and (ii) are controlled in Section 5.3.9.

The last requirement for the effective control of (ii) and (iii) is the existence of a deterministic lower-bound on  $D_n$  ensured by the following proposition. Its proof is postponed to Section 5.3.3.

**Proposition 70.** *Assume that assumptions **B3** and **B5** hold. Consider the families  $(b_n^\delta)_{n \in \mathbb{N}}$  defined in Equation (5.28). For any  $\delta > 0$  and for any  $n$  in  $\mathbb{N}$*

$$(5.36) \quad D_n \geq b_n^\delta.$$

In Section 5.3.10, thanks to the control of the first and second kind risks of the test  $\phi_n$  and to Equation (5.36), we choose a specific value for  $\delta$  so that conditions (i), (ii) and (iii) are simultaneously satisfied.

### 5.3.3 Minoration of the denominator $D_n$ : proof of Proposition 70

The denominator  $D_n$  being defined as the integral over  $\Theta$  of  $e^{l_n(\nu) - l_n(\nu^*)}$  with respect to  $\Pi$ , lower-bounding  $D_n$  is equivalent to upper-bounding the difference of the log-likelihoods  $l_n(\nu^*) - l_n(\nu)$ . Since Remark 65 ensures that  $l_n(\nu^*) - l_n(\nu) = m_n(\nu^*, \cdot) - m_n(\nu, \cdot)$ , we first look for a control of  $|m_n(\nu', \cdot) - m_n(\nu, \cdot)|$ .

**Lemma 71.** *Assume that **B2**, **B3** and **B4** hold. Then, for any  $\nu$  and  $\nu'$  in  $\text{supp}(\Pi)$ , any  $n$  in  $\mathbb{N}$ , any  $z_{0:n}$  in  $\mathbb{N}^{n+1}$  and any integer  $0 \leq i \leq n-1$*

$$(5.37) \quad \left| \log \left( \frac{\int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu'(dt)}{\int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu(dt)} \right) \right| \leq \frac{\|f_{\nu'} - f_\nu\|_\infty}{m}.$$

Therefore

$$(5.38) \quad |m_n(\nu', z_{0:n}) - m_n(\nu, z_{0:n})| \leq \frac{\|f_{\nu'} - f_\nu\|_\infty}{m} n.$$

*Proof.* Let  $n$  in  $\mathbb{N}$  and  $z_{0:n}$  in  $\mathbb{N}^{n+1}$ . Let  $m_n(\nu) = m_n(\nu, z_{0:n})$  for any  $\nu$  in  $\Theta$ . From Definition (5.24) of  $m_n(\nu, z_{0:n})$ , we derive, for any  $\nu$  and  $\nu'$  in  $\Theta$

$$m_n(\nu') - m_n(\nu) = \log \left( \prod_{i=0}^{n-1} \frac{K^{\nu'}(z_i, z_{i+1})}{K^\nu(z_i, z_{i+1})} \right) = \sum_{i=0}^{n-1} \log \left( \frac{\int_0^1 \binom{z_i + z_{i+1}}{z_i} t^{z_i+1} (1-t)^{z_{i+1}} \nu'(dt)}{\int_0^1 \binom{z_i + z_{i+1}}{z_i} t^{z_i+1} (1-t)^{z_{i+1}} \nu(dt)} \right).$$

Hence, by the triangular inequality

$$(5.39) \quad |m_n(\nu') - m_n(\nu)| \leq \sum_{i=0}^{n-1} \left| \log \left( \frac{\int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu'(dt)}{\int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu(dt)} \right) \right|.$$

By concavity of the logarithm,  $|\log(x) - \log(y)| \leq |x - y| / \min(x, y)$  for any  $x$  and  $y$  in  $\mathbb{R}_+^*$ . Hence, for any integer  $0 \leq i \leq n - 1$ , we have

$$\begin{aligned} \left| \log \left( \frac{\int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu'(dt)}{\int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu(dt)} \right) \right| &\leq \frac{\left| \int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu'(dt) - \int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu(dt) \right|}{\min \left( \int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu'(dt), \int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu(dt) \right)} \\ &= \frac{\left| \int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} (\nu' - \nu)(dt) \right|}{\min \left( \int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu'(dt), \int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu(dt) \right)}. \end{aligned}$$

The uniform lower bound on  $[\alpha, 1 - \alpha]$  of the densities in  $\text{supp}(\Pi)$  granted by **B3** yields

$$\begin{aligned} &\min \left( \int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} f_{\nu'}(t) dt, \int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} f_{\nu}(t) dt \right) \\ &\geq m \int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} dt. \end{aligned}$$

The upper-bound in  $\|\cdot\|_{\infty}$ -norm of the densities in  $\text{supp}(\Pi)$  granted by **B2** and **B4** yields

$$\begin{aligned} &\left| \int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} (f_{\nu'}(t) - f_{\nu}(t)) dt \right| \leq \int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} |f_{\nu'}(t) - f_{\nu}(t)| dt \\ &\leq \int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \|f_{\nu'} - f_{\nu}\|_{\infty} dt \leq \|f_{\nu'} - f_{\nu}\|_{\infty} \int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} dt. \end{aligned}$$

Hence

$$\begin{aligned} \left| \log \left( \frac{\int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu'(dt)}{\int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu(dt)} \right) \right| &\leq \frac{\|f_{\nu'} - f_{\nu}\|_{\infty} \int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} dt}{m \int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} dt} \\ &\leq \frac{\|f_{\nu'} - f_{\nu}\|_{\infty}}{m}. \end{aligned}$$

Combining Equations (5.39) and (5.37), we finally find that Inequality (5.38) holds for any  $\nu$  and  $\nu'$  in  $\text{supp}(\Pi)$ .  $\square$

From Remark 65 and Lemma 71, we derive

$$|l_n(\nu^*) - l_n(\nu)| = |m_n(\nu^*) - m_n(\nu)| \leq \frac{\|f_{\nu^*} - f_{\nu}\|_{\infty}}{m} n,$$

Hence, under **B5**, for any positive  $\delta$

$$D_n \geq \int_{\Theta} e^{-\frac{\|f_{\nu^*} - f_{\nu}\|_{\infty}}{m} n} \Pi(d\nu) \geq \int_{\{\nu / \|f_{\nu^*} - f_{\nu}\|_{\infty} \leq \delta\}} e^{-\frac{\delta}{m} n} \Pi(d\nu) \geq b_n^{\delta} = s_{\delta} e^{-\frac{\delta}{m} n},$$

where  $b_n^{\delta}$  and  $s_{\delta}$  are defined in Equations (5.28) and (5.29).

**Remark 72.** By Equation (5.38) of Lemma 71, under Assumptions **B2**, **B3** and **B4**, we derive in particular, for any  $\nu$  and  $\nu'$  in  $\text{supp}(\Pi)$ , any  $n$  in  $\mathbb{N}^*$  and any  $z_{0:n}$  in  $\mathbb{N}^{n+1}$

$$(5.40) \quad |d_n(\nu, \nu', z_{0:n})| = \left| \frac{m_n(\nu', z_{0:n}) - m_n(\nu'', z_{0:n})}{n} \right| \leq \frac{\|f_{\nu} - f_{\nu'}\|_{\infty}}{m}.$$

### 5.3.4 Some properties of $d^\nu$ and $d_n$

Let recall first the definition of the *bounded differences property*.

**Definition 73.** A function  $f : \mathcal{X}^n \rightarrow \mathbb{R}$  is said to have the bounded differences property if there exist nonnegative constants  $\gamma_{0:n-1} = (\gamma_0, \dots, \gamma_{n-1})$  in  $\mathbb{R}_+^n$  such that for all  $x_{0:n-1}$  in  $\mathcal{X}^n$  and all  $y_{0:n-1}$  in  $\mathcal{X}^n$

$$(5.41) \quad |f(x_{0:n-1}) - f(y_{0:n-1})| \leq \sum_{i=0}^{n-1} \gamma_i \mathbb{1}_{\{x_i \neq y_i\}}.$$

The class of all functions  $f : \mathcal{X}^n \rightarrow \mathbb{R}$  satisfying Inequality (5.41) is denoted  $\mathbb{BD}(\mathcal{X}^n, \gamma_{0:n-1})$ .

We can now give some auxiliary properties of functions  $d^\nu$  and  $d_n$  that will be used to prove Propositions 67 and 69.

**Proposition 74.** For any fixed positive integer  $n$  in  $\mathbb{N}$  and any fixed  $z_{0:n}$  in  $\mathbb{N}^{n+1}$ , the function  $(\nu', \nu'') \in \Theta^2 \mapsto d_n(\nu', \nu'', z_{0:n})$  is anti-symmetric and satisfy the triangular inequality. Moreover, under assumptions **B2**, **B3** and **B4**, for any  $\nu'$  and  $\nu''$  in  $\text{supp}(\Pi)$  and any  $n$  in  $\mathbb{N}^*$ , the function  $z_{0:n} \in \mathbb{N}^{n+1} \mapsto d_n(\nu', \nu'', z_{0:n})$  has the bounded differences property and belongs to  $\mathbb{BD}(\mathcal{X}^{n+1}, \gamma_{0:n}^{\nu', \nu''})$  where for any integer  $0 \leq i \leq n$

$$(5.42) \quad \gamma_i^{\nu', \nu''} = 4 \frac{\|f_{\nu''} - f_{\nu'}\|_\infty}{nm}.$$

*Proof.* Fix  $n$  in  $\mathbb{N}^*$ . For any  $\nu'$  and  $\nu''$  in  $\Theta$  and any  $z_{0:n}$  in  $\mathbb{N}^{n+1}$ ,

$$\begin{aligned} d_n(\nu', \nu'', z_{0:n}) &= m_n(\nu', z_{0:n}) - m_n(\nu'', z_{0:n}) \\ &= - (m_n(\nu'', z_{0:n}) - m_n(\nu', z_{0:n})) = -d_n(\nu'', \nu', z_{0:n}). \end{aligned}$$

Moreover,

$$\begin{aligned} d_n(\nu', \nu'', z_{0:n}) &= \frac{m_n(\nu', z_{0:n}) - m_n(\nu'', z_{0:n})}{n} \\ &= \frac{m_n(\nu', z_{0:n}) - m_n(\nu, z_{0:n})}{n} + \frac{m_n(\nu, z_{0:n}) - m_n(\nu'', z_{0:n})}{n} \\ &= d_n(\nu', \nu, z_{0:n}) + d_n(\nu, \nu'', z_{0:n}). \end{aligned}$$

Hence,  $(\nu', \nu'') \mapsto d_n(\nu', \nu'', z_{0:n})$  is anti-symmetric and satisfies the triangular inequality.

Consider  $\nu'$  and  $\nu''$  fixed in  $\text{supp}(\Pi)$ .

Under Assumptions **B2**, **B3** and **B4**, thanks to inequality (5.37) of Lemma 71, for any  $z_{0:n}$  and  $z'_{0:n}$  in  $\mathbb{N}^{n+1}$  and any integer  $0 \leq i \leq n-1$

$$\left| \log \left( \frac{\int_0^1 t^{z'_i+1} (1-t)^{z'_{i+1}} \nu'(dt)}{\int_0^1 t^{z'_i+1} (1-t)^{z'_{i+1}} \nu''(dt)} \right) \right| + \left| \log \left( \frac{\int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu'(dt)}{\int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu''(dt)} \right) \right| \leq 2 \frac{\|f_{\nu'} - f_{\nu''}\|_\infty}{m}.$$

Moreover, if  $(z_i, z_{i+1}) = (z'_i, z'_{i+1})$ , we directly have

$$\log \left( \frac{\int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu'(dt)}{\int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu''(dt)} \right) - \log \left( \frac{\int_0^1 t^{z'_i+1} (1-t)^{z'_{i+1}} \nu'(dt)}{\int_0^1 t^{z'_i+1} (1-t)^{z'_{i+1}} \nu''(dt)} \right) = 0.$$

Hence,

$$(5.43) \quad \left| \log \left( \frac{\int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu'(dt)}{\int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu''(dt)} \right) - \log \left( \frac{\int_0^1 t^{z'_i+1} (1-t)^{z'_{i+1}} \nu'(dt)}{\int_0^1 t^{z'_i+1} (1-t)^{z'_{i+1}} \nu''(dt)} \right) \right| \leq 2 \frac{\|f_{\nu'} - f_{\nu''}\|_\infty}{m} \mathbb{1}_{(z_i, z_{i+1}) \neq (z'_i, z'_{i+1})}.$$

Since  $(z_i, z_{i+1}) \neq (z'_i, z'_{i+1})$  implies either  $z_i \neq z'_i$  or  $z_{i+1} \neq z'_{i+1}$ , Inequality (5.43) leads to

$$\begin{aligned} & |d_n(\nu', \nu'', z_{0:n}) - d_n(\nu', \nu'', z'_{0:n})| \\ &= \frac{1}{n} |m_n(\nu', z_{0:n}) - m_n(\nu'', z_{0:n}) - (m_n(\nu', z'_{0:n}) - m_n(\nu'', z'_{0:n}))| \\ &= \frac{1}{n} \left| \sum_{i=0}^{n-1} \log \left( \frac{\int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu'(dt)}{\int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu''(dt)} \right) - \sum_{i=0}^{n-1} \log \left( \frac{\int_0^1 t^{z'_i+1} (1-t)^{z'_{i+1}} \nu'(dt)}{\int_0^1 t^{z'_i+1} (1-t)^{z'_{i+1}} \nu''(dt)} \right) \right| \\ &\leq \frac{1}{n} \sum_{i=0}^{n-1} \left| \log \left( \frac{\int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu'(dt)}{\int_0^1 t^{z_i+1} (1-t)^{z_{i+1}} \nu''(dt)} \right) - \log \left( \frac{\int_0^1 t^{z'_i+1} (1-t)^{z'_{i+1}} \nu'(dt)}{\int_0^1 t^{z'_i+1} (1-t)^{z'_{i+1}} \nu''(dt)} \right) \right| \\ &\leq \frac{1}{n} \sum_{i=0}^{n-1} 2 \frac{\|f_{\nu'} - f_{\nu''}\|_\infty}{m} \mathbb{1}_{(z_i, z_{i+1}) \neq (z'_i, z'_{i+1})} \leq \sum_{i=0}^{n-1} 2 \frac{\|f_{\nu'} - f_{\nu''}\|_\infty}{n m} (\mathbb{1}_{z_i \neq z'_i} + \mathbb{1}_{z_{i+1} \neq z'_{i+1}}) \\ &\leq \sum_{i=0}^n 4 \frac{\|f_{\nu'} - f_{\nu''}\|_\infty}{n m} \mathbb{1}_{z_i \neq z'_i}. \end{aligned}$$

Hence, for any fixed  $\nu'$  and  $\nu''$  in  $\Theta$ , the function  $z_{0:n} \mapsto d_n(\nu', \nu'', z_{0:n})$  has the bounded differences property and belongs to  $\mathbb{B}\mathbb{D}(\mathcal{X}^{n+1}, \gamma_{0:n}^{\nu', \nu''})$  where  $\gamma_{0:n}^{\nu', \nu''}$  is defined in Equation (5.42).  $\square$

**Proposition 75.** *For any  $\nu$  in  $\Theta$ , the function  $(\nu', \nu'') \mapsto d^\nu(\nu', \nu'')$  is anti-symmetric and satisfies the triangular inequality. Moreover, the function  $\nu' \mapsto d^\nu(\nu, \nu')$  is non negative.*

*Proof.* Fix  $\nu$  in  $\Theta$ . For any  $\nu', \nu''$  and  $\nu'''$  in  $\Theta$ ,

$$\begin{aligned} d^\nu(\nu', \nu'') &= m^\nu(\nu') - m^\nu(\nu'') \\ &= -(m^\nu(\nu'') - m^\nu(\nu')) = -d^\nu(\nu'', \nu'). \end{aligned}$$

Moreover,

$$\begin{aligned} d^\nu(\nu', \nu'') &= m^\nu(\nu') - m^\nu(\nu'') \\ &= m^\nu(\nu') - m^\nu(\nu''') + m^\nu(\nu''') - m^\nu(\nu'') \\ &= d^\nu(\nu', \nu''') + d^\nu(\nu''', \nu''). \end{aligned}$$

Hence,  $(\nu', \nu'') \mapsto d^\nu(\nu', \nu'')$  is anti-symmetric and satisfies the triangular inequality. For any  $\nu'$  and  $\nu''$  in  $\Theta$ , we have

$$\begin{aligned} (5.44) \quad d^\nu(\nu', \nu'') &= m^\nu(\nu') - m^\nu(\nu'') \\ &= \mathbf{E}_{\mu\nu}^\nu \left[ \log \left( \int_0^1 t^{Z_0+1} (1-t)^{Z_1} \nu'(dt) \right) \right] - \mathbf{E}_{\mu\nu}^\nu \left[ \log \left( \int_0^1 t^{Z_0+1} (1-t)^{Z_1} \nu''(dt) \right) \right] \\ &= \mathbf{E}_{\mu\nu}^\nu \left[ \log \left( \int_0^1 t^{Z_0+1} (1-t)^{Z_1} \nu'(dt) \right) - \log \left( \int_0^1 t^{Z_0+1} (1-t)^{Z_1} \nu''(dt) \right) \right] \\ &= \mathbf{E}_{\mu\nu}^\nu \left[ \log \left( \frac{\int_0^1 t^{Z_0+1} (1-t)^{Z_1} \nu'(dt)}{\int_0^1 t^{Z_0+1} (1-t)^{Z_1} \nu''(dt)} \right) \right] = \mathbf{E}_{\mu\nu}^\nu \left[ \log \left( \frac{K^{\nu'}(Z_0, Z_1)}{K^{\nu''}(Z_0, Z_1)} \right) \right] \\ &= \mathbf{E}_{\mu\nu}^\nu \left[ \mathbf{E}_{\mu\nu}^\nu \left[ \log \left( \frac{K^{\nu'}(Z_0, Z_1)}{K^{\nu''}(Z_0, Z_1)} \right) \middle| Z_0 \right] \right]. \end{aligned}$$

By Jensen's inequality applied with the conditional distribution  $\mathbf{P}_{\mu\nu}^\nu(\cdot | Z_0)$

$$(5.45) \quad \mathbf{E}_{\mu\nu}^\nu \left[ \log \left( \frac{K^{\nu'}(Z_0, Z_1)}{K^{\nu''}(Z_0, Z_1)} \right) \middle| Z_0 \right] \leq \log \left( \mathbf{E}_{\mu\nu}^\nu \left[ \frac{K^{\nu'}(Z_0, Z_1)}{K^{\nu''}(Z_0, Z_1)} \middle| Z_0 \right] \right).$$

As

$$\mathbf{E}_{\mu\nu}^\nu \left[ \frac{K^{\nu'}(Z_0, Z_1)}{K^{\nu''}(Z_0, Z_1)} \middle| Z_0 \right] = \sum_{z_1 \in \mathbb{N}} \frac{K^{\nu'}(Z_0, z_1)}{K^{\nu''}(Z_0, z_1)} K^\nu(Z_0, z_1),$$

for  $\nu'' = \nu$ , we derive

$$(5.46) \quad \mathbf{E}_{\mu\nu}^\nu \left[ \frac{K^{\nu'}(Z_0, Z_1)}{K^\nu(Z_0, Z_1)} \middle| Z_0 \right] = \sum_{z_1 \in \mathbb{N}} \frac{K^{\nu'}(Z_0, z_1)}{K^\nu(Z_0, z_1)} K^\nu(Z_0, z_1) = \sum_{z_1 \in \mathbb{N}} K^{\nu'}(Z_0, z_1) = K^{\nu'}(Z_0, \mathbb{N}) = 1.$$

Plugging Equations (5.45) and (5.46) in Equation (5.44), we get, for any  $\nu$  and  $\nu'$  in  $\Theta$ ,

$$d^\nu(\nu', \nu) \leq \mathbf{E}_{\mu\nu}^\nu [\log(1)] = 0.$$

Since  $d^\nu(\nu, \nu') = -d^\nu(\nu', \nu)$  by the first part of this proof, it follows that  $d^\nu(\nu, \nu')$  is non negative for any  $\nu$  and  $\nu'$  in  $\Theta$ .  $\square$

### 5.3.5 Link between $d^\nu$ and $d_n$ : proof of Proposition 66

Fix  $\nu$  in  $\mathbb{M}^1([0, 1])$  such that  $\int_0^1 \log(\frac{1-t}{t})\nu(dt)$ .

Since  $\mu^\nu$  is invariant under  $K^\nu$ ,  $(Z_0, Z_1)$  and  $(Z_i, Z_{i+1})$  have the same distribution under  $\mathbb{P}_{\mu^\nu}^\nu$  for any  $i$  in  $\mathbb{N}$ . Therefore, for any  $i$  in  $\mathbb{N}$  and any  $\nu'$  in  $\mathbb{M}^1([0, 1])$

$$\mathbf{E}_{\mu^\nu}^\nu \left[ \log \left( \int_0^1 t^{Z_0+1} (1-t)^{Z_1} \nu'(dt) \right) \right] = \mathbf{E}_{\mu^\nu}^\nu \left[ \log \left( \int_0^1 t^{Z_i+1} (1-t)^{Z_{i+1}} \nu'(dt) \right) \right].$$

Hence, by linearity of expectation

$$\mathbf{E}_{\mu^\nu}^\nu \left[ \log \left( \int_0^1 t^{Z_0+1} (1-t)^{Z_1} \nu'(dt) \right) \right] = \mathbf{E}_{\mu^\nu}^\nu \left[ \frac{1}{n} \sum_{i=0}^{n-1} \log \left( \int_0^1 t^{Z_i+1} (1-t)^{Z_{i+1}} \nu'(dt) \right) \right].$$

For any  $\nu, \nu'$  and  $\nu''$  in  $\Theta$ , the definitions of  $d^\nu(\nu', \nu'')$  and  $m^\nu(\nu')$  in Equations (5.12) and (5.10) then lead to

$$\begin{aligned} d^\nu(\nu', \nu'') &= m^\nu(\nu') - m^\nu(\nu'') \\ &= \mathbf{E}_{\mu^\nu}^\nu \left[ \log \left( \int_0^1 t^{Z_0+1} (1-t)^{Z_1} \nu'(dt) \right) \right] - \mathbf{E}_{\mu^\nu}^\nu \left[ \log \left( \int_0^1 t^{Z_0+1} (1-t)^{Z_1} \nu''(dt) \right) \right] \\ &= \mathbf{E}_{\mu^\nu}^\nu \left[ \frac{1}{n} \sum_{i=0}^{n-1} \log \left( \frac{\int_0^1 t^{Z_i+1} (1-t)^{Z_{i+1}} \nu'(dt)}{\int_0^1 t^{Z_i+1} (1-t)^{Z_{i+1}} \nu''(dt)} \right) \right] \\ &= \mathbf{E}_{\mu^\nu}^\nu \left[ \frac{1}{n} \sum_{i=0}^{n-1} \log \left( \frac{\binom{Z_i + Z_{i+1}}{Z_i} \int_0^1 t^{Z_i+1} (1-t)^{Z_{i+1}} \nu'(dt)}{\binom{Z_i + Z_{i+1}}{Z_i} \int_0^1 t^{Z_i+1} (1-t)^{Z_{i+1}} \nu''(dt)} \right) \right] \\ &= \mathbf{E}_{\mu^\nu}^\nu \left[ \frac{1}{n} \sum_{i=0}^{n-1} \log \left( \frac{K^{\nu'}(Z_i, Z_{i+1})}{K^{\nu''}(Z_i, Z_{i+1})} \right) \right] = \mathbf{E}_{\mu^\nu}^\nu \left[ \frac{1}{n} (m_n(\nu', Z_{0:n}) - m_n(\nu'', Z_{0:n})) \right] \\ &= \mathbf{E}_{\mu^\nu}^\nu [d_n(\nu', \nu'', Z_{0:n})]. \end{aligned}$$

Thus, Equation (5.30) holds.

### 5.3.6 Proof of Proposition 67

Fix  $\nu'$  and  $\nu''$  in  $\Theta$  and assume that **B1**, **B2** and **B3** hold. By Proposition 74, the application  $z_{0:n} \in \mathcal{X}^{n+1} \mapsto d_n(\nu', \nu'', z_{0:n})$  belongs to  $\mathbb{B}\mathbb{D}(\mathcal{X}^{n+1}, \gamma_{0:n}^{\nu', \nu''})$  and  $\gamma_{0:n}^{\nu', \nu''}$  is such that for any  $n$  in  $\mathbb{N}^*$

$$\sum_{i=0}^n (\gamma_i^{\nu', \nu''})^2 = \sum_{i=0}^n \left( 4 \frac{\|f_{\nu''} - f_{\nu'}\|_\infty}{nm} \right)^2 = 16 \frac{n+1}{n} \frac{\|f_{\nu''} - f_{\nu'}\|_\infty^2}{m^2 n} \leq 32 \frac{\|f_{\nu''} - f_{\nu'}\|_\infty^2}{m^2 n}.$$

Since any  $\nu$  in  $\text{supp}(\Pi)$  meets assumptions **F1** and **F2** of Chapter 4, Corollary 41 indeed applies and ensures that, for any  $\nu'$  and  $\nu''$  in  $\Theta$  and any  $t > 0$

$$\mathbf{P}_0^\nu (|d_n(\nu', \nu'', Z_{0:n}) - \mathbf{E}_0^\nu[d_n(\nu', \nu'', Z_{0:n})]| > t) \leq 2 \exp \left( - \frac{\beta(c_0, \alpha) m^2}{32 \|f_{\nu''} - f_{\nu'}\|_\infty^2} n t^2 \right),$$

where  $\beta(c_0, \alpha)$  is defined in Equation (4.9) of Chapter 4.



### 5.3.7 Proof of Proposition 68

For any  $\mu$  in  $\mathbb{M}^1(\mathbb{N})$ , the canonical process  $(Z_n)_{n \in \mathbb{N}}$  is a Markov chain under  $\mathbf{P}_\mu^\nu$  with initial distribution  $\mu$  and transition kernel  $K^\nu$ . By Proposition 83 of Appendix A, the process  $(Z_n, Z_{n+1})_{n \in \mathbb{N}}$  is also a Markov chain under  $\mathbf{P}_\mu^\nu$  with kernel  $\widetilde{K}^\nu$  and initial distribution  $\widetilde{\mu}^{K^\nu}$  respectively defined in Equations (A.6) and (A.7). We thus derive, for any  $\nu, \nu'$  and  $\nu''$  in  $\Theta$

$$\begin{aligned}
& \mathbf{E}_0^\nu [\mathrm{d}_n(\nu', \nu'', Z_{0:n})] - \mathbf{E}_{\mu^\nu}^\nu [\mathrm{d}_n(\nu', \nu'', Z_{0:n})] \\
&= \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{E}_0^\nu \left[ \log \left( \frac{K^{\nu'}(Z_i, Z_{i+1})}{K^{\nu''}(Z_i, Z_{i+1})} \right) \right] - \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{E}_{\mu^\nu}^\nu \left[ \log \left( \frac{K^{\nu'}(Z_i, Z_{i+1})}{K^{\nu''}(Z_i, Z_{i+1})} \right) \right] \\
&= \frac{1}{n} \sum_{i=0}^{n-1} \left( \mathbf{E}_0^\nu \left[ \log \left( \frac{K^{\nu'}(Z_i, Z_{i+1})}{K^{\nu''}(Z_i, Z_{i+1})} \right) \right] - \mathbf{E}_{\mu^\nu}^\nu \left[ \log \left( \frac{K^{\nu'}(Z_i, Z_{i+1})}{K^{\nu''}(Z_i, Z_{i+1})} \right) \right] \right) \\
(5.47) \quad &= \frac{1}{n} \sum_{i=0}^{n-1} \left( \widetilde{\delta}_0^{K^\nu} \widetilde{K}^{\nu^i} \left[ \log \left( \frac{K^{\nu'}}{K^{\nu''}} \right) \right] - \widetilde{\mu}^{\nu^{K^\nu}} \widetilde{K}^{\nu^i} \left[ \log \left( \frac{K^{\nu'}}{K^{\nu''}} \right) \right] \right).
\end{aligned}$$

By characterization of the total variation norm, we get, for any integer  $0 \leq i \leq n-1$

$$\begin{aligned}
(5.48) \quad & \left| \widetilde{\delta}_0^{K^\nu} \widetilde{K}^{\nu^i} \left( \log \left( \frac{K^{\nu'}}{K^{\nu''}} \right) \right) - \widetilde{\mu}^{\nu^{K^\nu}} \widetilde{K}^{\nu^i} \left( \log \left( \frac{K^{\nu'}}{K^{\nu''}} \right) \right) \right| \\
& \leq \left\| \log \left( \frac{K^{\nu'}}{K^{\nu''}} \right) \right\|_\infty \times \left\| \widetilde{\delta}_0^{K^\nu} \widetilde{K}^{\nu^i} - \widetilde{\mu}^{\nu^{K^\nu}} \widetilde{K}^{\nu^i} \right\|_{\mathrm{TV}}.
\end{aligned}$$

For any  $i$  and  $j$  in  $\mathbb{N}$ , by Equation (5.37) of Lemma 71, we know that

$$\log \left( \frac{K^{\nu'}(i, j)}{K^{\nu''}(i, j)} \right) = \log \left( \frac{\int_0^1 t^{i+1} (1-t)^j \nu'(dt)}{\int_0^1 t^{i+1} (1-t)^j \nu''(dt)} \right) \leq \frac{\|f_{\nu'} - f_{\nu''}\|_\infty}{m},$$

so that

$$(5.49) \quad \left\| \log \left( \frac{K^{\nu'}}{K^{\nu''}} \right) \right\|_\infty \leq \frac{\|f_{\nu'} - f_{\nu''}\|_\infty}{m}.$$

Depending whether integer  $i$  is equal or different from 0, we will have to proceed differently in the control of the factor  $\left\| \widetilde{\delta}_0^{K^\nu} \widetilde{K}^{\nu^i} - \widetilde{\mu}^{\nu^{K^\nu}} \widetilde{K}^{\nu^i} \right\|_{\mathrm{TV}}$ .

Consider the case where  $i \geq 1$ . Thanks to Remark 84, for any sets of integers  $A$  and  $B$  and any integer  $i \geq 1$ , we get

$$\begin{aligned}
& \left| \widetilde{\delta}_0^{K^\nu} \widetilde{K}^{\nu^i}(A \times B) - \widetilde{\mu}^{\nu^{K^\nu}} \widetilde{K}^{\nu^i}(A \times B) \right| \\
&= \left| \delta_0 (K^\nu)^{i-1} \left[ (K^\nu)^{\otimes 2}(\cdot, A \times B) \right] - \mu^\nu (K^\nu)^{i-1} \left[ (K^\nu)^{\otimes 2}(\cdot, A \times B) \right] \right| \\
&\leq \left\| (K^\nu)^{\otimes 2}(\cdot, A \times B) \right\|_\infty \left\| \delta_0 (K^\nu)^{i-1} - \mu^\nu (K^\nu)^{i-1} \right\|_{\mathrm{TV}}.
\end{aligned}$$

Since  $\mu^\nu$  is invariant under  $K^\nu$ ,  $\mu_\nu (K^\nu)^j = \mu_\nu$  for any integer  $j$ . By Equation (4.94), we thus have for any integer  $i \geq 1$

$$(5.50) \quad \left\| \delta_0 (K^\nu)^{i-1} - \mu^\nu (K^\nu)^{i-1} \right\|_{\mathrm{TV}} = \left\| \delta_0 (K^\nu)^{i-1} - \mu^\nu \right\|_{\mathrm{TV}} \leq \zeta_0(c_0, \alpha) r_0(c_0, \alpha)^{i-1},$$

where  $r_0(c_0, \alpha)$  and  $\zeta_0(c_0, \alpha)$  are respectively defined in Equation (4.95) and Equation (4.96).

Since  $K^\nu$  is a Markov kernel,  $0 \leq K^\nu(y, C) \leq 1$  for any  $y$  in  $\mathbb{N}$  and any  $C \subset \mathbb{N}$ . For any sets of integers  $A$  and  $B$  and any integer  $x$ , the definition of  $(K^\nu)^{\otimes 2}$  thus yields

$$(K^\nu)^{\otimes 2}(x, A \times B) = \int_A K^\nu(w, B) K^\nu(x, dw) \leq \int_A K^\nu(x, dw) = K^\nu(x, A) \leq 1,$$

so that

$$(5.51) \quad \left\| (K^\nu)^{\otimes 2}(\cdot, A \times B) \right\|_\infty \leq 1.$$

Let's then deal with the case where  $i = 0$ .

For any sets of integers  $A$  and  $B$ , by definition of  $\widetilde{K}^\nu$ ,  $\widetilde{\delta}_0^{K^\nu}$  and  $\widetilde{\mu}^{\nu, K^\nu}$ ,

$$\begin{aligned} & \widetilde{\delta}_0^{K^\nu} \left( \widetilde{K}^\nu \right)^0(A \times B) - \widetilde{\mu}^{\nu, K^\nu} \left( \widetilde{K}^\nu \right)^0(A \times B) \\ &= \sum_{x_0 \in \mathbb{N}} \mathbb{1}_A(x_0) K^\nu(x_0, B) \delta_0(x_0) - \sum_{x_0 \in \mathbb{N}} \mathbb{1}_A(x_0) K^\nu(x_0, B) \mu^\nu(x_0). \end{aligned}$$

Hence,

$$(5.52) \quad \left| \widetilde{\delta}_0^{K^\nu} \left( \widetilde{K}^\nu \right)^0(A \times B) - \widetilde{\mu}^{\nu, K^\nu} \left( \widetilde{K}^\nu \right)^0(A \times B) \right| \leq \|\delta_0 - \mu^\nu\|_{\text{TV}} \times \|\mathbb{1}_A(\cdot) K^\nu(\cdot, B)\|_\infty.$$

Equation (4.94) of Chapter 4 ensures

$$\|\delta_0 - \mu^\nu\|_{\text{T.V.}} \leq \zeta_0(c_0, \alpha) r_0(c_0, \alpha)^0 = \zeta_0(c_0, \alpha),$$

Since  $K^\nu$  is a Markov kernel,

$$\|\mathbb{1}_A(\cdot) K^\nu(\cdot, B)\|_\infty \leq 1.$$

Hence, Inequality (5.52) yields

$$(5.53) \quad \left\| \widetilde{\delta}_0^{K^\nu} \left( \widetilde{K}^\nu \right)^0 - \widetilde{\mu}^{\nu, K^\nu} \left( \widetilde{K}^\nu \right)^0 \right\|_{\text{TV}} \leq \zeta_0(c_0, \alpha).$$

Plugging Equations (5.48), (5.49), (5.50), (5.51) and Equation (5.53) in Equation (5.47), we thus have for any  $\nu$ ,  $\nu'$  and  $\nu''$  in  $\Theta$

$$\begin{aligned} \left| \mathbf{E}_0^\nu [d_n(\nu', \nu'', Z_{0:n})] - \mathbf{E}_{\mu_\nu}^\nu [d_n(\nu', \nu'', Z_{0:n})] \right| &\leq \sum_{i=1}^{n-1} 2 \frac{\|f_{\nu'}' - f_{\nu''}'\|_\infty}{n m} \zeta_0(c_0, \alpha) r_0(c_0, \alpha)^{i-1} \\ &\leq 2 \frac{\|f_{\nu'}' - f_{\nu''}'\|_\infty}{n m} \frac{\zeta_0(c_0, \alpha)}{1 - r_0(c_0, \alpha)}. \end{aligned}$$

### 5.3.8 Proof of Proposition 69

By Proposition 74, for any  $\nu_0$  in  $\Theta$

$$\begin{aligned} & \left| d_n(\nu', \nu'', Z_{0:n}) - \mathbf{E}_0^\nu [d_n(\nu', \nu'', Z_{0:n})] \right| \\ &= \left| d_n(\nu', \nu_0, Z_{0:n}) + d_n(\nu_0, \nu'', Z_{0:n}) - \mathbf{E}_0^\nu [d_n(\nu', \nu_0, Z_{0:n}) + d_n(\nu_0, \nu'', Z_{0:n})] \right| \\ &\leq \left| d_n(\nu', \nu_0, Z_{0:n}) - \mathbf{E}_0^\nu [d_n(\nu', \nu_0, Z_{0:n})] \right| + \left| d_n(\nu'', \nu_0, Z_{0:n}) - \mathbf{E}_0^\nu [d_n(\nu'', \nu_0, Z_{0:n})] \right|. \end{aligned}$$

Hence,

$$(5.54) \quad \sup_{(\nu', \nu'') \in \tilde{\Theta}^2} |\mathbf{d}_n(\nu', \nu'', Z_{0:n}) - \mathbf{E}_0^\nu[\mathbf{d}_n(\nu', \nu'', Z_{0:n})]| \\ \leq 2 \sup_{\nu' \in \tilde{\Theta}} |\mathbf{d}_n(\nu', \nu_0, Z_{0:n}) - \mathbf{E}_0^\nu[\mathbf{d}_n(\nu', \nu_0, Z_{0:n})]|.$$

Moreover, by Equation (5.33), for all  $\nu, \nu'$  and  $\nu''$  in  $\tilde{\Theta}$

$$(5.55) \quad |\mathbf{E}_0^\nu[\mathbf{d}_n(\nu', \nu'', Z_{0:n})] - \mathbf{E}_{\mu\nu}^\nu[\mathbf{d}_n(\nu', \nu'', Z_{0:n})]| \leq 2 \frac{\|f_{\nu'} - f_{\nu''}\|_\infty}{nm} \frac{\zeta_0(c_0, \alpha)}{1 - r_0(c_0, \alpha)} \\ \leq \frac{2 \zeta_0(c_0, \alpha) \Delta(\tilde{\Theta})}{nm (1 - r_0(c_0, \alpha))}.$$

Taking the supremum over  $(\nu', \nu'')$  in  $\tilde{\Theta}^2$  in Equation (5.31) and Inequalities (5.54) and (5.55) then lead to

$$(5.56) \quad \sup_{(\nu', \nu'') \in \tilde{\Theta}^2} |F_n(\nu', \nu'', Z_{0:n}, \nu)| \leq 2 \sup_{\nu' \in \tilde{\Theta}} |\mathbf{d}_n(\nu', \nu_0, Z_{0:n}) - \mathbf{E}_0^\nu[\mathbf{d}_n(\nu', \nu_0, Z_{0:n})]| \\ + \frac{2 \zeta_0(c_0, \alpha) \Delta(\tilde{\Theta})}{nm (1 - r_0(c_0, \alpha))}.$$

Since  $\tilde{\Theta}$  is compact for the topology induced by  $\|\cdot\|_\infty$ , it can be covered for any positive  $\tilde{\varepsilon}$  by a finite number of open balls of radius  $\tilde{\varepsilon}$  for  $\|\cdot\|_\infty$  and thus admits finite  $\tilde{\varepsilon}$ -nets (as defined in Appendix B). For fixed positive  $\tilde{\varepsilon}$ , let  $R(\tilde{\varepsilon})$  denote a finite  $\tilde{\varepsilon}$ -net of  $\tilde{\Theta}$  with respect to  $\|\cdot\|_\infty$  with minimal cardinality so that  $\text{card}(R(\tilde{\varepsilon})) = N(\tilde{\Theta}, \|\cdot\|_\infty, \tilde{\varepsilon})$ .

For any  $\nu'$  in  $\tilde{\Theta}$ , there exists  $\nu_1$  in  $R(\tilde{\varepsilon})$  such that  $\|f_{\nu'} - f_{\nu_1}\|_\infty \leq \tilde{\varepsilon}$  and so

$$\|f_{\nu'} - f_{\nu_0}\|_\infty = \|f_{\nu'} - f_{\nu_1} + f_{\nu_1} - f_{\nu_0}\|_\infty \leq \|f_{\nu'} - f_{\nu_1}\|_\infty + \|f_{\nu_1} - f_{\nu_0}\|_\infty \leq \tilde{\varepsilon} + \|f_{\nu_1} - f_{\nu_0}\|_\infty.$$

Since Remark 72 ensures that  $\mathbf{d}_n(\nu', \nu_1, Z_{0:n}) \leq \|f_{\nu'} - f_{\nu_1}\|_\infty / m$ , Proposition 74 leads to

$$(5.57) \quad |\mathbf{d}_n(\nu', \nu_0, Z_{0:n}) - \mathbf{E}_0^\nu[\mathbf{d}_n(\nu', \nu_0, Z_{0:n})]| \\ = |\mathbf{d}_n(\nu', \nu_1, Z_{0:n}) - \mathbf{E}_0^\nu[\mathbf{d}_n(\nu', \nu_1, Z_{0:n})] + \mathbf{d}_n(\nu_1, \nu_0, Z_{0:n}) - \mathbf{E}_0^\nu[\mathbf{d}_n(\nu_1, \nu_0, Z_{0:n})]| \\ \leq |\mathbf{d}_n(\nu', \nu_1, Z_{0:n})| + |\mathbf{E}_0^\nu[\mathbf{d}_n(\nu', \nu_1, Z_{0:n})]| + |\mathbf{d}_n(\nu_1, \nu_0, Z_{0:n}) - \mathbf{E}_0^\nu[\mathbf{d}_n(\nu_1, \nu_0, Z_{0:n})]| \\ \leq \frac{2\tilde{\varepsilon}}{m} + \sup_{\nu'' \in R(\tilde{\varepsilon})} |\mathbf{d}_n(\nu'', \nu_0, Z_{0:n}) - \mathbf{E}_0^\nu[\mathbf{d}_n(\nu'', \nu_0, Z_{0:n})]|.$$

By Inequalities (5.57) and (5.32), for any  $u$  in  $\mathbb{R}_+^*$

$$\mathbf{P}_0^\nu \left( \sup_{\nu' \in \tilde{\Theta}} |\mathbf{d}_n(\nu', \nu_0, Z_{0:n}) - \mathbf{E}_0^\nu[\mathbf{d}_n(\nu', \nu_0, Z_{0:n})]| \geq u + \frac{2\tilde{\varepsilon}}{m} \right) \\ \leq \mathbf{P}_0^\nu \left( \sup_{\nu'' \in R(\tilde{\varepsilon})} |\mathbf{d}_n(\nu'', \nu_0, Z_{0:n}) - \mathbf{E}_0^\nu[\mathbf{d}_n(\nu'', \nu_0, Z_{0:n})]| \geq u \right) \\ \leq \sum_{\nu'' \in R(\tilde{\varepsilon})} 2 \exp \left( -\frac{\beta(c_0, \alpha) m^2}{32 \|f_{\nu''} - f_{\nu_0}\|_\infty^2} n u^2 \right) \leq \sum_{\nu'' \in R(\tilde{\varepsilon})} 2 \exp \left( -\frac{\beta(c_0, \alpha) m^2}{32 \Delta(\tilde{\Theta})^2} n u^2 \right) \\ \leq \text{card}(R(\tilde{\varepsilon})) 2 \exp \left( -\frac{\beta(c_0, \alpha) m^2}{32 \Delta(\tilde{\Theta})^2} n u^2 \right).$$

By definition of the covering number, for any  $u$  in  $\mathbb{R}_+^*$ , with probability under  $\mathbf{P}_0^\nu$  larger than  $1 - e^{-u^2}$  it holds

$$\begin{aligned} & \sup_{\nu' \in \tilde{\Theta}} |d_n(\nu', \nu_0, Z_{0:n}) - \mathbf{E}_0^\nu [d_n(\nu', \nu_0, Z_{0:n})]| \\ & \leq \frac{4\sqrt{2} \Delta(\tilde{\Theta})}{m \sqrt{\beta(c_0, \alpha)}} \frac{1}{\sqrt{n}} \left( u + \sqrt{\log \left( 2N \left( \tilde{\Theta}, \|\cdot\|_\infty, \tilde{\varepsilon} \right) \right)} \right) + \frac{2\tilde{\varepsilon}}{m}. \end{aligned}$$

Moreover,

$$\begin{aligned} \frac{\frac{\zeta_0(c_0, \alpha) \sqrt{\beta(c_0, \alpha)}}{(1-r_0(c_0, \alpha)) \sqrt{2}}}{\sqrt{n}} & \leq \frac{\zeta_0(c_0, \alpha) \sqrt{\beta(c_0, \alpha)}}{(1-r_0(c_0, \alpha)) \sqrt{2}} \sqrt{\frac{\log \left( 2N \left( \tilde{\Theta}, \|\cdot\|_\infty, \tilde{\varepsilon} \right) \right)}{\log(2)}} \\ & \leq \frac{\zeta_0(c_0, \alpha) \sqrt{\beta(c_0, \alpha)}}{(1-r_0(c_0, \alpha)) \sqrt{2 \log(2)}} \sqrt{\log \left( 2N \left( \tilde{\Theta}, \|\cdot\|_\infty, \tilde{\varepsilon} \right) \right)}. \end{aligned}$$

Therefore,

$$\begin{aligned} & \frac{8\sqrt{2} \Delta(\tilde{\Theta})}{m \sqrt{\beta(c_0, \alpha)}} \frac{u + \sqrt{\log \left( 2N \left( \tilde{\Theta}, \|\cdot\|_\infty, \tilde{\varepsilon} \right) \right)}}{\sqrt{n}} + \frac{4\tilde{\varepsilon}}{m} + \frac{2 \frac{\zeta_0(c_0, \alpha)}{1-r_0(c_0, \alpha)} \frac{\Delta(\tilde{\Theta})}{m}}{n} \\ & = \frac{4\tilde{\varepsilon}}{m} + \frac{1}{\sqrt{n}} \frac{2\sqrt{2}}{m \sqrt{\beta(c_0, \alpha)}} \Delta(\tilde{\Theta}) \left( \frac{\frac{\zeta_0(c_0, \alpha) \sqrt{\beta(c_0, \alpha)}}{(1-r_0(c_0, \alpha)) \sqrt{2}}}{\sqrt{n}} + u + \sqrt{\log \left( 2N \left( \tilde{\Theta}, \|\cdot\|_\infty, \tilde{\varepsilon} \right) \right)} \right) \\ & \leq \frac{4\tilde{\varepsilon}}{m} + \frac{2\sqrt{2}}{m \sqrt{\beta(c_0, \alpha)}} \left( 1 + \frac{\zeta_0(c_0, \alpha) \sqrt{\beta(c_0, \alpha)}}{(1-r_0(c_0, \alpha)) \sqrt{2 \log(2)}} \right) \frac{\Delta(\tilde{\Theta})}{\sqrt{n}} \left( u + \sqrt{\log \left( 2N \left( \tilde{\Theta}, \|\cdot\|_\infty, \tilde{\varepsilon} \right) \right)} \right). \end{aligned}$$

By Equation (5.56), for any  $u$  in  $\mathbb{R}_+^*$ , it finally holds with probability under  $\mathbf{P}_0^\nu$  less than  $e^{-u^2}$

$$\sup_{(\nu', \nu'') \in \tilde{\Theta}^2} |F_n(\nu', \nu'', Z_{0:n}, \nu)| \geq C(c_0, \alpha, m) \left( \tilde{\varepsilon} + \Delta(\tilde{\Theta}) \frac{u + \sqrt{\log \left( 2N \left( \tilde{\Theta}, \|\cdot\|_\infty, \tilde{\varepsilon} \right) \right)}}{\sqrt{n}} \right),$$

where  $C(c_0, \alpha, m)$  is defined in Equation (5.35).

### 5.3.9 Control of the first and second kind risks

In Sections 5.3.9.1 and 5.3.9.2, we show that first and second kind risks of the test function  $\phi_n$  respectively appearing in conditions (i) and (ii) can be bounded by the same quantity. In Section 5.3.9.3, we show then how to control this common upper-bound tanks to Proposition 69.

#### 5.3.9.1 First kind risk

By definition of  $\phi_n$ , for any  $n$  in  $\mathbb{N}$

$$\mathbf{E}_0^{\nu^*} [\phi_n(Z_{0:n})] = \mathbf{P}_0^{\nu^*} \left( \sup_{\nu' \in \tilde{\Theta}_n} F_n(\nu', \nu^*, Z_{0:n}, \nu^*) > \frac{\varepsilon}{2} \right).$$

For any  $n$  in  $\mathbb{N}$ , let

$$(5.58) \quad \tilde{\Theta}_n = \Theta_n \cup \{\nu^*\}.$$

We have thus

$$(5.59) \quad \mathbf{E}_0^{\nu^*} [\phi_n(Z_{0:n})] \leq \mathbf{P}_0^{\nu^*} \left( \sup_{(\nu', \nu'') \in \tilde{\Theta}_n^2} |F_n(\nu', \nu'', Z_{0:n}, \nu^*)| > \frac{\varepsilon}{2} \right).$$

### 5.3.9.2 Second kind risk

By definition of  $\phi_n$ , for any  $n$  in  $\mathbb{N}$

$$\mathbf{E}_0^\nu [1 - \phi_n(Z_{0:n})] = \mathbf{P}_0^\nu \left( \sup_{\nu' \in \Theta_n} F_n(\nu', \nu^*, Z_{0:n}, \nu^*) \leq \frac{\varepsilon}{2} \right).$$

Since  $F_n(\nu, \nu^*, Z_{0:n}, \nu^*) \leq \sup_{\nu' \in \Theta_n} F_n(\nu', \nu^*, Z_{0:n}, \nu^*)$  for any  $\nu$  in  $\Theta_n$ , it follows

$$\mathbf{P}_0^\nu \left( \sup_{\nu' \in \Theta_n} F_n(\nu', \nu^*, Z_{0:n}, \nu^*) \leq \frac{\varepsilon}{2} \right) \leq \mathbf{P}_0^\nu \left( F_n(\nu, \nu^*, Z_{0:n}, \nu^*) \leq \frac{\varepsilon}{2} \right).$$

For any  $\nu$  in  $\Theta_n \cap A_\varepsilon$ , we have  $d^{\nu^*}(\nu^*, \nu) \geq \varepsilon$  (that is  $d^{\nu^*}(\nu, \nu^*) \leq -\varepsilon$ ), so

$$\mathbf{P}_0^\nu \left( d_n(\nu, \nu^*, Z_{0:n}) - d^{\nu^*}(\nu, \nu^*) \leq \frac{\varepsilon}{2} \right) \leq \mathbf{P}_0^\nu \left( d_n(\nu, \nu^*, Z_{0:n}) \leq \frac{\varepsilon}{2} - \varepsilon \right).$$

Since  $d^\nu(\nu, \nu^*) \geq 0$ , by Proposition 75, we have then

$$\mathbf{P}_0^\nu \left( d_n(\nu, \nu^*, z_{0:n}) \leq -\frac{\varepsilon}{2} \right) \leq \mathbf{P}_0^\nu \left( d_n(\nu, \nu^*, z_{0:n}) - d^\nu(\nu, \nu^*) \leq -\frac{\varepsilon}{2} \right).$$

Combining the last two inequalities, by the definition of  $F_n(\nu, \nu^*, Z_{0:n}, \nu)$ , we thus have, for any  $\nu$  in  $\Theta_n \cap A_\varepsilon$

$$\mathbf{E}_0^\nu [1 - \phi_n(Z_{0:n})] \leq \mathbf{P}_0^\nu \left( F_n(\nu, \nu^*, Z_{0:n}, \nu) \leq -\frac{\varepsilon}{2} \right) \leq \mathbf{P}_0^\nu \left( |F_n(\nu, \nu^*, Z_{0:n}, \nu)| \geq \frac{\varepsilon}{2} \right).$$

Hence, for any  $\nu$  in  $\Theta_n \cap A_\varepsilon$

$$\mathbf{E}_0^\nu [1 - \phi_n(Z_{0:n})] \leq \mathbf{P}_0^\nu \left( \sup_{(\nu', \nu'') \in \tilde{\Theta}_n^2} |F_n(\nu', \nu'', Z_{0:n}, \nu)| > \frac{\varepsilon}{2} \right),$$

where  $\tilde{\Theta}_n$  is defined in Equation (5.58).

### 5.3.9.3 Control of first and second kind risks : application of Proposition 69

Since the first and second kind risks are both upper-bounded by the same term, let's now prove that this term can be bounded thanks to Proposition 69.

By assumption **B6**, for any  $n$  in  $\mathbb{N}$ , the  $\varepsilon$ -covering number of the set  $\Theta_n$  is finite. Each of the set  $\Theta_n$  can thus be covered by a finite number of balls of radius  $\varepsilon$  for  $\|\cdot\|_\infty$  and is so a compact set for  $\|\cdot\|_\infty$ . Since  $\nu^*$  is a single point of  $\text{supp}(\Pi)$ , the family of sets  $(\tilde{\Theta}_n)_{n \in \mathbb{N}}$  where  $\tilde{\Theta}_n$  is defined in Equation (5.58) is also compact for  $\|\cdot\|_\infty$ . Proposition 69 can thus be

applied to  $(\tilde{\Theta}_n)_{n \in \mathbb{N}}$  and we get for any  $\nu$  in  $\tilde{\Theta}_n$ , any  $n$  in  $\mathbb{N}^*$ , any  $u$  in  $\mathbb{R}_+^*$  and any positive  $\tilde{\varepsilon}$  that with probability less than  $e^{-u^2}$  under  $\mathbf{P}_0^\nu$

$$\begin{aligned} & \sup_{(\nu', \nu'') \in \tilde{\Theta}_n^2} |F_n(\nu', \nu'', Z_{0:n}, \nu)| \\ & \geq C(c_0, \alpha, m) \left( \tilde{\varepsilon} + \Delta(\tilde{\Theta}_n) \frac{u + \sqrt{\log \left( 2N \left( \tilde{\Theta}_n, \|\cdot\|_\infty, \tilde{\varepsilon} \right) \right)}}{\sqrt{n}} \right), \end{aligned}$$

where  $C(c_0, \alpha, m)$  is defined in Equation (5.35).

For any  $\nu$  and  $\nu'$  in  $\text{supp}(\Pi)$ , we have

- by Assumption **B2**, for any  $t$  in  $[0, \alpha) \cup (1 - \alpha, 1]$

$$f_\nu(t) - f_{\nu'}(t) = 0,$$

- by Assumption **B3** and **B4**, for any  $t$  in  $[\alpha, 1 - \alpha]$

$$m - M \leq f_\nu(t) - f_{\nu'}(t) \leq M - m.$$

Hence, for any  $n$  in  $\mathbb{N}$

$$\Delta(\tilde{\Theta}_n) \leq \Delta(\text{supp}(\Pi)) \leq M - m,$$

since  $\tilde{\Theta}_n$  is a subset of  $\text{supp}(\Pi)$ .

Choose  $\tilde{\varepsilon} = \varepsilon / (8C(c_0, \alpha, m))$  and define  $u_n$ , for any  $n$  in  $\mathbb{N}$ , by

$$(5.60) \quad u_n = \frac{\varepsilon}{4C(c_0, \alpha, m)(M - m)} \sqrt{n}.$$

Since  $\log \left( 2N \left( \tilde{\Theta}_n, \|\cdot\|_\infty, \tilde{\varepsilon} \right) \right) \leq (\varepsilon / (M - m))^2 n$ , for any  $n$  in  $\mathbb{N}$ , by Inequality (5.13), we have

$$\begin{aligned} & C(c_0, \alpha, m) \left( \tilde{\varepsilon} + \Delta(\tilde{\Theta}_n) \frac{u_n + \sqrt{\log \left( 2N \left( \tilde{\Theta}_n, \|\cdot\|_\infty, \tilde{\varepsilon} \right) \right)}}{\sqrt{n}} \right) \\ & \leq C(c_0, \alpha, m) \left( 2\tilde{\varepsilon} + (M - m) \frac{u_n}{\sqrt{n}} \right) = C(c_0, \alpha, m) \left( 2 \frac{\varepsilon}{8C(c_0, \alpha, m)} + (M - m) \frac{\frac{\varepsilon}{4C(c_0, \alpha, m)(M - m)} \sqrt{n}}{\sqrt{n}} \right) \\ & \leq C(c_0, \alpha, m) \left( \frac{\varepsilon}{4C(c_0, \alpha, m)} + \frac{\varepsilon}{4C(c_0, \alpha, m)} \right) = \frac{\varepsilon}{2}. \end{aligned}$$

Thus, first and second kind risks are bounded :

$$\mathbf{E}_0^{\nu^*} [\phi_n^t(Z_{0:n})] \leq e^{-u_n^2} \quad \text{and} \quad \sup_{\nu \in \Theta_n \cap A_\varepsilon} (\mathbf{E}_0^\nu [1 - \phi_n^t(Z_{0:n})]) \leq e^{-u_n^2},$$

where  $A_\varepsilon$  is defined in Equation (5.16).

### 5.3.10 Final step of the proof of Theorem 64

On the one hand, since the sequence  $(u_n)_{n \in \mathbb{N}^*}$  defined in Equation (5.60) tends to  $\infty$  as  $n$  grows,  $e^{-u_n^2}$  tends to 0 as  $n$  grows to infinity.

On the other hand, if  $\delta$  is less than  $m \varepsilon^2 / (32 C(c_0, \alpha, m)^2 (M - m)^2)$ , then for any  $n$  in  $\mathbb{N}$

$$\frac{\delta}{m} n - u_n^2 = \frac{\delta}{m} n - \left( \frac{\varepsilon}{4 C(c_0, \alpha, m) (M - m)} \right)^2 n \leq - \frac{\varepsilon^2}{32 C(c_0, \alpha, m)^2 (M - m)^2} n$$

and thus when  $n$  grows to infinity,  $e^{-u_n^2} / b_n^\delta = e^{\frac{\delta}{m} n - u_n^2} / s_\delta$  tends to 0.

Simultaneously, by first Inequality of Assumption **B6**, we know that for any  $n$  in  $\mathbb{N}$

$$\Pi(\Theta_n^c) \leq e^{-\frac{C_\varepsilon}{m} n},$$

so that for any positive  $\delta$

$$\frac{\Pi(\Theta_n^c)}{b_n^\delta} = \frac{\Pi(\Theta_n^c)}{s_\delta e^{-\frac{\delta}{m} n}} \leq \frac{1}{s_\delta} e^{\frac{\delta - C_\varepsilon}{m} n}.$$

and if  $\delta$  is less than  $C_\varepsilon$ , the sequence  $\Pi(\Theta_n^c) / b_n^\delta$  tends to 0 as  $n$  grows.

As a conclusion, considering fixed positive  $\varepsilon$  and setting  $b_n = b_n^\delta$  where  $\delta$  is a positive real less than  $m \varepsilon^2 / (32 C(c_0, \alpha, m)^2 (M - m)^2)$  and  $C_\varepsilon$ , we proved that  $D_n$  is lower bounded by  $b_n$  and that the three terms

$$\mathbf{E}_0^{\nu^*} [\phi_n(Z_{0:n})] \quad \text{and} \quad \frac{\sup_{\nu \in A_\varepsilon \cap \Theta_n \cap \text{supp}(\Pi)} \mathbf{E}_0^\nu [1 - \phi_n(Z_{0:n})]}{b_n} \quad \text{and} \quad \frac{\Pi(\Theta_n^c)}{b_n},$$

all simultaneously tend to 0 as  $n$  grows so that requirements (i), (ii), (iii) and (iv) are satisfied.

According to Inequality (5.23), the quantity  $\mathbf{E}_0^{\nu^*} [\Pi(A_\varepsilon | Z_{0:n})]$  is bounded by the sum of this three terms and thus tends to 0.

Section 5.3.1 showing that posterior consistency of the Bayesian estimator of BPIRE is entailed by that later convergence, we thus indeed proved Theorem 64.

### 5.3.11 Proof of Theorem 63

For any  $\nu^*$  in  $\text{supp}(\Pi)$ ,  $(L_{n-k}^n)_{0 \leq k \leq n}$  and  $Z_{0:n}$  have the same distribution respectively under  $\mathbf{Q}_0^{\nu^*}$  and  $\mathbf{P}_0^{\nu^*}$ . Hence, for any measurable  $A \subset \Theta$ , the posterior distributions  $\Pi(A | L_{0:n}^n)$  defined by Equation (5.14) and  $\Pi(A | Z_{0:n})$  defined by Equation (5.15) also have the same distribution respectively under  $\mathbf{Q}_0^{\nu^*}$  and  $\mathbf{P}_0^{\nu^*}$ . For any  $\varepsilon > 0$  and for the very specific set  $A = A_\varepsilon = \{d^{\nu^*}(\nu^*, \nu) > \varepsilon\}$ , this is in particular valid.

Hence, if Assumptions of Theorem 64 hold, then, for any  $\varepsilon > 0$  and any  $\delta > 0$

$$\mathbf{Q}_0^{\nu^*} \left( \Pi \left( d^{\nu^*}(\nu^*, \nu) > \varepsilon \mid (L_{n-k}^n)_{0 \leq k \leq n} \right) \geq \delta \right) \xrightarrow{n \rightarrow +\infty} 0,$$

and Theorem 63 is thus proved.

# Appendices





## Appendix A

# Reminder on transition kernels and Markov chains

**Definition 76.** If  $(X, \mathcal{X})$  and  $(Y, \mathcal{Y})$  are two measurable spaces, then, a kernel  $N$  on  $X \times \mathcal{Y}$  is a mapping  $N : X \times \mathcal{Y} \rightarrow [0, +\infty]$  such that

- (i) for every  $x$  in  $X$ , the function  $N(x, \cdot) : A \in \mathcal{Y} \mapsto N(x, A)$  is a measure on  $\mathcal{Y}$ ,
- (ii) for every  $A$  in  $\mathcal{Y}$ , the function  $N(\cdot, A) : x \in X \mapsto N(x, A)$  is a measurable function from  $(X, \mathcal{X})$  to  $([0, +\infty], \mathcal{B}([0, +\infty]))$ .

If for all  $x$  in  $X$ , the kernel  $N$  satisfies  $N(x, Y) = 1$ , it is called a Markov kernel.

**Definition 77.** If  $(X, \mathcal{X})$ ,  $(Y, \mathcal{Y})$  and  $(Z, \mathcal{Z})$  are three measurable spaces and  $M$  a kernel on  $X \times \mathcal{Y}$  and  $N$  a kernel on  $Y \times \mathcal{Z}$ , then, the product  $MN$  of  $M$  and  $N$  on  $X \times \mathcal{Z}$  is defined for any  $x$  in  $X$  and any  $B$  in  $\mathcal{Z}$  by

$$(A.1) \quad MN(x, B) = \int_Y N(y, B)M(x, dy).$$

**Definition 78.** If  $(X, \mathcal{X})$ ,  $(Y, \mathcal{Y})$  and  $(Z, \mathcal{Z})$  are three measurable spaces and  $M$  a kernel on  $X \times \mathcal{Y}$  and  $N$  a kernel on  $Y \times \mathcal{Z}$ , then, the tensor kernel  $M \otimes N$  on  $X \times (\mathcal{Y} \otimes \mathcal{Z})$  is defined for any  $x$  in  $X$  and any  $A$  in  $\mathcal{Y}$  and any  $B$  in  $\mathcal{Z}$  by

$$(A.2) \quad M \otimes N(x, A \times B) = \int_A N(y, B)M(x, dy).$$

**Remark 79.** If  $M$  and  $N$  are both Markov, then,  $M \otimes N$  is also Markov.

Furthermore, " $\otimes$ " is associative : if  $(U, \mathcal{U})$  is an other measurable space and  $P$  a kernel on  $Z \otimes \mathcal{U}$ , then,  $(M \otimes N) \otimes P = M \otimes (N \otimes P)$ .

Hence, for any positive integer  $n$ , the  $n$ -th tensorial power of a kernel  $P$  on  $(X, \mathcal{X})$  is the kernel  $P^{\otimes n}$  defined on  $(X, \mathcal{X}^{\otimes n})$  defined for any  $x$  in  $X$  and any  $\mathcal{X}^{\otimes n}$ -measurable function  $f$  by

$$(A.3) \quad P^{\otimes n} f(x) = \int_{X^n} f(x_1, \dots, x_n)P(x, dx_1)P(x_1, dx_2) \dots P(x_{n-1}, dx_n).$$

For any  $x$  in  $X$  and any  $B$  in  $\mathcal{Z}$

$$MN(x, B) = M \otimes N(x, Y \times B).$$

**Definition 80.** If  $\nu$  is a  $\sigma$ -finite measure on  $(X, \mathcal{X})$  and  $N$  is a kernel on  $X \otimes \mathcal{Y}$ , then, the product  $\nu N$  of  $\nu$  and  $N$  is the measure on  $(Y, \mathcal{Y})$  defined for any  $B$  in  $\mathcal{Y}$  by

$$(A.4) \quad \nu N(B) = \int_X N(x, B) \nu(dx).$$

**Definition 81.** If  $\nu$  is a  $\sigma$ -finite measure on  $(X, \mathcal{X})$  and  $N$  is a kernel on  $X \otimes \mathcal{Y}$ , then, the tensor product  $\nu \otimes N$  of  $\nu$  and  $N$  is the measure on  $(X \times Y, \mathcal{X} \otimes \mathcal{Y})$  defined for any  $A$  in  $\mathcal{X}$  and any  $B$  in  $\mathcal{Y}$  by

$$(A.5) \quad \nu \otimes N(A \times B) = \int_A N(x, B) \nu(dx).$$

**Proposition 82.** Let  $P$  be a Markov kernel on a measurable space  $X \otimes \mathcal{X}$  and  $\nu$  a probability distribution on  $\mathcal{X}$ , then, an  $X$ -valued stochastic process  $(X_k)_{k \in \mathbb{N}}$  is a homogeneous Markov chain with kernel  $P$  and initial distribution  $\nu$ , if and only if, for any  $k$  in  $\mathbb{N}$ , the distribution of  $(X_0, \dots, X_k)$  is  $\nu \otimes P^{\otimes k}$ .

**Proposition 83.** If the random process  $(X_n)_{n \in \mathbb{N}}$  is a Markov chain on  $(X, \mathcal{X})$  with transition kernel  $P$  and initial distribution  $\mu$ , then, the random process  $((X_n, X_{n+1}))_{n \in \mathbb{N}}$  is also a Markov chain whose transition kernel  $\tilde{P}$  is defined for any  $x$  and  $y$  in  $X$  and any  $A$  and  $B$  in  $\mathcal{X}$  by

$$(A.6) \quad \tilde{P}((x, y), A \times B) = \delta_y \otimes P(A \times B) = \delta_y(A) P(y, B),$$

and initial distribution  $\tilde{\mu}^P$  defined for any  $A$  and  $B$  in  $\mathcal{X}$  by

$$(A.7) \quad \tilde{\mu}^P(A \times B) = \mu \otimes P(A \times B) = \int_A P(x, B) \mu(dx).$$

Furthermore, if the kernel  $P$  admits an invariant distribution  $\Pi$ , then,  $\tilde{P}$  admits also an invariant distribution  $\tilde{\Pi}^P$  defined for any  $A$  and  $B$  in  $\mathcal{X}$  by

$$(A.8) \quad \tilde{\Pi}^P(A \times B) = \Pi \otimes P(A \times B) = \int_A P(x, B) \Pi(dx)$$

*Proof.* For any  $A$  and  $B$  in  $\mathcal{X}$

$$\begin{aligned} & \tilde{\Pi}^P \tilde{P}(A \times B) \\ &= \int_{\mathcal{X}^2} \tilde{P}((x, y), A \times B) \tilde{\Pi}^P(d(x, y)) = \int_{\mathcal{X}^2} \delta_y \otimes P(A \times B) (\Pi \otimes P)(d(x, y)) \\ &= \int_{\mathcal{X}^2} \delta_y(A) P(y, B) \Pi(dx) P(x, dy) = \int_{\mathcal{X}^2} \int_{\mathcal{X}^2} \mathbb{1}_A(v) \mathbb{1}_B(w) P(y, dw) \delta_y(dv) \Pi(dx) P(x, dy) \\ &= \int_{\mathcal{X}} \left( \int_{\mathcal{X}} \mathbb{1}_A(y) \mathbb{1}_B(w) P(y, dw) \right) \Pi P(dy) = \int_{\mathcal{X}} \left( \int_{\mathcal{X}} \mathbb{1}_A(y) \mathbb{1}_B(w) P(y, dw) \right) \Pi(dy) \\ &= \Pi \otimes P(A \times B) = \tilde{\Pi}^P(A \times B). \end{aligned}$$

Hence, the probability distribution  $\tilde{\Pi}^P = \Pi \otimes P$  is invariant under  $\tilde{P}$ .  $\square$

**Remark 84.** For any  $A$  and  $B$  in  $\mathcal{X}$  and any  $n$  in  $\mathbb{N}^*$ , it follows from the definition of  $\tilde{P}$  and of the operation  $\otimes$

$$\tilde{\mu}^P \tilde{P}^n(A \times B) = \mu P^{n-1} [P^{\otimes 2}(\cdot, A \times B)].$$

**Lemma 85.** *If  $(\mathcal{X}, \mathcal{X}, \mathbb{P})$  is a discrete probability space,  $P$  a Markov kernel on  $\mathcal{X} \times \mathcal{X}$ ,  $(X_n)_{n \in \mathbb{N}}$  a Markov chain with transition kernel  $P$  under  $\mathbb{P}$  and  $(\mathcal{F}_n^X)_{n \in \mathbb{N}}$  the natural filtration (that is, for all  $n$  in  $\mathbb{N}$ ,  $\mathcal{F}_n^X = \sigma(X_0, \dots, X_n)$ ), then, for any  $n$  in  $\mathbb{N}$ , any  $A$  in  $\mathcal{F}_n$ , any  $x_n$  in  $\mathcal{X}$  such that  $\mathbb{P}(X_n = x_n, A) \neq 0$  and any  $y$  in  $\mathcal{X}$*

$$(A.9) \quad \mathbb{P}(X_{n+1} = y | X_n = x_n, A) = \mathbb{P}(X_{n+1} = y | X_n = x_n) = P(x_n, y).$$

*Proof.* Set  $n$  in  $\mathbb{N}$ ,  $A$  in  $\mathcal{F}_n$  and  $x_n$  in  $X$  such that  $\mathbb{P}(X_n = x_n, A) \neq 0$ . Since  $(X_n)_{n \in \mathbb{N}}$  is a Markov chain on discrete state space  $\mathcal{X}$  with transition kernel  $P$  under  $\mathbb{P}$ , whatever  $(x_0, \dots, x_{n-1})$  in  $\mathcal{X}^n$  such that  $\mathbb{P}((X_0, \dots, X_n) = (x_0, \dots, x_n)) \neq 0$ , we get

$$(A.10) \quad \begin{aligned} & \mathbb{P}(X_{n+1} = y, X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) \\ &= \mathbb{P}(X_{n+1} = y | X_n = x_n, \dots, X_0 = x_0) \mathbb{P}(X_n = x_n, \dots, X_0 = x_0) \\ &= P(y, x_n) \mathbb{P}(X_n = x_n, \dots, X_0 = x_0). \end{aligned}$$

Then, the event  $\{X_{n+1} = y, X_n = x_n\} \cap A$  can decompose as

$$(A.11) \quad \bigcup_{\substack{(x_0, \dots, x_{n-1}) \in \mathcal{X}^n \\ \{X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0\} \subset A}} \{X_{n+1} = y, X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0\}.$$

Thanks to Equation (A.10) and Equation (A.11), we thus finally get

$$\begin{aligned} & \mathbb{P}(X_{n+1} = y | X_n = x_n, A) \\ &= \frac{\mathbb{P}(\{X_{n+1} = y, X_n = x_n\} \cap A)}{\mathbb{P}(X_n = x_n, A)} \\ &= \frac{1}{\mathbb{P}(X_n = x_n, A)} \sum_{\substack{(x_0, \dots, x_{n-1}) \in \mathcal{X}^n \\ \{X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0\} \subset A}} \mathbb{P}(\{X_{n+1} = y, X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0\}) \\ &= \frac{1}{\mathbb{P}(X_n = x_n, A)} \sum_{\substack{(x_0, \dots, x_{n-1}) \in \mathcal{X}^n \\ \{X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0\} \subset A}} P(x_n, y) \mathbb{P}(X_n = x_n, \dots, X_0 = x_0) \\ &= \frac{1}{\mathbb{P}(X_n = x_n, A)} P(x_n, y) \mathbb{P}(X_n = x_n, A) = \mathbb{P}(X_{n+1} = y | X_n = x_n) = P(x_n, y). \end{aligned}$$

□

**Remark 86.** *Using notations and results given in Lemma 85, for any  $n$  in  $\mathbb{N}$ , any  $A$  in  $\mathcal{F}_n$  and any  $x_n$  in  $\mathcal{X}$  such that  $\mathbb{P}(X_n = x_n, A) \neq 0$  and any  $B$  in  $\mathcal{X}$*

$$\begin{aligned} \mathbb{P}(X_{n+1} \in B | X_n = x_n, A) &= \sum_{y \in B} \mathbb{P}(X_{n+1} = y | X_n = x_n, A) \\ &= \sum_{y \in B} P(x_n, y) = P(x_n, B). \end{aligned}$$



## Appendix B

# Reminder on covering numbers

Covering and bracketing numbers provide ways to measure the complexity of a metric space, or a set of functions.

In the following definitions, we consider that  $d$  is a metric on a given space  $E$ , that  $F$  is a subset of  $E$  and that  $\varepsilon$  is a scale parameter in  $(0, +\infty)$ .

**Definition 87.** An  $\varepsilon$ -net of  $F$  with respect to  $d$  is a subset  $V$  of  $F$  such that for all  $f$  in  $F$ , it exists  $v$  in  $V$  such that

$$(B.1) \quad d(f, v) \leq \varepsilon,$$

or equivalently  $V$  is an  $\varepsilon$ -net of  $F$ , if and only if

$$(B.2) \quad F \subset \bigcup_{v \in V} \{x \in E : d(x, v) \leq \varepsilon\} .$$

**Definition 88.** The covering numbers of  $F$  with respect to  $d$  is the infinite family of (possibly infinite) real numbers  $(N(F, d, \varepsilon))_{\varepsilon \in \mathbb{R}_+^*}$  where  $N(F, d, \varepsilon)$  is defined for any  $\varepsilon$  in  $\mathbb{R}_+^*$  by

$$(B.3) \quad N(F, d, \varepsilon) = \inf \{\text{card}(V) : V \text{ is an } \varepsilon\text{-net}\} .$$

The logarithm of the covering number is called the metric entropy. Unless  $F$  is finite, in which case covering number is bounded by the cardinality of  $F$ , the covering number  $N(F, d, \varepsilon)$  goes to  $\infty$  as  $n$  grows. The rate of the growth qualitatively measures the size of the space and is fundamental in Bayesian nonparametrics as well as in empirical process theory (to obtain control on the supremum of a stochastic process such as example Dudley's bound, for example in [Tal96a]) or in statistical learning theory (to obtain margin-based errors, see for example [STW99]) Covering numbers are useful to quantify somehow the complexity of a set. For example, it enables to obtain margin-based error bounds in statistical learning or to obtain control on the supremum of a stochastic process.



# Appendix C

## Reminder on stopping times

**Definition 89.** A random variable  $T : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$  is a stopping time on a filtered probability space  $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_k)_{k \in \mathbb{N}})$  if for any  $k$  in  $\mathbb{N}$ , the set  $\{T = k\}$  belongs to  $\mathcal{F}_k$ .

**Proposition 90.** Consider  $T : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$  be a stopping time on a filtered probability space  $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_k)_{k \in \mathbb{N}})$ .

(i) The subset  $\mathcal{F}_T$  of  $\mathcal{F}$  defined by

$$\mathcal{F}_T = \{A \in \mathcal{F} : \forall k \in \mathbb{N}, A \cap \{T = k\} \in \mathcal{F}_k\},$$

is a  $\sigma$ -field on  $\Omega$ .

(ii) If there exists  $i$  in  $\mathbb{N}$  such that  $T \geq i$  then

$$\mathcal{F}_i \subset \mathcal{F}_T.$$

(iii) If  $Z : \Omega \rightarrow \mathbb{R}$  is a  $\mathcal{F}_T$ -measurable random variable, for any  $k$  in  $\mathbb{N}$ , the random variables  $Z \mathbb{1}_{\{T=k\}}$  and  $Z \mathbb{1}_{\{T \geq k\}}$  are  $\mathcal{F}_k$ -measurable.

*Proof.* (i) Let us first prove that  $\mathcal{F}_T$  is a  $\sigma$ -field.

- $\Omega$  lies in  $\mathcal{F}$  and for any  $k$  in  $\mathbb{N}$ ,  $\{T = k\} \cap \Omega = \{T = k\}$  so that  $\Omega$  belongs to  $\mathcal{F}_T$ .
- If  $A$  in  $\mathcal{F}_T$ ,  $A$  belongs to  $\mathcal{F}$  and for any  $k$  in  $\mathbb{N}$ ,  $A \cap \{T = k\}$  belongs to  $\mathcal{F}_k$ . Hence, for any  $k$  in  $\mathbb{N}$ ,  $A^C \cap \{T = k\}$  decomposed as

$$A^C \cap \{T = k\} = \{T = k\} \cap (A \cap \{T = k\})^C,$$

is the intersection of  $\{T = k\}$  belonging to  $\mathcal{F}_k$  (because  $T$  is a stopping time) and of  $(A \cap \{T = k\})^C$  belonging also to  $\mathcal{F}_k$  (as the coset of a set belonging to the  $\sigma$ -field  $\mathcal{F}_k$ ) so that  $A^C \cap \{T = k\}$  belongs to  $\mathcal{F}_k$  and  $A^C$  thus belongs to  $\mathcal{F}_T$ .

- If  $(A_n)_{n \in \mathbb{N}}$  is a countable family of sets belonging to  $\mathcal{F}_T$ , then, for any  $n$  in  $\mathbb{N}$ ,  $A_n$  belongs to  $\mathcal{F}$  and for any  $k$  in  $\mathbb{N}$ ,  $A_n \cap \{T = k\}$  belongs to  $\mathcal{F}_k$ . Hence,  $\cup_{n \in \mathbb{N}} A_n$  belongs to  $\mathcal{F}$  and  $(\cup_{n \in \mathbb{N}} A_n) \cap \{T = k\} = \cup_{n \in \mathbb{N}} (A_n \cap \{T = k\})$  belongs to  $\mathcal{F}_k$  (as a countable union of sets of the  $\sigma$ -field  $\mathcal{F}_k$ ) and thus  $\cup_{n \in \mathbb{N}} A_n$  belongs to  $\mathcal{F}_T$ .

We thus proved that  $\mathcal{F}_T$  is a  $\sigma$ -field.



- (ii) Let us then prove that if  $i$  in  $\mathbb{N}$  is such that  $T \geq i$ , then,  $\mathcal{F}_i \subset \mathcal{F}_T$ .  
For any  $A$  in  $\mathcal{F}_i$  and any  $k$  in  $\mathbb{N}$ , it holds

$$A \cap \{T = k\} = A \cap \{T = k\} \cap \{T \geq i\} = \begin{cases} A \cap \emptyset = \emptyset & \text{if } k < i \\ A \cap \{T = k\} & \text{if } k \geq i \end{cases}.$$

Since  $T$  is a stopping time, for any  $k$  in  $\mathbb{N}$ ,  $\{T = k\}$  belongs to  $\mathcal{F}_k$  and since  $(\mathcal{F}_k)_{k \in \mathbb{N}}$  is a filtration, for any  $k$  in  $\mathbb{N}$ ,  $\emptyset$  belongs to  $\mathcal{F}_k$  and for any  $k \geq i$ ,  $\mathcal{F}_i$  is included in  $\mathcal{F}_k$ . For any  $k$  in  $\mathbb{N}$ ,  $A \cap \{T = k\}$  belongs to  $\mathcal{F}_k$  and  $A$  thus belongs to  $\mathcal{F}_T$  so that  $\mathcal{F}_i \subset \mathcal{F}_T$ .

- (iii) Let us then prove that for a  $\mathcal{F}_T$ -measurable random variable  $Z : \Omega \rightarrow \mathbb{R}$  and  $k$  in  $\mathbb{N}$  fixed,  $Z \mathbb{1}_{\{T=k\}}$  and  $Z \mathbb{1}_{\{T \geq k\}}$  are  $\mathcal{F}_k$ -measurable. To this end, we will use the definition of  $\mathcal{F}_k$ -measurability of a variable  $Y : (\Omega, \mathcal{F}) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , that is, for all  $B$  in  $\mathcal{B}(\mathbb{R})$

$$\{Y \in B\} = Y^{-1}(B) \in \mathcal{F}_k.$$

Fixing  $B$  in  $\mathcal{B}(\mathbb{R})$  and  $k$  in  $\mathbb{N}$ :

$$\{Z \mathbb{1}_{\{T=k\}} \in B\} = \begin{cases} \{Z \in B\} \cap \{T = k\} & \text{if } 0 \notin B \\ \{T \neq k\} \cup (\{Z \in B\} \cap \{T = k\}) & \text{if } 0 \in B \end{cases},$$

and

$$\{Z \mathbb{1}_{\{T \geq k\}} \in B\} = \begin{cases} C & \text{if } 0 \notin B \\ \{T < k\} \cup C & \text{if } 0 \in B \end{cases},$$

where  $C = \{Z \in B\} \cap \{T \geq k\} = \{Z \in B\} \cap \left( \bigcup_{j=0}^{k-1} \{Z \in B\} \cap \{T = j\} \right)^C$ .

Since  $Z$  is  $\mathcal{F}_T$ -measurable,  $\{Z \in B\}$  lies in  $\mathcal{F}_T$ .

By definition of  $\mathcal{F}_T$  itself,  $\{Z \in B\} \cap \{T = k\}$  lies in  $\mathcal{F}_k$  and since  $T$  stopping time ensures  $\{T \neq k\} = \{T = k\}^C \in \mathcal{F}_k$  and  $\mathcal{F}_k$   $\sigma$ -field. So in both cases ( $0 \in B$  or  $0 \notin B$ ),  $\{Z \mathbb{1}_{\{T=k\}} \in B\}$  belongs to  $\mathcal{F}_k$ .

By definition of  $\mathcal{F}_T$ ,  $\{Z \in B\} \cap \{T = j\}$  lies in  $\mathcal{F}_j$  for  $0 \leq j \leq k-1$  and using then the fact that  $(\mathcal{F}_j)_{j \in \mathbb{N}}$  is a filtration (so that for  $0 \leq j \leq k-1$ ,  $\mathcal{F}_j \subset \mathcal{F}_k$  and  $\mathcal{F}_k$  is a  $\sigma$ -field), we get that  $C$  belongs to  $\mathcal{F}_k$ . Since  $T$  is a stopping time (so that  $\{T = j\} \in \mathcal{F}_j$  for  $0 \leq j \leq k-1$ ) and  $(\mathcal{F}_j)_{j \in \mathbb{N}}$  is a filtration (so that  $\mathcal{F}_j \subset \mathcal{F}_k$  for  $0 \leq j \leq k-1$  and  $\mathcal{F}_k$  is a  $\sigma$ -field),  $\{T < k\} = \bigcup_{j=0}^{k-1} \{T = j\}$  lies in  $\mathcal{F}_k$  and thus  $C \cup \{T < k\}$  belongs to  $\mathcal{F}_k$ . So in both cases ( $0 \in B$  or  $0 \notin B$ ),  $\{Z \mathbb{1}_{\{T \geq k\}} \in B\}$  belongs to  $\mathcal{F}_k$ .

Hence, if  $Z : \Omega \rightarrow \mathbb{R}$  is a  $\mathcal{F}_T$ -measurable random variable and  $k \in \mathbb{N}$ , whatever  $B$  in  $\mathcal{B}(\mathbb{R})$ ,  $\{Z \mathbb{1}_{\{T=k\}} \in B\}$  and  $\{Z \mathbb{1}_{\{T \geq k\}} \in B\}$  belong to  $\mathcal{F}_k$  so that  $Z \mathbb{1}_{\{T=k\}}$  and  $Z \mathbb{1}_{\{T \geq k\}}$  are  $\mathcal{F}_k$ -measurable. □

**Lemma 91.** *If  $T : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$  is a stopping time on  $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_k)_{k \in \mathbb{N}})$  a filtered probability space and  $Y$  a random variable on  $\Omega$ , then  $\mathbb{P}$ -a.s. for any  $k$  in  $\mathbb{N}$*

$$(C.1) \quad \mathbb{E}[Y | \mathcal{F}_T] \mathbb{1}_{T=k} = \mathbb{E}[Y | \mathcal{F}_k] \mathbb{1}_{T=k} \quad \text{and} \quad \mathbb{E}[Y | \mathcal{F}_T] \mathbb{1}_{T \geq k} = \mathbb{E}[Y | \mathcal{F}_k] \mathbb{1}_{T \geq k}.$$

*Proof.* Let us fix  $k$  in  $\mathbb{N}$ .

By definition of conditional expectation given  $\mathcal{F}_k$ ,  $\mathbb{E}[Y \mathbb{1}_{\{T=k\}} | \mathcal{F}_k] = \mathbb{E}[Y | \mathcal{F}_k] \mathbb{1}_{\{T=k\}}$  and  $\mathbb{E}[Y \mathbb{1}_{\{T \geq k\}} | \mathcal{F}_k] = \mathbb{E}[Y | \mathcal{F}_k] \mathbb{1}_{\{T \geq k\}}$  are (up to a measurable set with  $\mathbb{P}$ -measure 0) the only  $\mathcal{F}_k$ -measurable random variables  $Z_1$  and  $Z_2$  that respectively satisfies for any  $B$  in  $\mathcal{F}_k$

$$\mathbb{E}[Y \mathbb{1}_{\{T=k\}} \mathbb{1}_B] = \mathbb{E}[Z_1 \mathbb{1}_B] \quad \text{and} \quad \mathbb{E}[Y \mathbb{1}_{\{T \geq k\}} \mathbb{1}_B] = \mathbb{E}[Z_2 \mathbb{1}_B] .$$

In order to prove Equation (C.1), it is thus necessary and sufficient to prove that  $\mathbb{E}[Y | \mathcal{F}_T] \mathbb{1}_{T=k}$  and  $\mathbb{E}[Y | \mathcal{F}_T] \mathbb{1}_{T \geq k}$  are  $\mathcal{F}_k$ -measurable random variable satisfying respectively for  $B$  in  $\mathcal{F}_k$

$$\mathbb{E}[Y \mathbb{1}_{\{T=k\}} \mathbb{1}_B] = \mathbb{E}[(\mathbb{E}[Y | \mathcal{F}_T] \mathbb{1}_{\{T=k\}}) \mathbb{1}_B] \quad \text{and} \quad \mathbb{E}[Y \mathbb{1}_{\{T \geq k\}} \mathbb{1}_B] = \mathbb{E}[(\mathbb{E}[Y | \mathcal{F}_T] \mathbb{1}_{\{T \geq k\}}) \mathbb{1}_B] .$$

First, by definition of conditional expectations with respect to  $\mathcal{F}_T$ ,  $\mathbb{E}[Y | \mathcal{F}_T]$  is  $\mathcal{F}_T$ -measurable and Proposition 90 applied to  $\mathbb{E}[Y | \mathcal{F}_T]$  ensures thus that  $\mathbb{E}[Y | \mathcal{F}_T] \mathbb{1}_{T=k}$  and  $\mathbb{E}[Y | \mathcal{F}_T] \mathbb{1}_{T \geq k}$  are  $\mathcal{F}_k$ -measurable.

Then, let us prove that  $B \cap \{T = k\}$  and  $B \cap \{T \geq k\}$  belong both to  $\mathcal{F}_T$ . We want to prove that  $B \cap \{T = k\} \cap \{T = i\}$  and  $B \cap \{T \geq k\} \cap \{T = i\}$  lie in  $\mathcal{F}_i$  for all  $i$  in  $\mathbb{N}$ .

Depending on the value of  $i$ , we get

$$B \cap \{T = k\} \cap \{T = i\} = \begin{cases} \emptyset & \text{if } i \neq k \\ B \cap \{T = i\} & \text{if } i = k \end{cases} ,$$

and

$$B \cap \{T \geq k\} \cap \{T = i\} = \begin{cases} \emptyset & \text{if } i < k \\ B \cap \{T = i\} & \text{if } i \geq k \end{cases} .$$

$\mathcal{F}_i$  being a  $\sigma$ -field,  $\emptyset$  always belong to  $\mathcal{F}_i$  whatever the value of  $i$ .

If  $i = k$ , then,  $B \cap \{T = i\} = B \cap \{T = k\}$  is the intersection of two sets of  $\mathcal{F}_k$  and since  $\mathcal{F}_k = \mathcal{F}_i$ ,  $B \cap \{T = i\}$  belongs to  $\mathcal{F}_i$ .

If  $i \geq k$ , then,  $\{T = i\}$  belongs to  $\mathcal{F}_i$  (since  $T$  stopping time) and  $B$  belongs also to  $\mathcal{F}_i$  (since  $B$  lies in  $\mathcal{F}_k$  and  $\mathcal{F}_k \subset \mathcal{F}_i$ ) so that  $B \cap \{T = i\}$  belongs to  $\mathcal{F}_i$ .

Hence, whatever the relative position of  $i$  and  $k$ , the sets  $B \cap \{T = k\} \cap \{T = i\}$  and  $B \cap \{T \geq k\} \cap \{T = i\}$  lie in  $\mathcal{F}_i$  and we thus proved that  $B \cap \{T = k\}$  and  $B \cap \{T \geq k\}$  both belong to  $\mathcal{F}_T$ .

Finally, since  $B \cap \{T = k\}$  and  $B \cap \{T \geq k\}$  both belong to  $\mathcal{F}_T$  so that  $\mathbb{1}_{\{T=k\}}$  and  $\mathbb{1}_{\{T \geq k\}}$  are  $\mathcal{F}_T$ -measurable and thanks to property of conditional expectation with respect to  $\mathcal{F}_T$

$$\mathbb{E}[Y \mathbb{1}_{\{T=k\}} \mathbb{1}_B] = \mathbb{E}[\mathbb{E}[Y \mathbb{1}_{\{T=k\}} \mathbb{1}_B | \mathcal{F}_T]] = \mathbb{E}[(\mathbb{E}[Y | \mathcal{F}_T] \mathbb{1}_{\{T=k\}}) \mathbb{1}_B] ,$$

as well as

$$\mathbb{E}[Y \mathbb{1}_{\{T \geq k\}} \mathbb{1}_B] = \mathbb{E}[\mathbb{E}[Y \mathbb{1}_{\{T \geq k\}} \mathbb{1}_B | \mathcal{F}_T]] = \mathbb{E}[(\mathbb{E}[Y | \mathcal{F}_T] \mathbb{1}_{\{T \geq k\}}) \mathbb{1}_B] .$$

□

A well-known classical property of conditional expectation is the so-called *tower property*: if  $\mathcal{G}_0$  and  $\mathcal{G}_1$  are two  $\sigma$ -fields such that  $\mathcal{G}_0 \subset \mathcal{G}_1 \subset \mathcal{F}$  then for any non-negative random variable  $Z$  defined on  $(\Omega, \mathcal{F}, \mathbb{P})$

$$(C.2) \quad \mathbb{E}[\mathbb{E}[Z | \mathcal{G}_0] | \mathcal{G}_1] = \mathbb{E}[\mathbb{E}[Z | \mathcal{G}_1] | \mathcal{G}_0] = \mathbb{E}[Z | \mathcal{G}_0] .$$

Defining from two already existing stopping times a new stopping time, it is possible to generalize somehow the tower property.

**Lemma 92.** Let  $T : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$  and  $S : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$  be two stopping times on  $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_k)_{k \in \mathbb{N}})$  a filtered probability space.

(i)  $T \wedge S$  is a stopping time on  $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_k)_{k \in \mathbb{N}})$  such that

$$(C.3) \quad \mathcal{F}_{T \wedge S} = \mathcal{F}_T \cap \mathcal{F}_S.$$

(ii) The sets  $\{T \leq S\}$ ,  $\{T < S\}$ ,  $\{S \leq T\}$  and  $\{S < T\}$  belong to  $\mathcal{F}_{T \wedge S}$ , for any  $A$  in  $\mathcal{F}_T$ , the sets  $A \cap \{T \leq S\}$  and  $A \cap \{T < S\}$  belong to  $\mathcal{F}_S$  and for any  $A$  in  $\mathcal{F}_S$ , the sets  $A \cap \{S \leq T\}$  and  $A \cap \{S < T\}$  belong to  $\mathcal{F}_T$ .

(iii) If  $X$  is an integrable random variable on  $(\Omega, \mathcal{F}, \mathbb{P})$  then  $\mathbb{P}$ -a.s.

$$(C.4) \quad \mathbb{E}[\mathbb{E}[X | \mathcal{F}_S] | \mathcal{F}_T] = \mathbb{E}[\mathbb{E}[X | \mathcal{F}_T] | \mathcal{F}_S] = \mathbb{E}[X | \mathcal{F}_{T \wedge S}].$$

**Remark 93.** From Equation (C.3), we derive that if  $T$  and  $S$  are two stopping times such that  $T \leq S$ , then  $\mathcal{F}_T$  is included in  $\mathcal{F}_S$ . This is the same property as (ii) but for stopping times.

*Proof.* (i) For any  $k$  in  $\mathbb{N}$ , we have decomposition

$$\{T \wedge S = k\} = (\{T = k\} \cap \{S \leq k - 1\}^c) \cup (\{S = k\} \cap \{T \leq k - 1\}^c),$$

but since  $S$  and  $T$  are stopping times,  $\{T = k\}$ ,  $\{S \leq k - 1\}$ ,  $\{S = k\}$  and  $\{T \leq k - 1\}$  all belong to  $\mathcal{F}_k$  and  $\mathcal{F}_k$  being a  $\sigma$ -field, we derive that  $\{T \wedge S = k\}$  belongs to  $\mathcal{F}_k$ . From Definition 89, we thus have that  $\mathcal{F}_{T \wedge S}$  is a stopping time. Furthermore, for any  $A$  in  $\mathcal{F}$ , since  $T \wedge S \leq T$ , for any  $k$  in  $\mathbb{N}$

$$A \cap \{T = k\} = A \cap \{T \wedge S \leq k\} \cap \{T = k\},$$

If  $A$  is in  $\mathcal{F}_{T \wedge S}$  then  $A \cap \{T \wedge S \leq k\}$  belongs to  $\mathcal{F}_k$  for any  $k$  in  $\mathbb{N}$  and since  $T$  is a stopping time,  $\{T = k\}$  also belongs to  $\mathcal{F}_k$  for any  $k$  in  $\mathbb{N}$  so that  $A \cap \{T = k\}$  belongs to  $\mathcal{F}_k$  for any  $k$  in  $\mathbb{N}$ . We thus have proved that any  $A$  in  $\mathcal{F}_{T \wedge S}$  also belongs to  $\mathcal{F}_T$  and we could have shown similarly that any  $A$  in  $\mathcal{F}_{T \wedge S}$  also belongs to  $\mathcal{F}_S$  so that

$$\mathcal{F}_{T \wedge S} \subset \mathcal{F}_T \cap \mathcal{F}_S.$$

Conversely, if  $A$  belongs to  $\mathcal{F}_T \cap \mathcal{F}_S$ , then for any  $k$  in  $\mathbb{N}$

$$A \cap \{T \wedge S = k\} = (A \cap \{T = k\} \cap \{S \leq k - 1\}^c) \cup (A \cap \{S = k\} \cap \{T \leq k - 1\}^c).$$

On the one hand, since  $A$  belongs to  $\mathcal{F}_T \cap \mathcal{F}_S$ , as well  $A \cap \{T = k\}$  as  $A \cap \{S = k\}$  lies in  $\mathcal{F}_k$ . On the other hand, since  $S$  and  $T$  are stopping times, as well  $\{S \leq k - 1\}$  as  $\{T \leq k - 1\}$  lies in  $\mathcal{F}_k$ . Since  $\mathcal{F}_k$  is a  $\sigma$ -field,  $A \cap \{T \wedge S = k\}$  lies in  $\mathcal{F}_k$ . We thus have proved that any  $A$  in  $\mathcal{F}_T \cap \mathcal{F}_S$  also belongs to  $\mathcal{F}_{T \wedge S}$  so that

$$\mathcal{F}_T \cap \mathcal{F}_S \subset \mathcal{F}_{T \wedge S}.$$

And we finally indeed proved that Identity (C.3) is true.

- (ii) We will prove (ii) for the events  $\{T \leq S\}$  and  $A \cap \{T \leq S\}$  with  $A$  in  $\mathcal{F}_T$  since other results are directly obtained switching roles of  $S$  and  $T$  or replacing strict by simple inequalities.

For any  $k$  in  $\mathbb{N}$ , we have

$$\{T \leq S\} \cap \{T = k\} = \{k \leq S\} \cap \{T = k\} = \{S \leq k - 1\}^c \cap \{T = k\},$$

and since  $\{S \leq k - 1\}$  and  $\{T = k\}$  both belong to  $\mathcal{F}_k$  (because  $S$  and  $T$  are stopping times),  $\{T \leq S\} \cap \{T = k\}$  belongs to  $\mathcal{F}_k$  for any  $k$  in  $\mathbb{N}$  and we thus proved that  $\{T \leq S\}$  belongs to  $\mathcal{F}_T$ . The same way, we get that  $\{T \leq S\}$  belongs to  $\mathcal{F}_S$  and so  $\{T \leq S\}$  belongs to  $\mathcal{F}_T \cap \mathcal{F}_S = \mathcal{F}_{T \wedge S}$ .

For any  $A$  in  $\mathcal{F}_T$ , for any  $k$  in  $\mathbb{N}$ , we have

$$A \cap \{T \leq S\} \cap \{S = k\} = A \cap \{T \leq k\} \cap \{S = k\},$$

and since  $A \cap \{T \leq k\}$  and  $\{S = k\}$  both belong to  $\mathcal{F}_k$  (because  $A$  lies in  $\mathcal{F}_T$  and because  $S$  is a stopping time),  $A \cap \{T \leq S\} \cap \{S = k\}$  belongs to  $\mathcal{F}_k$  for any  $k$  in  $\mathbb{N}$  and we thus proved that  $A \cap \{T \leq S\}$  belongs to  $\mathcal{F}_S$ .

- (iii) Consider  $Y$  a fixed random variable on  $\Omega$ .

Since  $\{T \leq S\}$  belongs to  $\mathcal{F}_{T \wedge S}$  and since for any  $A$  in  $\mathcal{F}_T$ ,  $A \cap \{T \leq S\}$  belongs to  $\mathcal{F}_S$  and  $\mathcal{F}_T$  so that it also belongs to  $\mathcal{F}_T \cap \mathcal{F}_S = \mathcal{F}_{T \wedge S}$ , we have

$$\begin{aligned} \mathbb{E} [\mathbb{E} [Y \mathbb{1}_{\{T \leq S\}} | \mathcal{F}_{T \wedge S}] \mathbb{1}_A] &= \mathbb{E} [\mathbb{E} [Y | \mathcal{F}_{T \wedge S}] \mathbb{1}_{A \cap \{T \leq S\}}] \\ &= \mathbb{E} [Y \mathbb{1}_{A \cap \{T \leq S\}}] = \mathbb{E} [\mathbb{E} [Y \mathbb{1}_{\{T \leq S\}} | \mathcal{F}_T] \mathbb{1}_A], \end{aligned}$$

which exactly means by definition of the expectation of  $Y \mathbb{1}_{\{T \leq S\}}$  conditionally to  $\mathcal{F}_T$  that  $\mathbb{P}$ -a.s.

$$(C.5) \quad \mathbb{E} [Y \mathbb{1}_{\{T \leq S\}} | \mathcal{F}_T] = \mathbb{E} [Y \mathbb{1}_{\{T \leq S\}} | \mathcal{F}_{T \wedge S}].$$

Since  $\{S < T\}$  belongs to  $\mathcal{F}_{T \wedge S}$  and since for any  $A$  in  $\mathcal{F}_S$ ,  $A \cap \{S < T\}$  belongs to  $\mathcal{F}_T$  and  $\mathcal{F}_S$  so that it also belongs to  $\mathcal{F}_T \cap \mathcal{F}_S = \mathcal{F}_{T \wedge S}$ , we have for any positive random variable  $Y$

$$\begin{aligned} \mathbb{E} [\mathbb{E} [Y \mathbb{1}_{\{S < T\}} | \mathcal{F}_{T \wedge S}] \mathbb{1}_A] &= \mathbb{E} [\mathbb{E} [Y | \mathcal{F}_{T \wedge S}] \mathbb{1}_{A \cap \{S < T\}}] \\ &= \mathbb{E} [Y \mathbb{1}_{A \cap \{S < T\}}] = \mathbb{E} [\mathbb{E} [Y \mathbb{1}_{\{S < T\}} | \mathcal{F}_S] \mathbb{1}_A], \end{aligned}$$

which exactly means by definition of the expectation of  $Y \mathbb{1}_{\{S < T\}}$  conditionally to  $\mathcal{F}_S$  that  $\mathbb{P}$ -a.s.

$$(C.6) \quad \mathbb{E} [Y \mathbb{1}_{\{S < T\}} | \mathcal{F}_S] = \mathbb{E} [Y \mathbb{1}_{\{S < T\}} | \mathcal{F}_{T \wedge S}].$$

Decomposing  $X$  as  $X \mathbb{1}_{\{T \leq S\}} + X \mathbb{1}_{\{S < T\}}$ , taking first expectation conditionally on  $\mathcal{F}_T$  and taking then expectation conditionally on  $\mathcal{F}_S$ , we have

$$\begin{aligned} (C.7) \quad \mathbb{E} [\mathbb{E} [X | \mathcal{F}_T] | \mathcal{F}_S] &= \mathbb{E} [\mathbb{E} [X \mathbb{1}_{\{T \leq S\}} + X \mathbb{1}_{\{S < T\}} | \mathcal{F}_T] | \mathcal{F}_S] \\ &= \mathbb{E} [\mathbb{E} [X \mathbb{1}_{\{T \leq S\}} | \mathcal{F}_T] | \mathcal{F}_S] + \mathbb{E} [\mathbb{E} [X \mathbb{1}_{\{S < T\}} | \mathcal{F}_T] | \mathcal{F}_S] \\ &= \mathbb{E} [\mathbb{E} [X \mathbb{1}_{\{T \leq S\}} | \mathcal{F}_T] | \mathcal{F}_S] + \mathbb{E} [\mathbb{E} [X | \mathcal{F}_T] \mathbb{1}_{\{S < T\}} | \mathcal{F}_S]. \end{aligned}$$

Particularizing on the one hand Equation (C.5) with  $Y = X$  and taking expectation conditionally on  $\mathcal{F}_S$  we have that  $\mathbb{P}$ -a.s.

$$(C.8) \quad \mathbb{E} \left[ \mathbb{E} \left[ X \mathbb{1}_{\{T \leq S\}} \mid \mathcal{F}_T \right] \mid \mathcal{F}_S \right] = \mathbb{E} \left[ \mathbb{E} \left[ X \mathbb{1}_{\{T \leq S\}} \mid \mathcal{F}_{T \wedge S} \right] \mid \mathcal{F}_S \right],$$

and particularizing on the second hand Equation (C.6) with  $Y = \mathbb{E}[X \mid \mathcal{F}_T]$ , we have that  $\mathbb{P}$ -a.s.

$$(C.9) \quad \begin{aligned} \mathbb{E} \left[ \mathbb{E} \left[ X \mathbb{1}_{\{S < T\}} \mid \mathcal{F}_T \right] \mid \mathcal{F}_S \right] &= \mathbb{E} \left[ \mathbb{E} \left[ X \mid \mathcal{F}_T \right] \mathbb{1}_{\{S < T\}} \mid \mathcal{F}_S \right] \\ &= \mathbb{E} \left[ \mathbb{E} \left[ X \mid \mathcal{F}_T \right] \mathbb{1}_{\{S < T\}} \mid \mathcal{F}_{T \wedge S} \right] = \mathbb{E} \left[ \mathbb{E} \left[ X \mathbb{1}_{\{S < T\}} \mid \mathcal{F}_T \right] \mid \mathcal{F}_{T \wedge S} \right]. \end{aligned}$$

Using the classical tower property as reminded in Equation (C.2) with  $\mathcal{G}_0 = \mathcal{F}_{T \wedge S} \subset \mathcal{G}_1 = \mathcal{F}_S$  in Equation (C.8) and with  $\mathcal{G}_0 = \mathcal{F}_{T \wedge S} \subset \mathcal{G}_1 = \mathcal{F}_T$  in Equation (C.9) and plugging the derived identities in Equation (C.7), we finally obtain

$$\begin{aligned} \mathbb{E} \left[ \mathbb{E} \left[ X \mid \mathcal{F}_T \right] \mid \mathcal{F}_S \right] &= \mathbb{E} \left[ \mathbb{E} \left[ X \mathbb{1}_{\{T \leq S\}} \mid \mathcal{F}_{T \wedge S} \right] \mid \mathcal{F}_S \right] + \mathbb{E} \left[ \mathbb{E} \left[ X \mid \mathcal{F}_T \right] \mathbb{1}_{\{S < T\}} \mid \mathcal{F}_{T \wedge S} \right] \\ &= \mathbb{E} \left[ X \mathbb{1}_{\{T \leq S\}} \mid \mathcal{F}_{T \wedge S} \right] + \mathbb{E} \left[ X \mathbb{1}_{\{S < T\}} \mid \mathcal{F}_{T \wedge S} \right] \\ &= \mathbb{E} \left[ X \mathbb{1}_{\{T \leq S\}} + X \mathbb{1}_{\{S < T\}} \mid \mathcal{F}_{T \wedge S} \right] = \mathbb{E} \left[ X \mid \mathcal{F}_{T \wedge S} \right]. \end{aligned}$$

□

# Bibliography

- [Ada08] R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electronic Journal of Probability*, 13:1000–1034, 2008.
- [AE04] O. Adelman and N. Enriquez. Random walks in random environment: what a single trajectory tells. *Israel Journal of Mathematics*, 142:205–220, 2004.
- [Akh65] N.I. Akhiezer. *The classical moment problem and some related questions in analysis*. Oliver and Boyd, Edinburgh, 1965.
- [Ali99a] S. Alili. Asymptotic behaviour for random walks in random environments. *Journal of Applied Probability*, 36(2):334–349, 1999.
- [Ali99b] S. Alili. Persistent random walks in stationary environment. *Journal of Statistical Physics*, 94(3–4):469–494, 1999.
- [ALM15] P. Andreoletti, D. Loukianova, and C. Matias. Hidden Markov model for parameter estimation of a random walk in a Markov environment. *ESAIM. Probability and Statistics*, 19:605–625, 2015.
- [AMJR12] A. Alemany, A. Mossa, I. Junier, and F. Ritort. Experimental free-energy measurements of kinetic molecular states using fluctuation theorems. *Nat Phys*, 8(9):688–694, 2012.
- [AS11] N. Alon and J. H. Spencer. *The Probabilistic Method*. John Wiley and Sons, Ltd, 2011.
- [Bar88] A. Barron. The exponential convergence of posterior probabilities with implications for bayes estimators of density functions. *Technical report, Department of Statistics, University of Illinois, Champaign, Illinois*, 1988.
- [BBC<sup>+</sup>06] V. Baldazzi, S. Bradde, S. Cocco, E. Marinari, and R. Manasson. Inference of DNA sequences from mechanical unzipping: an ideal-case study. *Phys. Rev. Lett.*, 96, 2006.
- [BBC<sup>+</sup>07] V. Baldazzi, S. Bradde, S. Cocco, E. Marinari, and R. Manasson. Inferring DNA sequences from mechanical unzipping data: the large-bandwidth case. *Phys. Rev. E*, 75, 2007.
- [BK10] K. Bertin and N. Klutchnikoff. Minimax properties of beta kernel density estimators. *Journal of Statistical Planning and Inference*, 141(7):2287–2297, 2010.

- [BK14] K. Bertin and N. Klutchnikoff. Adaptive estimation of a density function using beta kernels. *ESAIM. Probability and Statistics*, 18:400–417, 2014.
- [BLM13] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.
- [BR03] T. Bouezmarni and J.M. Rolin. Consistency of the beta kernel density function estimator. *Canadian Journal of Statistics*, 31(1):89–98, 2003.
- [BSW99] A. Barron, M. J. Schervish, and L. Wasserman. The consistency of posterior distributions in nonparametric problems. *The Annals of Statistics*, 27(2):536–561, 04 1999.
- [CFL<sup>+</sup>14] F. Comets, M. Falconnet, O. Loukianov, D. Loukianova, and C. Matias. Maximum likelihood estimator consistency for ballistic random walk in a parametric random environment. *Stochastic Processes and their Applications*, 124(1):268–288, 2014.
- [CFLL16] F. Comets, M. Falconnet, O. Loukianov, and D. Loukianova. Maximum likelihood estimator consistency for recurrent random walk in a parametric random environment with finite support. *Stochastic Processes and their Applications*, 126(11):3578–3604, 2016.
- [Che67] A.A. Chernov. Replication of a multicomponent chain by the lightning mechanism. *Biofizika*, 12:297–301, 1967.
- [Che99] S.X. Chen. Beta kernel estimators for density functions. *Computational Statistics and Data Analysis*, 31(2):131–145, 1999.
- [DF86] P. Diaconis and D. Freedman. On the consistency of bayes estimates. *The Annals of Statistics*, 14(1):1–26, 03 1986.
- [DG15] J. Dedecker and S. Gouëzel. Subgaussian concentration inequalities for geometrically ergodic Markov chains. *Electronic Communications in Probability*, 20:no. 64, 12, 2015.
- [DL18] R. Diel and M. Lerasle. Non parametric estimation for random walks in random environment. *Stochastic Processes and their Applications*, 128(1):132–155, 2018.
- [DMPS18] R. Douc, E. Moulines, P. Priouret, and P. Soulier. *Markov chains*. Springer, 2018.
- [Doo49] J.L. Doob. Application of the theory of martingales. In *Actes du Colloque International Le Calcul des Probabilités et ses applications*, pages 23–37. Paris CNRS, 1949. (Doob’s original paper can be found at the end of a historical account entitled “Doob at Lyon”, currently available at: [www.jehps.net/juin2009/Locker.pdf](http://www.jehps.net/juin2009/Locker.pdf)).
- [DOR19] R. Douc, J. Olsson, and F. Roueff. Posterior consistency for partially observed markov models. *Stochastic Processes and their Applications*, 2019.

- [DP09] D.P. Dubhashi and A. Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, 2009.
- [Dud02] R. M. Dudley. *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2002.
- [ESZ09] N. Enriquez, C. Sabot, and O. Zindy. Limit laws for transient random walks in random environment on  $\mathbb{Z}$ . *Annales de l'Institut Fourier*, 59:2469–2508, 2009.
- [FGL14] M. Falconnet, A. Gloter, and D. Loukianova. Maximum likelihood estimation in the context of a sub-ballistic random walk in a parametric random environment. *Mathematical Methods of Statistics*, 23(3):159–175, 2014.
- [FLM14] M. Falconnet, D. Loukianova, and C. Matias. Asymptotic normality and efficiency of the maximum likelihood estimator for the parameter of a ballistic random walk in a random environment. *Mathematical Methods of Statistics*, 23(1):1–19, 2014.
- [Fre63] D.A. Freedman. On the asymptotic behavior of bayes estimates in the discrete case. *Ann. Math. Statist.*, 34(4):1386–1403, 12 1963.
- [Fre65] D.A. Freedman. On the asymptotic behavior of bayes estimates in the discrete case ii. *Ann. Math. Statist.*, 36(2):454–456, 04 1965.
- [GGR99] S. Ghosal, J. K. Ghosh, and R. V. Ramamoorthi. Posterior consistency of dirichlet mixtures in density estimation. *The Annals of Statistics*, 27(1):143–158, 03 1999.
- [GGvdV00] S. Ghosal, J.K. Ghosh, and A. van der Vaart. Convergence rates of posterior distributions. *The Annals of Statistics*, 28(2):500–531, 04 2000.
- [GL08] A. Goldenshluger and O. Lepski. Universal pointwise selection rule in multivariate function estimation. *Bernoulli*, 14(4):1150–1190, 2008.
- [Gol83] A. O. Golosov. Limit distributions for random walks in a random environment. *Doklady Akademii Nauk SSSR*, 271(1):25–29, 1983.
- [GPS10] N. Gantert, Y. Peres, and Z. Shi. The infinite valley for a recurrent random walk in random environment. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 46(2):525–536, 05 2010.
- [GR03] J. K. Ghosh and R. V. Ramamoorthi. *Bayesian nonparametrics*. Springer Series in Statistics. Springer-Verlag, New York, 2003.
- [GS02] N. Gantert and Z. Shi. Many visits to a single site by a transient random walk in random environment. *Stochastic Processes and their Applications*, 99(2):159 – 176, 2002.
- [GvdV07a] S. Ghosal and A. van der Vaart. Convergence rates of posterior distributions for noniid observations. *Annals of Statistics*, 35(1):192–223, 08 2007.
- [GvdV07b] S. Ghosal and A. van der Vaart. Posterior convergence rates of dirichlet mixtures at smooth densities. *The Annals of Statistics*, 35(2):697–723, 04 2007.



- [GvdV17] S. Ghosal and A. van der Vaart. *Fundamentals of Nonparametric Bayesian Inference*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2017.
- [GW75] F. Galton and H. W. Watson. On the probability of the extinction of families. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 4:138–144, 1875.
- [HFR09] J. M. Huguët, N. Forns, and F. Ritort. Statistical properties of metastable intermediates in dna unzipping. *Phys. Rev. Lett.*, 103(24):248106, 2009.
- [HLM19] A. Havet, M. Lerasle, and É. Moulines. Density estimation for rwre. *Mathematical Methods of Statistics*, 28(1):18–38, Jan 2019.
- [Kes86] H. Kesten. The limit distribution of sinai’s random walk in random environment. *Physica A: Statistical Mechanics and its Applications*, 138(1):299 – 309, 1986.
- [KKS75] H. Kesten, M. Kozlov, and F. Spitzer. A limit law for random walk in a random environment. *Compositio Mathematica*, 30(2):145–168, 1975.
- [KSJW02] S. J. Koch, A. Shundrovsky, B. C. Jantzen, and M. D. Wang. Probing protein-dna interactions by unzipping a single dna double helix. *Biophysical Journal*, 83(2):1098–1105, 2002.
- [LLV18] S. Le Corff, M. Lerasle, and E. Vernet. A Bayesian nonparametric approach for generalized Bradley-Terry models in random environment. working paper or preprint, 2018.
- [LN02] D.K. Lubensky and D.R. Nelson. Single molecule statistics and the polynucleotide unzipping transition. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 65:031917, 04 2002.
- [Mau79] B. Maurey. Construction de suites symétriques. *Comptes Rendus Hebdomadaires des Séances de l’Académie des Sciences. Séries A et B*, 288(14):679–681, 1979.
- [McD89] C. McDiarmid. On the method of bounded differences. In *Surveys in Combinatorics, 1989: Invited Papers at the Twelfth British Combinatorial Conference*, volume 141 of *London Mathematical Society Lecture Note Series*, page 148–188. Cambridge University Press, 1989.
- [Mna08] R. Mnatsakanov. Hausdorff moment problem:reconstruction of probability density functions. *Statistics and probability letters*, 78(13):1869–1877, 2008.
- [MP07] P. Massart and J. Picard. *Concentration inequalities and model selection: Ecole d’Eté de Probabilités de Saint-Flour XXXIII - 2003*. Number 1896 in *Ecole d’Eté de Probabilités de Saint-Flour*. Springer-Verlag, 2007.
- [MR03] R. Mnatsakanov and F. H. Ruymgaart. Some properties of moment-empirical CDF’s with application to some inverse estimation problems. *Mathematical Methods of Statistics*, 12(4):478–495, 2003.
- [MS86] V.D. Milman and G. Schechtman. *Asymptotic Theory of Finite Dimensional Normed Spaces*. Springer-Verlag, 1986.

- [MWRZ04] E. Mayer-Wolf, A. Roitershtein, and O. Zeitouni. Limit theorems for one-dimensional transient random walks in markov environments. *Annales de l'I.H.P. Probabilités et statistiques*, 40(5):635–659, 2004.
- [Pau15] D. Paulin. Concentration inequalities for markov chains by marton couplings and spectral methods. *Electronic Journal of Probability*, 20:1–32, 2015.
- [PZ09] J. Peterson and O. Zeitouni. Quenched limits for transient, zero speed one-dimensional random walk in random environment. *The Annals of Probability*, 37(1):143–188, 01 2009.
- [Sch65] L. Schwartz. On bayes procedures. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 4(1):10–26, 1965.
- [Sin82] Y. G. Sinai. The limit behavior of a one-dimensional random walk in a random environment. *Akademiya Nauk SSSR. Teoriya Veroyatnostei i ee Primeneniya*, 27(2):247–258, 1982.
- [Sol75] F. Solomon. Random walks in a random environment. *The Annals of Probability*, 3(1):1–31, 1975.
- [STW99] J. Shawe-Taylor and R.C. Williamson. Generalization performance of classifiers in terms of observed covering numbers. In *Computational Learning Theory*, pages 274–285. Springer Berlin Heidelberg, 1999.
- [SvdV01] G. Subhashis and A.W. van der Vaart. Entropies and rates of convergence for maximum likelihood and bayes estimation for mixtures of normal densities. *The Annals of Statistics*, 29(5):1233–1263, 10 2001.
- [Tal95] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l’IHÉS*, 81:73–205, 1995.
- [Tal96a] M. Talagrand. Majorizing measures: the generic chaining. *The Annals of Probability*, 24(3):1049–1103, 07 1996.
- [Tal96b] M. Talagrand. A new look at independence. *The Annals of Probability*, 24(1):1–34, 01 1996.
- [Tem72] D.E. Temkin. One-dimensional random walks in a two-component chain. *Soviet Mathematics Doklady*, 13(4):1172–1176, 1972.
- [Tsy09] A.B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009.
- [Ver15] E. Vernet. Posterior consistency for nonparametric hidden markov models with finite state space. *Electronic Journal of Statistics*, 9(1):717–752, 2015.
- [Yur76] V.V. Yurinskii. Exponential inequalities for sums of random vectors. *Journal of Multivariate Analysis*, 6(4):473 – 499, 1976.
- [Zei12] O. Zeitouni. Random walks in random environment. In *Computational complexity. Vols. 1–6*, pages 2564–2577. Springer, New York, 2012.





**Titre :** Estimation de la loi du milieu d'une marche aléatoire en milieu aléatoire

**Mots Clefs :** Milieu aléatoire, Chaînes de Markov, Statistiques bayésiennes, Estimation non-paramétrique

**Résumé :** Introduit dans les années 1960, le modèle de la marche aléatoire en milieu aléatoire i.i.d. sur les entiers relatifs (ou MAMA) a récemment été l'objet d'un regain d'intérêt dans la communauté statistique. Divers travaux se sont en particulier intéressés à la question de l'estimation de la loi du milieu à partir de l'observation d'une unique trajectoire de la MAMA. Cette thèse s'inscrit dans cette dynamique. Dans un premier temps, nous considérons le problème d'estimation d'un point de vue fréquentiste. Lorsque la MAMA est transiente à droite ou récurrente, nous construisons le premier estimateur non paramétrique de la densité de la loi du milieu et obtenons une majoration du risque associé mesuré en norme infinie. Dans un deuxième temps, nous envisageons le problème d'estimation sous un angle Bayésien. Lorsque la MAMA est transiente à droite, nous démontrons la consistance à posteriori de l'estimateur Bayésien de la loi du milieu. La principale difficulté mathématique de la thèse a été l'élaboration des outils nécessaires à la preuve du résultat de consistance bayésienne. Nous démontrons pour cela une version quantitative de l'inégalité de concentration de type Mac Diarmid pour chaînes de Markov. Nous étudions également le temps de retour en 0 d'un processus de branchement en milieu aléatoire avec immigration. Nous montrons l'existence d'un moment exponentiel fini uniformément valable sur une classe de processus de branchement en milieu aléatoire. Le processus de branchement en milieu aléatoire constituant une chaîne de Markov, ce résultat permet alors d'explicitier la dépendance des constantes de l'inégalité de concentration en fonction des caractéristiques de ce processus.

**Title :** Estimation of the environment distribution of a random walk in random environment

**Keys words :** Random environment, Markov chains, Bayesian statistics, Non-parametric estimation

**Abstract :** Introduced in the 1960s, the model of random walk in i.i.d. environment on integers (or RWRE) raised only recently interest in the statistical community. Various works have in particular focused on the estimation of the environment distribution from a single trajectory of the RWRE. This thesis extends the advances made in those works and offers new approaches to the problem. First, we consider the estimation problem from a frequentist point of view. When the RWRE is transient to the right or recurrent, we build the first non-parametric estimator of the density of the environment distribution and obtain an upper-bound of the associated risk in infinite norm. Then, we consider the estimation problem from a Bayesian perspective. When the RWRE is transient to the right, we prove the posterior consistency of the Bayesian estimator of the environment distribution. The main difficulty of the thesis was to develop the tools necessary to the proof of Bayesian consistency. For this purpose, we demonstrate a quantitative version of a Mac Diarmid's type concentration inequality for Markov chains. We also study the return time to 0 of a branching process with immigration in random environment (or BPIRE). We show the existence of a finite exponential moment uniformly valid on a class of BPIRE. The BPIRE being a Markov chain, this result enables then to make explicit the dependence of the constants of the concentration inequality with respect to the characteristics of the BPIRE.

