

Mémoire de Magistère

*Vers une analyse statistique
des marches aléatoires en milieu aléatoire*

Antoine Havet-Morel



Département
de Mathématiques
d'Orsay



Un mémoire présenté en vue de l'obtention du diplôme de
Magistère de Mathématiques

1^{er} novembre 2016

Table des matières

1	Cursus au sein du Magistère	3
1.1	Année de Magistère 1	4
1.2	Année de Magistère 2	5
1.3	Année de Magistère 3	6
2	Présentation d'un domaine de recherche : <i>Statistique des marches aléatoires en milieu aléatoire</i>	7
2.1	Historique et présentation du modèle	8
2.2	Le problème de l'estimation du milieu	9
2.3	Résultats probabilistes sur la MAMA	10
2.4	Le problème de l'estimation de la loi du milieu	11
2.4.1	Le processus des sauts à gauche	12
2.4.2	Le cas paramétrique	14
2.4.3	Le cas non paramétrique	14
2.5	Perspectives et questions ouvertes	15
3	TER de Master 1	17
	Préambule	18
3.1	Introduction	19
3.1.1	Modélisation de la situation	19
3.1.2	Position du problème	19
3.2	La gestion de l'erreur stochastique	21
3.2.1	Le contrôle de l'erreur stochastique	21
3.2.1.1	L'inégalité de Mac Diarmid	21
3.2.1.2	Majoration de l'erreur stochastique	23
3.2.2	Le coefficient d'éclatement et la dimension de Vapnik-Chervonenkis	27
3.2.2.1	Le lemme de Sauer	28
3.2.2.2	Quelques calculs simple de VC-dimensions	32
3.3	Conclusion	34
4	Rapport de stage de Magistère	36
	Remerciements	37
4.1	Introduction	37
4.2	Développement	39
4.2.1	Présentation de l'Institut Max Planck pour la recherche démographique	39
4.2.1.1	La société Max Planck	39

4.2.1.2	L’Institut Max Planck pour la recherche démographique	40
4.2.2	Modélisation statistique	41
4.2.2.1	L’analyse de survie	41
4.2.2.2	Travaux réalisés	45
4.2.3	“29 th International Workshop on Statistical Modelling”	49
4.3	Conclusion	50
5	TER de Master 2 Agrégation	51
	Préambule	52
5.1	Introduction	53
5.1.1	Modélisation de la situation	53
5.1.2	Position du problème	55
5.2	La situation générale	56
5.3	La situation à deux types d’habitat	59
5.4	La situation avec un cycle de puits tous identiques	61
5.5	Conclusion	64
6	Mémoire de M2 Recherche	65
	Préambule	66
6.1	Probabilistic analysis of the RWRE	67
6.2	Combinatorial results	69
6.3	"Left jumps" process	72
6.4	Links with branching processes in random environment	75
6.4.1	Reminder on BPRE	75
6.4.2	Probabilistic and statistical properties of BPRE	77
6.4.3	Existence of a limiting distribution for the BPRE	80
6.4.4	Consistency	86
6.5	The parametric setting	89
6.5.1	The particular case of parametric ballistic RWRE	90
6.5.1.1	Consistency of $\widehat{\theta}_n$ for parametric ballistic RWRE	90
6.5.1.2	Asymptotic normality of parametric ballistic RWRE	91
6.5.2	The case of parametric sub-ballistic RWRE	97
6.5.3	Examples of parametric settings	100
6.5.4	Numerical simulations and effective estimation	102
6.5.4.1	Numerical simulations	102
6.5.4.2	Effective estimation	105
6.6	Conclusion	107

Chapitre 1

Cursus au sein du Magistère

Avant de décrire mon cursus au sein du Magistère de mathématiques d'Orsay, je juge indispensable d'expliquer un cours instant les raisons qui m'ont conduit à suivre ce parcours de formation pour et par la recherche en remontant le fil du temps un peu plus amont.

Ayant toujours eu un certain intérêt pour les Mathématiques, je ne me suis vraiment rendu compte de ce dernier qu'à l'occasion du stage en entreprise obligatoire de classe de Seconde. N'ayant pas pu effectuer ce stage dans un domaine étroitement lié aux Mathématiques, j'ai tout de même nourri le souhait d'orienter mon parcours vers des thématiques en lien avec les Mathématiques. C'est ainsi qu'après un baccalauréat général voie S spécialité Mathématiques et désirant a priori faire de la recherche en Mathématiques mais ignorant toutefois dans quel cadre, j'ai suivi les formations dispensées en classes préparatoires MPSI et MP à l'occasion desquelles j'ai pris connaissance de l'existence du Magistère de Mathématiques d'Orsay.

Après ces deux années de formation, toujours intéressé par la recherche en Mathématiques mais ayant obtenu des résultats plutôt satisfaisant à nombre de concours d'écoles d'ingénieurs, j'ai beaucoup hésité entre ces deux options : devant ce doute et suite à de nombreux conseils, j'ai jugé raisonnable de refaire une année en classe préparatoire MP pour mûrir mon choix.

Après cette année supplémentaire, mon envie de recevoir une formation en mathématiques fondamentales avait mûri et je posais donc ma candidature pour l'entrée au Magistère de Mathématiques d'Orsay en dépit de mes admissions au concours d'entrée de plusieurs écoles d'ingénieurs : c'est ainsi qu'à commencé mon parcours au sein du Magistère.

1.1 Année de Magistère 1

L'année de Magistère 1 a été pour moi extrêmement riche d'enseignements.

D'une part, cette année constituait la première où je me retrouvais nettement plus autonome quant à la gestion de mon temps et de mon travail.

Cette autonomie, d'abord difficile à gérer, m'a permis de mettre à l'épreuve la part de liberté me semblant être un élément déterminant du travail de chercheur et de commencer à apprendre la gestion de cette autonomie.

D'autre part, cette année dont la variété des enseignements était importante m'a permis de préciser quels domaines des Mathématiques m'intéresseraient plus ou moins.

En effet, ayant suivi au premier semestre les cours obligatoires intitulés "Algèbre", "Topologie et calcul différentiel", "Algorithmique" et "Intégration de Lebesgue" et le cours supplémentaire pour le Magistère "Mathématiques et biologie", j'ai pu prendre conscience à la fois que la vue de l'Algèbre que j'avais eu jusqu'alors était extrêmement partielle et que je ne souhaitais désormais plus m'y consacrer mais aussi que je ne souhaitais à priori pas faire des Mathématiques de façon aussi appliquées que cela l'était dans le cours "Mathématiques et biologie".

Les cours du second semestre suivis, à savoir les cours obligatoire intitulés "Équations différentielles", "Analyse de Fourier", "Probabilités" et "Fonctions

holomorphes", le cours supplémentaire choisi pour le Magistère "Informatique théorique" et le cours de Magistère "Topologie générale" (dispensé par Emmanuel Ullmo)m'ont quant à eux permis d'approfondir mes connaissances dans certains domaines qui sont entre autres d'usage constant dans le domaine des Probabilités que j'ai découvert à cette occasion et par lequel j'ai été particulièrement intéressé...

Finalement, ayant appris qu'il était possible d'effectuer l'apprentissage hors murs en deuxième année de Magistère et surtout jugeant peu intéressant d'effectuer mon apprentissage hors mur dans un laboratoire ou une entreprise d'un domaine quelconque (sans motivation qui pourrait m'aider quant à mon orientation), j'ai préféré repousser cet élément essentiel du parcours de Magistérien à l'année de Magistère 2.

1.2 Année de Magistère 2

L'année de Magistère 2 a été l'occasion de préciser les intérêts éveillés en première année.

Au premier semestre de la seconde année de Magistère, partant de mon attirance toute nouvelle pour les Probabilités, j'ai profité du cours d'approfondissement en Probabilités pour éventuellement confirmer cet intérêt. Lors de ce premier semestre, j'ai par ailleurs suivi le cours intitulé "Mathématiques générales I - Analyse I" qui m'a permis de renforcer mes connaissances générales aussi bien en Algèbre (avec la découverte de la théorie des représentations et des produits semi-directs de groupes) qu'en Analyse (découverte de l'intégrale de Kurzweil -Henstock et approfondissements en théorie de la mesure). J'ai par ailleurs suivi le cours de Magistère (dispensé par Frédéric Paulin) intitulé "Introduction à la théorie spectrale et à l'analyse harmonique" qui m'a permis de découvrir la théorie spectrale.

Mon vif intérêt pour les probabilités ayant été effectivement confirmé lors du premier semestre, j'ai profité du second semestre pour faire la découverte des Statistiques par le biais du cours du même nom et afin d'en avoir un aperçu plus applicatif, j'ai également suivi le cours "MAO Probabilités et Statistiques". J'y ai découvert les Statistiques sous un autre jour que celui répandu et en ai été enchanté! Dans la même optique qu'au premier semestre, j'ai poursuivi la consolidation de mes connaissances avec le cours "Mathématiques générales II - Analyse II". Le cours de Magistère (dispensé par Frédéric Bourgeois) intitulé "Introduction à la théorie des systèmes dynamiques" m'a donné des connaissances de base utiles sur les variétés ainsi qu'un aperçu des principaux aspects des problèmes de systèmes dynamiques. Afin d'investir les connaissances acquises dans les cours de Probabilités et Statistiques, j'ai demandé à être encadré pour le T.E.R. sur les fondements de l'apprentissage statistique par Christophe Giraud.

Finalement, ayant décidé à l'issue de cette deuxième année de m'orienter vers les Probabilités et les Statistiques, j'ai voulu effectuer mon apprentissage hors murs dans un organisme de recherche étranger qui appliquait les statistiques et probabilités à un domaine extérieur aux Mathématiques. J'ai ainsi passé 6

semaines sous la direction de Jutta Gampe dans l'équipe de démographie statistique de l'Institut Max Planck pour la recherche démographique à Rostock (Allemagne). Au cours de ce mois et demi, j'ai découvert la branche des Statistiques appelée *Analyse de survie* et après des travaux de modélisation, j'ai pu développer mes compétences avec le logiciel R en réalisant des simulations numériques et réaliser l'analyse statistique de données démographiques. J'ai par ailleurs eu la chance de me voir proposer et offrir par l'Institut Max Planck la participation au séminaire "International Workshop for Statistical Modelling" au cours duquel des utilisations très nombreuses et très variées des mathématiques pour la résolution de problèmes pratiques ont été présentées.

1.3 Année de Magistère 3

La dernière année de Magistère a été pour moi l'occasion de mûrir et de définitivement valider le choix d'orientation vers le domaine des Probabilités et Statistiques qui m'était apparu s'imposer lors de la deuxième année et de me spécialiser plus particulièrement aux Statistiques et Probabilités appliquées.

Cette dernière année s'est en effet déroulée dans le cadre du Master 2 Recherche "Mathématique de l'Aléatoire" de l'Université Paris-Saclay au cours duquel j'ai pu renforcer mes connaissances mais surtout découvrir de nouveaux domaines aussi bien en Statistiques qu'en Probabilités appliquées comme théoriques au cours des 2 périodes de cours. Pour me permettre de choisir le plus justement le domaine vers lequel m'orienter tout en ayant un aperçu des divers spécialités qui s'offraient et élargir ma culture, j'ai suivi plus de cours que normalement requis.

Intrigué et intéressé par l'apparition récurrente de matrices aléatoires dans divers domaines des Statistiques, j'ai décidé de suivre et de valider en tant que cours de M2 supplémentaire pour le Magistère le cours intitulé "Matrices aléatoires" (dispensé par Edouard Maurel-Segala) qui proposait une introduction au domaine de recherche des matrices aléatoire en présentant divers résultats de convergence sur le spectre de ces matrices (théorèmes de Wigner).

Cette troisième année de formation par et pour la recherche m'a ainsi donné des éléments importants pour le choix du domaine de mon stage de recherche de M2 effectué sous la direction d'Éric Moulines et Matthieu Lerasle au CMAP de l'École Polytechnique qui devient désormais le sujet d'une thèse de doctorat en Mathématiques débutant au 1 octobre 2016.

Chapitre 2

Présentation d'un domaine
de recherche :

*Statistique des marches
aléatoires en milieu
aléatoire*

Cette partie du mémoire a pour dessein de présenter le domaine de recherche qui se consacre aux marches aléatoires en milieu aléatoire (ce qui s'abrègera désormais MAMA) sur \mathbb{Z} et plus particulièrement à l'estimation de la loi du milieu d'une MAMA sur \mathbb{Z} mais elle ne saurait être un état de l'art dans ce domaine. L'estimation de la loi du milieu d'une MAMA est le domaine de recherche sur lequel s'est concentré mon stage de M2 dans un cadre paramétrique et autour duquel doit s'organiser mon travail de thèse de doctorat en mathématiques débutant le 1 octobre 2016 au CMAP de l'École Polytechnique sous la direction d'Éric Moulines et Matthieu Lerasle.

2.1 Historique et présentation du modèle

Historique

La marche aléatoire en milieu aléatoire (MAMA) est un modèle introduit en 1967 par A.A. Chernov ([Che67]) et en 1972 par D.E. Temkin ([Tem72]).

En biologie, Chernov fait intervenir les MAMA dans la modélisation de procédures de dégrafage des deux brins d'une molécule d'ADN, l'environnement aléatoire étant celui constitué par les quatre bases nucléiques A,C,G et T et les forces assurant leur appariement.

En physique, Temkin utilise le modèle pour étudier la cinétique des transitions de phase dans les alliages de métaux. C'est ce champs d'application qui est d'ailleurs à l'origine d'une partie des terminologies consacrées à l'étude des milieux aléatoires.

Présentation du modèle

Soit ν une mesure de probabilité sur $]0, 1[$ appelée *loi de l'environnement* (ou du milieu).

L'ensemble $E =]0, 1[^\mathbb{Z}$ appelé *ensemble des environnements* de la MAMA est muni de la tribu engendrée par les cylindres $\mathcal{E} = \mathcal{B}(]0, 1[)^{\otimes \mathbb{Z}}$ et on y définit la probabilité $\mathbb{P}^\nu = \nu^{\otimes \mathbb{Z}}$.

On note $(X_n)_{n \in \mathbb{N}}$ le processus canonique sur l'espace canonique $S = \mathbb{Z}^\mathbb{N}$ muni de la tribu $\mathcal{S} = (2^\mathbb{Z})^{\otimes \mathbb{N}}$ engendrée par les cylindres :

$$\forall n \in \mathbb{N}, X_n : s = (s_k)_{k \in \mathbb{N}} \in S \mapsto X_n(s) = s_n \in \mathbb{Z}.$$

On peut alors définir une famille de probabilités $(\mathbf{P}_{x,\omega})_{(x,\omega) \in \mathbb{Z} \times E}$ sur \mathcal{S} appelées lois *trempees* (traduit du terme anglais *annealed* rappelant l'origine métallurgique du problème) par :

$$\forall (x, \omega, n) \in \mathbb{Z} \times E \times \mathbb{N}, \quad \begin{cases} \mathbf{P}_{x,\omega}(X_0 = x) = 1 \\ \mathbf{P}_{x,\omega}(X_{n+1} = X_n + 1 | X_0, \dots, X_n) = \omega_{X_n} \\ \mathbf{P}_{x,\omega}(X_{n+1} = X_n - 1 | X_0, \dots, X_n) = 1 - \omega_{X_n}. \end{cases}$$

En d'autres mots, pour (x, ω) fixé dans $\mathbb{Z} \times E$, la loi $\mathbf{P}_{x,\omega}$ est l'unique probabilité sur S sous laquelle $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov homogène partant de x dont le noyau de transition Q_ω est donné par :

$$Q_\omega(d, f) = \omega_d \mathbb{1}_{\{f=d+1\}} + (1 - \omega_d) \mathbb{1}_{\{f=d-1\}}.$$

Définition 1

Pour (x, ω) fixé dans $\mathbb{Z} \times E$, la marche aléatoire en milieu ω partant de x est le couple constitué du processus canonique $(X_n)_{n \in \mathbb{N}}$ et de la probabilité $\mathbf{P}_{x, \omega}$.

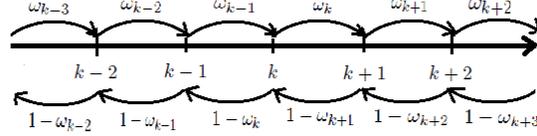


FIGURE 2.1 – Transitions de la marche aléatoire en environnement fixé

L'application $\omega \in E \mapsto \mathbf{P}_{x, \omega}(B) \in [0, 1]$ étant \mathcal{E} -mesurable pour tout $B \in \mathcal{S}$, on définit une nouvelle famille de lois de probabilité $(\mathbf{P}_x^\nu)_{x \in \mathbb{Z}}$ sur \mathcal{S} appelées lois *recuites* (traduit du terme anglais *annealed*) par :

$$\forall x \in \mathbb{Z}, \forall B \in \mathcal{S}, \quad \mathbf{P}_x^\nu(B) = \int_E \mathbf{P}_{x, \omega}(B) \mathbb{P}^\nu(d\omega).$$

Définition 2

La marche aléatoire en milieu aléatoire ν partant de x est le couple constitué du processus canonique $(X_n)_{n \in \mathbb{N}}$ et de la probabilité \mathbf{P}_x^ν .

Ainsi, pour construire une trajectoire de la MAMA partant de x , il s'agit de d'abord tirer des variables aléatoires i.i.d. $\omega = (\omega_k)_{k \in \mathbb{Z}}$ de loi ν et ensuite de réaliser une trajectoire selon une chaîne de Markov de noyau Q_ω .

Cette hypothèse d'indépendance et d'égalité des distributions permet de restreindre l'étude des MAMA aux MAMA partant de 0 : *c'est cette hypothèse que l'on fera désormais* et on omettra l'indice x dans les notations précédemment introduites, ce dernier valant constamment 0.

Toutefois, le double aléa consécutif au caractère aléatoire du milieu ne permet pas d'envisager une caractérisation si directe du comportement de la MAMA en temps long.

2.2 Le problème de l'estimation du milieu

Lorsque l'on étudie la MAMA d'un point de vue statistique, un premier problème peut être celui de l'estimation du milieu. En effet, la véritable loi du milieu notée ν^* est a priori inconnue et étant donné le procédé de construction de la MAMA, il est manifeste que l'on ne peut directement à partir d'une trajectoire déterminer quelle est la loi de l'environnement puisqu'à l'aléa du milieu s'ajoute l'aléa markovien des transitions. Pour alléger les notations, tous les objets fonctions de ν^* seront désormais désignés avec un $*$ en indice supérieur : $\mathbb{P}_0^{\nu^*}$ sera noté \mathbb{P}_0^* , $\mathbb{E}_0^{\nu^*}$ sera noté \mathbb{E}_0^* , ...

Néanmoins, une idée naturelle pour estimer à partir d'une trajectoire $x = (x_0, x_1, \dots, x_n)$ en un site $k \in \mathbb{Z}$ donné la valeur de ω_k serait de définir un

estimateur $\widehat{\omega}_k^n$ comme le quotient du nombre de fois que la trajectoire "saute de k à $k + 1$ " par le nombre de fois que la trajectoire occupe le site k :

$$\widehat{\omega}_k^n = \frac{\sum_{j=0}^{n-1} \mathbb{1}_{\{X_j=k, X_{j+1}=k+1\}}}{\sum_{j=0}^{n-1} \mathbb{1}_{\{X_j=k\}}}.$$

Étant donné le comportement Markovien de $(X_k)_{k \in \mathbb{N}}$, *sous certaines hypothèses d'ergodicité et de récurrence de la chaîne*, $\widehat{\omega}_k^n$ serait alors un estimateur consistant de ω_k , i.e., converge en probabilité ou presque sûrement vers ω_k lorsque n tend vers $+\infty$.

Il resterait alors à préciser quelles sont ces hypothèses et sous quelle(s) probabilité(s) (loi trempée ou recuite), ces résultats sont vérifiés.

2.3 Résultats probabilistes sur la MAMA

Dès son introduction, le modèle de la MAMA a été intensivement étudié par les probabilistes. Une quantité s'avérant alors centrale dans l'étude des MAMA est la variable aléatoire ρ_k définie pour tout k dans \mathbb{Z} par :

$$\rho_k = \frac{1 - \omega_k}{\omega_k}.$$

Ainsi, Solomon donne en 1975 un critère permettant de décrire simplement les comportements asymptotiques de la MAMA, reposant sur certains "moments" de ρ_0 .

Théorème 1 (Solomon, [Sol75])

Si $\nu \in \mathcal{M}_1(\mathcal{B}([0, 1]))$ est telle que $\mathbb{E}^\nu [|\log(\rho_0)|] < +\infty$, alors :

$$\begin{aligned} \mathbb{E}^\nu [\log(\rho_0)] < 0 &\Rightarrow X_n \xrightarrow[n \rightarrow +\infty]{\mathbf{P}_0^{\nu\text{-p.s.}}} +\infty, \text{ i.e., la MAMA est transiente à droite.} \\ \mathbb{E}^\nu [\log(\rho_0)] = 0 &\Rightarrow \forall k \in \mathbb{N}, \mathbf{P}_0^\nu (\overline{\lim} \{X_n = k\}) = 1, \text{ i.e., la MAMA est récurrente.} \\ \mathbb{E}^\nu [\log(\rho_0)] > 0 &\Rightarrow X_n \xrightarrow[n \rightarrow +\infty]{\mathbf{P}_0^{\nu\text{-a.s.}}} -\infty, \text{ i.e., la MAMA est transiente à gauche.} \end{aligned}$$

Théorème 2 (Solomon, [Sol75])

Si $\nu \in \mathcal{M}_1(\mathcal{B}([0, 1]))$ est telle que $\mathbb{E}^\nu [|\log(\rho_0)|] < +\infty$, alors :

$$\begin{aligned} \mathbb{E}^\nu [\rho_0] < 1 &\Rightarrow \frac{X_n}{n} \xrightarrow[n \rightarrow +\infty]{\mathbf{P}_0^{\nu\text{-p.s.}}} \frac{1 - \mathbb{E}^\nu [\rho_0]}{1 + \mathbb{E}^\nu [\rho_0]}. \\ \mathbb{E}^\nu \left[\frac{1}{\rho_0} \right] < 1 &\Rightarrow \frac{X_n}{n} \xrightarrow[n \rightarrow +\infty]{\mathbf{P}_0^{\nu\text{-p.s.}}} -\frac{1 - \mathbb{E}^\nu \left[\frac{1}{\rho_0} \right]}{1 + \mathbb{E}^\nu \left[\frac{1}{\rho_0} \right]}. \\ \frac{1}{\mathbb{E}^\nu [\rho_0]} \leq 1 \leq \mathbb{E}^\nu \left[\frac{1}{\rho_0} \right] &\Rightarrow \frac{X_n}{n} \xrightarrow[n \rightarrow +\infty]{\mathbf{P}_0^{\nu\text{-p.s.}}} 0. \end{aligned}$$

Théorème 3 (Solomon, [Sol75])

Si la MAMA est telle que $\mathbb{E}^\nu [|\log(\rho_0)|] < +\infty$ et $\mathbb{E}^\nu [\log(\rho_0)] < 0$, alors, pour tout n dans \mathbb{N} , le temps d'arrêt $T_n = \inf(\{k \in \mathbb{N} / X_k = n\})$ est $\mathbf{P}_0^{\nu\text{-p.s.}}$ bien défini et fini, et il vient :

$$\begin{aligned} \mathbb{E}^\nu [\rho_0] < 1 &\Rightarrow \frac{T_n}{n} \xrightarrow[n \rightarrow +\infty]{\mathbf{P}_0^{\nu\text{-p.s.}}} \mathbf{E}_0^\nu [T_1] = \frac{1 + \mathbb{E}^\nu [\rho_0]}{1 - \mathbb{E}^\nu [\rho_0]}. \\ \mathbb{E}^\nu [\rho_0] \geq 1 &\Rightarrow \frac{T_n}{n} \xrightarrow[n \rightarrow +\infty]{\mathbf{P}_0^{\nu\text{-p.s.}}} \mathbf{E}_0^\nu [T_1] = +\infty. \end{aligned}$$

Définition 3

Si $\frac{X_n}{n} \xrightarrow[n \rightarrow +\infty]{P_0^{p.s.}} v \in \mathbb{R}$, alors, la MAMA est dite :

ballistique si $v \neq 0$ et sous-ballistique si $v = 0$.

2.4 Le problème de l'estimation de la loi du milieu

Le problème de l'estimation de la loi du milieu est un problème qui n'a été envisagé que depuis peu : d'abord dans un cadre paramétrique dans [FC14], [CFL14], [FLG14] et [FLM13], et bien plus récemment dans un cadre non paramétrique dans [Die16].

L'estimation de la loi du milieu consiste à fournir un estimateur de la loi ν (c'est-à-dire pour tout entier n fixé, une application mesurable de l'espace des trajectoires de longueur n dans un espace mesurable) et à étudier les propriétés de cet estimateur (biais et variance, efficacité, consistance, vitesse de convergence, ...).

Une stratégie naturelle pour estimer la loi du milieu serait d'estimer le milieu dans un premier temps et de se servir de cette estimation du milieu pour en estimer la loi dans un second temps. Cette approche semble ne pas avoir encore été abordée à moins qu'elle n'ait été jugée totalement sous-optimale en raison de l'estimation elle-même faite à partir d'estimations. Dans tous les cas, cette stratégie nécessiterait d'abord une estimation du milieu qui d'après la partie 2.2 est encore à développer : le cas où la marche est récurrente semble a priori ne pas être un obstacle mais les situations de transience s'avèrent d'emblée être bien plus difficilement abordables...

Les stratégies envisagées jusqu'alors pour résoudre le problème de l'estimation du milieu ne sont donc pas d'estimer le milieu puis sa loi mais directement la loi par des procédures plus ou moins variées.

Ainsi, dans la première série d'articles citée ci-dessus ([FC14], [CFL14], [FLG14] et [FLM13]), les auteurs ont décidé de s'intéresser au problème dans un cadre paramétrique par une procédure d'estimation par maximum de vraisemblance et ont obtenu des résultats présentant une généralité importante "seulement" dans la situation de transience et encore ils sont contraints de modifier la procédure d'estimation selon que le régime de la MAMA est ballistique ou sous-ballistique.

Les auteurs s'y étant intéressés très récemment ([Die16]) sont quant à eux parvenus à obtenir une procédure générique quelle que soit le régime de la MAMA en s'affranchissant en plus du caractère paramétrique du modèle : ils estiment la fonction de répartition de la loi du milieu.

Néanmoins, dans ces deux approches, une statistique (i.e. *une fonction mesurable des observations*) particulière est au centre de toutes les investigations. On s'attachera dans la suite de cette présentation à présenter cette statistique et à montrer en quoi elle paraît indispensable dans le problème de l'estimation de la loi du milieu.

2.4.1 Le processus des sauts à gauche

Définition 4

Pour $l \in \mathbb{N}$, un chemin au plus proche voisin de longueur l (partant de $p \in \mathbb{Z}$) est une suite $x = (x_0, \dots, x_l) \in \mathbb{Z}^{l+1}$ vérifiant :

$$\forall i \in \{1, \dots, l\}, |x_i - x_{i-1}| = 1 \quad (\text{et } x_0 = p).$$

Définition 5

Pour tout chemin au plus proche voisin $x = (x_0, \dots, x_l)$ de longueur l et tout $k \in \mathbb{Z}$, le nombre de sauts à gauche $L(k, x)$ du chemin x à partir de k et le nombre de sauts à droite $R(k, x)$ du chemin x à partir de k sont définis par :

$$L(k, x) = \sum_{i=0}^{l-1} \mathbb{1}_{\{x_i=k, x_{i+1}=k-1\}} \quad \text{et} \quad R(k, x) = \sum_{i=0}^{l-1} \mathbb{1}_{\{x_i=k, x_{i+1}=k+1\}}.$$

Pour tout chemin au plus proche voisin $x = (x_0, \dots, x_l)$, l'ensemble des entiers visités par le chemin x noté $v(x)$ est l'ensemble des entiers $k \in \mathbb{Z}$ tels qu'il existe $i \in \{0, \dots, l\}$ pour lequel $x_i = k$.

Pour s'assurer que T_n est bien défini pour tout $n \in \mathbb{N}$ (i.e. $T_n < +\infty$ \mathbf{P}_0^ν -p.s.), on suppose l'hypothèse suivante vérifiée :

$$\forall \nu \in \mathcal{P}, \mathbb{E}^\nu [|\log(\rho_0)|] < +\infty \quad \text{et} \quad \mathbb{E}^\nu [\log(\rho_0)] < 0.$$

La MAMA atteignant \mathbf{P}_0^ν -p.s. n'importe quel $n \in \mathbb{N}$, il est possible de considérer le plus proche chemin aléatoire $X_{[0, T_n]} = (X_0, \dots, X_{T_n})$ de longueur T_n et de définir pour tout $k \in \mathbb{Z}$ les trois objets aléatoires suivants :

$$L_k^n = L(k, X_{[0, T_n]}) \quad \text{et} \quad R_k^n = R(k, X_{[0, T_n]}) \quad \text{et} \quad V_{T_n} = v(X_{[0, T_n]}).$$

Pour tout $n \in \mathbb{N}$ et tout chemin au plus proche voisin $x_{[0, t_n]} = (x_0, \dots, x_{t_n})$ de longueur t_n , à environnement fixé ω , les propriétés Markoviennes de la MAMA mènent à :

$$\mathbf{P}_{0, \omega}(X_{[0, T_n]} = x_{[0, t_n]}) = \prod_{k \in v_{x_{[0, t_n-1]}}} \omega_k^{R(k, x_{[0, t_n]})} (1 - \omega_k)^{L(k, x_{[0, t_n]})}.$$

En intégrant par rapport à la loi de l'environnement, il vient donc :

$$\mathbf{P}_0^\nu(X_{[0, T_n]} = x_{[0, t_n]}) = \prod_{k \in v_{x_{[0, t_n-1]}}} \int_{]0, 1[} t^{R(k, x_{[0, t_n]})} (1 - t)^{L(k, x_{[0, t_n]})} \nu(dt).$$

Une propriété combinatoire de marches au plus proche voisin permet de ne plus faire dépendre l'expression de cette probabilité que de $L(k, x_{[0, t_n]})$ car :

$$\forall k \in \mathbb{Z} \setminus \llbracket 0, n-1 \rrbracket, R(k, x_{[0, t_n]}) = L(k+1, x_{[0, t_n]})$$

et

$$\forall k \in \{0, \dots, n-1\}, R(k, x_{[0, t_n]}) = 1 + L(k+1, x_{[0, t_n]}).$$

La famille de statistiques $(L_k^n)_{k \in \mathbb{Z}}$ apparaît donc comme centrale dans l'étude des trajectoires de la MAMA puisqu'elle résume toute l'information disponible

sur la vraisemblance des trajectoires.

Dès les premières études probabilistes (voir [Kes75]), un parallèle fondamental a été établi entre le processus des sauts à gauche tronqué $(L_k^n)_{k \in \{0, \dots, n\}}$ d'une MAMA et les processus de branchement en milieu aléatoire.

Définition 6

Un processus de branchement inhomogène en environnement ω avec immigration et loi de reproduction géométrique est une suite de variables aléatoires $(Z_k)_{k \in \mathbb{N}}$ définies par :

$$Z_0 = 0 \quad \text{et} \quad \forall k \in \{0, \dots, n-1\}, \quad Z_{k+1} = \sum_{j=0}^{Z_k} \xi_{k+1,j}$$

où $(\xi_{i,j})_{(i,j) \in \mathbb{N}^* \times \mathbb{N}}$ sont des variables aléatoires indépendantes telles que pour tout $i \in \mathbb{N}^*$ et tout $j \in \mathbb{N}$ $\xi_{i,j} \sim \mathcal{G}_{\mathbb{N}}(\omega_i)$ sous $\mathbf{P}_{0,\omega}$.

De même que pour les MAMA, on peut considérer que l'environnement est aléatoire et regarder la loi des variables aléatoires du processus de branchement $(Z_k)_{k \in \mathbb{N}}$ sous la loi recuite \mathbf{P}_0^* qui est appelé *processus de branchement en environnement aléatoire*. On peut alors énoncer un résultat d'abord mis en évidence dans [Kes75] qui permet de ramener l'étude de la loi de la suite $(L_k^n)_{k \in \{0, \dots, n\}}$ à celle de l'étude de la suite $(Z_k)_{k \in \mathbb{N}}$.

Théorème 4 (Kesten, Kozlov, Spitzer, [Kes75])

Pour tout n dans \mathbb{N} , les vecteurs aléatoires $(L_{n-k}^n)_{k \in \{0, \dots, n\}}$ et $(Z_k)_{k \in \{0, \dots, n\}}$ ont la même distribution sous la loi recuite \mathbf{P}_0^* .

En raison du parallèle établi, il s'avère indispensable de connaître les propriétés les plus importantes des processus de branchement en environnement aléatoire.

Théorème 5

Si le milieu est i.i.d., i.e. $\omega = (\omega_k)_{k \in \mathbb{Z}} \sim \mathbb{P}^\nu = \nu^{\otimes \mathbb{Z}}$, alors, le processus de branchement inhomogène en environnement aléatoire avec immigration et loi de reproduction géométrique $(Z_k)_{k \in \mathbb{N}}$ est :

- une chaîne de Markov inhomogène sous la loi trempée $\mathbf{P}_{0,\omega}$ pour tout milieu fixé $\omega \in E$ dont les noyaux de transition Q_k sont donnés par :

$$Q_k(m, n) = \binom{m+n}{m} \omega_{k+1}^{m+1} (1 - \omega_{k+1})^n.$$

- une chaîne de Markov homogène sous la loi recuite \mathbf{P}_0^ν dont le noyau de transition Q^ν est donné par :

$$Q^\nu(m, n) = \binom{m+n}{m} \int_{]0,1[} t^{m+1} (1-t)^n \nu(dt) = \binom{m+n}{m} e^{\phi_\nu(m,n)}.$$

Ainsi, de nombreux outils de la théorie des chaînes de Markov (sur un espace d'état dénombrable) sont utilisables pour l'étude de $(Z_k)_{k \in \mathbb{N}}$ tels que les théorèmes d'existence de probabilité invariante et les théorèmes ergodiques.

Par ailleurs, le caractère discret de l'espace d'état de la chaîne $(Z_k)_{k \in \mathbb{N}}$ permet d'utiliser l'outil puissant que représentent les fonctions génératrices pour l'étude des lois et des convergences en loi.

Quoiqu'il en soit, la famille de statistiques $(L_k^n)_{k \in \{0, \dots, n\}}$ et l'identité en loi donnée par le théorème 4 permettant de "réduire" l'étude des MAMA à celle des processus de branchement en milieu aléatoire semblent être jusqu'à présent au coeur de l'étude statistique des MAMA.

2.4.2 Le cas paramétrique

Le problème de l'estimation du milieu de la MAMA a principalement été étudié d'un point de vue paramétrique au travers des articles [FC14], [FLG14] et [FLM13] dans le cadre d'un régime transiant à droite et ballistique ou sous-ballistique.

Ainsi, sous des hypothèses de régularité classiques (continuité, classe \mathcal{C}^2 , inversibilité de la matrice d'information de Fisher) et en construisant un estimateur $\widehat{\theta}_n^+$ du véritable paramètre $\theta^* \in \Theta$ par maximum de vraisemblance :

$$\widehat{\theta}_n^+ \in \underset{\theta \in \Theta}{\operatorname{Argmax}}(l_n^+(\theta)) \quad \text{où} \quad l_n^+(\theta) = \sum_{k=0}^{n-1} \log \left(\int_{]0,1[} t^{1+L_{k+1}^n} (1-t)^{L_k^n} \nu(dt) \right),$$

la consistance et la normalité asymptotique de l'estimateur sont obtenues grâce à divers résultats de la théorie des statistiques asymptotiques (voir [Vaa98]).

Dans [CFLL14], le régime récurrent est abordé pour le cas d'une loi du milieu à support fini et malgré le cadre très restrictif, d'intéressants phénomènes apparaissent comme le fait que l'on peut facilement et rapidement estimer le support de la loi du milieu mais pas les poids de chacun des points supports. Néanmoins, la procédure d'estimation envisagée est très spécifique et paraît difficilement extensible au cas général.

2.4.3 Le cas non paramétrique

L'article [Die16] utilisant de outils issus des théories probabilistes et statistiques les plus récentes (concentration de mesures pour des chaînes de Markov et des martingales, sélection de modèles, estimation adaptative, ...) appréhende le cas où la loi du milieu n'est plus supposée appartenir à un modèle paramétrique (on parle de *cadre non paramétrique*) et construit un estimateur de la fonction de répartition de la loi du milieu qui est générique en cela qu'il ne dépend pas de la nature du régime.

À la fois des résultats de convergence uniforme en moyenne et de concentration sont ainsi obtenus mais aussi un contrôle non asymptotique des erreurs d'approximation et des vitesses de convergence dépendant de la nature des régimes sont mises en évidence.

Toutefois, des hypothèses fortes de régularité sur la fonction de répartition de la loi du milieu sont faites et les vitesses d'estimation obtenues sont moins bonnes que dans le cadre paramétrique.

2.5 Perspectives et questions ouvertes

Comme il a été souligné dans cette présentation du domaine de recherche de l'estimation de la loi du milieu d'une marche aléatoire en milieu aléatoire, le problème a été peu abordé jusqu'à présent et qui plus est les cadres envisagés sont très restrictifs (modèle paramétrique, procédure d'estimation différente selon le régime de la MAMA, seules des lois à supports finis étudiées dans le régime récurrent) à l'exception peut-être de celui de l'article [Die16].

Ainsi, il semble que beaucoup de questions courantes en statistiques (efficacité des estimateurs, optimalité des vitesses de convergence,...) n'ont pas été envisagées et comme paraît l'illustrer l'article [Die16], les théories probabilistes et statistiques les plus récentes (concentration de mesures pour des chaînes de Markov, sélection de modèles, estimation adaptative) sembleraient permettre de répondre à nombreuses d'entre elles.

Finalement, un aspect non négligeable de l'estimation est celui de l'estimation effective. Les articles [FC14], [FLG14], [CFLL14], [FLM13] et [Die16] illustrent certes leur résultats par des simulations et calculs d'estimateurs mais seulement pour des lois du milieu très particulières et ne semblent pas s'attarder sur les problèmes computationnels ouverts lors de la simulation de trajectoires et de la résolution numérique de problèmes d'optimisation. L'utilisation d'outils de parallélisation des calculs et de techniques d'optimisation plus réfléchies et récentes (optimisation non linéaire, optimisation stochastique, ...) pourraient peut-être rendre plus rapides et efficaces les étapes de simulation et d'estimation.

Bibliographie

- [CFL14] 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. *ArXiv e-prints*, April 2014.
- [Che67] A.A. Chernov. Replication of a multicomponent chain by the lightning mechanism. *Biofizika*, 12 :297–301, 1967.
- [Die16] M. Diel, R. et Lerasle. Non parametric estimation for random walks in random environment. preprint sur <https://arxiv.org/abs/1606.03848>, 2016.
- [FC14] O.Loukianov D.Loukianova F.Comets, M.Falconnet and C.Matias. Maximum likelihood estimator consistency for ballistic random walk in a parametric random environment. *Stochastic Processes and their Applications*, 124(1) :166–188, 2014.
- [FLG14] M. Falconnet, D. Loukianova, and A. Gloter. Maximum likelihood estimation in the context of a sub-ballistic random walk in a parametric random environment. *ArXiv e-prints*, May 2014.
- [FLM13] 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. *ArXiv e-prints*, February 2013.
- [Kes75] Kozlov M. V. Spitzer F. Kesten, H. A limit law for random walk in a random environment. *Compositio Mathematica*, 30(2) :145–168, 1975.
- [Sol75] Fred Solomon. Random walks in a random environment. *Ann. Probab.*, 3(1) :1–31, 02 1975.
- [Tem72] D.E. Temkin. One-dimensional random walks in a two-component chain. *SovietMathematics Doklady*, 13 :1172—1176, 1972.
- [Vaa98] A. W. van der Vaart. *Asymptotic statistics*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge (UK), New York (N.Y.), 1998. Autre tirage : 2000 (édition brochée), 2005, 2006, 2007.

Chapitre 3

TER de Master 1

Préambule

Ce mémoire intitulé *Introduction aux fondements mathématiques de l'apprentissage statistique* a été réalisé en binôme dans le cadre du module d'enseignement *T.E.R (Travail d'étude et de recherche)* du Master 1 de Mathématiques Fondamentales et Appliquées du département de mathématiques de l'université Paris-Sud 11.

L'ensemble du travail a été réalisé sous la tutelle de Christophe GIRAUD, professeur à l'université Paris-Sud 11 et chargé de cours à l'École Polytechnique de février à juin 2014. L'ensemble de ce mémoire se base d'ailleurs sur le chapitre rédigé par ce dernier à l'occasion des journées mathématiques X-UPS 2013 et intitulé « Fondements mathématiques de l'apprentissage statistique » paru dans l'ouvrage Aléatoire aux *Éditions de l'École Polytechnique*.

Ce mémoire ne se veut nullement exhaustif quant aux fondements mathématiques de l'apprentissage statistique ni quant à la classification automatique dont il traite plus spécifiquement. L'ensemble du travail réalisé avait en effet pour principal dessein de confronter les étudiants à la démarche de la recherche en mathématiques et aux outils dont celle-ci dispose tant pour sa réalisation que pour sa diffusion.

3.1 Introduction

Dans bien des situations, il est fréquent de devoir attribuer à un certain objet une classe sur la base d'un certain nombre d'attributions de ce même objet déjà effectuées : on désire que l'objet se retrouve dans la catégorie qui est la plus vraisemblable quant aux observations déjà faites.

La démarche statistique semble donc être appropriée pour traiter la situation : on dispose d'un certain nombre d'observations d'un même phénomène et l'on désire trouver l'information fournie par ces dernières afin de commettre le moins d'erreur possible si l'on devait « prédire » la classe d'une observation supplémentaire.

L'exemple type d'une telle situation est celle de la classification automatique des courriers électroniques entre courriers désirables et spams : on cherche, à partir des courriers reçus et déjà classés, à prévoir le mieux possible la catégorie d'un courrier futur. Par la suite, on s'intéressera à une telle situation, c'est à dire où l'objet que l'on cherche à classer doit l'être parmi deux catégories. Néanmoins, bien que cela soit réalisable, on ne cherchera pas à déterminer explicitement un tel moyen de classification mais bien plus à estimer la qualité d'une telle procédure de classification automatique.

3.1.1 Modélisation de la situation

En tenant compte du fait que $(h(X), Y) \in \{-1, 1\}$ et que par conséquent $|Y - h(X)| = 2 \iff [h(X) \neq Y]$ et $|Y - h(X)| = 0 \iff [h(X) = Y]$, on peut réécrire $L(h)$ de cette manière :

$$\begin{aligned} L(h) &= \mathbb{P}(\cdot | h(X) \neq Y) = \mathbb{E} [\mathbb{1}_{h(X) \neq Y}] \\ &= \frac{1}{4} \mathbb{E} [(Y - h(X))^2 \mathbb{1}_{h(X) \neq Y}] = \frac{1}{4} \mathbb{E} [(Y - h(X))^2] \\ &= \frac{1}{4} \mathbb{E} [(Y - \mathbb{E}[Y|X]) + (\mathbb{E}[Y|X] - h(X))^2] \\ &= \frac{1}{4} \mathbb{E} [(Y - \mathbb{E}[Y|X])^2] + \frac{1}{4} \mathbb{E} [(\mathbb{E}[Y|X] - h(X))^2] \end{aligned}$$

$L(h)$ est donc minimal pour le classifieur h_* tel que $h_*(X) = \text{sign}(\mathbb{E}[Y|X])$ p.s. où $\text{sign}(x) = \mathbb{1}_{x>0} - \mathbb{1}_{x \leq 0}$.

3.1.2 Position du problème

Le classifieur h_* n'est néanmoins calculable que si l'on connaît \mathbb{P} , ce qui n'est justement pas le cas.

En réalité, dans le problème pratique, on dispose de n couples « objet-classe » issus de $((X_1, Y_1), \dots, (X_n, Y_n))$ un n -échantillon de loi \mathbb{P} et l'on cherche un classifieur \hat{h} qui soit le plus proche possible de h_* dans le sens où l'on cherche à minimiser $L(\hat{h}) - L(h_*)$.

Comme on ne dispose pas de la loi \mathbb{P} mais seulement des n observations « objet-classe », il est naturel de considérer la probabilité empirique de mauvaise classification de h , $L_n(h)$ qui si on note $\hat{\mathbb{P}}_n$ la probabilité empirique s'exprime :

$$L_n(h) = \hat{\mathbb{P}}_n(Y \neq h(X)) = \hat{\mathbb{E}}_n [\mathbb{1}_{Y \neq h(X)}] = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{Y_i \neq h(X_i)}$$

On se fixe alors un ensemble de classifieur $\mathcal{H} \subset \mathcal{P}(\{-1, 1\}^{\mathcal{X}})$ qu'on appelle dictionnaire sur lequel on cherche de façon plutôt naturelle à minimiser $L_n(h)$, c'est pourquoi on introduit le classifieur de minimisation du risque empirique :

$$\hat{h}_{\mathcal{H}} = \underset{h \in \mathcal{H}}{\operatorname{argmin}}(L_n(h))$$

Néanmoins, deux questions surgissent :

- Quelle est l'influence de \mathcal{H} sur $\hat{h}_{\mathcal{H}}$?
- Comment $\hat{h}_{\mathcal{H}}$ se comporte-t-il comparativement à h_* en terme de probabilité de mauvaise classification ?

Par un jeu d'écriture, on peut interpréter la différence entre les probabilités de mauvaise classification $L(\hat{h}_{\mathcal{H}})$ et $L(h_*)$:

$$0 \leq L(\hat{h}_{\mathcal{H}}) - L(h_*) = \underbrace{\left(\hat{h}_{\mathcal{H}} - \min_{h \in \mathcal{H}}(L(h)) \right)}_{\text{erreur stochastique}} + \underbrace{\left(\min_{h \in \mathcal{H}}(L(h)) - L(h_*) \right)}_{\text{erreur d'approximation}}$$

D'une part, le second terme de cette inégalité $\min_{h \in \mathcal{H}}(L(h)) - L(h_*)$ désigne l'erreur entre la probabilité minimale de mauvaise classification lorsque l'on se restreint aux classificateurs de \mathcal{H} et la probabilité de mauvaise classification de l'estimateur bayésien h_* : elle résulte de l'approximation consistant à limiter l'ensemble des classifieurs possibles au dictionnaire \mathcal{H} , c'est pour cette raison qu'elle porte le nom « d'erreur d'approximation ». Le membre $\min_{h \in \mathcal{H}}(L(h))$ ne peut que diminuer lorsque l'on augmente le cardinal du dictionnaire \mathcal{H} puisqu'il fait intervenir un \min sur \mathcal{H} . Il est également purement déterministe puisqu'il ne dépend pas des observations.

D'autre part, le premier terme de cette inégalité $\hat{h}_{\mathcal{H}} - \min_{h \in \mathcal{H}}(L(h))$ dépend clairement des observations puisqu'il désigne l'erreur $\hat{h}_{\mathcal{H}}$ réalisée en minimisant sur \mathcal{H} la probabilité empirique de mauvaise classification par rapport au minimum sur \mathcal{H} de la probabilité véritable de mauvaise classification. Son appellation « d'erreur stochastique » lui vient donc de cette dépendance stochastique qui tend à le faire augmenter lorsque \mathcal{H} grandit. Néanmoins, comme on peut l'envisager intuitivement et comme on le justifiera par la suite dans la section 3.2.2, ce n'est pas tant le nombre d'éléments du dictionnaire \mathcal{H} qui fait croître ce terme mais plutôt sa « flexibilité », c'est à dire la capacité des classifieurs de \mathcal{H} à pouvoir classifier un nouveau couple « objet-classe » : si on peut classifier n couples « objet-classe » de toutes les manières possibles avec un dictionnaire, sera-t-il également capable de classifier de toutes les manières possibles les n couples précédents ainsi qu'un nouveau couple ? Aussi, plus le dictionnaire sera « flexible », plus la probabilité que l'erreur stochastique associée soit grande car il aura tellement bien pu « épouser » la répartition de ces n couples que le classifieur trouvé ne classifiera pas bien de nouveaux couples.

C'est justement cette approche intuitive que va légitimer le travail qui suit avec une gestion de l'erreur stochastique en fonction de cette « flexibilité » puis une véritable étude de cette notion.

3.2 La gestion de l'erreur stochastique

3.2.1 Le contrôle de l'erreur stochastique

3.2.1.1 L'inégalité de Mac Diarmid

Soit \mathcal{X} un ensemble mesurable et $F : \mathcal{X}^n \rightarrow \mathbb{R}$ une application mesurable telle que $\exists(\delta_1, \dots, \delta_n) \in (\mathbb{R}^+)^n$ pour lesquels :

$$\forall i \in \{0, \dots, n\}, \forall (x_1, \dots, x_n, x'_i) \in \mathcal{X}^{n+1}, |F(x_1, \dots, x_i, \dots, x_n) - F(x_1, \dots, x'_i, \dots, x_n)| \leq \delta_i.$$

Soit $(\mathcal{F}_k)_{0 \leq k \leq n}$ la filtration définie par :

$$\mathcal{F}_0 = \{\emptyset, \Omega\} \text{ et } \forall k \in \{1, \dots, n\}, \mathcal{F}_k = \sigma(X_1, \dots, X_k)$$

Pour $k \in \{1, \dots, n\}$, on pose :

$$\Delta_k = \mathbb{E}[F(X_1, \dots, X_n) | \mathcal{F}_k] - \mathbb{E}[F(X_1, \dots, X_n) | \mathcal{F}_{k-1}].$$

On conservera ces notations tout au long de cette section. Pour démontrer l'inégalité de Mac Diarmid, on va s'appuyer sur le lemme suivant :

Lemme 1

Pour tout $\lambda \in \mathbb{R}^{*+}$ et tout $k \in \{1, \dots, n\}$, il vient :

$$\mathbb{E}[e^{\lambda \Delta_k} | \mathcal{F}_{k-1}] \leq e^{\lambda^2 \frac{\delta_k^2}{8}}.$$

Démo 1

Pour tout entier $k \in \{1, \dots, n\}$, on considère les applications :

- $F_k : \mathcal{X}^k \rightarrow \mathbb{R}$ telle que $F_k(X_1, \dots, X_k) = \mathbb{E}[F(X_1, \dots, X_n) | \mathcal{F}_k]$ p.s.
- $S_k : \mathcal{X}^{k-1} \rightarrow \mathbb{R}$ telle que $S_k(X_1, \dots, X_{k-1}, x) = \sup_{x \in \mathcal{X}} \{F_k(X_1, \dots, X_{k-1}, x) - \mathbb{E}[F(X_1, \dots, X_n) | \mathcal{F}_{k-1}]\}$ p.s.
- $I_k : \mathcal{X}^{k-1} \rightarrow \mathbb{R}$ telle que $I_k(X_1, \dots, X_{k-1}, x) = \inf_{x \in \mathcal{X}} \{F_k(X_1, \dots, X_{k-1}, x) - \mathbb{E}[F(X_1, \dots, X_n) | \mathcal{F}_{k-1}]\}$ p.s.

Par définition de ces applications, on a p.s. $\forall k \in \{1, \dots, n\}$:

$$I_k \leq \Delta_k \leq S_k \text{ et } 0 \leq S_k - I_k \leq \delta_k$$

Par convexité de la fonction $x \in \mathbb{R} \mapsto e^{\lambda x}$, il vient alors :

$$e^{\lambda \Delta_k} \leq \frac{\Delta_k - I_k}{S_k - I_k} e^{\lambda S_k} + \frac{S_k - \Delta_k}{S_k - I_k} e^{\lambda I_k}$$

De plus, S_k et I_k sont \mathcal{F}_{k-1} -mesurable et $\mathbb{E}[\Delta_k | \mathcal{F}_{k-1}] = 0$ (car $\mathcal{F}_{k-1} \subset \mathcal{F}_k$, $(\mathcal{F}_k)_{0 \leq k \leq n}$ étant une filtration) donc en notant $f : s \in \mathbb{R} \mapsto \frac{s I_k}{S_k - I_k} + \log(\frac{S_k - I_k e^s}{S_k - I_k})$, on a :

$$\mathbb{E}[e^{\lambda \Delta_k} | \mathcal{F}_{k-1}] \leq \frac{-I_k}{S_k - I_k} e^{\lambda S_k} + \frac{S_k}{S_k - I_k} e^{\lambda I_k} = e^{f(\lambda(S_k - I_k))}$$

L'étude de la fonction f nous donne alors que :

$$f(0) = 0$$

$$f'(0) = 0 \text{ car } f'(s) = \frac{I_k}{S_k - I_k} + \frac{-I_k}{S_k - I_k} \frac{S_k - I_k}{S_k - I_k e^s}$$

$$f''(s) = \frac{-I_k S_k e^s}{(S_k - I_k e^s)^2}$$

La fonction f étant de classe \mathcal{C}^2 sur \mathbb{R} , la formule de Taylor-Lagrange de nous assure alors que :

$$\forall s \geq 0, \quad \exists \tilde{s} \in]0, s[\quad f(s) = f(0) + f'(0)(s - 0) + f''(\tilde{s})(s - 0)^2.$$

$$\text{Or } [\forall (u, v) \in \mathbb{R}, (u + v)^2 - (u - v)^2 = 4uv] \implies [\forall (u, v) \in \mathbb{R}, uv \leq \frac{(u+v)^2}{4}] \blacksquare$$

Théorème 6 (inégalité de concentration de Mac Diarmid)

Si \mathcal{X} est un ensemble mesurable et $F : \mathcal{X}^n \rightarrow \mathbb{R}$ une application mesurable telle que $\exists (\delta_1, \dots, \delta_n) \in (\mathbb{R}^+)^n$ pour lesquels :

$$\forall i \in \{0, \dots, n\}, \forall (x_1, \dots, x_n, x'_i) \in \mathcal{X}^{n+1}, |F(x_1, \dots, x_i, \dots, x_n) - F(x_1, \dots, x'_i, \dots, x_n)| \leq \delta_i.$$

Alors, pour tout $t \geq 0$, pour toute variable aléatoire $(X_1, \dots, X_n) : \Omega \rightarrow \mathcal{X}^n$, il vient :

$$\mathbb{P}(F(X_1, \dots, X_n) > \mathbb{E}[F(X_1, \dots, X_n)] + t) \leq e^{-\frac{2t^2}{\delta_1^2 + \dots + \delta_n^2}}.$$

Démo 2

Soit $(\lambda, t) \in (\mathbb{R}^{*+})^2$. Comme l'application $t \in \mathbb{R} \mapsto e^{\lambda t}$ est croissante, l'inégalité de Markov donne :

$$\mathbb{P}(F(X_1, \dots, X_n) > \mathbb{E}[F(X_1, \dots, X_n)] + t) \leq e^{-\lambda t} \mathbb{E} \left[e^{\lambda(F(X_1, \dots, X_n) - \mathbb{E}[F(X_1, \dots, X_n)])} \right] = e^{-\lambda t} \mathbb{E} \left[\prod_{k=1}^n e^{\lambda \Delta_k} \right]$$

et par somme télescopique, il vient :

$$\begin{aligned} F(X_1, \dots, X_n) - \mathbb{E}[F(X_1, \dots, X_n)] &= \mathbb{E}[F(X_1, \dots, X_n) | \mathcal{F}_n] - \mathbb{E}[\mathbb{E}[F(X_1, \dots, X_n) | \mathcal{F}_0]] \\ &= \sum_{k=1}^n \mathbb{E}[F(X_1, \dots, X_n) | \mathcal{F}_k] - \mathbb{E}[\mathbb{E}[F(X_1, \dots, X_n) | \mathcal{F}_{k-1}]] \\ &= \sum_{k=1}^n \Delta_k; \end{aligned}$$

$$\text{d'où } \mathbb{P}(F(X_1, \dots, X_n) > \mathbb{E}[F(X_1, \dots, X_n)] + t) \leq e^{-\lambda t} \mathbb{E} \left[\prod_{k=1}^n e^{\lambda \Delta_k} \right].$$

Le lemme combiné avec une récurrence immédiate nous assure ensuite que :

$$\mathbb{E} \left[\prod_{k=1}^{n-1} e^{\lambda \Delta_k} \right] = \mathbb{E} \left[\left(\prod_{k=1}^{n-1} e^{\lambda \Delta_k} \right) \mathbb{E} [e^{\lambda \Delta_n} | \mathcal{F}_{n-1}] \right] \leq e^{\frac{\lambda^2 \delta_n^2}{8}} \mathbb{E} \left[\prod_{k=1}^{n-1} e^{\lambda \Delta_k} \right] \leq \prod_{k=1}^n e^{\frac{\lambda^2 \delta_k^2}{8}} = e^{\lambda^2 \sum_{k=1}^n \frac{\delta_k^2}{8}}.$$

Finalement, il vient alors que :

$$\forall t > 0, \quad \mathbb{P}(F(X_1, \dots, X_n) > \mathbb{E}[F(X_1, \dots, X_n)] + t) \leq e^{-\lambda t} e^{\lambda^2 \sum_{k=1}^n \frac{\delta_k^2}{8}}.$$

Ainsi pour $\lambda = \frac{4t}{\sum_{k=1}^n \delta_k^2}$, on obtient :

$$\begin{aligned} \mathbb{P}(F(X_1, \dots, X_n) > \mathbb{E}[F(X_1, \dots, X_n)] + t) &\leq \exp\left(-t \frac{4t}{\sum_{k=1}^n \delta_k^2}\right) \exp\left(\left(\frac{4t}{\sum_{k=1}^n \delta_k^2}\right)^2 \sum_{k=1}^n \frac{\delta_k^2}{8}\right) \\ &= \exp\left(\frac{-2t^2}{\sum_{k=1}^n \delta_k^2}\right) \end{aligned}$$

■

3.2.1.2 Majoration de l'erreur stochastique

Théorème 7

Pour tout $t > 0$, on a, avec probabilité au moins $1 - e^{-t}$:

$$L(\hat{h}_{\mathcal{H}}) - \min_{h \in \mathcal{H}} L(h) \leq 4\sqrt{\frac{2 \log(2\mathbb{S}_n(\mathcal{H}))}{n}} + \sqrt{\frac{2t}{n}}$$

et

$$\left|L(\hat{h}_{\mathcal{H}}) - \hat{L}_n(\hat{h}_{\mathcal{H}})\right| \leq 2\sqrt{\frac{2 \log(2\mathbb{S}_n(\mathcal{H}))}{n}} + \sqrt{\frac{t}{2n}}$$

Démo 3

Pour prouver ce théorème on va avoir besoin de 3 lemmes. On note :

$$\hat{\Delta}_n(\mathcal{H}) = \sup_{h \in \mathcal{H}} \left| \hat{L}_n(h) - L(h) \right|$$

Le premier lemme permet de majorer les termes de gauche dans les inégalités du théorème grâce à $\hat{\Delta}_n(\mathcal{H})$.

Lemme 2

Pour tout dictionnaire \mathcal{H} , on a :

$$L(\hat{h}_{\mathcal{H}}) - \min_{h \in \mathcal{H}} L(h) \leq 2\hat{\Delta}_n(\mathcal{H}) \quad \text{et} \quad \left|L(\hat{h}_{\mathcal{H}}) - \hat{L}_n(\hat{h}_{\mathcal{H}})\right| \leq \hat{\Delta}_n(\mathcal{H})$$

Démo 4

On a par définition, $\hat{L}_n(\hat{h}_{\mathcal{H}}) \leq \hat{L}_n(h)$ donc,

$$\begin{aligned} L(\hat{h}_{\mathcal{H}}) - L(h) &= L(\hat{h}_{\mathcal{H}}) - \hat{L}_n(\hat{h}_{\mathcal{H}}) + \hat{L}_n(\hat{h}_{\mathcal{H}}) - L(h) \\ &\leq L(\hat{h}_{\mathcal{H}}) - \hat{L}_n(\hat{h}_{\mathcal{H}}) + \hat{L}_n(h) - L(h) \\ &\leq 2\hat{\Delta}_n(\mathcal{H}) \end{aligned}$$

Ce qui nous donne la première borne du lemme, car toutes les inégalités sont vraies pour tout $h \in \mathcal{H}$. La deuxième inégalité découle de la définition de $\hat{\Delta}_n(\mathcal{H})$. ■

Les deuxième et troisième lemmes nous donnent une majoration de $\hat{\Delta}_n(\mathcal{H})$.

Lemme 3

Avec probabilité $1 - e^{-t}$, on a :

$$\widehat{\Delta}_n(\mathcal{H}) \leq \mathbb{E} \left[\widehat{\Delta}_n(\mathcal{H}) \right] + \sqrt{\frac{t}{2n}}.$$

Démo 5

On définit

$$F : (\mathcal{X} \times \{-1, 1\})^n \longrightarrow \mathbb{R}$$

$$((x_1, y_1), \dots, (x_n, y_n)) \longmapsto \frac{1}{n} \sup_{h \in \mathcal{H}} \left| \sum_{i=1}^n n \mathbb{1}_{y_i \neq h(x_i)} - L(h) \right|.$$

Pour tout $(x_1, y_1), \dots, (x_n, y_n), (x'_i, y'_i) \in \mathcal{X} \times \{-1, 1\}$, on a :

$$|F((x_1, y_1), \dots, (x'_i, y'_i), \dots, (x_n, y_n)) - F((x_1, y_1), \dots, (x_i, y_i), \dots, (x_n, y_n))| \leq \frac{1}{n}.$$

Cette inégalité est évidente, on peut procéder par l'absurde.

Ensuite, l'inégalité de McDiarmid, vue au paragraphe précédent, nous donne qu'avec probabilité au moins $1 - e^{-2ns^2}$, on a

$$\widehat{\Delta}_n(\mathcal{H}) \leq \mathbb{E} \left[\widehat{\Delta}_n(\mathcal{H}) \right] + s.$$

Un simple changement de variable $s = \sqrt{\frac{t}{2n}}$ permet de conclure la démonstration du deuxième lemme. \blacksquare

Ce deuxième lemme nous permet de voir très clairement, que pour avoir les majorants du théorème, il reste à montrer que :

Lemme 4

Pour tout dictionnaire \mathcal{H} , on a

$$\mathbb{E} \left[\widehat{\Delta}_n(\mathcal{H}) \right] \leq 2 \sqrt{\frac{2 \log(2\mathbb{S}_n(\mathcal{H}))}{n}}.$$

Démo 6

On commence par représenter $L(h)$ comme une espérance de probabilité empirique. Soit les $(\tilde{X}_i, \tilde{Y}_i)$ iid, de même loi que les (X_i, Y_i) et indépendants de ces derniers, on notera $\tilde{\mathbb{E}}$ l'espérance par rapport aux $(\tilde{X}_i, \tilde{Y}_i)$ et \mathbb{E} l'espérance par rapport aux (X_i, Y_i) . On a, en effet :

$$\tilde{\mathbb{E}} \left[\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\tilde{Y}_i \neq h(\tilde{X}_i)} \right] = \frac{1}{n} \sum_{i=1}^n \tilde{\mathbb{E}} \left[\mathbb{1}_{\tilde{Y}_i \neq h(\tilde{X}_i)} \right] = \mathbb{P}(\tilde{Y}_1 \neq h(\tilde{X}_1)) = \mathbb{P}(X \neq h(Y)).$$

D'où

$$L(h) = \tilde{\mathbb{E}} \left[\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\tilde{Y}_i \neq h(\tilde{X}_i)} \right],$$

$$\begin{aligned}
\mathbb{E} \left[\widehat{\Delta}_n(\mathcal{H}) \right] &= \mathbb{E} \left[\sup_{h \in \mathcal{H}} \widehat{L}_n(h) - L(h) \right] \\
&= \mathbb{E} \left[\sup_{h \in \mathcal{H}} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{Y_i \neq h(X_i)} \right| \right] - \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\tilde{Y}_i \neq h(\tilde{X}_i)} \right] \\
&\leq \mathbb{E} \mathbb{E} \left[\sup_{h \in \mathcal{H}} \left| \frac{1}{n} \sum_{i=1}^n \left(\mathbb{1}_{Y_i \neq h(X_i)} - \mathbb{1}_{\tilde{Y}_i \neq h(\tilde{X}_i)} \right) \right| \right].
\end{aligned}$$

Jensen nous permet de rentrer la valeur absolue dans \mathbb{E} , la somme des indicatrices de $X_i \neq h(Y_i)$ étant une constante pour \mathbb{E} . Ensuite le lemme de Fatou permet à son tour de rentrer le sup dans \mathbb{E} .

Pour la suite de la démonstration, on introduit des variables $(\sigma_i)_{i=1, \dots, n}$ indépendantes des $(X_i, Y_i, \tilde{X}_i, \tilde{Y}_i)_{i=1, \dots, n}$ et de loi $\mathbb{P}^\sigma(\sigma_i = 1) = \mathbb{P}^\sigma(\sigma_i = -1) = \frac{1}{2}$.

On va démontrer rapidement que $(\sigma_i(\mathbb{1}_{Y_i \neq h(X_i)} - \mathbb{1}_{\tilde{Y}_i \neq h(\tilde{X}_i)}))$ a même loi que $\mathbb{1}_{Y_i \neq h(X_i)} - \mathbb{1}_{\tilde{Y}_i \neq h(\tilde{X}_i)}$

$$\begin{aligned}
\mathbb{P} \left(\sigma_i \left(\mathbb{1}_{Y_i \neq h(X_i)} - \mathbb{1}_{\tilde{Y}_i \neq h(\tilde{X}_i)} \right) = k \right) &= \mathbb{P} \left(\left(\sigma_i \left(\mathbb{1}_{Y_i \neq h(X_i)} - \mathbb{1}_{\tilde{Y}_i \neq h(\tilde{X}_i)} \right) = k \right) \cap (\sigma_i = 1) \right) \\
&\quad + \mathbb{P} \left(\left(\sigma_i \left(\mathbb{1}_{Y_i \neq h(X_i)} - \mathbb{1}_{\tilde{Y}_i \neq h(\tilde{X}_i)} \right) = k \right) \cap (\sigma_i = -1) \right) \\
&= \frac{1}{2} \left(\mathbb{P} \left(\mathbb{1}_{Y_i \neq h(X_i)} - \mathbb{1}_{\tilde{Y}_i \neq h(\tilde{X}_i)} = k \right) \right. \\
&\quad \left. + \mathbb{P} \left(\mathbb{1}_{\tilde{Y}_i \neq h(\tilde{X}_i)} - \mathbb{1}_{Y_i \neq h(X_i)} = k \right) \right).
\end{aligned}$$

Or, comme les $(\tilde{X}_i, \tilde{Y}_i)$ et les (X_i, Y_i) sont iid, on peut échanger leurs rôles dans \mathbb{P} sans conséquences. On a donc :

$$\frac{1}{2} \left(\mathbb{P} \left(\mathbb{1}_{Y_i \neq h(X_i)} - \mathbb{1}_{\tilde{Y}_i \neq h(\tilde{X}_i)} = k \right) + \mathbb{P} \left(\mathbb{1}_{\tilde{Y}_i \neq h(\tilde{X}_i)} - \mathbb{1}_{Y_i \neq h(X_i)} = k \right) \right) = \mathbb{P} \left(\mathbb{1}_{Y_i \neq h(X_i)} - \mathbb{1}_{\tilde{Y}_i \neq h(\tilde{X}_i)} = k \right).$$

Revenons à la démonstration, d'après ce qui précède on peut remplacer $\mathbb{1}_{Y_i \neq h(X_i)} - \mathbb{1}_{\tilde{Y}_i \neq h(\tilde{X}_i)}$ par $\sigma_i(\mathbb{1}_{Y_i \neq h(X_i)} - \mathbb{1}_{\tilde{Y}_i \neq h(\tilde{X}_i)})$ dans l'expression de $\mathbb{E} \left[\widehat{\Delta}_n(\mathcal{H}) \right]$. On a alors :

$$\begin{aligned}
\mathbb{E} \left[\widehat{\Delta}_n(\mathcal{H}) \right] &\leq \mathbb{E} \mathbb{E} \mathbb{E}_\sigma \left[\left| \sum_{h \in \mathcal{H}} \left| \frac{1}{n} \sum_{i=1}^n \sigma_i \left(\mathbb{1}_{Y_i \neq h(X_i)} - \mathbb{1}_{\tilde{Y}_i \neq h(\tilde{X}_i)} \right) \right| \right| \right] \\
&\leq 2 \mathbb{E} \mathbb{E}_\sigma \left[\sup_{h \in \mathcal{H}} \left| \frac{1}{n} \sum_{i=1}^n \sigma_i \left(\mathbb{1}_{Y_i \neq h(X_i)} \right) \right| \right] \\
&\leq 2 \max_{y \in \{-1, +1\}^n} \max_{x \in \mathcal{X}^n} \mathbb{E}_\sigma \left[\sup_{h \in \mathcal{H}} \left| \frac{1}{n} \sum_{i=1}^n \sigma_i \left(\mathbb{1}_{Y_i \neq h(X_i)} \right) \right| \right].
\end{aligned}$$

Pour tout $(x, y) \in \mathcal{X}^n \times \{-1, 1\}^n$, on considère l'ensemble :

$$\mathcal{V}_{\mathcal{H}}(x, y) = \left\{ \left(\mathbb{1}_{y_1 \neq h(x_1)}, \dots, \mathbb{1}_{y_n \neq h(x_n)} \right) / h \in \mathcal{H} \right\}.$$

On peut donc majorer $\mathbb{E} [\widehat{\Delta}_n(\mathcal{H})]$ par

$$\frac{2}{n} \times \max_{y \in \{-1,1\}^n} \max_{c \in \mathcal{X}^n} \mathbb{E}_\sigma \left[\sup_{v \in \mathcal{V}_{\mathcal{H}}(x,y)} |\langle \sigma | v \rangle| \right].$$

avec $\langle x | y \rangle$ le produit scalaire canonique sur \mathbb{R}^n .

De plus, pour $y \in \{-1,1\}^n$, on a une bijection entre $\mathcal{V}_{\mathcal{H}}(x,y)$ et l'ensemble $\{(h(x_1), \dots, h(x_n)) / h \in \mathcal{H}\}$. On a donc la majoration par :

$$\max_{y \in \{-1,1\}^n} \max_{x \in \mathcal{X}^n} (|\mathcal{V}_{\mathcal{H}}(x,y)|) \leq \mathbb{S}_n(\mathcal{H}).$$

Pour avoir la formule qu'on recherche, il ne nous reste donc plus qu'à montrer que

$$\mathbb{E}_\sigma \left[\sup_{v \in \mathcal{V}} |\langle \sigma | v \rangle| \right] \leq \sqrt{2n \log(2 \text{card} \mathcal{V})}$$

pour tout ensemble \mathcal{V} inclus dans $\{-1,0,1\}$. On aura alors :

$$\begin{aligned} \mathbb{E} \left[\widehat{\Delta}_n(\mathcal{H}) \right] &\leq \frac{2}{n} \times \max_{y \in \{-1,1\}^n} \max_{c \in \mathcal{X}^n} \mathbb{E}_\sigma \left[\sup_{v \in \mathcal{V}_{\mathcal{H}}(x,y)} |\langle \sigma | v \rangle| \right] \\ &\leq \frac{2}{n} \times \max_{y \in \{-1,1\}^n} \max_{c \in \mathcal{X}^n} \sqrt{2n \log(2 |\mathcal{V}|)} \\ &\leq \frac{2}{n} \sqrt{2n \log(2 |\mathbb{S}_n(\mathcal{H})|)} \\ &\leq \frac{2}{\sqrt{n}} \sqrt{2 \log(2 |\mathbb{S}_n(\mathcal{H})|)}, \end{aligned}$$

ce qui est bien ce qu'on cherche à montrer.

Soit $\mathcal{V} \in \{-1,0,1\}^n$, on note $\mathcal{V}^\# = \mathcal{V} \cup -\mathcal{V}$, on a l'inégalité de Jensen, pour tout $s > 0$:

$$\begin{aligned} \mathbb{E}_\sigma \left[\sup_{v \in \mathcal{V}} |\langle \sigma | v \rangle| \right] &= \mathbb{E}_\sigma \left[\sup_{v \in \mathcal{V}^\#} \langle \sigma | v \rangle \right] \\ &\leq \frac{1}{s} \log \left(\mathbb{E}_\sigma \left[\sup_{v \in \mathcal{V}^\#} \exp(s \langle \sigma | v \rangle) \right] \right) \\ &\leq \frac{1}{s} \log \left(\sum_{v \in \mathcal{V}^\#} \mathbb{E}_\sigma [\exp(s \langle \sigma | v \rangle)] \right). \end{aligned}$$

Or les σ_i sont indépendants, on a donc :

$$\mathbb{E}_\sigma [\exp(s \langle \sigma | v \rangle)] = \prod_{i=1}^n \mathbb{E}_\sigma [\exp(sv_i \sigma_i)] = \prod_{i=1}^n \frac{1}{2} (\exp(sv_i) + \exp(-sv_i)).$$

On rappelle que, pour tout $v \in \mathcal{V}^\#$, $v_i^2 \in \{0,1\}$. On va montrer que $(\exp(x) + \exp(-x)) \leq 2 \exp(\frac{x^2}{2})$. Pour ça on va décomposer les deux termes de l'inégalité

en séries :

$$\begin{aligned} (\exp(x) + \exp(-x)) &= \sum_{k \in \mathbb{N}} \frac{x^k + (-x)^k}{k!} = \sum_{k \in \mathbb{N}} \frac{2x^{2k}}{(2k)!} \\ &= 2 \sum_{k \in \mathbb{N}} \frac{2(x^2)^k}{(2k)!} \leq 2 \sum_{k \in \mathbb{N}} \frac{2(x^2)^k}{2^k (k)!} = 2 \exp \frac{x^2}{2}. \end{aligned}$$

On a donc :

$$\mathbb{E}_\sigma [\exp(s \langle \sigma | v \rangle)] \leq \prod_{i=1}^n \exp \left(\frac{s^2 v_i^2}{2} \right) \leq \exp \left(\frac{ns^2}{2} \right).$$

Si on reprend les calculs précédents, cela nous donne pour tout $s > 0$:

$$\begin{aligned} \mathbb{E}_\sigma \left[\sup_{v \in \mathcal{V}} |\langle \sigma | v \rangle| \right] &\leq \frac{1}{s} \log \left(\sum_{v \in \mathcal{V}^\#} \mathbb{E}_\sigma [\exp(s \langle \sigma | v \rangle)] \right) \\ &\leq \frac{\log(|\mathcal{V}^\#|)}{s} + \frac{ns}{2}. \end{aligned}$$

Le dernier terme est minimal pour $s = \sqrt{\frac{2 \log(|\mathcal{V}^\#|)}{n}}$, ce qui donne finalement

$$\mathbb{E}_\sigma \left[\sup_{v \in \mathcal{V}} |\langle \sigma | v \rangle| \right] \leq \sqrt{2n \log(|\mathcal{V}^\#|)}.$$

ce qui permet de conclure la démonstration du dernier lemme. ■

Grâce aux trois lemmes démontrés plus haut, on obtient très simplement les inégalités voulues dans le théorème en injectant successivement les différentes expressions trouvées. ■

3.2.2 Le coefficient d'éclatement et la dimension de Vapnik-Chervonenkis

Comme on a pu le constater lors de la majoration de l'erreur stochastique, la qualité de classification obtenue avec un dictionnaire dépend du nombre maximal de classifications d'un ensemble de n points permises par ce dernier. C'est pour cette raison qu'on introduit le « *coefficient d'éclatement* » :

Définition 7

On définit pour tout entier naturel non nul $n \in \mathbb{N}^*$ et tout dictionnaire $\mathcal{H} \subset \mathcal{P}(\{-1, 1\}^{\mathcal{X}})$ le coefficient d'éclatement d'ordre n de \mathcal{H} par :

$$S_n(\mathcal{H}) = \max_{(x_1, \dots, x_n) \in \mathcal{X}^n} \{ \#\{(h(x_1), \dots, h(x_n)) / h \in \mathcal{H}\} \}.$$

Comme voulu, $\mathbb{S}_n(\mathcal{H})$ donne le nombre maximal d'étiquetages par $\{-1, 1\}$ de n points de \mathcal{X} par les classifieurs de \mathcal{H} .

Aussi, pour tout dictionnaire \mathcal{H} et tout entier $n \in \mathbb{N}^*$, comme $\mathcal{H} \subset (\mathcal{X} \rightarrow \{-1, 1\})$ et que $\mathcal{H} \neq \emptyset$, on a nécessairement que :

$$\forall h \in \mathcal{H}, \quad (h(x_1), \dots, h(x_n)) \in \{-1, 1\}^n$$

$$\text{donc } \forall (x_1, \dots, x_n) \in \mathcal{X}^n, \quad 1 \leq \#\{(h(x_1), \dots, h(x_n))/h \in \mathcal{H}\} \leq 2^n$$

$$\text{d'où } 1 \leq \mathbb{S}_n(\mathcal{H}) \leq 2^n$$

En raison de cette double inégalité, on convient alors de définir également pour tout dictionnaire \mathcal{H} le coefficient d'éclatement d'ordre 0 de \mathcal{H} par : $\mathbb{S}_0(\mathcal{H}) = 1$.

Néanmoins, cette quantité ne présente guère d'intérêt dans notre recherche de gestion de l'erreur stochastique si on ne peut savoir comment elle évolue quand on augmente n . Toutefois, on remarque que :

$$\text{si } \exists n_0 / \mathbb{S}_{n_0}(\mathcal{H}) < 2^{n_0} \quad , \text{ alors, } \forall n \in \mathbb{N}, \quad n > n_0, \quad \mathbb{S}_n(\mathcal{H}) < 2^n.$$

D'où la définition suivante :

Définition 8

On appelle *dimension de Vapnik-Chervonenkis* (abrégé en “*V.C.-dimension*”) d'un dictionnaire \mathcal{H} la quantité notée $d_{\mathcal{H}}$ ou $\text{dim}_{V.C.}(\mathcal{H})$ définie par :

$$d_{\mathcal{H}} = \text{dim}_{V.C.}(\mathcal{H}) = \sup_{d \in \mathbb{N}} \{d / \mathbb{S}_d(\mathcal{H}) = 2^d\} \in \mathbb{N} \cup \{+\infty\}$$

La V.C.-dimension d'un dictionnaire donne alors la quantité maximale de points de \mathcal{X} qui peuvent être étiquetés de toutes les façons possibles par les classifieurs de \mathcal{H} . Apparaît alors une élégante propriété combinatoire des coefficients d'éclatement qui fournit une majoration polynomiale de chacun d'entre eux, c'est ce que nous assure le lemme de Sauer !

3.2.2.1 Le lemme de Sauer

Lemme 5

Si \mathcal{H} un dictionnaire de V.C.-dimension finie $d_{\mathcal{H}}$, alors, en notant pour $k \in \mathbb{N}$ et $j \in \{0, \dots, k\}$, $\binom{k}{j} = \frac{k!}{(k-j)!j!}$, il vient :

$$\forall n \in \mathbb{N}, \quad \mathbb{S}_n(\mathcal{H}) \leq \sum_{i=0}^{d_{\mathcal{H}}} \binom{n}{i} \leq (n+1)^{d_{\mathcal{H}}}$$

Démo 7

On remarque tout d'abord que la deuxième inégalité est une conséquence de la première.

En effet, puisque l'on a que :

$$\forall i \in \{0, \dots, \min(n, d_{\mathcal{H}})\}, \quad \frac{n!}{(n-i)!} \leq \prod_{j=0}^{i-1} (n-j) \leq n^i,$$

mais aussi :

$$\forall i \in \{0, \dots, d_{\mathcal{H}}\}, \quad \frac{d_{\mathcal{H}}!}{(d_{\mathcal{H}}-i)!} \geq \prod_{j=0}^{i-1} (d_{\mathcal{H}}-j) \geq \prod_{j=0}^{i-1} 1 = 1$$

on déduit immédiatement, à partir de la première inégalité du lemme, la suite d'inégalité :

$$\begin{aligned} \mathbb{S}_n(\mathcal{H}) &\leq \sum_{i=0}^{d_{\mathcal{H}}} \binom{n}{i} \leq \sum_{i=0}^{d_{\mathcal{H}}} \frac{n!}{(n-i)!i!} \leq \sum_{i=0}^{d_{\mathcal{H}}} \frac{n^i}{i!} \\ &\leq \sum_{i=0}^{d_{\mathcal{H}}} \left(\prod_{j=0}^{i-1} (d_{\mathcal{H}} - j) \right) \frac{n^i}{i!} \leq \sum_{i=0}^{d_{\mathcal{H}}} \frac{d_{\mathcal{H}}!}{(d_{\mathcal{H}} - i)!i!} n^i \leq (n+1)^{d_{\mathcal{H}}} \end{aligned}$$

Établissons donc la première inégalité. A cette fin on se propose de démontrer par récurrence pour tout $k \in \mathbb{N}$ la propriété :

$$P(n) : \forall \mathcal{H} / \dim_{V.C.}(\mathcal{H}) < +\infty, \quad \mathbb{S}_n(\mathcal{H}) \leq \sum_{i=0}^{d_{\mathcal{H}}} \binom{n}{i}$$

Initialisation :

Pour tout dictionnaire \mathcal{H} tel que $d_{\mathcal{H}} = 0$, on a par convention :

$$\mathbb{S}_0(\mathcal{H}) = 1 \leq \sum_{i=0}^0 \binom{0}{i}$$

d'où $P(0)$ est vraie.

Pour tout dictionnaire \mathcal{H} , on a :

- Si $d_{\mathcal{H}} = 0$, alors, par définition de la V.C.-dimension, $\mathbb{S}_1(\mathcal{H}) = 1$ et par définition même du coefficient d'éclatement d'ordre 1, cela signifie que :

$$\forall x \in \mathcal{X}, \quad \forall (h, h') \in \mathcal{H}^2, \quad h(x) = h'(x)$$

d'où $\mathbb{S}_1(\mathcal{H}) = 1 \leq \binom{1}{0} = \sum_{i=0}^0 \binom{1}{i}$ et $P(1)$ est vraie.

- $d_{\mathcal{H}} = 0$, alors, cette fois-ci par définition de la V.C.-dimension, cela donne :

$$\mathbb{S}_1(\mathcal{H}) = 2 = \binom{1}{0} + \binom{1}{1} = \sum_{i=0}^1 \binom{1}{i} \leq \sum_{i=0}^{d_{\mathcal{H}}} \binom{1}{i}$$

d'où $P(1)$ est vraie.

Hérédité :

Soit n un entier naturel non nul quelconque tel que $\forall i \in \{0, \dots, n\}$, $P(i)$ est vraie. Soit alors \mathcal{H} un dictionnaire de V.C.-dimension finie.

- Si $d_{\mathcal{H}} = 0$, alors, par définition de la V.C.-dimension, $\mathbb{S}_1(\mathcal{H}) = 1$ et par définition même du coefficient d'éclatement d'ordre 1, cela signifie que :

$$\forall x \in \mathcal{X}, \quad \forall (h, h') \in \mathcal{H}^2, \quad h(x) = h'(x)$$

d'où on a que : $\forall j \in \mathbb{N}$, $\mathbb{S}_j(\mathcal{H}) = 1 \leq \binom{j}{0} = \sum_{i=0}^0 \binom{j}{i}$ et en particulier $P(n+1)$ est vraie.

- Sinon $d_{\mathcal{H}} \geq 1$ et on fixe alors $(x_1, \dots, x_n, x_{n+1}) \in \mathcal{X}^{n+1}$.
L'ensemble qui intervient dans le calcul d'un coefficient d'éclatement d'ordre $j \in \mathbb{N}^*$ est :

$$\mathcal{H}(x_1, \dots, x_j) \{ (h(x_1), \dots, h(x_j)) / h \in \mathcal{H} \}.$$

Comme $\forall h \in$

$\mathcal{H}, (h(x_1), \dots, h(x_n), h(x_{n+1}))$ dépend seulement des valeurs de h aux points x_1, \dots, x_n, x_{n+1} , on se ramène à considérer le dictionnaire $\mathcal{G} = \{h|_{\{x_1, \dots, x_n, x_{n+1}\}} / h \in \mathcal{H}\}$.

Ce nouveau dictionnaire \mathcal{G} vérifie en outre $\forall j \in \{0, \dots, n\}$ que :

$$\begin{aligned} \mathbb{S}_{j+1}(\mathcal{G}) &= \#\{(g(x_1), \dots, g(x_j), g(x_{j+1})) / h \in \mathcal{G}\} \\ &= \#\{(g(x_1), \dots, g(x_j), g(x_{j+1})) / h \in \mathcal{H}\} \\ &\leq \max_{(y_1, \dots, y_j, y_{j+1}) \in \mathcal{X}^{j+1}} \{\#\{(h(y_1), \dots, h(y_j), h(y_{j+1})) / h \in \mathcal{H}\}\} \leq \mathbb{S}_{j+1}(\mathcal{H}). \end{aligned}$$

Par définition de la V.C.-dimension, \mathcal{G} vérifie donc également que :

$$d_{\mathcal{G}} \leq d_{\mathcal{H}}.$$

On s'est ainsi ramené à considérer l'ensemble $\mathcal{X} = \{x_1, \dots, x_n, x_{n+1}\}$ et le dictionnaire \mathcal{G} .

On envisage alors le dictionnaire auxiliaire $\tilde{\mathcal{G}}$ défini par

$$\tilde{\mathcal{G}} = \{h \in \mathcal{G} / h(x_{n+1}) = 1 \text{ et } (h - 2 \times \mathbb{1}_{\{x_{n+1}\}}) \in \mathcal{G}\}.$$

Au passage, on remarque que :

$$h \in \mathcal{G} \setminus \tilde{\mathcal{G}} \iff [(h(x_{n+1}) = 1 \text{ et } (h - 2 \times \mathbb{1}_{\{x_{n+1}\}}) \notin \mathcal{G}) \text{ ou } (h(x_{n+1}) = -1)]$$

On a alors que $\mathcal{G} = \tilde{\mathcal{G}} \amalg (\mathcal{G} \setminus \tilde{\mathcal{G}})$ et donc

$$\begin{aligned} \mathcal{G}(x_1, \dots, x_n, x_{n+1}) &= \tilde{\mathcal{G}}(x_1, \dots, x_n, x_{n+1}) \amalg (\mathcal{G} \setminus \tilde{\mathcal{G}})(x_1, \dots, x_n, x_{n+1}) \quad \text{d'où} \\ \#(\mathcal{G}(x_1, \dots, x_n, x_{n+1})) &= \\ \#(\tilde{\mathcal{G}}(x_1, \dots, x_n, x_{n+1})) &+ \#((\mathcal{G} \setminus \tilde{\mathcal{G}})(x_1, \dots, x_n, x_{n+1})) \quad . \end{aligned}$$

Il nous reste alors seulement à trouver un ma-

ajorant de chacun des termes du membre de gauche de cette dernière égalité.

1. Comme $\forall g \in \tilde{\mathcal{G}}, g(x_{n+1}) = 1$ les ensembles $\tilde{\mathcal{G}}(x_1, \dots, x_n)$ et $\tilde{\mathcal{G}}(x_1, \dots, x_n, x_{n+1})$ sont en bijection et on a l'égalité :
 $\#(\tilde{\mathcal{G}}(x_1, \dots, x_n)) = \#(\tilde{\mathcal{G}}(x_1, \dots, x_n, x_{n+1}))$.
De plus, $\forall p \in \{1, \dots, n\}, \forall (x_{i_1}, \dots, x_{i_p}) \in \{x_1, \dots, x_n, x_{n+1}\}^p$ pouvant être étiquetés de toutes les façons possibles par $\tilde{\mathcal{G}}$, on a forcément $x_{n+1} \notin \{x_1, \dots, x_n, x_{n+1}\}$ puisque $\forall h \in \tilde{\mathcal{G}}, x_{n+1}$ ne peut être étiqueté que d'une seule manière : $h(x_{n+1}) = 1$.
De ce constat, il vient donc : $p \leq d_{\tilde{\mathcal{G}}} < (n+1)$.
Mais également, $\forall p \in \{1, \dots, n\}, \forall (x_{i_1}, \dots, x_{i_p}) \in \{x_1, \dots, x_n\}^p$ pouvant être étiquetés de toutes les façons possibles par \mathcal{G} , par définition de $\tilde{\mathcal{G}}$, l'ensemble de points $\{x_{i_1}, \dots, x_{i_p}, x_{n+1}\}$ peut être

étiqueté de toutes les façons possible par \mathcal{G} d'où : $p < d_{\mathcal{G}} + 1$.
Ainsi, comme on a vu que $d_{\tilde{\mathcal{G}}} \in \{1, \dots, n\}$, cette dernière inégalité nous assure que :

$$d_{\tilde{\mathcal{G}}} < d_{\mathcal{G}} + 1 \iff d_{\tilde{\mathcal{G}}} \leq d_{\mathcal{G}} - 1 \leq d_{\mathcal{H}} - 1$$

Aussi, par hypothèse de récurrence, on obtient la première majoration :

$$\#(\tilde{\mathcal{G}}(x_1, \dots, x_n)) = \#(\tilde{\mathcal{G}}(x_1, \dots, x_n, x_{n+1})) \leq \sum_{i=0}^{d_{\tilde{\mathcal{G}}}} \binom{n}{i} \leq \sum_{i=0}^{d_{\mathcal{H}}-1} \binom{n}{i}$$

2. Par la remarque faite sur les éléments de $\mathcal{G} \setminus \tilde{\mathcal{G}}$, il vient $\forall (g, g') \in (\mathcal{G} \setminus \tilde{\mathcal{G}})^2 / \forall i \in \{1, \dots, n\}, g(x_i) = g'(x_i) :$

$$[g(x_{n+1}) = g'(x_{n+1}) = 1 \quad \text{ou} \quad g(x_{n+1}) = g'(x_{n+1}) = -1].$$

En effet, si par exemple on avait h et h' vérifiant toutes les conditions suscitées tels que $h(x_{n+1}) = 1$ et $h'(x_{n+1}) = -1$, on aurait $g' = (g - 2 \times \mathbb{1}_{\{x_{n+1}\}}) \in \tilde{\mathcal{G}}$ ce qui est absurde puisque $g \in (\mathcal{G} \setminus \tilde{\mathcal{G}})$.

Cela nous assure donc que les ensembles $(\mathcal{G} \setminus \tilde{\mathcal{G}})(x_1, \dots, x_n)$ et $(\mathcal{G} \setminus \tilde{\mathcal{G}})(x_1, \dots, x_n, x_{n+1})$ sont en bijection et on a l'égalité :

$$\#((\mathcal{G} \setminus \tilde{\mathcal{G}})(x_1, \dots, x_n)) = \#((\mathcal{G} \setminus \tilde{\mathcal{G}})(x_1, \dots, x_n, x_{n+1})).$$

Par ailleurs, la relation $(\mathcal{G} \setminus \tilde{\mathcal{G}}) \subset \mathcal{G}$ et la définition des coefficients d'éclatement nous donne par propriété du *max* que : $d_{(\mathcal{G} \setminus \tilde{\mathcal{G}})} \leq d_{\tilde{\mathcal{G}}}$

Aussi, par hypothèse de récurrence, on obtient la deuxième majoration :

$$\#((\mathcal{G} \setminus \tilde{\mathcal{G}})(x_1, \dots, x_n)) = \#((\mathcal{G} \setminus \tilde{\mathcal{G}})(x_1, \dots, x_n, x_{n+1})) \leq \sum_{i=0}^{d_{(\mathcal{G} \setminus \tilde{\mathcal{G}})}} \binom{n}{i} \leq \sum_{i=0}^{d_{\mathcal{H}}} \binom{n}{i}$$

La combinaison de ces 2 majorations et le triangle de Pascal nous assurant que $\forall i \in \{1, \dots, n\} \binom{n}{i} + \binom{n}{i-1} = \binom{n+1}{i}$, nous donne alors :

$$\begin{aligned} \#(\mathcal{H}(x_1, \dots, x_n, x_{n+1})) &= \#(\mathcal{G}(x_1, \dots, x_n, x_{n+1})) \\ &\leq \sum_{i=0}^{d_{\mathcal{H}}-1} \binom{n}{i} + \sum_{i=0}^{d_{\mathcal{H}}} \binom{n}{i} \leq \sum_{i=1}^{d_{\mathcal{H}}} \binom{n}{i-1} + \sum_{i=0}^{d_{\mathcal{H}}} \binom{n}{i} \\ &\leq \binom{n}{0} + \sum_{i=1}^{d_{\mathcal{H}}} \binom{n+1}{i} \leq \sum_{i=0}^{d_{\mathcal{H}}} \binom{n+1}{i} \end{aligned}$$

En passant au sup sur $\{(x_1, \dots, x_n, x_{n+1}) \in \mathcal{X}^{n+1}\}$, on obtient donc :

$$\mathbb{S}_{n+1}(\mathcal{H}) \leq \sum_{i=0}^{d_{\mathcal{H}}} \binom{n+1}{i},$$

d'où $P(n+1)$ est vraie et la propriété est héréditaire.

En conclusion, on a donc démontré la première inégalité du lemme de Sauer :

$$\forall \mathcal{H} / \dim_{V.C.}(\mathcal{H}) < \infty, \quad \forall n \in \mathbb{N}, \quad \mathbb{S}_n(\mathcal{H}) \leq \sum_{i=0}^{d_{\mathcal{H}}} \binom{n}{i},$$

■

3.2.2.2 Quelques calculs simple de VC-dimensions

Dans cette section, on va montrer comment il est possible de déterminer la V.C.-dimension de dictionnaires de classifieurs définis sur \mathbb{R}^p assez simples : le dictionnaire des classifieurs linéaires, le dictionnaire des classifieurs affines et celui des classifieurs polygonaux convexes.

Théorème 8

Le dictionnaire des classifieurs linéaires $\mathcal{H}_{lin} = h_w : x \in \mathbb{R}^p \mapsto h_w(x) = \text{sign}(\langle w|x \rangle) / w \in \mathbb{R}^p$ admet p pour V.C.-dimension.

Démo 8

On va assez naturellement procéder par double inégalité : d'abord on va montrer qu'il existe un p -uplet de points et un classifieur linéaire (c'est à dire un vecteur $w \in \mathbb{R}^p$ tel que le classifieur soit h_w) qui peut être étiqueté de toutes les façons possibles à savoir de 2^p façons, puis on montrera par l'absurde que n'importe quel $(p+1)$ -uplet ne peut être classifié de 2^{p+1} façons.

- * D'une part, on considère le p -uplet constitué des vecteurs de la base canonique (e_1, \dots, e_p) de \mathbb{R}^p et montrons que ce dernier peut être étiqueté de 2^p façons.

Pour cela, $\forall \delta = (\delta_1, \dots, \delta_p) \in \{-1, 1\}^p$, on définit : $w_\delta = (\delta_1, \dots, \delta_p) \in \mathbb{R}^p$ et on constate alors que :

$$\forall \delta = (\delta_1, \dots, \delta_p) \in \{-1, 1\}^p, \forall j \in \{1, \dots, p\}, h_{w_\delta}(e_j) = \text{sign}(\langle w_\delta | e_j \rangle) = \text{sign}(\delta_j \times 1) = \text{sign}(\delta_j) = \delta_j$$

Par conséquent, le p -uplet (e_1, \dots, e_p) peut être étiqueté de 2^p manières d'où $\mathbb{S}_p(\mathcal{H}_{lin}) = 2^p$ et donc $\dim_{V.C.}(\mathcal{H}_{lin}) \geq p$

- * D'autre part, un $(p+1)$ -uplet $(x_1, \dots, x_p, x_{p+1}) \in (\mathbb{R}^p)^{p+1}$ constitue un système de $p+1$ vecteurs en dimension p donc on a que : $\exists (\lambda_1, \dots, \lambda_p, \lambda_{p+1}) \in (\mathbb{R}^{p+1} \setminus \{(0, \dots, 0)\}) / \sum_{i=1}^{p+1} \lambda_i x_i = 0$.

Sans perte de généralité, comme les λ_i sont non tous nuls, on peut supposer qu'il existe j_0 tel que $\lambda_{j_0} > 0$.

On considère alors le $(n+1)$ -uplet suivant

$$\delta = (\delta_1, \dots, \delta_p, \delta_{p+1}) \in \{-1, 1\}^{p+1} / \forall i \in \{1, \dots, p, (p+1)\}, \delta_i = \text{sign}(\lambda_i)$$

et supposons par l'absurde que :

$$\exists \tilde{w}_\delta / \forall j \in \{1, \dots, p, (p+1)\}, h_{\tilde{w}_\delta}(x_j) = \text{sign}(\langle \tilde{w}_\delta | x_j \rangle) = \text{sign}(\delta_j) = \delta_j$$

On aurait alors que :

$$h_{\tilde{w}_\delta}(\sum_{i=1}^{p+1} \lambda_i x_i) = \text{sign}(\langle \tilde{w}_\delta | 0 \rangle) = \text{sign}(0) = -1$$

mais par linéarité du produit scalaire et définition des δ_i et la stricte

positivité de λ_{j_0} on a que :

$$\begin{aligned} \left\langle \tilde{w}_\delta \left| \sum_{i=1}^{p+1} \lambda_i x_i \right. \right\rangle &= \sum_{i=1}^{p+1} \langle \tilde{w}_\delta | \lambda_i x_i \rangle = \sum_{i=1}^{p+1} \lambda_i \langle \tilde{w}_\delta | x_i \rangle \\ &= \sum_{i=1}^{p+1} \lambda_i \delta_i |\langle \tilde{w}_\delta | x_i \rangle| = \sum_{i=1}^{p+1} |\lambda_i| |\langle \tilde{w}_\delta | x_i \rangle| > 0 \end{aligned}$$

on aurait donc également :

$$h_{\tilde{w}_\delta} \left(\sum_{i=1}^{p+1} \lambda_i x_i \right) = \text{sign} \left(\left\langle \tilde{w}_\delta \left| \sum_{i=1}^{p+1} \lambda_i x_i \right. \right\rangle \right) = 1$$

ce qui est absurde.

On en déduit donc que le $(p+1)$ -uplet $(x_1, \dots, x_p, x_{p+1}) \in (\mathbb{R}^p)^{p+1}$ ne peut être étiqueté par δ donc ne peut pas être étiqueté de 2^{p+1} façons.

Comme le $(p+1)$ -uplet $(x_1, \dots, x_p, x_{p+1}) \in (\mathbb{R}^p)^{p+1}$ était quelconque, on en déduit que :

$\forall (x_1, \dots, x_p, x_{p+1}) \in (\mathbb{R}^p)^{p+1}$, $(x_1, \dots, x_p, x_{p+1})$ ne peut pas être étiqueté de 2^{p+1} façons.

Ainsi, par définition du coefficient d'éclatement d'ordre $(p+1)$, on a : $\mathbb{S}_p(\mathcal{H}_{\text{lin}}) < 2^{p+1}$.

Finalement, grâce à cette double inégalité, on a donc bien montré que le dictionnaire des classifieurs linéaires \mathcal{H}_{lin} admet p pour V.C.-dimension. \blacksquare

Théorème 9

Le dictionnaire des classifieurs affines

$$\mathcal{H}_{\text{aff}} = h_{w,b} : x \in \mathbb{R}^p \mapsto h_{w,b}(x) = \text{sign}(\langle w | x \rangle - b) / (w, b) \in \mathbb{R}^p \times \mathbb{R}$$

admet $(p+1)$ pour V.C.-dimension.

Démo 9

Comme avec les classifieurs linéaires, on va procéder par double inégalité : d'abord on va montrer qu'il existe un $(p+1)$ -uplet de points et un classifieur affine (c'est à dire un vecteur $w \in \mathbb{R}^p$ et un b réel tel que le classifieur soit $h_{w,b}$) qui peut être étiqueté de toutes les façons possibles à savoir de 2^{p+1} façons, puis on montrera par l'absurde que n'importe quel $(p+2)$ -uplet ne peut être classifié de 2^{p+2} façons.

* D'une part, on considère le $(p+1)$ -uplet constitué des vecteurs de la base canonique (e_1, \dots, e_p) de \mathbb{R}^p et du vecteur nul $(0, \dots, 0)$ et montrons que ce dernier peut être étiqueté de 2^{p+1} façons.

Pour cela, $\forall \delta = (\delta_1, \dots, \delta_p, \delta_{p+1}) \in \{-1, 1\}^{p+1}$, on définit :

$$— w_\delta = (w_1, \dots, w_p) \in \mathbb{R}^p \text{ où } \forall j \in \{1, \dots, p\}, w_j = \begin{cases} +2 & \text{si } \delta_j = 1 \\ -2 & \text{si } \delta_j = -1 \end{cases}$$

— $b_\delta = \langle w_\delta | a \rangle = \delta_{j+1} \in \mathbb{R}$ où $a = \delta_{j+1} \frac{w_\delta}{\|w_\delta\|_2} \in \mathbb{R}^p$

et on constate alors que $\forall \delta = (\delta_1, \dots, \delta_p) \in \{-1, 1\}^p, \forall j \in \{1, \dots, p\}$:

$$\begin{aligned} h_{w_\delta, b_\delta}(e_j) &= \text{sign}(\langle w_\delta | e_j \rangle - b) = \text{sign} \left(w_j \times 1 - \left\langle w_\delta \left| \delta_{j+1} \frac{w_\delta}{\|w_\delta\|_2} \right\rangle \right) \\ &= \text{sign}(w_j - \delta_{j+1}) = \begin{cases} \text{sign}(+2 - \delta_{j+1}) = 1 & \text{si } \delta_j = 1 \\ \text{sign}(-2 - \delta_{j+1}) = -1 & \text{si } \delta_j = -1 \end{cases} = \delta_j, \end{aligned}$$

et $h_{w_\delta, b_\delta}((0, \dots, 0)) = \text{sign}(\langle w_\delta | (0, \dots, 0) \rangle - \delta_{p+1}) = \delta_{p+1}$ Par conséquent, le $(p+1)$ -uplet $(e_1, \dots, e_p, (0, \dots, 0))$ peut être étiqueté de 2^{p+1} manières d'où $\mathbb{S}_{p+1}(\mathcal{H}_{lin}) = 2^{p+1}$ et donc $\dim_{V.C.}(\mathcal{H}_{lin}) \geq p+1$

* D'autre part, supposons qu'il existe un $(p+2)$ -uplet de vecteurs $(x_1, \dots, x_{p+1}, x_{p+2}) \in (\mathbb{R}^p)^{p+2}$ qui peut-être étiqueté de toutes les façons possibles. On a donc :

■

Théorème 10

Le dictionnaire des classifieurs polygonaux convexes

$$\mathcal{H}_{poly} = \{h_P : x \in \mathbb{R}^p \mapsto h_P(x) = 2 \times \mathbb{1}_P / P \in \{\text{polygone de } \mathbb{R}^p\}\}$$

admet $+\infty$ pour V.C.-dimension.

Démo 10

Il est nécessaire et suffisant de montrer que pour tout $n \in \mathbb{N}^*$ il existe un n -uplet de vecteurs distincts de \mathbb{R}^p qui peut être étiqueté de 2^n façons différentes.

Soit donc un n -uplet de vecteurs distincts (x_1, \dots, x_n) de la sphère unité de \mathbb{R}^p . On définit alors pour tout $\delta = (\delta_1, \dots, \delta_n) \in \{-1, 1\}^n$ le polygone P_δ comme l'enveloppe convexe des points x_j tels que $\delta_j = 1$ et le classifieur associé $h_P = \mathbb{1}_{P_\delta}$. On constate alors que :

$$\forall \delta = (\delta_1, \dots, \delta_n) \in \{-1, 1\}^n, \forall j \in \{1, \dots, n\}, h_{P_\delta}(x_j) = \delta_j$$

Ainsi, le n -uplet de vecteurs distincts (x_1, \dots, x_n) peut être étiqueté de 2^n manières différentes donc par définition du coefficient d'éclatement, $\mathbb{S}_n(\mathcal{H}_{poly}) = 2^n$. Comme n est quelconque, on en déduit donc finalement, par définition de la V.C.-dimension, que le dictionnaire \mathcal{H}_{poly} admet $+\infty$ pour V.C.-dimension.

■

3.3 Conclusion

Ce TER nous aura permis de mieux comprendre la démarche d'un chercheur. En effet, une grande partie du travail de ce dernier consiste à lire et comprendre des travaux ou articles déjà effectués ou rédigés par d'autres afin d'en tirer de nouveaux axes de recherche. Notre but était donc de comprendre l'article qui nous avait été donné par Christophe GIRAUD, de le restituer avec

nos mots, tout en essayant de développer certains points des démonstrations et de répondre à des questions plus ou moins simples en rapport avec le sujet.

Ce projet nous a fait découvrir un autre aspect des statistiques que celui qui nous est enseigné en cours de Statistiques en M1, ainsi qu'une des nombreuses applications de ce domaine des mathématiques. Avec plus de temps, il aurait été intéressant de se pencher sur les applications pratiques en informatique et essayer d'écrire un programme permettant des estimation plus efficaces comme vu en cours de MAO Proba Stats.

Nous avons appris par la même occasion à utiliser le langage \LaTeX dans le cadre d'un article de recherche, ce qui se révélera très utile par la suite.

Chapitre 4

Rapport de stage de Magistère

Remerciements

Dans un premier temps, je souhaite exprimer d'immenses remerciements à mon maître de stage Madame le Docteur Jutta Gampe pour l'ensemble de ce stage aussi bien dans la phase de candidature qu'au cours du stage. Sans son extrême bienveillance, j'aurais en effet eu de grandes difficultés tout d'abord à obtenir en si peu de temps une réponse positive pour un stage si enrichissant et ensuite à avoir la possibilité de découvrir autant dans le domaine des statistiques appliquées. De surcroît, malgré ses nombreuses tâches professionnelles, elle s'est efforcée de me consacrer du temps et m'apporter une aide personnalisée, ne me délaissant pas face à ma méconnaissance de beaucoup de notions.

Mes remerciements s'adressent également à l'Institut Max Planck pour la recherche démographique qui m'a fourni le cadre pour réaliser un tel stage et a proposé de financer le voyage et l'hébergement à Rostock ainsi que la participation au vingt-neuvième "International Workshop on Statistical Modelling".

Par ailleurs, je désire remercier l'ensemble du personnel de l'Institut Max Planck pour la recherche démographique pour leur chaleureux accueil et leur sympathie à mon égard. Plus particulièrement, je tiens à remercier Madame Annett Döpke pour les efforts déployés afin que le voyage et le séjour à Rostock se passent pour le mieux ainsi que pour son extrême gentillesse.

Finalement, je suis reconnaissant aux autres étudiants supervisés par le Docteur Gampe qui m'ont accueilli, accompagné, aidé et offert bien plus qu'une simple collaboration professionnelle : Kristen Cyffka, Marie Böhnstedt, Inês Raquel Campos Rodrigues et Silvia Rizzi.

4.1 Introduction

Ce rapport de stage a été réalisé à l'occasion de l'apprentissage hors-mur du Magistère de Mathématiques de l'université Paris-Sud 11. Le stage a été effectué du 23 juin au 1 août 2014 dans le département de démographie statistique de l'Institut Max Planck pour la recherche démographique sous la direction de Madame le Docteur Jutta Gampe.

A l'inverse des recommandations usuelles, j'ai préféré réaliser cette étape de ma formation à l'issue de la deuxième année de Magistère plutôt que de la première pensant que le niveau de compétences alors acquis me permettrait une meilleure compréhension des éventuels outils mathématiques abordés et une plus grande possibilité d'action au sein de l'organisme d'accueil.

Ma volonté de réaliser un stage à l'étranger et en particulier en Allemagne s'explique par la volonté de découvrir un environnement professionnel différent de celui français et par la volonté de perfectionner ma pratique de la langue allemande.

Finalement, les raisons de ma candidature à un stage pour cette institution étaient doubles : premièrement, je désirais découvrir le milieu de la recherche

institutionnelle en Allemagne et deuxièmement, ayant éprouvé un fort intérêt pour les domaines des Probabilités et des Statistiques pendant les deux dernières années de ma formation universitaire, je souhaitais découvrir la recherche appliquée dans l'un de ces domaines, critères que le département de démographie statistique de l'institut Max Planck pour la recherche démographique me semblait concilier.

La totalité du stage devait initialement être dédiée à des travaux de modélisation statistique et d'analyse de données. Néanmoins, suite à une discussion avec le maître de stage concernant les enjeux de cette partie de ma formation, ce dernier m'a proposé d'assister à une école d'été sur la modélisation statistique organisée annuellement par la SMS (Statistical Modelling Society) afin que je puisse découvrir un plus large éventail de domaines d'application des théories statistiques et rencontrer des chercheurs venant aussi bien d'institutions de recherche que d'entreprises internationales.

4.2 Développement

4.2.1 Présentation de l'Institut Max Planck pour la recherche démographique

4.2.1.1 La société Max Planck

La société Max Planck (*Max-Planck-Gesellschaft*) est une société scientifique allemande qui concrétise ses objectifs au travers de 80 instituts de recherche implantés aussi bien en Allemagne qu'à l'étranger. Fidèle à l'esprit de la société de l'empereur Guillaume (*Kaiser-Wilhelm-Gesellschaft*) qu'elle remplaça en 1948, sa politique consiste à soutenir les chercheurs les plus prometteurs dans tous les domaines de la recherche fondamentale en leur apportant le soutien financier, matériel et humain dont ils ont besoin.

Un des principes fondateurs de la société Max Planck est de soutenir au travers de ses divers instituts la recherche fondamentale au service de la communauté. Elle soutient plus particulièrement les sujets innovants gourmands en temps et en argent. Dans cette mesure, la société Max Planck est en évolution constante : elle ouvre des Instituts dans les domaines en marge mais avec des questions scientifiques prometteuses et en contrepartie en ferme d'autres, par exemple dans les domaines ayant trouvé un large intérêt auprès d'autres institutions plus académiques.

Pour autant, la société Max Planck n'omet pas la nécessité d'une recherche durable et s'implique beaucoup dans la formation des futures générations de chercheurs en sciences fondamentales. Entre autres, elle offre à de nombreux doctorants et post-doctorants allemands et étrangers de réaliser leurs travaux au sein des Instituts profitant ainsi d'un environnement exceptionnel pour l'issue de leur formation initiale. Mais la société Max Planck s'implique également dans le milieu académique en proposant à de jeunes professeurs d'université de diriger les groupes de recherche Max Planck et entretient d'étroits liens avec la société nationale de recherche scientifique allemande (*DFG : Deutsche Forschungsgemeinschaft*).

4.2.1.2 L'Institut Max Planck pour la recherche démographique

L'institut Max Planck pour la recherche démographique est un des nombreux instituts de la société Max Planck. Créé en 1995 par la société Max Planck avec la volonté de donner un nouvel élan à la démographie en Allemagne et de fortifier ses liens avec la communauté de recherche internationale, l'Institut Max Planck pour la recherche démographique démarre ses activités en 1996 avec la fondation du laboratoire de recherche sur le vieillissement et la longévité par le chercheur de renommée internationale James W. Vaupel. L'actuelle implantation physique de l'Institut avec ses locaux principaux à Rostock, ville de Poméranie occidentale située au Nord-Est de l'Allemagne, remonte quant à elle à 2003. La direction de l'Institut est actuellement confiée à son membre fondateur James W. Vaupel.

La recherche dans cet institut est axée sur la dynamique et la structure des populations et s'organise au sein de quelques groupes de recherche et de différents départements de recherche dont celui de démographie statistique où le stage donnant lieu à ce rapport a été effectué. Entre autres, conformément à la politique de la société Max Planck, les différents départements de recherche de cet institut s'intéressent à des thèmes importants pour la société actuelle tels que le changement démographique, le vieillissement, la dynamique des naissances, la répartition du temps de travail au cours de la vie aussi bien qu'à des aspects du vieillissement du point de vue de la biologie évolutive ou de celui de la médecine.

Dans ce contexte, l'Institut emploie environ 170 personnes dont 130 scientifiques et accueille de nombreux doctorants, post-doctorants et chercheurs invités. Ainsi, l'Institut Max Planck pour la recherche démographique est une des plus grosses institutions en recherche démographique d'Europe et compte parmi les instituts mondiaux de pointe dans cette discipline.

4.2.2 Modélisation statistique

4.2.2.1 L'analyse de survie

L'analyse de survie est l'étude du temps d'occurrence d'un certain événement d'intérêt. Dans le domaine démographique par exemple, on peut étudier l'âge d'une femme à sa première grossesse ou la durée de vie d'un individu après son passage à la retraite. Cela permet alors d'établir une distribution des temps de survie et éventuellement d'expliquer quantitativement la modification de ces distributions entre différentes situations au moyen de variables explicatives.

Les quelques lignes précédentes invitent donc à considérer des variables aléatoires réelles positives ou nulles pour la modélisation de cette durée de survie (aussi appelée temps de survie). On fait alors l'hypothèse supplémentaire qu'il s'agit d'une variable aléatoire de loi absolument continue. Dans la suite de cette partie, on notera X la variable aléatoire représentant ce temps de survie. Ainsi, déterminer la distribution du temps de survie revient à connaître la fonction de répartition du temps de survie ou encore à déterminer la densité du temps de vie qui en est sa dérivée presque partout. Dans le cadre de l'analyse de survie, afin d'avoir de meilleures interprétations des distributions, on introduit deux autres fonctions également caractéristiques de la distribution en probabilité du temps de survie :

- la fonction de survie S définie par :

$$\forall t \geq 0, S(t) = \mathbb{P}(t < X)$$

qui donne la probabilité que l'événement d'intérêt survienne après l'instant t .

- la fonction de risque h définie par :

$$\forall t \geq 0, h(t) = \lim_{\substack{h \rightarrow 0 \\ h > 0}} \frac{\mathbb{P}(t \leq X \leq t + h | t \leq X)}{h}$$

qui correspond au risque instantané d'occurrence de l'événement sachant qu'il n'a pas encore eu lieu avant le temps t .

Néanmoins, lors de la collecte pratique de données de survie, on n'obtient pas toujours des variables aléatoires indépendantes et identiquement distribuées de loi celle choisie pour le modèle. En effet, la prise de résultat étant soumise à diverses contraintes expérimentales, on observe des réalisations « perturbées » de la variable aléatoire que l'on s'est fixée : il peut y avoir *censure* et/ou *troncature*.

D'une part, dans une population étudiée fixée, il se peut que le temps de survie ne soit pas directement observable pour tous les individus car l'événement d'intérêt a déjà eu lieu, on parle de *censure à gauche*, ou n'a pas encore eu lieu, on parle de *censure à droite*. Par exemple, si on s'intéresse aux humains nés entre 1960 et 1970, en commençant en 2000 mais en arrêtant l'étude en 2002, il y aura eu des personnes qui seront déjà décédées dont on ne connaîtra pas l'âge de décès, des personnes qui seront décédées durant l'expérience dont on connaîtra l'âge de décès et des personnes qui ne seront pas encore décédées et dont on ne connaîtra pas l'âge de décès. Ainsi, si le temps de survie étudié

était l'âge de décès, pour les personnes décédées avant 2000, on ne disposerait que de l'âge maximal jusqu'au quel elles ont potentiellement vécu, pour les personnes décédées dans la période 2000-2002, on disposerait à la fin de l'étude du véritable âge de décès (ces réalisations seraient non censurées) tandis que pour celles toujours en vie, on ne disposerait que de l'âge atteint à l'issue de l'étude (ces réalisations seraient censurées à droite).

D'autre part, dans une population fixée, il se peut que l'on ne considère un individu et son temps de survie que si l'événement d'intérêt a lieu au sein du temps d'observation. En reprenant l'exemple précédent, si on s'intéressait aux humains nés entre 1960 et 1970, en ne considérant que les individus dont le décès a eu lieu entre 2000 et 2002, on ignorerait tous les individus décédés avant 2000 et tout ceux qui décèderont après. En ne considérant pas ceux déjà décédés, on aurait fait une "*troncature à gauche*" et en ne considérant pas ceux dont la mort est à venir, on aurait procédé à une "*troncature à droite*".

Dans tout ce rapport, on considérera que les temps de survie, de censure et de troncature sont indépendants entre eux, ce qui permettra d'utiliser les méthodes classiques d'estimation d'analyse de survie.

Ces « perturbations » des données que sont la censure et la troncature introduisent bien sûr une certaine perte d'information avec laquelle on doit judicieusement procéder lors de l'analyse de données pour utiliser au mieux l'information que chaque donnée fournit.

Afin de faciliter la tâche, on introduit d'autres variables auxiliaires qui indiquent si les données ont été censurées et fournissent d'autres éléments utiles.

Dans le cas de la censure à droite par exemple, pour l'individu d'indice i , en notant X_i son temps de survie et C le temps de censure, on introduit les deux quantités suivantes :

- la durée réellement observée $T_i = \min(X_i, C)$
- l'indicatrice de censure à droite $\delta_i = \mathbb{1}_{X_i \leq C}$

Ainsi, si l'observation a été censurée à droite δ_i vaudra 0 et on aura $T_i = C < X_i$ tandis que si elle ne l'a pas été, δ_i vaudra 1 et on aura $T_i = X_i \leq C$.

Dans le contexte statistique, on collecte l'information fournie par les observations au moyen des fonctions de vraisemblance au travers desquelles on peut tenter d'estimer les paramètres du modèle. Le bon fonctionnement de ces procédures étant donc conditionné par une expression exacte de ces fonctions de vraisemblance, il est raisonnable d'exposer ici cette problématique.

On considère toujours des observations résultant d'une réalisation de variables aléatoires réelles positives ou nulles indépendantes et identiquement distribuées de même loi que X de densité f ayant été soumises à la censure et à la troncature sous l'hypothèse que les temps de survie et les temps de censure sont indépendants. Dans chacun des cas de censure et de troncature, on va justifier l'expression de la contribution de l'observation à la fonction de vraisemblance l .

Pour une observation correspondant effectivement à un temps de survie, on reçoit l'information sur la probabilité que l'événement d'intérêt a lieu à un moment t , qui est la valeur de la densité f en cet instant t . La contribution à la fonction de vraisemblance d'une observation x_i ni censurée ni tronquée sera donc la valeur de la fonction densité en x_i :

$$l = f(x_i).$$

Pour une observation d_i censurée à droite, ce que l'on sait est seulement que le moment d'occurrence de l'événement est plus grand que ce temps observé, il s'agit donc d'information sur la probabilité que l'événement soit ultérieur à ce temps d_i : $\mathbb{P}(d_i < X)$. La contribution à la fonction de vraisemblance d'une observation d_i seulement censurée à droite sera donc la valeur de la fonction de survie S en d_i :

$$l = S(d_i).$$

Pour une observation g_i censurée à gauche, ce que l'on sait est seulement que l'événement a déjà eu lieu et par conséquent le moment d'occurrence de l'événement est plus petit que le temps observé, il s'agit donc d'information sur la probabilité que l'événement soit antérieur à ce temps g_i : $\mathbb{P}(X \leq g_i)$. La contribution à la fonction de vraisemblance d'une observation g_i seulement censurée à gauche sera donc la valeur de la fonction de répartition F en g_i :

$$l = F(g_i) = 1 - S(g_i).$$

Pour une observation r_i donnée lorsqu'il y a un phénomène de troncature à droite, ce que l'on connaît est le temps de survie effectif sachant que l'événement est antérieur au temps de troncature à droite, l'information apportée est donc la probabilité que l'événement a eu lieu au temps r_i conditionnellement au fait qu'il a eu lieu avant le temps de troncature à droite. La contribution à la fonction de vraisemblance d'une observation r_i alors qu'il existe un phénomène de troncature à droite sera donc la valeur de la fonction densité en r_i divisée par la probabilité que l'événement ai eu lieu avant le temps de troncature à droite R :

$$l = \frac{f(r_i)}{\mathbb{P}(X < R)} = \frac{f(r_i)}{1 - S(R)}.$$

Pour une observation l_i donnée lorsqu'il y a un phénomène de troncature à gauche, ce que l'on connaît est le temps de survie effectif sachant que l'événement est ultérieur au temps de troncature à gauche, l'information apportée est donc la probabilité que l'événement a eu lieu au temps l_i conditionnellement au fait qu'il a eu lieu après le temps de troncature à gauche. La contribution à la fonction de vraisemblance d'une observation l_i alors qu'il existe un phénomène de troncature à gauche sera donc la valeur de la fonction densité en l_i divisée par la probabilité que l'événement ai eu lieu après le temps de troncature à droite L :

$$l = \frac{f(l_i)}{\mathbb{P}(L < X)} = \frac{f(l_i)}{S(L)}.$$

Pour une observation x_i donnée lorsqu'il y a simultanément un phénomène de troncature à droite et un phénomène de troncature à gauche, ce que l'on connaît est le temps de survie effectif sachant que l'événement est antérieur au

temps de troncature à droite et ultérieur au temps de troncature à gauche, l'information apportée est donc la probabilité que l'événement a eu lieu au temps x_i conditionnellement au fait qu'il a eu lieu avant le temps de troncature à droite et après le temps de troncature à gauche. La contribution à la fonction de vraisemblance d'une observation x_i alors qu'il existe un phénomène de troncature à gauche sera donc la valeur de la fonction densité en x_i divisée par la probabilité que l'événement ai eu lieu entre les deux temps de troncature R et L :

$$l = \frac{f(x_i)}{\mathbb{P}(L < X < R)} = \frac{f(x_i)}{S(L) - S(R)}.$$

Ainsi, disposant de l'expression de la contribution d'une observation à la fonction de vraisemblance dans chacun des cas de figure et les observations étant supposées indépendantes, l'expression finale de la vraisemblance est le produit de la contribution en fonction de sa « perturbation » de chacune des observations.

4.2.2.2 Travaux réalisés

Au fur et à mesure du stage, mon maître de stage m'a confié des tâches de programmation d'une complexité progressive. Celles-ci consistaient principalement à rédiger des programmes en langage R ayant pour but d'estimer les paramètres d'un modèle démographique et de déterminer le type de données requises pour une bonne estimation en ayant préalablement simulé des données. Toutes ces étapes constituent les préliminaires d'un travail d'analyse de données sur données réelles à des fins démographiques auxquelles je n'ai pas été confronté pendant le stage.

Comme les bases de la théorie en ont été exposées ci-dessus, j'ai été confronté dans chacune des différentes tâches que j'ai dû exécuter à des situations dont la résolution nécessitait l'application de l'analyse de survie. En effet, le cadre dans lequel je me trouvais était celui de durée de vie pour des êtres humains. En général, on considérait que ces derniers au nombre de n avaient des dates de naissances dont la distribution probabiliste était uniforme sur un intervalle de temps $[t_0, t_1]$ où $t_0 < t_1$ et que leurs durées de vie étaient des variables aléatoires réelles positives ou nulles indépendantes et identiquement distribuées de loi de Gompertz, la loi de Gompertz étant la loi de probabilité caractérisée par une fonction de risque définie par :

$$\forall x \in \mathbb{R}, \quad h(x) = ae^{bx} \mathbb{1}_{\mathbb{R}^+}(x).$$

Les étapes premières de chaque programme consistaient alors à générer n telles données pour $n \in \{200, 2000, 10000\}$ et ensuite à les traiter de telle sorte à se placer dans des situations où l'on ajoutait des phénomènes de censure et/ou de troncature.

L'étape suivante consistait alors à programmer une procédure estimant les paramètres du modèle choisi à partir des données. C'est à ce point que la distinction entre observations censurées ou non et tronquées ou non était cruciale puisque l'on cherchait à estimer les paramètres d'un modèle par la méthode du maximum de vraisemblance. En effet, cette méthode d'estimation des paramètres d'un modèle se base sur l'assomption que les paramètres les plus plausibles sont ceux pour lesquels la densité a la valeur la plus forte. Pour l'appliquer, il était donc indispensable de calculer la densité du modèle choisi pour n'importe quelle observation de la variable aléatoire modélisée (censurée ou pas et tronquée ou pas) et pour n'importe quel paramètre, le problème se réduisant alors à une recherche de maximum : il suffisait, le jeu de données étant posé, de maximiser la fonction de vraisemblance en le paramètre.

Néanmoins, les recherches de maximum se sont à plus d'une reprise révélées problématiques pour moi, bien que j'eusse utilisé une procédure d'optimisation procédant par descente du gradient déjà implémentée sous R pour les recherches de ce maximum.

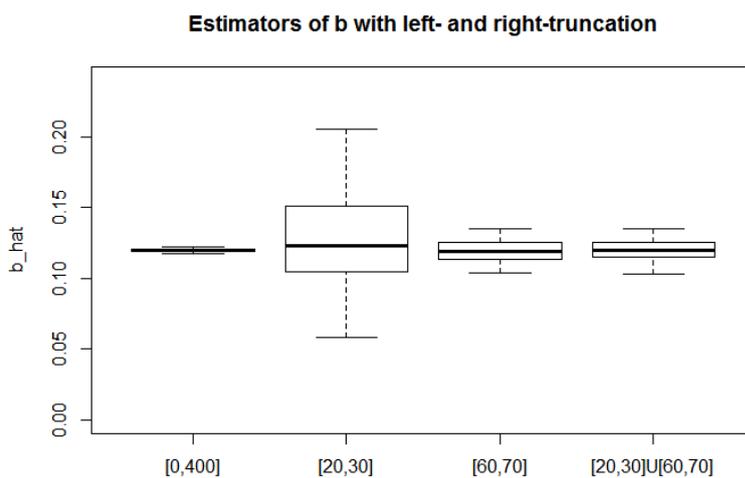
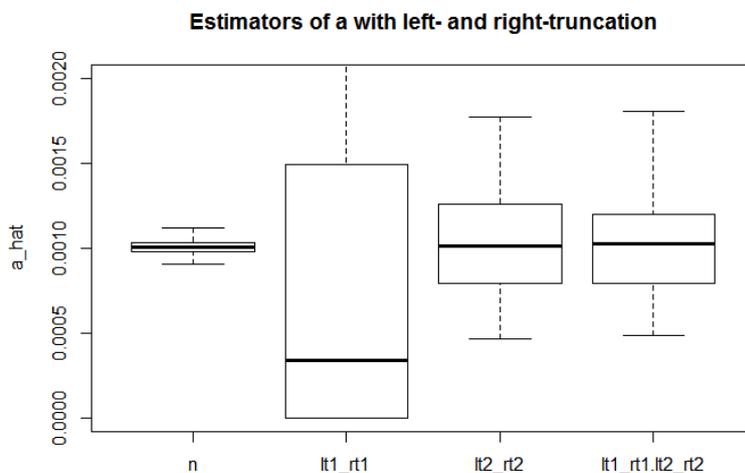
La première difficulté rencontrée a été celle du domaine des paramètres. Le modèle de Gompertz étant un modèle paramétré par un couple de réels positifs, il était nécessaire au cours des différentes étapes de l'algorithme calculant à chaque fois la valeur de la vraisemblance en le jeu de données fixé et le couple de paramètres actualisés à chaque itération que cette positivité des paramètres soit vérifiée. Cependant, cela n'était pas le cas en raison de la méthode d'op-

timisation, ce qui conduisait à des erreurs de procédure. Aussi, il a été nécessaire de changer l'espace des paramètres en procédant à une correspondance bijective entre \mathbb{R}^{*+} et \mathbb{R} au moyen de la fonction logarithme. Cela permettait d'obtenir une estimation des paramètres du nouveau domaine maximisant la vraisemblance et de retourner au domaine initial des paramètres par application de la fonction exponentielle mais contraignait à l'utilisation de résultats plus complexes tels que la méthode Delta si l'on souhaitait obtenir des résultats asymptotiques comme par exemple l'information de Fisher du modèle.

La deuxième difficulté pour moi résultait d'un manque de méthode et de pratique en programmation. Lorsque je devais illustrer la qualité de mes approximations par des graphiques, je ne stockais préalablement pas les données sous forme "brute" si bien que les calculs de maximum étant souvent long, je devais attendre plusieurs heures avant d'obtenir une représentation de mes résultats sans même avoir la possibilité de reproduire le graphique, ni de contrôler si certains des maxima calculés n'étaient pas anormaux. A mes dépens, j'ai donc appris qu'il était plus judicieux de stocker dans des fichiers de données externes au code les données générées et les résultats obtenus afin de pouvoir reproduire à l'identique les simulations et les réutiliser rapidement pour des illustrations sans avoir à tout recalculer.

Quant à la dernière difficulté majeure que j'ai eu, je ne suis pas certain que son origine première soit la procédure de maximisation : c'est une absence de convergence de l'algorithme d'optimisation qui m'est apparue la plus suspecte alors que je cherchais une erreur dans le dernier programme réalisé durant mon stage. Malgré une semaine de recherche de mon erreur, je ne suis parvenu à déterminer l'origine de cette absence de convergence.

Dans tous les cas, je suis parvenu à construire des procédures donnant des estimateurs des paramètres du modèle dans les différentes situations de « perturbation » des observations et à voir la perte d'information qu'induisaient celles-ci. L'observation de ces pertes d'informations m'a été possible grâce à l'utilisation d'une représentation graphique à laquelle je n'étais pas habitué : *le diagramme en boîte* ou *boîte à moustaches*. Il s'agissait en effet de répéter la procédure d'estimation 100 fois pour un échantillon de n individus et de réaliser le diagramme des estimateurs calculés.



La figure ci-dessus représente les boîtes à moustaches obtenues pour les 2 paramètres a et b du modèle de Gompertz avec 100 répétitions pour un jeu de données de $n = 10000$ individus nés dans l'intervalle $[t_0, t_1] = [0, 80]$ dans quatre situations :

1. aucune troncature ;
2. troncatures gauche et droite sur l'intervalle $[20, 30]$;
3. troncatures gauche et droite sur l'intervalle $[60, 70]$;
4. troncatures gauche et droite sur les deux intervalles $[20, 30]$ et $[60, 70]$.

Une telle figure permet de faire ressortir directement la plus ou moins bonne qualité de l'estimation des paramètres en fonction du type de données dont on dispose : on voit à la fois si l'estimateur est biaisé ou pas et si la variance empirique de celui-ci semble élevée ou pas.

Sur la figure ci-dessus, on constate que l'estimateur du maximum de vraisemblance a un biais quasiment identique dans toutes les situations. Cependant, en comparant avec la 1^{re} situation, on remarque que dans les 3^e et 4^e situations la variance empirique est un peu plus grande tandis que dans la 2^e situation, la variance empirique est beaucoup plus importante !

Mais ces différences sont tout à fait justifiables. Comme déjà évoqué, lors de « perturbations » telles que la censure ou la troncature, il existe une perte d'information par rapport à la réalité. Le problème modélisé ainsi que le modèle choisi permettent également de justifier les divergences entre les diverses situations de troncature : il s'agit des durées de vie d'individus nés dans l'intervalle $[0, 80]$ et le modèle de Gompertz ainsi que les paramètres a et b choisis pour simuler les données assurent une espérance de vie de l'ordre de 50 ans.

Ainsi, la différence entre la 1^{re} situation, qui est celle où l'on dispose de toutes les données non perturbées pour faire l'estimation, et les autres situations est imputable à la perte d'information introduite par les troncatures gauche et droite tandis que celle observée entre la 2^e et les 3^e et 4^e situations est due au fait que la troncature de la 2^e situation introduit une perte d'information nettement plus importante que dans les 3^e et 4^e en raison du problème modélisé et du choix du modèle (très peu d'individus nés dans l'intervalle de temps $[t_0, t_1]$ décèdent dans l'intervalle de temps $[20, 30]$).

La simulation donnée en exemple ci-dessus donne donc un aperçu du type de « perturbation » des données qui peut influencer sur la qualité de l'estimation : dans le cas précis, il s'agit de la localisation de l'intervalle de troncature des données. Mais j'ai également pu observer d'autres types de « perturbation » des données tels que la largeur de l'intervalle de troncature ou la position du temps de censure.

Ces pertes d'informations dans les données dont on dispose, introduites ici pour le besoin de l'étude de façon artificielle, sont récurrentes dans les situations pratiques de par la méthode d'observation, la durée de la période d'observation ou encore la réception de données incomplètes en raison de clause de confidentialités.

C'est à ce problème que se sont donc attachés mes travaux pendant le stage et j'aurais apprécié pouvoir poursuivre ce travail de manière plus avancée afin de voir les retombées pratiques d'une telle recherche.

4.2.3 “29th International Workshop on Statistical Modelling”

Au cours de la quatrième semaine de mon stage grâce à la proposition de Madame le Docteur Gampe et au financement de l’Institut Max Planck envers lesquels je suis extrêmement reconnaissant, j’ai eu l’opportunité d’assister au “29th International Workshop on Statistical Modelling” organisé par la SMS. Cet événement était le premier rassemblement scientifique auquel je participais et a par conséquent été riche d’enseignements.

Tout d’abord, la découverte du déroulement d’un séminaire scientifique était une expérience totalement nouvelle pour moi. Entre autres, j’y ai pris conscience de l’importance des temps plus informels entrecoupant les sessions de conférences : ils m’ont paru être un vif milieu d’échange d’idées bénéfique pour les futurs travaux de recherche de chaque participant.

Par ailleurs, le grand nombre de présentations orales au cours du séminaire m’a permis, de par mon impression personnelle puis de par l’intérêt suscité chez les autres auditeurs lors du temps de question, de constater l’importance de la présentation des travaux de recherche au cours du travail d’un scientifique même auprès de ses semblables ! Une excellente maîtrise de la langue anglaise, une bonne capacité de synthèse et d’illustration, un support visuel agréable à l’auditeur m’ont ici paru être des éléments cruciaux pour susciter l’intérêt du plus grand nombre d’auditeurs d’un milieu scientifique amenant ainsi des questions potentiellement riches pour un travail futur.

En outre, le caractère appliqué des problèmes auxquels certains orateurs s’étaient intéressés, m’a donné à découvrir de nombreux nouveaux problèmes auxquels je n’avais jamais été confronté lors de la résolution de problèmes plus théoriques : nécessité d’avoir un domaine de recherche des paramètres stable sous l’utilisation de méthodes d’approximation numériques, nécessité de pouvoir apporter une interprétation simple de la valeur des paramètres d’un modèle, nécessité de connaître dans le détail les conditions de mesure des données expérimentales à modéliser afin d’expliquer certains résultats curieux, ...

Finalement, malgré une compréhension lacunaire des diverses présentations orales, j’ai pu découvrir de nombreux domaines appliqués où la modélisation statistique et les Statistiques jouaient un rôle crucial : la prévision de la pollution de l’air en milieu urbain, la modélisation du rendement de production de viande porcine en Union européenne, la prévision météorologique quand les méthodes numériques sont trop instables, les études géologiques, la modélisation des résultats académiques des étudiants en fonction de facteurs socioéconomiques, ...

4.3 Conclusion

Conformément à l'esprit de l'option *Apprentissage hors-mur* du Magistère de Mathématiques de l'université Paris-Sud 11, ce stage m'aura permis de découvrir sous un angle nouveau les mathématiques et de constater l'utilisation effective des théories mathématiques dispensées au sein des enseignements de l'université.

Néanmoins, j'ai l'impression que ce stage aura également eu d'autres retombées bénéfiques pour moi.

Tout d'abord, le nombre de concepts mathématiques nouveaux que j'ai été amené à rencontrer et utiliser lors de ce stage m'a non seulement apporté de nouvelles connaissances mais aussi appris à ne pas vouloir nécessairement comprendre les choses d'emblée dans le détail tant la quantité de choses nouvelles était importantes.

Par ailleurs, ayant été amené au cours de ce stage à devoir commenter mes programmes, à la fois pour une utilisation future par d'autres personnes et pour demander de l'aide à mon maître de stage ou à un autre membre de l'équipe pour les tâches de programmation, j'ai pris conscience de la difficulté que je pouvais éprouver à expliquer certaines de mes idées de manière claire et concise. Cette constatation combinée à l'importance de la présentation limpide du travail d'un chercheur que j'ai pu remarquer lors du "*29th International Workshop on Statistical Modelling*", m'ont apporté un élément de décision crucial pour mon orientation. Hésitant en effet jusqu'alors entre réaliser d'emblée un Master 2 Recherche ou réaliser au préalable un Master 2 de préparation à l'agrégation de Mathématiques, j'ai désormais l'impression qu'une année de préparation à l'agrégation pourrait me permettre de développer une plus grande aisance à exprimer simplement et clairement mes propres idées, ce que je considère comme utile pour la suite de ma vie professionnelle.

Finalement, ce stage a également répondu à mes attentes. J'ai pu y perfectionner mes compétences linguistiques et, de manière surprenante, nettement plus celles en langue anglaise que celle en langue allemande. Mais surtout, il m'a d'une certaine manière permis de découvrir la recherche institutionnelle en Allemagne et d'aborder un domaine où les Probabilité et Statistique étaient utilisées de façon relativement appliquée.

Chapitre 5

TER de Master 2 Agrégation

Préambule

Ce mémoire intitulé *Survie d'une population soumise à hétérogénéité spatiale* a été réalisé en binôme dans le cadre du module d'enseignement « Stage Recherche » de la filière dédiée à la formation des professeurs agrégés du Master 2 de Mathématiques Fondamentales et Appliquées du département de mathématiques de l'université Paris-Sud 11.

L'ensemble du travail a été réalisé sous la tutelle de Christophe GIRAUD, professeur à l'université Paris-Sud 11 et chargé de cours à l'École Polytechnique d'octobre 2014 à mai 2015. L'ensemble de ce mémoire se base principalement sur un article intitulé *Past, growth and persistence of source-sink metapopulations* de Vincent BANSAYE et Amaury LAMBERT publié en novembre 2011.

Ce mémoire ne se veut nullement exhaustif quant à l'étude de la survie d'une population dans un contexte d'hétérogénéité spatiale. L'ensemble du travail réalisé ayant en effet pour principal dessein de confronter les étudiants à la démarche de la recherche en mathématiques et aux outils dont celle-ci dispose tant pour sa réalisation que pour sa diffusion afin de conforter les étudiants dans leur désir de s'orienter vers la recherche dans ce domaine.

5.1 Introduction

Lorsque l'on s'intéresse à l'évolution de la taille d'une population, il est légitime de se demander si celle-ci va s'éteindre à terme ou non et plus généralement de savoir si l'un ou l'autre de ces deux événements a une probabilité positive de survenir. Étant donné le caractère spatial introduit par les déplacements des individus de la population et l'influence de l'habitat sur la capacité de reproduction de chacun d'eux, il semble judicieux de tenir compte de cette hétérogénéité spatiale introduite par les migrations et les capacités de reproduction variables. On aimerait alors connaître une manière simple de savoir si une population soumise à une telle hétérogénéité spatiale a une chance de se perpétuer. La connaissance de tels critères a notamment son intérêt dans le cadre de la protection des espèces ou de la gestion du réchauffement climatique.

5.1.1 Modélisation de la situation

On considère un ensemble d'habitats appelés "*lieux d'habitats*" qui peuvent être de différents types appelés "*types d'habitats*". Les types d'habitats correspondent aux différentes lois de reproduction qui existent selon la qualité du lieu d'habitat.

On attribue à chaque lieu d'habitat un entier différent $i \in \{1, \dots, K\}$ où $K \in \mathbb{N}^*$ de sorte que le modèle spatial peut se représenter par un graphe fini à arêtes orientées et pondérées dont un sommet symbolise un lieu d'habitat et dont une arête orientée du sommet i vers le sommet j et pondérée par le poids $M_{i,j}$ symbolise la possibilité de migration du lieu d'habitat i vers le lieu d'habitat j avec une probabilité $M_{i,j}$.

On suppose que les individus des populations étudiées ont un mode de reproduction asexué et que les moments de reproduction de tous les individus ont lieu simultanément à des instants discrets que l'on indexera par $n \in \mathbb{N}$. A chaque instant $n \in \mathbb{N}$, chaque individu de la génération n donne naissance à un nombre aléatoire d'individu selon une loi qui dépend uniquement du type d'habitat dans lequel cet individu "*parent*" vit à cet instant et est indépendant des autres individus que ceux soient ceux des générations antérieures ou ceux nés en même temps que lui et ce, tous lieux d'habitat confondus. Par ailleurs, chacun des individus "*nouveaux-nés*" survit jusque l'instant qui suit directement son instant de naissance et, après avoir donné naissance à sa descendance (éventuellement vide), cet individu devenu parent décède, de sorte qu'en chaque instant discret, on observe seulement les individus de la nouvelle génération.

Finalement, on suppose que chacun des nouveaux-nés d'un même lieu d'habitat i migre instantanément après sa naissance et indépendamment des autres individus vers le lieu d'habitat j avec probabilité $M_{i,j}$ de sorte qu'en chaque instant discret, on observe les individus de la nouvelle génération dans leur nouveau (mais néanmoins définitif) lieu d'habitat.

Pour tout $i \in \{1, \dots, K\}$, on note μ_i la loi de reproduction d'un individu résidant dans le lieu d'habitat indexé par i et m_i la moyenne de μ_i qui est donc le nombre moyen de descendants d'un individu du lieu d'habitat i . On

dit alors qu'un lieu d'habitat est un puits si $m_i \leq 1$ et que c'est une source si $m_i > 1$. Pour des raisons de simplification, on fera l'hypothèse additionnelle que $m_1 \geq m_2 \geq \dots \geq m_{K-1} \geq m_K$. Les situations dignes d'intérêt étant celles où l'on ne considère pas exclusivement des sources ou exclusivement des puits, c'est à dire celles où l'on a au moins $m_1 > 1$ et $m_K \leq 1$.

Définition 9

Un processus de branchement (ou de Galton-Watson) multitype à K types ($K \in \mathbb{N}$) est une suite de variables aléatoires $(Z_n)_{n \in \mathbb{N}}$ à valeurs dans \mathbb{N}^K défini par :

$$\forall n \in \mathbb{N}, Z_{n+1} = \sum_{i=1}^K \sum_{j=1}^{Z_n^i} \xi_{n,j}(i)$$

où $(\xi_{n,j}(i))_{(i,j,n) \in \{1, \dots, K\} \times \mathbb{N}^* \times \mathbb{N}} \in \mathbb{N}^K$ est une suite de variables aléatoires indépendantes telles que :

$$\forall (j, n) \in \mathbb{N}^* \times \mathbb{N}, \xi_{n,j}(i) \text{ a pour loi } \nu_i \text{ où } \nu_i \text{ est une loi sur } \mathbb{N}^K.$$

La variable aléatoire Z_n d'un processus de branchement multitype peut s'interpréter comme les nombres des descendants de types différents à la génération n d'une population d'individus de K types distincts dont le mécanisme de reproduction est tel qu'à chaque génération, un individu de la génération précédente donne naissance à des individus des K types selon la loi ν_i indépendamment des générations précédentes et des autres individus de la même génération qu'ils soient du même type ou non.

Ainsi, dans notre modèle, si pour tout $l \in \{1, \dots, K\}$, on désigne par Z_n^l le nombre d'individus résidant dans les lieux d'habitat de type l à la n -ième génération et Z_n le vecteur (Z_n^1, \dots, Z_n^K) qui décrit exactement la constitution de la population de la n -ième génération en terme de types d'habitat, on a :

$$Z_{n+1} = \sum_{i=1}^K \sum_{j=1}^{Z_n^i} \xi_{n,j}(i)$$

où $(\xi_{n,j}(i))_{(i,j,n) \in \{1, \dots, K\} \times \mathbb{N}^* \times \mathbb{N}} \in \mathbb{N}^K$ est une suite de variables aléatoires indépendantes telles que pour tout $(j, n) \in \mathbb{N}^* \times \mathbb{N}$, $\xi_{n,j}(i)$ a pour loi ν_i où ν_i est la loi sur \mathbb{N}^K de "reproduction-migration" d'un individu résidant dans le lieu d'habitat de type i , c'est à dire, la loi après migration du nombre des K types de descendants d'un parent résidant dans le lieu d'habitat de type i . Par conséquent, la suite $(Z_n)_{n \in \mathbb{N}}$ est un processus de branchement multitype à K types.

Définition 10

On appelle marche aléatoire sur un graphe fini orienté à K sommets dont les arêtes sont pondérées par des probabilités la chaîne de Markov à valeurs dans l'ensemble des sommets du graphe de noyau de transition la matrice M , où M est la matrice stochastique dont le coefficient $M_{i,j}$ d'indice (i, j) est le poids sur l'arête $i \rightarrow j$ si celle-ci existe et 0 sinon.

Dans notre modèle, on considérera une marche aléatoire $(X_n)_{n \in \mathbb{N}}$ sur le graphe modélisant la situation, c'est à dire que $(X_n)_{n \in \mathbb{N}}$ sera une chaîne de Markov de probabilité de transition donnée par :

$$\forall (i, j) \in \{1, \dots, K\}^2, \mathbb{P}(X_{n+1} = j | X_n = i) = M_{i,j}.$$

On supposera en outre qu'il existe un entier $i_0 \in \{1, \dots, K\}$ tel que $M_{i_0, i_0} > 0$ et que pour tout (i, j) dans $\{1, \dots, K\}^2$, il existe $n_{i,j} \in \mathbb{N}^*$ tel que le coefficient d'indices (i, j) de la matrice $M^{n_{i,j}}$ est strictement positif. Les conséquences de ces hypothèses sur la marche aléatoire $(X_n)_{n \in \mathbb{N}}$ en étant que celle-ci est irréductible et apériodique.

5.1.2 Position du problème

Dans ce mémoire, nous allons nous intéresser à un critère permettant de déterminer si la probabilité qu'une population soumise à hétérogénéité spatiale se perpétue est strictement positive.

Inversement à des approches qui s'appuient sur la théorie des processus de branchement multitype et se basent sur la localisation de la plus grande valeur propre d'une matrice appelée *matrice de descendance* définie à partir des lois de reproduction et de migration, nous adopterons ici une approche certes également fondée sur les propriétés de "branchement" du processus $(Z_n)_{n \in \mathbb{N}}$ mais qui donne un "critère de survie" pouvant trouver une interprétation simple à partir de certaines quantités propres au déplacement de la marche aléatoire $(X_n)_{n \in \mathbb{N}}$ sur le graphe modélisant la situation.

Dans un premier temps, nous démontrerons un théorème établissant un critère dans le cas général où K -types d'habitats interviennent en fonction du temps de retour de la marche aléatoire en le lieu d'habitat de la première génération. Puis, dans un deuxième temps, nous appliquerons ce résultat à deux situations particulières où des conditions sur les migrations et sur les types d'habitats en jeu sont imposées et relierons alors la survie de la population à d'autres quantités propres à la marche aléatoire : d'abord, on s'intéressera à une situation où il y a deux types d'habitats - un seul type de source et un seul type de puits - et où la source est uniquement connectée à des puits, et ensuite, on s'intéressera à une situation où il y a deux types d'habitats - un seul type de source et un seul type de puits - et où la source est reliée à deux puits et où l'ensemble des puits forme un cycle.

5.2 La situation générale

Quitte à regrouper les différents lieux d'habitat de même type et à modifier la "matrice de migration" $M = (M_{i,j})_{(i,j) \in \{1,\dots,K\}^2}$, on pourra supposer qu'il y a autant de lieux que de types d'habitat à savoir K .

Théorème 11

Si on suppose que $X_0 = 1$ p.s. et que l'on définit le temps d'arrêt $T = \inf\{n \in \mathbb{N}^ / X_n = 1\}$, alors, la population survit avec une probabilité non nulle, si et seulement si, on a : $m_1 \mathbb{E} \left[\prod_{i=1}^{T-1} m_{X_i} \right] > 1$.*

Démo 11

On considère un individu vivant dans le lieu d'habitat de type 1 au temps $n = 0$ que l'on note a et on va suivre ses descendants de telle sorte à ne considérer que ceux qui retournent pour la première fois de la lignée dans le lieu d'habitat de type 1.

Réduction à un processus de branchement simple :

On définit alors la variable aléatoire Y_1 comme le nombre de descendants de a qui ne quittent pas le lieu d'habitat de type 1 à la première génération et pour tout $n \geq 2$, on définit la variable aléatoire Y_n comme le nombre d'individus nés à la génération n descendants de a qui migrent dans le lieu d'habitat de type 1 dont les ancêtres aux générations $1, \dots, n-1$ n'ont pas vécu dans le lieu d'habitat de type 1.

A partir de la suite $(Y_n)_{n \geq 1}$, on définit alors $Y = \sum_{n \geq 1} Y_n$ qui représente le nombre d'individus descendant de a qui retournent pour la première fois dans leur lignée depuis a dans le lieu d'habitat de type 1.

Les hypothèses d'indépendance faites sur les lois de reproduction des divers individus contemporains ou pas et sur les règles de migration, nous assurent que pour tout $m \in \mathbb{N}^*$ le nombre total de descendants de chacun de n quelconques individus distincts a_1, a_2, \dots, a_m comptés dans Y retournant pour la première fois dans leur lignée depuis a_1, a_2, \dots, a_m dans le lieu d'habitat de type 1 sont indépendants dans leur ensemble et identiquement distribués de même loi que Y .

Le nombre total d'individus descendants de a qui retournent dans le lieu d'habitat de type 1 (pas nécessairement pour la première fois depuis a dans leur lignée) peut alors être vu comme un processus de branchement simple de loi de reproduction la loi de Y . Mais le nombre total d'individus descendant de a qui retournent dans le lieu d'habitat de type 1 est fini, si et seulement si, la population des individus descendant de a qui retournent dans le lieu d'habitat de type 1 s'éteint.

On est donc ramené à un critère d'extinction de la population issue d'un processus de branchement simple de loi de reproduction la loi de Y que l'on connaît sous la forme : la population s'éteint presque sûrement, si et seulement si, on a $\mathbb{E}[Y] < 1$. Il faut donc désormais donner l'expression de $\mathbb{E}[Y]$ en fonction de

quantités propres à la marche aléatoire $(X_n)_{n \in \mathbb{N}}$ et le résultat sera prouvé.

Calcul de $\mathbb{E}[Y]$:

Pour ce calcul, on rappelle que, pour tout $l \in \{1, \dots, K\}$, on désigne par Z_n^l le nombre d'individus résidant dans le lieu d'habitat de type l à la n -ième génération et par Z_n le vecteur (Z_n^1, \dots, Z_n^K) et on constate alors que :

$$Z_{n+1} = \sum_{i=1}^K \sum_{j=1}^{Z_n^i} \xi_{n,j}(i)$$

où $(\xi_{n,j}(i))_{(i,j,n) \in \{1, \dots, K\} \times \mathbb{N}^* \times \mathbb{N}} \in \mathbb{N}^K$ est une suite de variables aléatoires indépendantes telles que pour tout $(j, n) \in \mathbb{N}^* \times \mathbb{N}$, $\xi_{n,j}(i)$ a pour loi ν_i où ν_i est la loi sur \mathbb{N}^K de "reproduction-migration" d'un individu résidant dans le lieu d'habitat de type i .

Fixons alors $i \in \{1, \dots, K\}$ et déterminons la loi ν_i de "reproduction-migration" d'un individu résidant dans le lieu d'habitat de type i . Par construction du modèle, un individu résidant dans le lieu d'habitat de type i se reproduit selon la loi μ_i et ses descendants migrent indépendamment de leurs ancêtres et indépendamment les uns des autres vers les autres habitats selon la loi donnée par le vecteur $(M_{i,1}, \dots, M_{i,K})$: on connaît donc la loi de "reproduction-migration" d'un individu résidant dans le lieu d'habitat de type i conditionnellement au nombre de descendants de cet individu. En d'autres termes, si on note R_i la variable aléatoire de loi μ_i représentant le nombre de descendant d'un individu du type i , la loi de "reproduction-migration" de cet individu conditionnellement à R_i sera une loi multinomiale de paramètre $(M_{i,1}, \dots, M_{i,K})$ et d'ordre R_i .

Ainsi, on peut calculer l'espérance de Z_{n+1} en utilisant l'identité de Wald puisque par hypothèse pour tout $i \in \{1, \dots, K\}$ et tout n dans \mathbb{N} , les variables aléatoires $\xi_{n,1}(i), \dots, \xi_{n,p}(i), \dots$ sont indépendantes entre elles et identiquement distribuées :

$$\mathbb{E}[Z_{n+1}] = \mathbb{E} \left[\sum_{i=1}^K \sum_{j=1}^{Z_n^i} \xi_{n,j}(i) \right] = \sum_{i=1}^K \mathbb{E} \left[\sum_{j=1}^{Z_n^i} \xi_{n,j}(i) \right] = \sum_{i=1}^K \mathbb{E}[Z_n^i] \mathbb{E}[\xi_{n,1}(i)].$$

Or on sait que si une variable aléatoire $S = (S_1, \dots, S_K)$ à valeurs dans \mathbb{N}^K suit une loi mutinomiale de paramètre (p_1, \dots, p_K) et d'ordre r alors pour tout $l \in \{1, \dots, K\}$, la loi de S_l est une loi binomiale de paramètres r et p_l et la

moyenne d'une telle loi binomiale est rp_l donc il vient :

$$\begin{aligned}
\mathbb{E} [\xi_{n,1}(i)] &= \mathbb{E} [\mathbb{E} [\xi_{n,1}(i)|R_i]] = \sum_{k=0}^{+\infty} \mathbb{E} [\xi_{n,1}(i)|R_i = k] \mathbb{P}((R_i = k)) \\
&= \sum_{k=0}^{+\infty} \mathbb{P}((R_i = k)) (kM_{i,1}, \dots, kM_{i,K}) \\
&= \left(\sum_{k=0}^{+\infty} k\mathbb{P}((R_i = k)) \right) (M_{i,1}, \dots, M_{i,K}) \\
&= \mathbb{E} [R_i] (M_{i,1}, \dots, M_{i,K}) = m_i(M_{i,1}, \dots, M_{i,K}).
\end{aligned}$$

Pour récapituler, on a donc trouvé que :

$$\mathbb{E} [Z_{n+1}] = \sum_{i=1}^K \mathbb{E} [Z_n^i] \mathbb{E} [\xi_{n,1}(i)] = \sum_{i=1}^K \mathbb{E} [Z_n^i] m_i(M_{i,1}, \dots, M_{i,K}).$$

Par ailleurs, en désignant par Y_n^l pour tout $n \in \mathbb{N}^*$ le nombre d'individus résidant dans le lieu d'habitat de type l à la n -ième génération qui vivent pour la première fois dans leur lignée depuis le tout premier ancêtre dans le lieu d'habitat de type l et par Y_n le vecteur (Y_n^1, \dots, Y_n^K) , on constate que :

$$Y_{n+1} = \sum_{i=2}^K \sum_{j=1}^{Y_n^i} \xi_{n,j}(i)$$

et par un raisonnement analogue au précédent, on obtient que pour tout $n \in \mathbb{N}$ et pour tout $l \in \{1, \dots, K\}$:

$$\mathbb{E} [Y_{n+1}^l] = \sum_{i=2}^K \mathbb{E} [Y_n^i] \mathbb{E} [\xi_{n,1}^l(i)] = \sum_{i=2}^K \mathbb{E} [Y_n^i] m_i M_{i,l}.$$

Puis par une récurrence immédiate, on trouve pour tout $l \in \{1, \dots, K\}$:

$$\begin{aligned}
\mathbb{E} [Y_{n+1}^l] &= \sum_{i=2}^K \mathbb{E} [Y_n^i] m_i M_{i,l} = \sum_{i_n=2}^K \left(\sum_{i_{n-1}=2}^K \mathbb{E} [Y_{n-1}^{i_{n-1}}] m_{i_{n-1}} M_{i_{n-1}, i_n} \right) m_{i_n} M_{i_n, l} \\
&= \sum_{(i_1, \dots, i_n) \in \{2, \dots, K\}^n} \mathbb{E} [Y_1^{i_1}] M_{1, i_2} M_{i_2, i_3} \dots M_{i_{n-1}, i_n} M_{i_n, l} m_{i_1} m_{i_2} \dots m_{i_{n-1}} m_{i_n} \\
&= \sum_{(i_1, \dots, i_n) \in \{2, \dots, K\}^n} m_1 M_{1, i_1} M_{i_1, i_2} M_{i_2, i_3} \dots M_{i_{n-1}, i_n} M_{i_n, l} m_{i_1} m_{i_2} \dots m_{i_{n-1}} m_{i_n} \\
&= m_1 \left(\sum_{(i_1, \dots, i_n) \in \{2, \dots, K\}^n} M_{1, i_1} M_{i_1, i_2} M_{i_2, i_3} \dots M_{i_{n-1}, i_n} M_{i_n, l} m_{i_1} m_{i_2} \dots m_{i_{n-1}} m_{i_n} \right).
\end{aligned}$$

Au final, par définition même de la suite $(Y_n^1)_{n \in \mathbb{N}^*}$ comme le nombre d'individus résidant dans le lieu d'habitat de type 1 à la n -ième génération qui vivent pour

la première fois dans leur lignée dans le lieu d'habitat de type 1 depuis le tout premier ancêtre, il vient $Y_n^1 = Y_n$ pour tout $n \in \mathbb{N}^*$ d'où :

$$\begin{aligned} \mathbb{E}[Y_n] &= m_1 \left(\sum_{(i_1, \dots, i_{n-1}) \in \{2, \dots, K\}^n} M_{1, i_1} M_{i_1, i_2} M_{i_2, i_3} \dots M_{i_{n-2}, i_{n-1}} M_{i_{n-1}, 1} m_{i_1} m_{i_2} \dots m_{i_{n-2}} m_{i_{n-1}} \right) \\ &= m_1 \mathbb{E} [\mathbb{1}_{T=n} m_{X_1} m_{X_2} \dots m_{X_{T-2}} m_{X_{T-1}}] \end{aligned}$$

et puisque par définition $Y = \sum_{n \geq 1} Y_n$, on a :

$$\begin{aligned} \mathbb{E}[Y] &= \mathbb{E} \left[\sum_{n=1}^{+\infty} Y_n \right] = \sum_{n=1}^{+\infty} \mathbb{E}[Y_n] = \sum_{n=1}^{+\infty} m_1 \mathbb{E} [\mathbb{1}_{T=n} m_{X_1} m_{X_2} \dots m_{X_{T-2}} m_{X_{T-1}}] \\ &= m_1 \mathbb{E} \left[\sum_{n=1}^{+\infty} \mathbb{1}_{T=n} m_{X_1} m_{X_2} \dots m_{X_{T-2}} m_{X_{T-1}} \right] = m_1 \mathbb{E} \left[\prod_{i=1}^{T-1} m_{X_i} \right] \end{aligned}$$

car l'hypothèse d'irréductibilité de la chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ nous assure que p.s. $T < +\infty$ donc p.s. $\sum_{n=1}^{+\infty} \mathbb{1}_{T=n} = 1$.

En conclusion, on a donc bien montré que la population survit avec une probabilité non nulle, si et seulement si, on a : $m_1 \mathbb{E} \left[\prod_{i=1}^{T-1} m_{X_i} \right] > 1$. ■

5.3 La situation à deux types d'habitat

Concentrons nous désormais sur le cas particulier où il n'y a que deux types d'habitat et où la source est connecté seulement à des puits. On pose :

$$M = m_1 > 1, m = m_2 = \dots = m_K < 1$$

On note $p = \sum_{j=2}^K d_{1j}$ la probabilité de sortir de l'habitat de type 1. Le nombre moyen d'enfants par tête envoyé à l'extérieur de la source est Mp . En effet, on a :

$$\mathbb{E}[\text{Nombre d'enfants envoyés à l'extérieur par tête}] = \mathbb{E} \left[\sum_{i=1}^{Z_n^j} \chi_i \right] = \mathbb{E}[Z_n^1] \mathbb{E}[\chi_1] = Mp.$$

Soit σ le temps d'attente défini comme le temps de première visite d'un puits par un marcheur aléatoire :

$$\sigma = \inf \{n \geq 0 : X_n \neq 1\},$$

si bien que σ est une variable aléatoire de loi géométrique avec probabilité de succès p .

Ensuite, soit S le temps d'attente défini comme le temps postérieur à σ avant que le marcheur aléatoire ne revienne dans une source, que ce soit la source originelle ou pas :

$$S = \inf \{n \geq 0 : X_{\sigma+n} \neq 1\}.$$

La durée S peut être vue comme le temps passé dans les puits entre deux visites consécutives dans les sources. Alors le théorème précédent se lit :

Théorème 12

On a comme critère de survie de la population avec probabilité positive

$$M(1 - p) + eMp > 1,$$

où e est le taux de décroissance dû aux puits dans le graphe, défini par

$$e = \mathbb{E} [m^S] = \sum_{k \geq 1} m^k \mathbb{P} (S = k).$$

Démo 12

Le théorème nous donne comme condition :

$$m_1 \mathbb{E} \left[\prod_{i=1}^{T-1} m_{X_i} \right] > 1.$$

Or, en sachant que $S = T - 1$ si on sait que $T \geq 2$, on a :

$$\begin{aligned} m_1 \mathbb{E} \left[\prod_{i=1}^{T-1} m_{X_i} \right] &= M \sum_{k=1}^{+\infty} \mathbb{E} [\mathbb{1}_{T=k} m^{k-1}] \\ &= M \mathbb{E} [\mathbb{1}_{T=1}] + M \sum_{k=2}^{+\infty} \mathbb{E} [\mathbb{1}_{T=k} m^{k-1}] \\ &= M(1 - p) + M \sum_{k=1}^{+\infty} \mathbb{E} [\mathbb{1}_{S=k} m^k] \mathbb{P} (T \geq 2) \\ &= M(1 - p) + M \sum_{k=1}^{+\infty} m^k \mathbb{E} [\mathbb{1}_{S=k}] p \\ &= M(1 - p) + M \sum_{k=1}^{+\infty} m^k \mathbb{P} (S = k) p \\ &= M(1 - p) + eMp \end{aligned}$$

On a bien le résultat que l'on cherche. ■

Cela nous donne une condition suffisante :

Théorème 13

Si le temps moyen passé dans les puits a comme propriété

$$\mathbb{E} [S] < \frac{M - 1}{Mp(1 - m)},$$

alors la population survit avec probabilité positive.

Démo 13

On considère la fonction $f : x \mapsto f(x) = \mathbb{E} [x^S]$. Cette fonction est convexe car elle peut se développer en série entière à coefficients positifs. Alors

$$e = f(m) \geq 1 + f'(1)(m - 1) = 1 - (1 - m)\mathbb{E} [S] > 1 - \frac{M - 1}{Mp} = \frac{1 - M(1 - p)}{Mp},$$

ce qui donne $eMp + M(1 - p) > 1$. ■

Une autre condition équivalente est :

Théorème 14

Le critère est équivalent à

$$\mathbb{E} [M^\sigma m^S] > 1.$$

Démo 14

On a

$$\begin{aligned} \mathbb{E} [M^\sigma m^S] &= \mathbb{E} [M^\sigma] \times \mathbb{E} [m^S] \\ &= \sum_{k=1}^{+\infty} M^k (1-p)^{k-1} p \times e \\ &= Mp \sum_{k=0}^{+\infty} M^k (1-p)^k \times e \\ &= Mp \frac{1}{1 - M(1-p)} e \end{aligned}$$

Donc si on injecte ce résultat dans l'inégalité on a

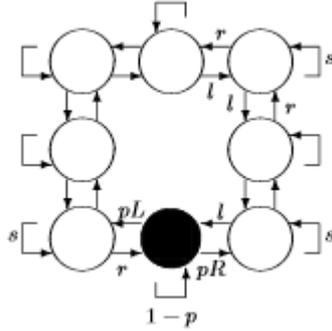
$$\begin{aligned} \mathbb{E} [M^\sigma m^S] &> 1 \\ Mp \frac{1}{1 - M(1-p)} e &> 1 \\ Mpe &> 1 - M(1-p) \\ M(1-p) + eMp &> 1 \end{aligned}$$

On a la réciproque en remontant les égalités. ■

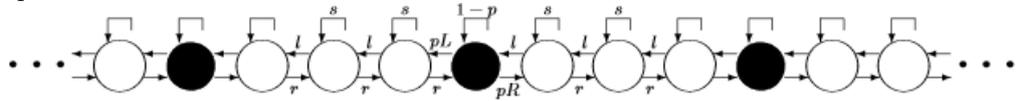
On a donc des résultats de persistance de la population qui découpent les rôles des nombres moyens d'enfants M et m de la structure de l'espace. Sans cela, il faudrait calculer la valeur propre maximale d'une matrice, ce qui peut se révéler compliqué.

5.4 La situation avec un cycle de puits tous identiques

Supposons que la source est un sommet de degré deux dans le graphe, uniquement connecté à un puits gauche et à un puits droit. La probabilité de rester sur place dans la source est $1-p$, la probabilité de passer au puits gauche est Lp et au puits droit Rp , avec $L+R=1$. Les puits forment une chaîne de n puits adjacents reliant des sources adjacentes. La probabilité de rester sur un puits donné est systématiquement s , celle d'aller à droite systématiquement r et celle d'aller à gauche systématiquement l .



On définit un graphe *source-transitif* par le fait qu'il existe un isomorphisme de graphe du graphe sur lui même qui envoie une source a sur une source b , pour toute paire de sources a et b .



On peut calculer exactement le taux de décroissance e pour les graphes *source-transitifs* du type ci-dessus.

Soit $\lambda > 1 > \mu$ les deux racines de l'équation :

$$mr x^2 - (1 - ms)x + ml = 0 \quad \text{où } x \geq 0.$$

Alors, par les relations entre coefficients et racines, on a :

$$\lambda\mu = \frac{l}{r} \quad \text{et} \quad \lambda + \mu = \frac{1 - ms}{mr}.$$

Théorème 15

Le taux de décroissance e admet pour expression :

$$e = \frac{\lambda^n - \mu^n}{\lambda^{n+1} - \mu^{n+1}} (R + L\lambda\mu) + \frac{\lambda - \mu}{\lambda^{n+1} - \mu^{n+1}} (L + R(\lambda\mu)^n).$$

Démo 15

On considère une marche aléatoire Y sur $\{0, 1, \dots, n + 1\}$, où l'on se déplace à chaque étape de -1 avec probabilité l , 0 avec probabilité s et $+1$ avec probabilité r . Pour les numéros 0 et $n + 1$, si on devait sortir du cadre, on reste immobile à la place. Soit T_i le temps de première entrée sur le numéro i par Y , et posons T le temps de première arrivée en 0 ou $n + 1$:

$$T_i = \min(n \geq 1 | Y_n = i \text{ et } \forall m < n, Y_m \neq i) \quad \text{et} \quad T = \min(T_0, T_{n+1}).$$

Pour tout k dans $\{0, 1, \dots, n + 1\}$, on définit :

$$a_k = \mathbb{E} [m^T | Y_0 = k].$$

Sur le graphe originel et son marcheur aléatoire X , on a que si lors de la première sortie de la source par X , on est sorti à gauche, alors le parcours de retour à une source peut être transposé sur un mouvement de Y à partir du

nombre 1 jusqu'au retour au nombre 0, qui représente la source atteinte depuis la droite, ou $n+1$ qui représente la source atteinte depuis la gauche. On a donc :

$$\mathbb{E} [m^S \mid X_\sigma \text{ est un voisin gauche}] = a_1.$$

De même, on a :

$$\mathbb{E} [m^S \mid X_\sigma \text{ est un voisin droit}] = a_n.$$

Comme la première sortie de la source se fait forcément à droite ou à gauche, on a exhaustion des cas. Ainsi, il vient :

$$\begin{aligned} e &= \mathbb{E} [m^S] \\ &= \mathbb{P}(X_\sigma \text{ est un voisin gauche}) \mathbb{E} [m^S \mid X_\sigma \text{ est un voisin gauche}] \\ &\quad + \mathbb{P}(X_\sigma \text{ est un voisin droit}) \mathbb{E} [m^S \mid X_\sigma \text{ est un voisin droit}] \\ &= Ra_1 + La_n. \end{aligned}$$

Pour calculer a_1 et a_n , on établit la relation de récurrence suivante :

$$a_k = msa_k + mla_{k-1} + mra_{k+1} \quad \text{où } k \in \{1, \dots, n\},$$

avec pour conditions aux frontières $a_0 = a_{n+1} = 1$. En effet, on a :

$$\begin{aligned} \mathbb{E} [m^T \mid Y_0 = k] &= \mathbb{E} [m^T \mathbb{1}_{Y_1=k-1} \mid Y_0 = k] + \mathbb{E} [m^T \mathbb{1}_{Y_1=k} \mid Y_0 = k] + \mathbb{E} [m^T \mathbb{1}_{Y_1=k+1} \mid Y_0 = k] \\ &= m\mathbb{E} [m^{T-1} \mathbb{1}_{Y_1=k-1} \mid Y_0 = k] + m\mathbb{E} [m^{T-1} \mathbb{1}_{Y_1=k} \mid Y_0 = k] + m\mathbb{E} [m^{T-1} \mathbb{1}_{Y_1=k+1} \mid Y_0 = k] \\ &= m\mathbb{E} [m^{T-1} \mid Y_1 = k-1] \mathbb{P}(Y_1 = k-1 \mid Y_0 = k) + m\mathbb{E} [m^{T-1} \mid Y_1 = k] \mathbb{P}(Y_1 = k \mid Y_0 = k) \\ &\quad + m\mathbb{E} [m^{T-1} \mid Y_1 = k+1] \mathbb{P}(Y_1 = k+1 \mid Y_0 = k) \\ &= mla_{k-1} + msa_k + mra_{k+1}. \end{aligned}$$

On résout ensuite cette récurrence : on retrouve comme polynôme caractéristique le polynôme défini avant le théorème, de racine λ et μ .

On a alors comme solution générale $a_k = C\lambda^k + D\mu^k$ et les conditions aux bords imposent :

$$\begin{array}{ll} C + D = 1 & \text{grâce à } a_0 = 1 \\ C\lambda^{n+1} + D\mu^{n+1} = 1 & \text{grâce à } a_{n+1} = 1 \end{array}$$

Donc $C = 1 - D$ et en remplaçant dans l'autre équation, on obtient :

$$\begin{aligned} (1 - D)\lambda^{n+1} + D\mu^{n+1} &= 1 \\ D(\mu^{n+1} - \lambda^{n+1}) &= 1 - \lambda^{n+1} \\ D &= \frac{1 - \lambda^{n+1}}{\mu^{n+1} - \lambda^{n+1}} \\ C &= \frac{\mu^{n+1} - 1}{\mu^{n+1} - \lambda^{n+1}} \end{aligned}$$

On remplace dans la formule de e :

$$\begin{aligned} e &= R \left(\frac{\mu^{n+1} - 1}{\mu^{n+1} - \lambda^{n+1}} \lambda + \frac{1 - \lambda^{n+1}}{\mu^{n+1} - \lambda^{n+1}} \mu \right) + L \left(\frac{\mu^{n+1} - 1}{\mu^{n+1} - \lambda^{n+1}} \lambda^n + \frac{1 - \lambda^{n+1}}{\mu^{n+1} - \lambda^{n+1}} \mu^n \right) \\ &= \frac{1}{\mu^{n+1} - \lambda^{n+1}} (R\mu^{n+1}\lambda - R\lambda + R\mu - R\lambda^{n+1}\mu + L\mu^{n+1}\lambda^n - L\lambda^n + L\mu^n - L\lambda^{n+1}\mu^n) \end{aligned}$$

Si on développe l'expression de e que l'on cherche à obtenir, on obtient la même expression, ce qui nous donne l'égalité recherchée. \blacksquare

5.5 Conclusion

Ce TER nous aura permis de comprendre d'avantage la démarche d'un chercheur. En effet, une partie du travail de ce dernier consiste à lire et comprendre des travaux ou articles déjà effectués ou rédigés par d'autres afin d'en tirer de nouveaux axes de recherche. Notre but était donc de comprendre l'article qui nous avait été donné par Christophe GIRAUD, de le restituer par nous-même, tout en essayant de développer certains points des démonstrations et de nous consacrer à des questions en rapport plus ou moins direct avec le sujet. Par la même occasion, nous avons pu réinvestir et approfondir notre pratique du langage \LaTeX .

Ce projet nous aura également fait aborder divers domaine de la théorie des probabilités qui nous étaient inconnus, celui des processus de branchement multitype et celui des arbres aléatoires. Avec plus de temps, il aurait été intéressant de s'intéresser plus en détail à la théorie des arbres aléatoires (en particulier à la notion de ligne d'arrêt) et surtout de pouvoir attester des avantages apportés par l'*approche par marche aléatoire* concernant l'étude de situations où le nombre de types d'habitat est toujours fini mais le nombre de lieux d'habitat est infini.

Chapitre 6

Mémoire de M2 Recherche

Préambule

Ce mémoire intitulé *Statistical analysis of the one-dimensional random walk in random environment* a été réalisé dans le cadre du stage de Recherche du Master 2 Recherche "Mathématiques de l'Aléatoire" de l'université Paris-Saclay.

L'ensemble du travail a été réalisé sous la tutelle d'Éric MOULINES, Professeur à l'École Polytechnique, et Matthieu LERASLE, Professeur chargé de cours à l'École Polytechnique et chargé de recherche CNRS au laboratoire de Mathématiques d'Orsay, du 1 avril au 31 juillet 2016.

Ce mémoire ne se veut nullement exhaustif quant à l'étude des marches aléatoires en milieu aléatoire ni quant à l'estimation de la loi du milieu de ces marches dans un cadre paramétrique dont il traite plus spécifiquement.

6.1 Probabilistic analysis of the RWRE

The mathematical model we are going to work with is the i.i.d. random walk in random environment in one dimension, that's a random walk on \mathbb{Z} whose transition probabilities from one site to the previous or the next depend on the current state of the walk, are random and independent from one another and follow the same probability law ν on $]0, 1[$.

The letter E will be the set of environments that is $E =]0, 1[^\mathbb{Z}$ endowed with $\mathcal{E} = \mathcal{B}(]0, 1[)^{\otimes \mathbb{Z}}$ the σ -field generated by the cylinder and we denote by \mathbb{P}^ν the probability $\nu^{\otimes \mathbb{Z}}$ on (E, \mathcal{E}) .

Let denote $(X_n)_{n \in \mathbb{N}}$ the canonical process of the canonical space $S = \mathbb{Z}^\mathbb{N}$ endowed with the canonical σ -field $\mathcal{S} = (2^\mathbb{Z})^{\otimes \mathbb{N}}$ generated by the cylinder. Define a family of probability distributions $(\mathbf{P}_{x,\omega})_{(x,\omega) \in \mathbb{Z} \times E}$ on \mathcal{S} called *quenched law* by :

$$\forall (x, \omega, n) \in \mathbb{Z} \times E \times \mathbb{N}, \quad \begin{cases} \mathbf{P}_{x,\omega}(X_0 = x) = 1. \\ \mathbf{P}_{x,\omega}(X_{n+1} = X_n + 1 | X_0, \dots, X_n) = \omega_{X_n}. \\ \mathbf{P}_{x,\omega}(X_{n+1} = X_n - 1 | X_0, \dots, X_n) = 1 - \omega_{X_n}. \end{cases}$$

In other words, for (x, ω) in $\mathbb{Z} \times E$, $\mathbf{P}_{x,\omega}$ is the unique probability on S under which $(X_n)_{n \in \mathbb{N}}$ is a Markov chain starting at x and whose transition kernel Q_ω is given by :

$$Q_\omega(d, f) = \omega_d \mathbb{1}_{\{f=d+1\}} + (1 - \omega_d) \mathbb{1}_{\{f=d-1\}}.$$

It can be shown by use of the monotone class' lemma that for every $B \in \mathcal{S}$, the function $\omega \in E \mapsto \mathbf{P}_{x,\omega}(B) \in [0, 1]$ is \mathcal{E} -measurable. Therefore, we can define an other family of probability distributions $(\mathbf{P}_x^\nu)_{x \in \mathbb{Z}}$ named *annealed law* by :

$$\forall x \in \mathbb{Z}, \forall B \in \mathcal{S}, \quad \mathbf{P}_x^\nu(B) = \int_E \mathbf{P}_{x,\omega}(B) \mathbb{P}^\nu(d\omega).$$

The theoretical study of the i.i.d. RWRE relies on the ratio ρ_x defined for all $\omega \in E$ and for all $x \in \mathbb{N}$ by :

$$\rho_x = \frac{1 - \omega_x}{\omega_x}.$$

Théorème 16 (Solomon,, [Sol75])

If x in \mathbb{Z} and $\nu \in \mathcal{M}_1(\mathcal{B}(]0, 1[))$ are such that $\mathbb{E}^\nu[|\log(\rho_x)|] < +\infty$, then :

$$\begin{aligned} \mathbb{E}^\nu[\log(\rho_x)] < 0 &\Rightarrow X_n \xrightarrow[n \rightarrow +\infty]{\mathbf{P}_x^\nu\text{-a.s.}} +\infty, \text{ i.e., the RWRE is transient to the right.} \\ \mathbb{E}^\nu[\log(\rho_x)] = 0 &\Rightarrow \forall k \in \mathbb{N}, \mathbf{P}_x^\nu(\overline{\lim}\{X_n = k\}) = 1, \text{ i.e., the RWRE is recurrent.} \\ \mathbb{E}^\nu[\log(\rho_x)] > 0 &\Rightarrow X_n \xrightarrow[n \rightarrow +\infty]{\mathbf{P}_x^\nu\text{-a.s.}} -\infty, \text{ i.e., the RWRE is transient to the left.} \end{aligned}$$

The ratio ρ_x allows also to qualify the “speed“ of the walk in different cases.

Théorème 17 (Solomon, [Sol75])

If x in \mathbb{Z} and $\nu \in \mathcal{M}_1(\mathcal{B}([0, 1]))$ are such that $\mathbb{E}^\nu [|\log(\rho_x)|] < +\infty$, then :

$$\begin{aligned} \mathbb{E}^\nu [\rho_x] < 1 &\Rightarrow \frac{X_n}{n} \xrightarrow[n \rightarrow +\infty]{\mathbf{P}_x^\nu\text{-a.s.}} \frac{1 - \mathbb{E}^\nu [\rho_x]}{1 + \mathbb{E}^\nu [\rho_x]}. \\ \mathbb{E}^\nu \left[\frac{1}{\rho_x} \right] < 1 &\Rightarrow \frac{X_n}{n} \xrightarrow[n \rightarrow +\infty]{\mathbf{P}_x^\nu\text{-a.s.}} -\frac{1 - \mathbb{E}^\nu \left[\frac{1}{\rho_x} \right]}{1 + \mathbb{E}^\nu \left[\frac{1}{\rho_x} \right]}. \\ \frac{1}{\mathbb{E}^\nu [\rho_x]} \leq 1 \leq \mathbb{E}^\nu \left[\frac{1}{\rho_x} \right] &\Rightarrow \frac{X_n}{n} \xrightarrow[n \rightarrow +\infty]{\mathbf{P}_x^\nu\text{-a.s.}} 0. \end{aligned}$$

Assume that $\frac{X_n}{n} \xrightarrow[n \rightarrow +\infty]{\mathbf{P}_x^\nu\text{-a.s.}} v \in \mathbb{R}$. If $v \neq 0$, the walk is said to be ballistic whereas if $v = 0$, it is said to be sub-ballistic.

Théorème 18 (Solomon, [Sol75])

If the RWRE is such that $\mathbb{E}^\nu [|\log(\rho_0)|] < +\infty$ and $\mathbb{E}^\nu [\log(\rho_0)] < 0$ then, for all n in \mathbb{N} , the stopping time $T_n = \inf(\{k \in \mathbb{N} / X_k = n\})$ is \mathbf{P}_0^ν -a.s. well defined and finite and it holds :

$$\begin{aligned} \mathbb{E}^\nu [\rho_0] < 1 &\Rightarrow \frac{T_n}{n} \xrightarrow[n \rightarrow +\infty]{\mathbf{P}_0^\nu\text{-a.s.}} \mathbf{E}_0^\nu [T_1] = \frac{1 + \mathbb{E}^\nu [\rho_0]}{1 - \mathbb{E}^\nu [\rho_0]}. \\ \mathbb{E}^\nu [\rho_0] \geq 1 &\Rightarrow \frac{T_n}{n} \xrightarrow[n \rightarrow +\infty]{\mathbf{P}_0^\nu\text{-a.s.}} \mathbf{E}_0^\nu [T_1] = +\infty. \end{aligned}$$

Another also very important result is this of Kesten, Kozlov and Spitzer describing the non-degenerate asymptotic behaviour of T_n and X_t .

Théorème 19 (Kesten, Kozlov and Spitzer, [Kes75])

If the RWRE is such that $\mathbb{E}^\nu [\log(\rho_0)] < 0$, if there exists $\kappa > 0$ such that $\mathbb{E}^\nu [\rho_0^\kappa] = 1$ and $\mathbb{E}^\nu [\rho_0^\kappa \log^+(\rho_0)] < +\infty$ and if the group generated by the support of $\log(\rho_0)$ is dense in \mathbb{R} then, there exist $(B_1, B_2, B_3) \in \mathbb{R}^{*+3}$, $(A_t)_{t \in \mathbb{R}^{*+}} \in \mathbb{R}^{*+ \mathbb{R}^{*+}}$ and a stable probability distribution L_κ such that :

$$\begin{aligned} \cdot \text{ If } \kappa < 1, \text{ it comes : } &\left\{ \begin{array}{l} \frac{1}{n^\kappa} T_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} L_\kappa \\ \text{and} \\ \lim_{t \rightarrow +\infty} (\mathbb{P}^\nu \left(\frac{X_t}{t^\kappa} \leq x \right)) = 1 - L_\kappa \left(\frac{1}{x^\kappa} \right). \end{array} \right. \\ \cdot \text{ If } \kappa = 1, \text{ there exist } D_n \underset{n \rightarrow +\infty}{\sim} \log(n) \text{ and } \delta_t \underset{t \rightarrow +\infty}{\sim} \frac{t}{A_1 \log(t)} \text{ so that :} & \\ &\left\{ \begin{array}{l} \frac{1}{n} \left(T_n - A_1 n D_{\frac{n}{\mu}} \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} L_1 \\ \text{and} \\ \lim_{t \rightarrow +\infty} \left(\mathbb{P}^\nu \left(\frac{\log(t)^2}{t} (X_t - \delta_t) \leq x \right) \right) = 1 - L_1(-A_1^2 x). \end{array} \right. \\ \cdot \text{ If } 1 < \kappa < 2, \text{ it comes :} &\left\{ \begin{array}{l} \frac{1}{n^\kappa} (T_n - A_\kappa n) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} L_\kappa \\ \text{and} \\ \lim_{t \rightarrow +\infty} \left(\mathbb{P}^\nu \left(\frac{1}{t^\kappa} (X_t - \frac{t}{A_\kappa}) \leq x \right) \right) = 1 - L_\kappa \left(-A_\kappa^{1 + \frac{1}{\kappa}} x \right). \end{array} \right. \\ \cdot \text{ If } \kappa = 2, \text{ it comes :} &\left\{ \begin{array}{l} \frac{1}{B_1 \sqrt{n \log(n)}} (T_n - A_2 n) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1) \\ \text{and} \\ \frac{A_2^{\frac{3}{2}}}{B_1 \sqrt{n \log(n)}} \left(X_t - \frac{t}{A_2} \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1). \end{array} \right. \end{aligned}$$

$$\cdot \text{ If } 2 < \kappa, \text{ it comes : } \left\{ \begin{array}{l} \frac{1}{B_2\sqrt{n}} (T_n - B_3n) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1) \\ \text{and} \\ \frac{B_3^{\frac{3}{2}}}{B_2\sqrt{n}} \left(X_t - \frac{t}{B_3} \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1). \end{array} \right.$$

Remarque 1

If the RWRE is such that $\mathbb{E}^\nu [\log(\rho_0)] < 0$ and if there exists $\kappa > 0$ such that $\mathbb{E}^\nu [\rho_0^\kappa] = 1$ and $\mathbb{E}^\nu [\rho_0^\kappa \log^+(\rho_0)] < +\infty$ are met, then, $\kappa > 0$ is the unique strictly positive real number satisfying those conditions since $u \in \mathbb{R}^{*+} \mapsto \mathbb{E}^\nu [\rho_0^u]$ is convex with negative derivative in the vicinity of 0 and value in 0 equal to 1.

For further results on RWRE, we refer to the lecture notes of O.Zeitouni in [Tav04].

In this setting, we would like to be able to provide some generic estimator of ν from the observation of one trajectory of the RWRE starting from 0 (henceforth, we will assume that the RWRE starts almost surely from zero) and reaching for the first time a fixed distant site in \mathbb{N}^* . Aiming at it, we may introduce several additional assumptions on the model.

Finally, we will write ν^* the true distribution under which the random process $(X_n)_{n \in \mathbb{N}}$ has been generated and shorten \mathbb{P}^{ν^*} into \mathbb{P}^* , $\mathbf{P}_0^{\nu^*}$ into \mathbf{P}_0^* and accordingly the associated expectation symbols.

6.2 Combinatorial results

Définition 11

For $l \in \mathbb{N}$, a nearest neighbour path of length l (starting from $p \in \mathbb{Z}$) is any sequence $x = (x_0, \dots, x_l) \in \mathbb{Z}^{l+1}$ such that :

$$\forall i \in \{1, \dots, l\}, |x_i - x_{i-1}| = 1 \quad (\text{and } x_0 = p).$$

Définition 12

For any nearest neighbour path $x = (x_0, \dots, x_l)$ of length l and any $k \in \mathbb{Z}$, the number of left steps $L(k, x)$ of the walk x from k and the number of right steps $R(k, x)$ of the walk x from k are defined by :

$$L(k, x) = \sum_{i=0}^{l-1} \mathbb{1}_{\{x_i=k, x_{i+1}=k-1\}} \quad \text{and} \quad R(k, x) = \sum_{i=0}^{l-1} \mathbb{1}_{\{x_i=k, x_{i+1}=k+1\}}.$$

For any nearest neighbour path $x = (x_0, \dots, x_l)$, the set of integers visited by the path x denoted by $v(x)$ is the set of integers $k \in \mathbb{Z}$ such that there exists $i \in \{0, \dots, l\}$ for which $x_i = k$.

As nearest neighbour paths are crucial in our study, we will give some results on their combinatorial properties and in particular results about the links existing between $L(k, x)$ and $R(k, x)$ for a given $k \in \mathbb{N}$.

Lemme 6

If $x = (x_0, \dots, x_l)$ is a nearest neighbour path of length l such that $x_0 = x_l$ (i.e. the path ends where it started) then it holds :

$$R(x_0, x) = L(x_0 + 1, x).$$

Démo 16

Let's note $a = x_0 = x_l$ and define the two sets of integers \mathcal{R} and \mathcal{L} by :

$$\mathcal{R} = \{i \in \{0, \dots, l-1\} / x_i = a \text{ and } x_i = a + 1\}$$

and

$$\mathcal{L} = \{i \in \{0, \dots, l-1\} / x_i = a + 1 \text{ and } x_i = a\}$$

We want to prove that :

$$R(x_0, x) = |\mathcal{R}| = |\mathcal{L}| = L(x_0 + 1, x).$$

In the case where $|\mathcal{R}| = 0$ (that is $\mathcal{R} = \emptyset$), as $x = (x_0, \dots, x_l)$ is a nearest neighbour path, we must have $\forall i \in \{0, \dots, l-1\}, |x_{i+1} - x_i| = 1$ so that x_1 lies in $\{x_0 - 1, \dots, x_0 + 1\}$ and \mathcal{R} being empty, we necessarily have $x_1 = x_0 - 1 \leq x_0$ and by induction, it comes easily $\forall i \in \{0, \dots, l-1\}, x_i \leq x_0$ and hence $\mathcal{L} = \emptyset$ or equivalently $|\mathcal{L}| = 0$ and $|\mathcal{R}| = |\mathcal{L}|$.

Otherwise $\mathcal{R} \neq \emptyset$ and writting $d_a = |\mathcal{R}|$ and $\mathcal{R} = \{r_1, \dots, r_{d_a}\}$ where $r_1 < \dots < r_{d_a}$, we will show that $\mathcal{L} = \{l_1, \dots, l_{d_a}\}$ where for all i in $\{1, \dots, d_a - 1\}$, $r_i < l_i < r_{i+1}$ and $r_{d_a} < l_{d_a} \leq l - 1$ so that $|\mathcal{R}| = d_a = |\mathcal{L}|$.

An integer $i \in \{1, \dots, d_a - 1\}$ being fixed, assume by contradiction that for all $j \in \{0, \dots, l-1\}$ such that $r_i < j < r_{i+1}$, integer j isn't in \mathcal{L} . Using that $x_{r_{i+1}} = x_0 + 1$ and the fact that $x = (x_0, \dots, x_l)$ is a nearest neighbour path, we easily show by induction for all j in $\{0, \dots, l-1\}$ such that $r_i + 1 \leq j \leq r_{i+1}$, that $x_0 + 1 \leq x_j$ holds and in particular $x_0 + 1 \leq x_{r_{i+1}}$ which contradicts the definition of r_{i+1} since $x_{r_{i+1}}$ must be equal to $x_0 \neq x_0 + 1$. Hence, it necessarily exists $l_i \in \{0, \dots, l-1\}$ such that $r_i < l_i < r_{i+1}$ and $l_i \in \mathcal{L}$.

Moreover, for $i \in \{1, \dots, d_a - 1\}$, l_i is the only integer $k \in \{0, \dots, l-1\}$ such that $r_i < k < r_{i+1}$ and $k \in \mathcal{L}$. Otherwise, if k was such an other integer, it would come $x_{k+1} = x_0$ and then using the same reasoning, there should exist an integer \tilde{r}_i being in \mathcal{R} and strictly between k and l_i (and so between r_i and r_{i+1}) which is impossible by definitions of \mathcal{R} , r_i and r_{i+1} .

Using again the fact that $x = (x_0, \dots, x_l)$ is a nearest neighbour path and the definitions of \mathcal{R} and r_{d_a} , we finally show that an integer $r_{d_a} \in \mathcal{L}$ being such that $r_{d_a} < l_{d_a} \leq l - 1$ must exist and that it is the only one in $\{0, \dots, l-1\}$ satisfying that condition.

In conclusion, we have finally just proved that in any case $|\mathcal{R}| = |\mathcal{L}|$ or equivalently $R(x_0, x) = L(x_0 + 1, x)$. ■

Lemme 7

If $l \in \mathbb{N}^*$ and $x = (x_0, \dots, x_l)$ is a nearest neighbour path, it holds :

- $x_0 = x_l \Rightarrow \forall k \in \mathbb{Z}, R(k, x) = L(k + 1, x)$.
- $x_0 < x_l \Rightarrow \begin{cases} \forall k \in \llbracket x_0, x_l - 1 \rrbracket, R(k, x) = L(k + 1, x) + 1, \\ \forall k \in \mathbb{Z} \setminus \llbracket x_0, x_l - 1 \rrbracket, R(k, x) = L(k + 1, x). \end{cases}$

$$\cdot x_0 > x_l \Rightarrow \begin{cases} \forall k \in \llbracket x_l, x_0 - 1 \rrbracket, R(k, x) + 1 = L(k + 1, x), \\ \forall k \in \mathbb{Z} \setminus \llbracket x_l, x_0 - 1 \rrbracket, R(k, x) = L(k + 1, x). \end{cases}$$

Démo 17

The first thing to notice is that the two last points of lemma 7 are equivalent. Indeed, if $x = (x_0, \dots, x_l)$ is a nearest neighbour path of length l with $x_0 > x_l$, then, $\tilde{x} = (\tilde{x}_0, \dots, \tilde{x}_l) = (x_l, \dots, x_0)$ is a nearest neighbour path of length l with $\tilde{x}_0 < \tilde{x}_l$ and for any k in \mathbb{Z} , it comes :

$$\begin{aligned} R(k, \tilde{x}) &= \sum_{i=0}^{l-1} \mathbb{1}_{\{\tilde{x}_i=k, \tilde{x}_{i+1}=k+1\}} = \sum_{i=0}^{l-1} \mathbb{1}_{\{x_{l-i}=k, x_{l-(i+1)}=k+1\}} \\ &= \sum_{i'=0}^{l-1} \mathbb{1}_{\{x_{i'+1}=k, x_{i'}=k+1\}} = L(k, x) \end{aligned}$$

and

$$\begin{aligned} L(k + 1, \tilde{x}) &= \sum_{i=0}^{l-1} \mathbb{1}_{\{\tilde{x}_i=k+1, \tilde{x}_{i+1}=k\}} = \sum_{i=0}^{l-1} \mathbb{1}_{\{x_{l-i}=k+1, x_{l-(i+1)}=k\}} \\ &= \sum_{i'=0}^{l-1} \mathbb{1}_{\{x_{i'+1}=k+1, x_{i'}=k\}} = R(k, x). \end{aligned}$$

Hence, the two last points of lemma 7 are equivalent and we will prove the second one.

The second thing to notice is that whatever x_0 and x_l are, for any k in $\mathbb{Z} \setminus \{x_0, \dots, x_l\}$, it holds :

$$\{i \in \{0, \dots, l\} / x_i = k\} = \emptyset$$

but

$$\{i \in \{0, \dots, l-1\} / x_i = k\} \supset \{i \in \{0, \dots, l-1\} / x_i = k, x_{i+1} = k+1\}$$

and

$$\{i \in \{0, \dots, l-1\} / x_{i+1} = k\} \supset \{i \in \{0, \dots, l-1\} / x_i = k+1, x_{i+1} = k\}$$

so that for any k in $\mathbb{Z} \setminus \{x_0, \dots, x_l\}$, it comes : $R(k, x) = L(k + 1, x) = 0$. And it then remains to establish the link between $R(k, x)$ and $L(k + 1, x)$ for any k in $\{x_0, \dots, x_l\}$.

In order to prove the results, for all $k \in \{x_0, \dots, x_l\}$ we will decompose the path x according to peculiar "times" adapted to x and the site k :

$$\begin{aligned} \cdot t_k &= \min(\{j \in \{0, \dots, l\} / x_j = k\}) \quad (\text{first entry time in } k), \\ \cdot d_k &= \max(\{j \in \{0, \dots, l\} / x_j = k\}) \quad (\text{last exit time from } k). \end{aligned}$$

We are now going to establish the link between $R(k, x)$ and $L(k + 1, x)$ for any k in $\{x_0, \dots, x_l\}$ accordingly to the relative position of k with respect to x_0 and x_l .

From the definition of the "times" t_k and d_k , for any $k \in \{x_0, \dots, x_l\}$, it always holds :

$$x_{t_k} = x_{d_k} = k \quad \text{and} \quad 0 \leq t_k \leq d_k \leq l$$

and more specifically, it holds :

- if $k < x_0$, then, $[\forall j \leq t_k - 1, x_j > k]$ and $[\forall j \geq d_k + 1, x_j > k]$.
- if $x_l \leq k$, then, $[\forall j \leq t_k - 1, x_j < k]$ and $[\forall j \geq d_k + 1, x_j < k]$.
- if $x_0 \leq k \leq x_l - 1$, then, $[\forall j \leq t_k - 1, x_j < k]$ and $[\forall j \geq d_k + 1, x_j > k]$.

So that in any of those three situations, we can write the two following decompositions :

$$\begin{aligned}
R(k, x) &= \sum_{i=0}^{l-1} \mathbb{1}_{\{x_i=k, x_{i+1}=k+1\}} = \sum_{i=t_k}^{d_k \wedge (l-1)} \mathbb{1}_{\{x_i=k, x_{i+1}=k+1\}} \\
&= \sum_{i=t_k}^{d_k-1} \mathbb{1}_{\{x_i=x_{t_k}, x_{i+1}=x_{t_k}+1\}} + \mathbb{1}_{\{x_{d_k}=k, x_{d_k+1}=k+1\}} \mathbb{1}_{k \neq x_l} \\
&= R(x_{t_k}, (x_{t_k}, \dots, x_{d_k})) + \mathbb{1}_{\{x_{d_k}=k, x_{d_k+1}=k+1\}} \mathbb{1}_{k \neq x_l}
\end{aligned}$$

and

$$\begin{aligned}
L(k+1, x) &= \sum_{i=0}^{l-1} \mathbb{1}_{\{x_i=k+1, x_{i+1}=k\}} = \sum_{i=(t_k-1) \vee 0}^{d_k-1} \mathbb{1}_{\{x_i=k+1, x_{i+1}=k\}} \\
&= \mathbb{1}_{\{x_{t_k-1}=k+1, x_{t_k}=k\}} \mathbb{1}_{k \neq x_0} + \sum_{i=t_k}^{d_k-1} \mathbb{1}_{\{x_i=x_{t_k}+1, x_{i+1}=x_{t_k}\}} \\
&= \mathbb{1}_{\{x_{t_k-1}=k+1, x_{t_k}=k\}} \mathbb{1}_{k \neq x_0} + L(x_{t_k} + 1, (x_{t_k}, \dots, x_{d_k})).
\end{aligned}$$

But $(x_{t_k}, \dots, x_{d_k})$ is a nearest neighbour path of length $d_k - t_k$ such that $x_{t_k} = x_{d_k} = k$ so that the lemma 6 ensures that $L(x_{t_k} + 1, (x_{t_k}, \dots, x_{d_k})) = R(x_{t_k}, (x_{t_k}, \dots, x_{d_k}))$ and it comes that :

$$\begin{aligned}
L(k+1, x) - R(k, x) &= \mathbb{1}_{\{x_{t_k-1}=k+1, x_{t_k}=k\}} \mathbb{1}_{k \neq x_0} - \mathbb{1}_{\{x_{d_k}=k, x_{d_k+1}=k+1\}} \mathbb{1}_{k \neq x_l} \\
&= \mathbb{1}_{\{x_{t_k-1}=k+1\}} \mathbb{1}_{k \neq x_0} - \mathbb{1}_{\{x_{d_k+1}=k+1\}} \mathbb{1}_{k \neq x_l}.
\end{aligned}$$

Finally, we get :

- when $k < x_0$, $x_{t_k-1} = k+1$ and $x_{d_k+1} = k+1$, thus, $L(k+1, x) = R(k, x)$.
- when $x_l \leq k$, $x_{t_k-1} \neq k+1$ and $x_{d_k+1} \neq k+1$, thus, $L(k+1, x) = R(k, x)$.
- when $x_0 \leq k \leq x_l - 1$, $x_{t_k-1} \neq k+1$ and $x_{d_k+1} = k+1$, thus, $L(k+1, x) + 1 = R(k, x)$.

■

6.3 "Left jumps" process

To ensure that T_n is well defined for any $n \in \mathbb{N}$ (i.e. $T_n < +\infty$ \mathbf{P}_0^ν -a.s.), we will first define the assumption :

$$\boxed{\mathbf{A1} : \forall \nu \in \mathcal{P}, \mathbb{E}^\nu [|\log(\rho_0)|] < +\infty \text{ and } \mathbb{E}^\nu [\log(\rho_0)] < 0.}$$

Since for any $n \in \mathbb{N}$ (thanks to assumption **A1**) the RWRE reaches \mathbf{P}_0^ν -a.s. the site n , it is possible to consider the random nearest neighbour path

$X_{[0,T_n]} = (X_0, \dots, X_{T_n})$ of length T_n and to define for any $k \in \mathbb{Z}$ the three random quantities :

$$L_k^n = L(k, X_{[0,T_n]}) \quad \text{and} \quad R_k^n = R(k, X_{[0,T_n]}) \quad \text{and} \quad V_{T_n} = v(X_{[0,T_n]}).$$

Thus for any $n \in \mathbb{N}$, any nearest neighbour path $x_{[0,t_n]} = (x_0, \dots, x_{t_n})$ of length t_n , as $(X_n)_{n \in \mathbb{N}}$ is a Markov chain starting at 0 under the quenched probability $\mathbf{P}_{0,\omega}$ for $\omega \in E$, it comes by the strong Markov property :

$$\begin{aligned} & \mathbf{P}_{0,\omega} (X_{[0,T_n]} = x_{[0,t_n]}) \\ &= \mathbf{P}_{0,\omega} (X_{T_n} = x_{t_n} \mid X_{[0,T_n-1]} = x_{[0,t_n-1]}) \\ & \quad \times \mathbf{P}_{0,\omega} (X_{[0,T_n-1]} = x_{[0,t_n-1]}) \\ &= Q_\omega(x_{t_n-1}, x_{t_n}) \mathbf{P}_{0,\omega} (X_{[0,T_n-1]} = x_{[0,t_n-1]}) \\ &= (\omega_{x_{t_n-1}} \mathbb{1}_{\{x_{t_n}=x_{t_n-1}+1\}} + (1 - \omega_{x_{t_n-1}}) \mathbb{1}_{\{x_{t_n}=x_{t_n-1}-1\}}) \\ & \quad \times \mathbf{P}_{0,\omega} (X_{[0,T_n-1]} = x_{[0,t_n-1]}) \\ &= \prod_{i=0}^{t_n-1} (\omega_{x_i} \mathbb{1}_{\{x_{i+1}=x_i+1\}} + (1 - \omega_{x_i}) \mathbb{1}_{\{x_{i+1}=x_i-1\}}) \\ &= \prod_{k \in v_{x_{[0,t_n-1]}}} \omega_k^{\sum_{i=0}^{t_n-1} \mathbb{1}_{\{x_i=k, x_{i+1}=k+1\}}} (1 - \omega_k)^{\sum_{i=0}^{t_n-1} \mathbb{1}_{\{x_i=k, x_{i+1}=k-1\}}} \\ &= \prod_{k \in v_{x_{[0,t_n-1]}}} \omega_k^{R(k, x_{[0,t_n]})} (1 - \omega_k)^{L(k, x_{[0,t_n]})}. \end{aligned}$$

And since the latter equality holds for every $\omega \in E$ by integrating the last equality with respect to the probability $\mathbb{P}^\nu = \nu^{\otimes \mathbb{Z}}$ on (E, \mathcal{E}) , it comes :

$$\begin{aligned} \mathbf{P}_0^\nu (X_{[0,T_n]} = x_{[0,t_n]}) &= \int_{]0,1[\mathbb{Z}} \left(\prod_{k \in v_{x_{[0,t_n-1]}}} \omega_k^{R(k, x_{[0,t_n]})} (1 - \omega_k)^{L(k, x_{[0,t_n]})} \right) \nu^{\otimes \mathbb{Z}}(d\omega) \\ &= \prod_{k \in v_{x_{[0,t_n-1]}}} \int_{]0,1[} t^{R(k, x_{[0,t_n]})} (1 - t)^{L(k, x_{[0,t_n]})} \nu(dt). \end{aligned}$$

Hence, when the RWRE starts from 0, the log-likelihood of the observations reads :

$$\begin{aligned} l_n(\nu) &= \log(\mathbf{P}_0^\nu (X_{[0,T_n]} = x_{[0,t_n]})) = \sum_{k \in v_{x_{[0,t_n-1]}}} \log \left(\int_{]0,1[} t^{R_k^n} (1 - t)^{L_k^n} \nu(dt) \right) \\ &= \sum_{k=0}^{n-1} \log \left(\int_{]0,1[} t^{R_k^n} (1 - t)^{L_k^n} \nu(dt) \right) + \sum_{k=\min(v_{x_{[0,t_n-1]}})}^{-1} \log \left(\int_{]0,1[} t^{R_k^n} (1 - t)^{L_k^n} \nu(dt) \right). \end{aligned}$$

Taking into account the link between R_k^n and L_k^n established in the lemma 7, it comes :

$$\sum_{k=0}^{n-1} \log \left(\int_{]0,1[} t^{R_k^n} (1 - t)^{L_k^n} \nu(dt) \right) = \sum_{k=0}^{n-1} \log \left(\int_{]0,1[} t^{L_{k+1}^n + 1} (1 - t)^{L_k^n} \nu(dt) \right)$$

and

$$\sum_{k=\min(v_{x_{[0,t_{n-1}]})}^{-1}} \log \left(\int_{]0,1[} t^{L_k^n} (1-t)^{L_k^n} \nu(dt) \right) = \sum_{k=\min(v_{x_{[0,t_{n-1}]})}^{-1}} \log \left(\int_{]0,1[} t^{L_{k+1}^n} (1-t)^{L_k^n} \nu(dt) \right)$$

so that we finally get :

$$\begin{aligned} l_n(\nu) &= \sum_{k=0}^{n-1} \log \left(\int_{]0,1[} t^{L_{k+1}^n+1} (1-t)^{L_k^n} \nu(dt) \right) + \sum_{k=\min(v_{x_{[0,t_{n-1}]})}^{-1}} \log \left(\int_{]0,1[} t^{L_{k+1}^n} (1-t)^{L_k^n} \nu(dt) \right) \\ &= \sum_{k=0}^{n-1} \phi_\nu(L_{k+1}^n, L_k^n) + \sum_{k=\min(v_{x_{[0,t_{n-1}]})}^{-1}} \phi_\nu(L_{k+1}^n - 1, L_k^n). \end{aligned}$$

We will now consider the situation where ν are probability distributions in a compact subset \mathcal{P} of the metric space of probability distributions on $\mathcal{B}(]0,1[)$ endowed with the total variation norm :

$$\boxed{\mathcal{P} \subset (\mathcal{M}_1(\mathcal{B}(]0,1[)), \|\cdot\|_{V.T.}) \text{ is compact}}$$

Defining $\phi(m, n) : \nu \in \mathcal{B}(]0,1[) \mapsto \phi_\nu(m, n) = \log \left(\int_{]0,1[} t^{m+1} (1-t)^n \nu(dt) \right)$ for any (m, n) in \mathbb{N}^2 , we automatically get that this application is continuous since the function \log is continuous and $t \in]0,1[\mapsto |t^{m+1}(1-t)^n|$ is bounded by 1 hence ensuring that $\nu \in \mathcal{P} \mapsto \int_{]0,1[} t^{m+1}(1-t)^n \nu(dt)$ is 1-lipschitz and so continuous by characterization of the total variation norm of a signed measure :

$$\begin{aligned} & \left| \int_{]0,1[} t^{m+1} (1-t)^n \nu(dt) - \int_{]0,1[} t^{m+1} (1-t)^n \mu(dt) \right| \\ &= \left| \int_{]0,1[} t^{m+1} (1-t)^n (\nu - \mu)(dt) \right| \\ &\leq \sup(\{(\nu - \mu)(f) / \|f\|_\infty \leq 1\}) \\ &= \|\nu - \mu\|_{V.T.}. \end{aligned}$$

Thanks to the compactness of the set \mathcal{P} and the continuity for all (m, n) in \mathbb{N}^2 of $\nu \in \mathcal{P} \mapsto \phi_\nu(m, n)$, the ‘‘Extreme value theorem’’ ensures that $\nu \in \mathcal{P} \mapsto l_n(\nu)$ is bounded and achieves its maximum on \mathcal{P} .

It is thus legitimate to consider any maximizer of this function :

Définition 13

We call maximum likelihood estimator of ν any M-estimator $\widehat{\nu}_n$ defined as :

$$\widehat{\nu}_n \in \underset{\nu \in \mathcal{P}}{\text{Argmax}}(l_n(\nu)).$$

Denoting $l_n^-(\nu) = \sum_{k=\min(v_{x_{[0,t_{n-1}]})}^{-1}} \phi_\nu(L_{k+1}^n - 1, L_k^n)$ and $l_n^+(\nu) = \sum_{k=0}^{n-1} \phi_\nu(L_{k+1}^n, L_k^n)$,

we can similarly define :

$$\widehat{\nu}_n^- \in \underset{\nu \in \mathcal{P}}{\text{Argmax}}(l_n^-(\nu)) \quad \text{and} \quad \widehat{\nu}_n^+ \in \underset{\nu \in \mathcal{P}}{\text{Argmax}}(l_n^+(\nu))$$

6.4 Links with branching processes in random environment

In the present work, the phrase “branching processes in random environment” will be abbreviated to BPRE.

6.4.1 Reminder on BPRE

Définition 14

If $(\zeta_{i,j})_{(i,j) \in \mathbb{N}^* \times \mathbb{N}}$ is a sequence of integer valued independent random variables such that for any $i \in \mathbb{N}^*$ the random variable $(\zeta_{i,j})_{j \in \mathbb{N}}$ are i.i.d. , then, the sequence of random variable $(Z_n)_{k \in \mathbb{N}}$ defined recursively by :

$$Z_0 = 0 \quad \text{and} \quad \forall k \in \mathbb{N}, Z_{k+1} = \sum_{j=0}^{Z_k} \zeta_{k+1,j}$$

is an inhomogeneous branching process with immigration.

Let $\omega \in E$ be a fixed environment and $(\zeta_{i,j})_{(i,j) \in \{1, \dots, n\} \times \mathbb{N}}$ be a sequence of integer valued independent random variables such that for any $i \in \{1, \dots, n\}$ and any $j \in \mathbb{N}$, $\zeta_{i,j}$ is a random variable with geometric distribution on \mathbb{N} with parameter ω_{n-i} (abbreviated $\zeta_{i,j} \sim \mathcal{G}_{\mathbb{N}}(\omega_{n-i})$ now) under $\mathbf{P}_{0,\omega}$ i.e. :

$$\forall k \in \mathbb{N}, \mathbf{P}_{0,\omega}(\zeta_{i,j} = k) = (1 - \omega_{n-i})^k \omega_{n-i}.$$

We define a finite sequence of random variable $(Z_k^n)_{k \in \{0, \dots, n\}}$ recursively by :

$$Z_0^n = 0 \quad \text{and} \quad \forall k \in \{0, \dots, n-1\}, Z_{k+1}^n = \sum_{j=0}^{Z_k^n} \zeta_{k+1,j}.$$

Lemme 8

If $\omega \in E$ is a fixed environment and $(Z_k^n)_{k \in \{0, \dots, n\}}$ the associated finite sequence of random variable defined above then the random vectors (L_n^n, \dots, L_0^n) and (Z_0^n, \dots, Z_n^n) have the same distribution under $\mathbf{P}_{0,\omega}$.

Démo 18

By definition of L_n^n as the number of left steps from site n of the walk until time T_n and since T_n is the first entry time (from the right side) in site n (what in particular implies that $\forall k \in \{0, \dots, T_n - 1\}$, $x_k < n$), it comes : $\mathbf{P}_{x,\omega} - a.s. L_n^n = 0$.

By definition, since k is an integer in $\{0, \dots, n-1\}$, we get $L_{n-(k+1)}^n = L(n-k-1, X_{[0, T_n]}) = \sum_{i=0}^{T_n-1} \mathbb{1}_{\{x_i = n-k-1, x_{i+1} = n-k-2\}}$, that is L_{n-k-1}^n is the number of left steps from site $n-k-1$ of the walk until time T_n .

Reminding that each time the walk is in site $n-k-1$, it either goes to site $n-k$ by a right step or it goes to site $n-k-2$ by a left step and using Markov property several times, if we know that $L_{n-k}^n = h \Leftrightarrow R_{n-k-1}^n = h+1$ where $h \in \mathbb{N}$, the number of left steps from site $n-k-1$ of the walk until time T_n can be seen as the sum of $h+1$ terms, each corresponding to the number of times

the walk having already made exactly j right steps from $n - k - 1$ to $n - k$ for $j \in \{0, \dots, h\}$ and now being in $n - k - 1$ fails going to site $n - k$ by a right step and instead goes to site $n - k - 1$ by a left step.

Using again the Markov property of the RWRE, for every $j \in \{0, \dots, h\}$ the number of times the walk having already made exactly j right steps from $n - k - 1$ to $n - k$ and now being in $n - k - 1$ fails going to site $n - k$ by a right step and instead goes to site $n - k - 1$ by a left step is a random variable with distribution $\mathcal{G}_{\mathbb{N}}(\omega_{n-k-1})$ since it corresponds to the time one has to wait for getting a success when he repeats independent trials with probability of success ω_{n-k-1} minus 1.

So k being an integer in $\{0, \dots, n - 1\}$, we get that $L_{n-(k+1)}^n$ is the sum of $L_{n-k}^n + 1$ i.i.d. random variables with distribution $\mathcal{G}_{\mathbb{N}}(\omega_{n-(k+1)})$ that is :

$$\forall k \in \{0, \dots, n - 1\}, L_{n-(k+1)}^n = \sum_{j=0}^{L_{n-k}^n} \zeta_{k+1,j}$$

where for any $(\zeta_{i,j})_{(i,j) \in \{1, \dots, n\} \times \mathbb{N}}$ are independent random variables such that for any $i \in \{1, \dots, n\}$ and any $j \in \mathbb{N}$, $\zeta_{i,j} \sim \mathcal{G}_{\mathbb{N}}(\omega_{n-i})$.

The proof is thus complete. \blacksquare

Remarque 2

Similarly, defining $m_n = \min(X_0, \dots, X_n)$ the (random) smallest integer reached by the R.W.R.E., we can ensure that given m_n the random vectors $(L_0^n, \dots, L_{m_n}^n)$ and $(\tilde{Z}_0^n, \dots, \tilde{Z}_{m_n}^n)$ have the same distribution under $\mathbf{P}_{0,\omega}$ where the finite sequence of random variable $(\tilde{Z}_k^n)_{k \in \{0, \dots, m_n - 1\}}$ is recursively defined by :

$$\tilde{Z}_0^n = Z_n^n \quad \text{and} \quad \forall k \in \{0, \dots, m_n + 2\}, \tilde{Z}_{k-1}^n = \sum_{j=1}^{Z_k^n} \tilde{\zeta}_{k-1,j}$$

where $(\tilde{\zeta}_{i,j})_{(i,j) \in \{m_n + 1, \dots, 0\} \times \mathbb{N}^*}$ is a sequence of integer valued independent random variables such that for any $(i, j) \in \{m_n + 2, \dots, 0\} \times \mathbb{N}^*$, $\tilde{\zeta}_{i,j} \sim \mathcal{G}_{\mathbb{N}}(\omega_i)$ under $\mathbf{P}_{0,\omega}$ and for any $j \in \mathbb{N}^*$, $\zeta_{m_n+1,j} = 0$ $\mathbf{P}_{0,\omega}$ - a.s..

If we furthermore assume that the environment ω is random, that is $\omega = (\omega_k)_{k \in \mathbb{Z}} \sim \mathbb{P}^\nu$, since $\mathbb{P}^\nu = \nu^{\otimes \mathbb{Z}}$, the random vectors $(\omega_0, \dots, \omega_n)$ and $(\omega_n, \dots, \omega_0)$ have the same distribution. And this allows to say that the random sequence $(L_{n-k}^n)_{k \in \{0, \dots, n\}}$ has the same distribution under the annealed law \mathbf{P}_0^* as the first $n+1$ generations of an inhomogeneous branching process in random environment with immigration and geometric offspring distribution $(Z_k)_{k \in \mathbb{N}}$ defined by :

$$Z_0 = 0 \quad \text{and} \quad \forall k \in \{0, \dots, n - 1\}, Z_{k+1} = \sum_{j=0}^{Z_k} \xi_{k+1,j}$$

where $(\xi_{i,j})_{(i,j) \in \mathbb{N}^* \times \mathbb{N}}$ are independent random variables such that for any $i \in \mathbb{N}^*$ and any $j \in \mathbb{N}$ $\xi_{i,j} \sim \mathcal{G}_{\mathbb{N}}(\omega_i)$ under $\mathbf{P}_{0,\omega}$.

Using that $(L_{n-k}^n)_{k \in \{0, \dots, n\}}$ and $(Z_k)_{k \in \{0, \dots, n\}}$ have the same distribution under annealed law \mathbf{P}_0^* in the definition of $l_n^+(\nu)$, it comes that $l_n^+(\nu)$ and $\sum_{k=0}^{n-1} \phi_\nu(Z_k, Z_{k+1})$ have the same distribution under \mathbf{P}_0^* .

Hence the study of the statistical property of $l_n^+(\nu)$ is equivalent to those of the BPRE $(Z_k)_{k \in \mathbb{N}}$. For this reason, the probabilistic and statistical properties of $(Z_k)_{k \in \mathbb{N}}$ will be developed in the next sections.

6.4.2 Probabilistic and statistical properties of BPRE

In order to prove many results, we will use the fact that the distribution of a random variable X valued in \mathbb{N} under a probability measure \mathbb{Q} is characterized by its generating function under the probability measure \mathbb{Q} defined as :

$$u \in [0, 1] \mapsto \mathbb{E}_{\mathbb{Q}} [u^X] = \sum_{k=0}^{+\infty} u^k \mathbb{Q}(X = k).$$

Lemme 9

If X is a random variable whose distribution is $\mathcal{G}_{\mathbb{N}}(p)$ under a probability measure \mathbb{Q} (that is for all k in \mathbb{N}^* , $\mathbb{Q}(X = k) = (1-p)^k p$), the generating function G_X of X under \mathbb{Q} reads for any u in $[0, 1]$:

$$G_X(u) = \frac{p}{1 - u(1-p)}$$

Démo 19

For any u in $[0, 1]$, it comes :

$$\begin{aligned} G_X(u) &= \mathbb{E}_{\mathbb{Q}} [u^X] = \sum_{k=0}^{+\infty} u^k \mathbb{Q}(X = k) \\ &= \sum_{k=0}^{+\infty} u^k (1-p)^k p = p \sum_{k=0}^{+\infty} (u(1-p))^k = \frac{p}{1 - u(1-p)} \end{aligned}$$

■

Hence, the generating function of a random variable X with distribution $\mathcal{G}_{\mathbb{N}}(p)$ reads $G_X(u) = \frac{p}{1 - u(1-p)}$ and conversely any random variable X whose generating function reads $G_X(u) = \frac{p}{1 - u(1-p)}$ has distribution $\mathcal{G}_{\mathbb{N}}(p)$.

Lemme 10

If X_1, \dots, X_n are n i.i.d. random variables with distribution $\mathcal{G}_{\mathbb{N}}(p)$ under a probability measure \mathbb{Q} where $p \in]0, 1[$ and $n \in \mathbb{N}^*$, then, the random variable $S_n = \sum_{k=1}^n X_k$ is a \mathbb{N} -valued random variable with distribution under the probability measure \mathbb{Q} given by :

$$\forall l \in \mathbb{N}, \mathbb{Q}(S_n = l) = \binom{l+n-1}{n-1} p^n (1-p)^l.$$

Démo 20

For any s in $[0, 1]$, using that X_1, \dots, X_n are n i.i.d. random variables with distribution $\mathcal{G}_{\mathbb{N}}(p)$ combined with lemma 9 it comes :

$$\begin{aligned}
 \mathbb{E} [s^{S_n}] &= \mathbb{E} \left[s^{\sum_{k=1}^n X_k} \right] = \prod_{k=1}^n \mathbb{E} [s^{X_k}] \\
 &= \mathbb{E} [s^{X_1}]^n = \left(\frac{p}{1 - s(1-p)} \right)^n = p^n \frac{1}{(1 - s(1-p))^n} \\
 &= p^n \frac{1}{(n-1)!} \left(u \mapsto \frac{1}{(1-u)} \right)^{(n-1)} [s(1-p)] \\
 &= p^n \frac{1}{(n-1)!} \left(u \mapsto \sum_{l=0}^{+\infty} u^l \right)^{(n-1)} [s(1-p)] \\
 &= p^n \frac{1}{(n-1)!} \left(u \mapsto \sum_{l=n-1}^{+\infty} \frac{l!}{(l-(n-1))!} u^{l-(n-1)} \right) [s(1-p)] \\
 &= \sum_{l=n-1}^{+\infty} \frac{l!}{(l-(n-1))!(n-1)!} p^n (s(1-p))^{l-(n-1)} \\
 &= \sum_{l=0}^{+\infty} \binom{l+n-1}{n-1} p^n (1-p)^l s^l
 \end{aligned}$$

And since the generating function $\mathbb{E} [s^{S_n}] = \sum_{l=0}^{+\infty} \mathbb{Q}(S_n = l) s^l$ characterizes the distribution of S_n , it follows that :

$$\forall l \in \mathbb{N}, \mathbb{Q}(S_n = l) = \binom{l+n-1}{n-1} p^n (1-p)^l.$$

■

We now state the markovian property of the process $(Z_k)_{k \in \mathbb{N}}$ which is the key result for all the statistical properties of the estimators, we will later investigate.

Théorème 20

If the environment is i.i.d. , that is $\omega = (\omega_k)_{k \in \mathbb{Z}} \sim \mathbb{P}^\nu = \nu^{\otimes \mathbb{Z}}$, then the inhomogeneous branching process in random environment with immigration and geometric offspring distribution (just defined) $(Z_k)_{k \in \mathbb{N}}$ is :

- an inhomogeneous Markov chain under quenched law $\mathbf{P}_{0,\omega}$ for each fixed environment $\omega \in E$ whose transition kernels Q_k are given by :

$$Q_k(m, n) = \binom{m+n}{m} \omega_{k+1}^{m+1} (1 - \omega_{k+1})^n.$$

- an homogeneous Markov chain under annealed law \mathbf{P}_0^ν whose transition kernel Q^ν is given by :

$$Q^\nu(m, n) = \binom{m+n}{m} \int_{]0,1[} t^{m+1} (1-t)^n \nu(dt) = \binom{m+n}{m} e^{\phi_\nu(m,n)}.$$

Démo 21

We will first prove that for each environment $\omega \in E$, $(Z_k)_{k \in \mathbb{N}}$ is an inhomogeneous Markov chain under quenched law $\mathbf{P}_{0,\omega}$.

By definition of $(Z_k)_{k \in \mathbb{N}}$ using the independence of $(\xi_{i,j})_{(i,j) \in \mathbb{N}^* \times \mathbb{N}}$, if $\omega \in E$ is a fixed environment, for $n \in \mathbb{N}$ and integers $(z_0, \dots, z_n, z_{n+1}) \in \mathbb{N}^{n+2}$ such that $\mathbf{P}_{0,\omega}(Z_n = z_n, \dots, Z_0 = z_0) \neq 0$, it comes :

$$\begin{aligned} \mathbf{P}_{0,\omega} \left(Z_{n+1} = z_{n+1} \middle| \bigcap_{r=0}^{n+1} \{Z_r = z_r\} \right) &= \frac{\mathbf{P}_{0,\omega} \left(\left\{ \sum_{j=0}^{z_n} \xi_{n+1,j} = z_{n+1} \right\} \cap \bigcap_{r=0}^n \{Z_r = z_r\} \right)}{\mathbf{P}_{0,\omega} \left(\bigcap_{r=0}^n \{Z_r = z_r\} \right)} \\ &= \frac{\mathbf{P}_{0,\omega} \left(\sum_{j=0}^{z_n} \xi_{n+1,j} = z_{n+1} \right) \mathbf{P}_{0,\omega} \left(\bigcap_{r=0}^n \{Z_r = z_r\} \right)}{\mathbf{P}_{0,\omega} \left(\bigcap_{r=0}^n \{Z_r = z_r\} \right)} \\ &= \mathbf{P}_{0,\omega} \left(\sum_{j=0}^{z_n} \xi_{n+1,j} = z_{n+1} \right). \end{aligned}$$

Since the variable $\xi_{n+1,0}, \dots, \xi_{n+1,z_n}$ are i.i.d. random variables with distribution $\mathcal{G}_{\mathbb{N}}(\omega_{n+1})$ under $\mathbb{P}_{0,\omega}$, lemma 10 ensures that the random variable $\sum_{j=0}^{z_n} \xi_{n+1,j}$ is integer-valued with distribution under $\mathbf{P}_{0,\omega}$ given by :

$$\mathbf{P}_{0,\omega} \left(\sum_{j=0}^{z_n} \xi_{n+1,j} = z_{n+1} \right) = \binom{z_n + z_{n+1}}{z_n} \omega_{n+1}^{z_n+1} (1 - \omega_{n+1})^{z_{n+1}}.$$

Thus, for each fixed environment $\omega \in E$, $(Z_k)_{k \in \mathbb{N}}$ is an inhomogeneous Markov chain under quenched law $\mathbf{P}_{0,\omega}$ whose transition kernels Q_k are given by :

$$Q_k(m, n) = \binom{m+n}{m} \omega_{k+1}^{m+1} (1 - \omega_{k+1})^n.$$

It then only remains to integrate with respect to the probability $\mathbb{P} = \nu^{\otimes \mathbb{Z}}$ on (E, \mathcal{E}) , to obtain for any $n \in \mathbb{N}$ and any integers $(z_0, \dots, z_n, z_{n+1}) \in \mathbb{N}^{n+2}$ such that $\mathbf{P}'_0(Z_n = z_n, \dots, Z_0 = z_0) \neq 0$:

$$\begin{aligned} \mathbf{P}'_0 \left(Z_{n+1} = z_{n+1} \middle| \bigcap_{r=0}^{n+1} \{Z_r = z_r\} \right) &= \int_{]0,1]^{\mathbb{Z}}} \left(\mathbf{P}_{0,\omega} \left(Z_{n+1} = z_{n+1} \middle| \bigcap_{r=0}^{n+1} \{Z_r = z_r\} \right) \right) \nu^{\otimes \mathbb{Z}}(d\omega) \\ &= \int_{]0,1]^{\mathbb{Z}}} Q_k^\nu(z_n, z_{n+1}) \nu^{\otimes \mathbb{Z}}(d\omega) \\ &= \int_{]0,1]^{\mathbb{Z}}} \left(\binom{z_n + z_{n+1}}{z_n} \omega_{n+1}^{z_n+1} (1 - \omega_{n+1})^{z_{n+1}} \right) \nu^{\otimes \mathbb{Z}}(d\omega) \\ &= \binom{z_n + z_{n+1}}{z_n} \int_{]0,1]^{\mathbb{Z}}} t^{z_n+1} (1-t)^{z_{n+1}} \nu(dt). \end{aligned}$$

So that we finally get that $(Z_k)_{k \in \mathbb{N}}$ is an homogeneous Markov chain under annealed law \mathbf{P}_0^ν whose transition kernel Q^ν is given by :

$$Q^\nu(m, n) = \binom{m+n}{m} \int_{]0,1[} t^{m+1}(1-t)^n \nu(dt) = \binom{m+n}{m} e^{\phi_\nu(m,n)}.$$

■

6.4.3 Existence of a limiting distribution for the BPRE

Lemme 11

If $(X_k)_{k \in \mathbb{N}}$ is a Markov chain with transition kernel P and initial distribution μ on a countable state space D and if π is a measure on D such that $\forall d \in D$, $\mathbb{P}_\mu(X_k = d) \xrightarrow[k \rightarrow +\infty]{} \pi(d)$, then, π is an invariant measure (not necessarily a probability) with respect to P .

Démo 22

Writting $D = \{d_1, d_2, d_3, \dots\}$, for (k, n) in $\mathbb{N} \times \mathbb{N}^*$, $\mathbb{P}_\mu(X_{k+1} \in \{d_1, \dots, d_n\}) \leq 1$ and $\pi(D)$ being the limit of the increasing sequence $(\pi(\{d_1, \dots, d_n\}))_{n \in \mathbb{N}^*}$ whose terms are itself the limit of the increasing sequence $(\mathbb{P}_\mu(X_k \in \{d_1, \dots, d_n\}))_{k \in \mathbb{N}}$, we first get that : $\pi(D) \leq 1 < +\infty$.

For any d in D and any k in \mathbb{N} , using the Markov property, it comes :

$$\mathbb{P}_\mu(X_{k+1} = d) = \sum_{c \in D} P(c, d) \mathbb{P}_\mu(X_k = c).$$

Nevertheless, thanks to Fatou's lemma applied to the counting measure on D , we knows that :

$$\underline{\lim} \left(\sum_{c \in D} P(c, d) \mathbb{P}_\mu(X_k = c) \right) \geq \sum_{c \in D} P(c, d) \underline{\lim} (\mathbb{P}_\mu(X_k = c)).$$

Combining those, it yields for all d in D :

$$\pi(d) = \underline{\lim} (\mathbb{P}_\mu(X_{k+1} = d)) \geq \sum_{c \in D} P(c, d) \pi(d) = \pi P(d).$$

Using the fact that for all d in D , $\pi(d) \geq \pi P(d)$ and that $\pi(D) = \pi P(D)$ (P being a Markov kernel), it then comes :

$$\pi(D) = \sum_{c \in D} \pi(d) \geq \sum_{c \in D} \pi P(d) = \pi P(D) = \pi(D).$$

Since $\pi(D) < +\infty$, this implies that all inequalities used for obtaining this upper bound are in fact equalities : for all d in D , $\pi(d) = \pi P(d)$. Hence π is an invariant measure with respect to P (with total mass less than 1). ■

Théorème 21

When assumption **A1** is satisfied, it holds for all $\nu \in \mathcal{P}$:

- The Markov chain $(Z_k)_{k \in \mathbb{N}}$ is irreducible, positive and recurrent under annealed law \mathbf{P}'_0 and its unique invariant probability measure π_ν satisfies :

$$\forall k \in \mathbb{N}, \pi_\nu(k) = \lim_{n \rightarrow +\infty} (\mathbf{P}'_0(Z_n = k)).$$

- Writing $S = \left(\sum_{i=0}^{+\infty} \prod_{j=1}^i \rho_j \right)^{-1}$, it holds :

$$\forall k \in \mathbb{N}, \pi_\nu(k) = \mathbb{E}^\nu [S(1 - S)^k].$$

In particular, $\forall j \in \mathbb{N}$, $\sum_{k=j+1}^{+\infty} \frac{k!}{(k-(j+1))!} \pi_\nu(k) = (j+1)! \mathbb{E}^\nu \left[\left(\sum_{n=1}^{+\infty} \prod_{k=1}^n \rho_k \right)^{j+1} \right]$

and thus $\sum_{k=0}^{+\infty} k \pi_\nu(k) = \sum_{n=1}^{+\infty} (\mathbb{E}^\nu [\rho_0])^n$ so that π_ν has a finite first order moment, if and only if $\mathbb{E}^\nu [\rho_0] < 1$.

Démo 23

We introduce the *quenched probability generating function* of the random variable $\xi_{n,0}$ written H_n and the one of the random variable Z_n written F_n that is their generating functions under the quenched probability $\mathbf{P}_{0,\omega}$.

We also introduce for any n in \mathbb{N}^* the sums S_n and \widetilde{S}_n defined by :

$$S_n^{-1} = \sum_{k=0}^n \prod_{i=0}^{k-1} \rho_{n-i} \quad \text{and} \quad \widetilde{S}_n^{-1} = \sum_{k=0}^n \prod_{i=1}^k \rho_i.$$

When i.i.d. random environment is assumed and when the assumption **A1** holds then for all ν in \mathcal{P} the random variable $(\log(\rho_n))_{n \in \mathbb{N}^*}$ are also i.i.d. in $\mathbb{L}^1(\mathbb{P}^\nu)$ and such that $\mathbb{E}^\nu [\log(\rho_1)] < 0$ so that the strong law of large numbers yields :

$$\frac{1}{n} \log \left(\prod_{k=1}^n \rho_k \right) = \frac{1}{n} \sum_{k=1}^n \log(\rho_k) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}'_0\text{-a.s.}} \mathbb{E}^\nu [\log(\rho_1)] = \mathbb{E}^\nu [\log(\rho_0)] < 0.$$

Hence by definition of the \mathbb{P}'_0 -a.s.-convergence, it implies in particular that it exists $\widetilde{\Omega} \in \mathcal{E}$ with $\mathbf{P}'_0(\widetilde{\Omega}) = 1$ such that :

$$\forall \omega \in \widetilde{\Omega}, \exists N(\omega) / \forall n \geq N(\omega), \prod_{k=1}^n \rho_k \leq e^{n \frac{\mathbb{E}^\nu[\log(\rho_0)]}{2}} = \left(e^{\frac{\mathbb{E}^\nu[\log(\rho_0)]}{2}} \right)^n.$$

As $\mathbb{E}^\nu [\log(\rho_0)] < 0$, it holds $e^{\frac{\mathbb{E}^\nu[\log(\rho_0)]}{2}} < 1$ so that for all ω in $\widetilde{\Omega}$ the geometric series $\sum_{n \geq N(\omega)} \left(e^{\frac{\mathbb{E}^\nu[\log(\rho_0)]}{2}} \right)^n$ converges and hence by comparison-test for

positive-term series, the positive-term series $\sum_{n \geq N(\omega)} \left(\prod_{k=1}^n \rho_k \right)$ also converges. So

the positive-term series $\sum_{n \geq 0} \left(\prod_{k=1}^n \rho_k \right)$ converges \mathbb{P}'_0 -a.s. towards a random variable which is strictly greater than 1 (since for all k in \mathbb{Z} , $\rho_k \neq 0$) and by taking the inverse we get that the sequence of random variables $\left(\widetilde{S}_n \right)_{n \in \mathbb{N}}$ converges

\mathbb{P}'_0 -a.s. towards a random variable S verifying $\mathbb{P}'(0 < S < 1) = 1$.

Using the fact that $Z_0 = 0$ so that $Z_1 = \xi_{1,0} \sim \mathcal{G}_{\mathbb{N}}(\omega_1)$ under $\mathbf{P}_{0,\omega}$ it first comes :

$$F_1(u) = H_1(u) = \frac{\omega_1}{1 - u(1 - \omega_1)}$$

and since for all n in \mathbb{N}^* , $\xi_{n,0} \sim \mathcal{G}_{\mathbb{N}}(\omega_n)$, lemma 9 yields :

$$H_n(u) = \frac{\omega_n}{1 - u(1 - \omega_n)}$$

and so by definition of the BPRE $(Z_n)_{n \in \mathbb{N}}$, we compute for all n in \mathbb{N} :

$$\begin{aligned} F_{n+1}(u) &= \mathbf{E}_{0,\omega} [u^{Z_{n+1}}] = \mathbf{E}_{0,\omega} \left[u^{\sum_{j=0}^{Z_n} \xi_{n+1,j}} \right] \\ &= \mathbf{E}_{0,\omega} \left[u^{\sum_{j=0}^{Z_n} \xi_{n+1,j}} \right] = \mathbf{E}_{0,\omega} \left[\sum_{k=0}^{+\infty} u^{\sum_{j=0}^{Z_n} \xi_{n+1,j}} \mathbb{1}_{Z_n=k} \right] \\ &= \sum_{k=0}^{+\infty} \mathbf{E}_{0,\omega} \left[u^{\sum_{j=0}^{Z_n} \xi_{n+1,j}} \mathbb{1}_{\{Z_n=k\}} \right] = \sum_{k=0}^{+\infty} \mathbf{E}_{0,\omega} \left[\prod_{j=0}^k u^{\xi_{n+1,j}} \right] \mathbf{E}_{0,\omega} [\mathbb{1}_{\{Z_n=k\}}] \\ &= \sum_{k=0}^{+\infty} \mathbf{E}_{0,\omega} [u^{\xi_{n+1,0}}]^{k+1} \mathbf{P}_{0,\omega}(Z_n = k) = H_{n+1}(u) \sum_{k=0}^{+\infty} H_{n+1}(u)^k \mathbf{P}_{0,\omega}(\mathbb{1}_{Z_n=k}) \\ &= H_{n+1}(u) F_n(H_{n+1}(u)). \end{aligned}$$

Using those two intermediate results, we obtain by induction that for all $n \in \mathbb{N}^*$:

$$F_n(u) = \frac{\prod_{k=1}^n \omega_k}{A_n(\omega) - uB_n(\omega)} \text{ where } \begin{cases} A_{n+1}(\omega) = A_n(\omega) - \omega_{n+1}B_n(\omega), \\ B_{n+1}(\omega) = (1 - \omega_{n+1})A_n(\omega), \end{cases}$$

with $A_1(\omega) = 1$ and $B_1(\omega) = 1 - \omega_1$.

Indeed, we just showed that the result holds for F_1 and thanks to the recursion formula $F_{n+1}(u) = H_{n+1}(u)F_n(H_{n+1}(u))$ assuming it holds for a positive

integer n , it comes :

$$\begin{aligned}
F_{n+1}(u) &= H_{n+1}(u)F_n(H_{n+1}(u)) \\
&= \frac{\omega_{n+1}}{1-u(1-\omega_{n+1})} F_n\left(\frac{\omega_{n+1}}{1-u(1-\omega_{n+1})}\right) \\
&= \frac{\omega_{n+1}}{1-u(1-\omega_{n+1})} \frac{\prod_{k=1}^n \omega_k}{A_n(\omega) - \frac{\omega_{n+1}}{1-u(1-\omega_{n+1})} B_n(\omega)} \\
&= \frac{\prod_{k=1}^{n+1} \omega_k}{A_n(\omega)(1-u(1-\omega_{n+1})) - \omega_{n+1} B_n(\omega)} \\
&= \frac{\prod_{k=1}^{n+1} \omega_k}{[A_n(\omega) - \omega_{n+1} B_n(\omega)] - u[(1-\omega_{n+1})A_n(\omega)]} \\
&= \frac{\prod_{k=1}^{n+1} \omega_k}{A_{n+1}(\omega) - uB_{n+1}(\omega)}.
\end{aligned}$$

By induction, we also obtain by induction for all n in \mathbb{N}^* :

$$A_n(\omega) = S_n^{-1} \prod_{k=1}^n \omega_k \quad \text{and} \quad B_n(\omega) = (S_n^{-1} - 1) \prod_{k=1}^n \omega_k.$$

Indeed, by definition, we first get :

$$A_1(\omega) = 1 = \omega_1 \left(1 + \frac{1-\omega_1}{\omega_1}\right) = \sum_{k=0}^1 \prod_{i=0}^{k-1} \rho_{1-i} \prod_{k=1}^1 \omega_k = S_1^{-1} \prod_{k=1}^1 \omega_k$$

and

$$B_1(\omega) = 1 - \omega_1 = \omega_1 \left(\frac{1-\omega_1}{\omega_1}\right) = \sum_{k=1}^1 \prod_{i=0}^{k-1} \rho_{1-i} \prod_{k=1}^1 \omega_k = (S_1^{-1} - 1) \prod_{k=1}^1 \omega_k.$$

And then assuming that the result holds for $n \geq 1$, we get :

$$\begin{aligned}
A_{n+1}(\omega) &= A_n(\omega) - \omega_{n+1} B_n(\omega) \\
&= S_n^{-1} \prod_{k=1}^n \omega_k - \omega_{n+1} (S_n^{-1} - 1) \prod_{k=1}^n \omega_k = \prod_{k=1}^n \omega_k (\omega_{n+1} + (1 - \omega_{n+1}) S_n^{-1}) \\
&= (1 + \rho_{n+1} S_n^{-1}) \prod_{k=1}^{n+1} \omega_k = \left(1 + \rho_{n+1} \sum_{k=0}^n \prod_{i=0}^{k-1} \rho_{n-i}\right) \prod_{k=1}^{n+1} \omega_k \\
&= \left(1 + \sum_{k=0}^n \prod_{i=-1}^{k-1} \rho_{n-i}\right) \prod_{k=1}^{n+1} \omega_k = \left(\sum_{k=-1}^n \prod_{i=0}^k \rho_{(n+1)-i}\right) \prod_{k=1}^{n+1} \omega_k \\
&= \left(\sum_{k=0}^{n+1} \prod_{i=0}^k \rho_{(n+1)-i}\right) \prod_{k=1}^{n+1} \omega_k = S_{n+1}^{-1} \prod_{k=1}^{n+1} \omega_k,
\end{aligned}$$

and

$$\begin{aligned}
B_{n+1}(\omega) &= (1 - \omega_{n+1})A_n(\omega) = S_n^{-1} \frac{(1 - \omega_{n+1})}{\omega_{n+1}} \prod_{k=1}^{n+1} \omega_k \\
&= \left(\sum_{k=0}^n \prod_{i=0}^{k-1} \rho_{n-i} \right) \rho_{n+1} \prod_{k=1}^{n+1} \omega_k = \left(\sum_{k=0}^n \prod_{i=-1}^{k-1} \rho_{n-i} \right) \prod_{k=1}^{n+1} \omega_k \\
&= \left(\sum_{k=0}^n \prod_{i=0}^k \rho_{(n+1)-i} \right) \prod_{k=1}^{n+1} \omega_k = \left(\sum_{k=-1}^n \prod_{i=0}^k \rho_{(n+1)-i} - 1 \right) \prod_{k=1}^{n+1} \omega_k \\
&= \left(\sum_{k=0}^{n+1} \prod_{i=0}^k \rho_{(n+1)-i} - 1 \right) \prod_{k=1}^{n+1} \omega_k = (S_{n+1}^{-1} - 1) \prod_{k=1}^{n+1} \omega_k.
\end{aligned}$$

Finally, we get that for all n in \mathbb{N}^* and all u in $[0, 1]$, $F_n(u) = \frac{S_n}{1-u(1-S_n)}$ so that under the quenched probability $\mathbf{P}_{0,\omega}$, the random variable Z_n has distribution $\mathcal{G}_{\mathbb{N}}(S_n)$.

Since for all n in \mathbb{N}^* the random vectors $(\omega_0, \dots, \omega_n)$ and $(\omega_n, \dots, \omega_0)$ have the same distribution under \mathbb{P}^ν when i.i.d. random environment is assumed, the sums S_n and \widetilde{S}_n have the same distribution under \mathbb{P}^ν .

In particular, it directly follows for all u in $[0, 1]$ that the functions F_n and \widetilde{F}_n evaluated in u that is $F_n(u)$ and $\widetilde{F}_n(u) = \frac{\widetilde{S}_n}{1-u(1-\widetilde{S}_n)}$ have the same distribution under \mathbb{P}^ν and since $Z_n \sim \mathcal{G}_{\mathbb{N}}(S_n)$, it comes for all k in \mathbb{N} that $\mathbf{P}_0^\nu(Z_n = k) = \mathbb{E}^\nu[\mathbf{P}_{0,\omega}(Z_n = k)] = \mathbb{E}^\nu[S_n(1 - S_n)^k] = \mathbb{E}^\nu[\widetilde{S}_n(1 - \widetilde{S}_n)^k]$.

On one hand, since \mathbb{P}^ν -a.s.-convergence is stronger than convergence in distribution under \mathbb{P}^ν and since for all u in $[0, 1]$ the random variables $F_n(u)$ and $\widetilde{F}_n(u)$ have the same distribution, the last two results ensure that $(\widetilde{F}_n(u))_{n \in \mathbb{N}}$ converges in distribution under \mathbb{P}^ν towards $F(u) = \frac{S}{1-u(1-S)}$ the generating function evaluated in u of a random variable with probability distribution $\mathcal{G}_{\mathbb{N}}(S)$.

On the other hand, since $\widetilde{S}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}^\nu\text{-a.s.}} S$ and \mathbb{P}^ν -a.s. $0 < \widetilde{S}_n < 1$ and since for all k in \mathbb{N} the function $x \in \mathbb{R} \mapsto x^k$ is continuous and hence bounded on the compact interval $[0, 1]$, the dominated convergence theorem implies that :

$$\forall k \in \mathbb{N}, \mathbf{P}_0^\nu(Z_n = k) = \mathbb{E}^\nu[\widetilde{S}_n(1 - \widetilde{S}_n)^k] \xrightarrow[n \rightarrow +\infty]{} \mathbb{E}^\nu[S(1 - S)^k].$$

And so, since $(Z_n)_{n \in \mathbb{N}}$ is a Markov chain under annealed law \mathbf{P}_0^ν with transition kernel Q^ν and initial distribution δ_0 on the countable state space \mathbb{Z} , lemma 11 ensures that the probability measure π_ν is invariant with respect to Q^ν .

Writting π_ν the measure on \mathbb{N} characterized by $\pi_\nu(k) = \mathbb{E}^\nu[S(1 - S)^k]$ and using Fubini-Tonelli theorem we get :

$$\sum_{k=0}^{+\infty} \pi_\nu(k) = \sum_{k=0}^{+\infty} \mathbb{E}^\nu[S(1 - S)^k] = \mathbb{E}^\nu \left[\sum_{k=0}^{+\infty} S(1 - S)^k \right] = \mathbb{E}^\nu[1] = 1$$

and for any j in \mathbb{N}

$$\begin{aligned}
\sum_{k=j+1}^{+\infty} \frac{k!}{(k-(j+1))!} \pi_\nu(k) &= \sum_{k=j+1}^{+\infty} \frac{k!}{(k-(j+1))!} \mathbb{E}^\nu [S(1-S)^k] \\
&= \mathbb{E}^\nu \left[S \sum_{k=j+1}^{+\infty} \frac{k!}{(k-(j+1))!} (1-S)^k \right] \\
&= \mathbb{E}^\nu \left[S \left(u \mapsto \sum_{k=j+1}^{+\infty} \frac{k!}{(k-(j+1))!} u^k \right) [1-S] \right] \\
&= \mathbb{E}^\nu \left[S \left(u \mapsto \sum_{k=0}^{+\infty} u^k \right)^{(j+1)} [1-S] \right] \\
&= \mathbb{E}^\nu \left[S \left(u \mapsto \frac{1}{1-u} \right)^{(j+1)} [1-S] \right] \\
&= \mathbb{E}^\nu \left[S \left(u \mapsto \frac{(j+1)!}{(1-u)^{j+2}} \right) [1-S] \right] \\
&= \mathbb{E}^\nu \left[\frac{S}{S^{j+2}} \right] = \mathbb{E}^\nu \left[(S^{-1})^{j+1} \right] = (j+1)! \mathbb{E}^\nu \left[\left(\sum_{n=1}^{+\infty} \prod_{k=1}^n \rho_k \right)^{j+1} \right]
\end{aligned}$$

and

$$\sum_{k=0}^{+\infty} k \pi_\nu(k) = \mathbb{E}^\nu \left[\sum_{k=0}^{+\infty} k S(1-S)^k \right] = \mathbb{E}^\nu [S^{-1} - 1] = \mathbb{E}^\nu \left[\sum_{k=1}^{+\infty} \prod_{i=1}^k \rho_i \right] = \sum_{k=1}^{+\infty} (\mathbb{E}^\nu [\rho_0])^k.$$

This shows that π_ν is a probability measure on \mathbb{N} equivalent to the counting measure (indeed, $\mathbb{P}^\nu - a.s.$ $0 < S < 1$ so that $\forall k \in \mathbb{N}$, $\mathbb{E}^\nu [S(1-S)^k] > 0$) whose mean is finite, if and only if, $\mathbb{E}^\nu [\rho_0] < 1$.

Furthermore, since $\forall (m, n) \in \mathbb{Z}^2 / m \neq n$, $Q^\nu(m, n) > 0$, the Markov chain $(Z_k)_{k \in \mathbb{N}}$ is irreducible what ensures that the existence of the invariant probability measure π_ν for Q^ν is equivalent to its positive recurrence. Finally, $(Z_k)_{k \in \mathbb{N}}$ being an irreducible recurrent positive Markov chain, it admits a unique invariant probability measure which is thus π_ν . \blacksquare

Thanks to this last theorem, we can now legitimately consider the Markov chain $(\widetilde{Z}_k)_{k \in \mathbb{N}}$ whose transition kernel is the kernel $Q^* = Q^{\nu^*}$ and initial distribution $\pi^* = \pi_{\nu^*}$ and we define the function :

$$l^+ : \nu \in \mathcal{P} \mapsto \mathbf{E}^* \left[\phi_\nu(\widetilde{Z}_0, \widetilde{Z}_1) \right] \in [-\infty, 0].$$

Théorème 22

If assumption **A1** is met, then for all ν in \mathcal{P} , it comes :

$$\frac{1}{n} \sum_{k=0}^{n-1} \phi_\nu(Z_k, Z_{k+1}) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^* - a.s.} l^+(\nu).$$

Démo 24

Thanks to theorem 21, $(Z_k)_{k \in \mathbb{N}}$ is an irreducible recurrent positive Markov chain with kernel Q^* under \mathbf{P}^* and unique invariant probability measure π^* so that $(Z_k, Z_{k+1})_{k \in \mathbb{N}}$ is also an irreducible recurrent positive Markov chain under \mathbf{P}^* with kernel $\tilde{Q}^*((x_1, x_2), (y_1, y_2)) = Q^*(y_1, y_2) \mathbb{1}_{\{x_2=y_1\}}$ hence admitting a unique invariant probability measure $\tilde{\pi}^*$ defined by $\tilde{\pi}^*(x_1, x_2) = \pi^*(x_1)Q^*(x_1, x_2)$. By definition of $(\tilde{Z}_k)_{k \in \mathbb{N}}$, the ergodic theorem applied to the latter ensures :

$$\frac{1}{n} \sum_{k=0}^{n-1} \phi_\nu(Z_k, Z_{k+1}) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^* \text{-a.s.}} \mathbf{E}^* \left[\phi_\nu(\tilde{Z}_0, \tilde{Z}_1) \right] = l^+(\nu).$$

■

6.4.4 Consistency**Théorème 23**

If assumption **A1** is met then for all ν in \mathcal{P} , it comes :

$$\frac{1}{n} l_n^+(\nu) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^*} l^+(\nu).$$

Démo 25

The random variables $l_n^+(\nu)$ and $\sum_{k=0}^{n-1} \phi_\nu(Z_k, Z_{k+1})$ having the same distribution under \mathbf{P}^* and theorem 22 giving the \mathbf{P}^* -a.s. convergence of $\frac{1}{n} \sum_{k=0}^{n-1} \phi_\nu(Z_k, Z_{k+1})$ which is stronger than the convergence in \mathbf{P}^* -probability, it is straightforward that :

$$\frac{1}{n} l_n^+(\nu) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^*} l^+(\nu).$$

■

Théorème 24

Provided **A1** is met, for any open subset $\mathcal{C} \subset \mathcal{P}$, there is a local uniform convergence :

$$\frac{1}{n} \sum_{k=0}^{n-1} \sup(\{\phi_\nu(L_{k+1}^n, L_k^n)/\nu \in \mathcal{C}\}) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^*} \mathbf{E}^* \left[\sup(\{\phi_\nu(\tilde{Z}_0, \tilde{Z}_1)/\nu \in \mathcal{C}\}) \right].$$

Démo 26

Let $\mathcal{C} \subset \mathcal{P}$ be an open subset of $(\mathcal{M}_1(\mathcal{B}([0, 1])), \|\cdot\|_{V.T.})$. Because of the equality in distribution of (Z_n) and L_k^n , we have :

$$\frac{1}{n} \sum_{k=0}^{n-1} \sup(\{\phi_\nu(L_{k+1}^n, L_k^n)/\nu \in \mathcal{C}\}) \sim \frac{1}{n} \sum_{k=0}^{n-1} \sup(\{\phi_\nu(Z_k, Z_{k+1})/\nu \in \mathcal{C}\}).$$

The function $(x, y) \in \mathbb{N}^2 \mapsto \sup(\{\phi_\nu(x, y)/\nu \in \mathcal{C}\})$ being of constant sign (negative), $\mathbf{E}^* \left[\sup(\{\phi_\nu(\tilde{Z}_0, \tilde{Z}_1)/\nu \in \mathcal{C}\}) \right]$ is well-defined and so, applying again

the ergodic theorem to the multivariate Markov chains $(Z_k, Z_{k+1})_{k \in \mathbb{N}}$, it comes by the same reasoning as before that :

$$\frac{1}{n} \sum_{k=0}^{n-1} \sup \left(\{ \phi_\nu(L_{k+1}^n, L_k^n) / \nu \in \mathcal{C} \} \right) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^*} \mathbf{E}^* \left[\sup \left(\{ \phi_\nu(\widetilde{Z}_0, \widetilde{Z}_1) / \nu \in \cdot \} \right) \right].$$

■

Lemme 12

Under assumption **A1**, if $\nu^* \in \mathcal{P}$ is such that $\mathbb{E}^{\nu^*} [\rho_0] < 1$, then, it holds :

$$\forall \nu \in \mathcal{P}, \quad l^+(\nu) \in]-\infty, 0].$$

Démo 27

For any $(m, n) \in \mathbb{N}^2$, by Jensen's inequality, we have for all ν in \mathcal{P} :

$$\begin{aligned} \phi_\nu(m, n) &= \log \left(\int_{]0,1[} t^{m+1} (1-t)^n \nu(dt) \right) \\ &\geq \int_{]0,1[} \log(t^{m+1} (1-t)^n) \nu(dt) \\ &= (m+1) \int_{]0,1[} \log(t) \nu(dt) + n \int_{]0,1[} \log(1-t) \nu(dt) \\ &= (m+1) \mathbb{E}^\nu [\log(\omega_0)] + n \mathbb{E}^\nu [\log(1-\omega_0)]. \end{aligned}$$

So that for any k in \mathbb{N} , it holds :

$$\phi_\nu(Z_k, Z_{k+1}) \geq (Z_k + 1) \mathbb{E}^\nu [\log(\omega_0)] + Z_{k+1} \mathbb{E}^\nu [\log(1-\omega_0)]$$

and summing this inequality for k ranging from 0 to $n-1$, it comes :

$$\begin{aligned} \frac{1}{n} \sum_{k=0}^{n-1} \phi_\nu(Z_k, Z_{k+1}) &\geq \frac{1}{n} \sum_{k=0}^{n-1} ((Z_k + 1) \mathbb{E}^\nu [\log(\omega_0)] + Z_{k+1} \mathbb{E}^\nu [\log(1-\omega_0)]) \\ &= \mathbb{E}^\nu [\log(\omega_0)] \frac{1}{n} \sum_{k=0}^{n-1} (Z_k + 1) + \mathbb{E}^\nu [\log(1-\omega_0)] \frac{1}{n} \sum_{k=0}^{n-1} Z_{k+1}. \end{aligned}$$

Since $(\widetilde{Z}_k)_{k \in \mathbb{N}}$ has $\pi^* = \pi_{\nu^*}$ as initial distribution and $\mathbb{E}^{\nu^*} [\rho_0] < 1$, it holds :

$$0 \leq \mathbb{E}^{\nu^*} [\widetilde{Z}_0] = \sum_{k=0}^{+\infty} k \pi_\nu(k) = \sum_{k=1}^{+\infty} (\mathbb{E}^{\nu^*} [\rho_0])^k < +\infty.$$

The Markov chain $(Z_k)_{k \in \mathbb{N}}$ being irreducible recurrent positive with kernel Q^* under \mathbf{P}^* and unique invariant probability measure π^* by definition of $(\widetilde{Z}_k)_{k \in \mathbb{N}}$ and thanks to theorem 21, the ergodic theorem ensures :

$$\frac{1}{n} \sum_{k=0}^{n-1} (Z_k + 1) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^* \text{-a.s.}} \mathbb{E}^{\nu^*} [\widetilde{Z}_0] + 1 \quad \text{and} \quad \frac{1}{n} \sum_{k=0}^{n-1} Z_{k+1} \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^* \text{-a.s.}} \mathbb{E}^{\nu^*} [\widetilde{Z}_0].$$

Since $\mathbb{E}^{\nu^*}[\rho_0] < 1 \iff \mathbb{E}^{\nu^*}\left[\frac{1}{\omega_0}\right] < 2$, Jensen's inequality gives :

$$0 < \mathbb{E}^{\nu^*}\left[\log\left(\frac{1}{\omega_0}\right)\right] < \log(2) \iff 0 > \mathbb{E}^{\nu^*}[\log(\omega_0)] > -\log(2).$$

Hence, it also comes :

$$-\infty < \mathbb{E}^{\nu^*}[\log(\omega_0)] + \mathbb{E}^{\nu^*}[\log(\rho_0)] = \mathbb{E}^{\nu^*}[\log(\omega_0\rho_0)] = \mathbb{E}^{\nu^*}[\log(1 - \omega_0)] < 0.$$

Finally, combining the latter \mathbf{P}^* -a.s. convergence, the \mathbf{P}^* -a.s. convergence ensured by theorem 22 and the above inequalities, we get that for any ν in \mathcal{P} , it holds \mathbf{P}^* -a.s. :

$$l^+(\nu) \geq \mathbb{E}^{\nu}[\log(\omega_0)] \left(1 + \mathbb{E}^{\nu^*}\left[\widetilde{Z}_0\right]\right) + \mathbb{E}^{\nu}[\log(1 - \omega_0)] \mathbb{E}^{\nu^*}\left[\widetilde{Z}_0\right] > -\infty.$$

Since $l^+(\nu)$ is a deterministic quantity, it yields $l^+(\nu) \in]-\infty, 0]$. \blacksquare

Lemme 13

Under assumption **A1**, if $\nu^* \in \mathcal{P}$ is such that $\mathbb{E}^{\nu^*}[\rho_0] < 1$, then, ν^* is the unique maximizer of the criterion function l^+ .

Démo 28

For any $\nu \in \mathcal{P}$, since it is now legitimate to consider $l^+(\nu) \in]-\infty, 0]$ thanks to lemma 12, using the fact that $(\widetilde{Z}_k)_{k \in \mathbb{N}}$ is a Markov chain with transition kernel $Q^* = Q^{\nu^*}$ with $Q^{\nu}(m, n) = \binom{m+n}{m} e^{\phi_{\nu}(m, n)}$ and initial distribution $\pi^* = \pi_{\nu^*}$, we get :

$$\begin{aligned} l^+(\nu) - l^+(\nu^*) &= \mathbf{E}^*[\phi_{\nu}(\widetilde{Z}_0, \widetilde{Z}_1)] - \mathbf{E}^*[\phi_{\nu^*}(\widetilde{Z}_0, \widetilde{Z}_1)] = \mathbf{E}^*[\phi_{\nu}(\widetilde{Z}_0, \widetilde{Z}_1) - \phi_{\nu^*}(\widetilde{Z}_0, \widetilde{Z}_1)] \\ &= \sum_{m=0}^{+\infty} \sum_{n=0}^{+\infty} [\phi_{\nu}(m, n) - \phi_{\nu^*}(m, n)] Q^*(m, n) \pi^*(m) \\ &= \sum_{m=0}^{+\infty} \sum_{n=0}^{+\infty} \left[\log\left(\frac{Q^{\nu}(m, n)}{\binom{m+n}{m}}\right) - \log\left(\frac{Q^*(m, n)}{\binom{m+n}{m}}\right) \right] Q^*(m, n) \pi^*(m) \\ &= \sum_{m=0}^{+\infty} \pi^*(m) \sum_{n=0}^{+\infty} \log\left(\frac{Q^{\nu}(m, n)}{Q^*(m, n)}\right) Q^*(m, n). \end{aligned}$$

For every $m \in \mathbb{N}$, Jensen's inequality applied with the probability measure $Q^*(m, \cdot)$ on \mathbb{N} yields :

$$\begin{aligned} \sum_{n=0}^{+\infty} \log\left(\frac{Q^{\nu}(m, n)}{Q^*(m, n)}\right) Q^*(m, n) &\leq \log\left(\sum_{n=0}^{+\infty} \frac{Q^{\nu}(m, n)}{Q^*(m, n)} Q^*(m, n)\right) \\ &= \log\left(\sum_{n=0}^{+\infty} Q^{\nu}(m, n)\right) = \log(Q^{\nu}(m, \mathbb{N})) = \log(1) = 0 \end{aligned}$$

with equality, if and only if, $n \in \mathbb{N} \mapsto \frac{Q^{\nu}(m, n)}{Q^*(m, n)}$ is $Q^*(m, \cdot)$ -a.s. constant and equals 1, if and only if, $n \in \mathbb{N} \mapsto \frac{Q^{\nu}(m, n)}{Q^*(m, n)}$ is constant and equals 1 (since for all

n in \mathbb{N} , $Q^*(m, n) > 0$), if and only if, $\forall n \in \mathbb{N}$, $Q^\nu(m, n) = Q^*(m, n)$, if and only if, $Q^\nu(m, \cdot) = Q^*(m, \cdot)$.

It thus comes $l^+(\nu) - l^+(\nu^*) \leq 0$ with equality, if and only if, for all $m \in \mathbb{N}$, $Q^\nu(m, \cdot) = Q^*(m, \cdot)$. Hence ν^* is a maximizer of the criterion function l^+ and moreover $l^+(\nu) = l^+(\nu^*)$, if and only if, for all $m \in \mathbb{N}$, $Q^\nu(m, \cdot) = Q^*(m, \cdot)$.

In particular, if $l^+(\nu) = l^+(\nu^*)$, then, for all $m \in \mathbb{N}$, $Q^\nu(m, 0) = Q^*(m, 0)$, that is, for all $m \in \mathbb{N}$, $\int_{]0,1[} t^{m+1} \nu(dt) = \phi_\nu(m, 0) = \phi_{\nu^*}(m, 0) = \int_{]0,1[} t^{m+1} \nu^*(dt)$, that is, the probability distributions ν and ν^* have the same moments and since their supports are bounded, it implies that they are the same : $\nu = \nu^*$. Hence, ν^* is the unique maximizer of the criterion function l^+ . ■

When restricting ourselves to the ballistic case, since we don't know the true distribution, it is natural to consider an other assumption :

$$\boxed{\mathbf{A2} : \forall \nu \in \mathcal{P}, \mathbb{E}^\nu [\rho_0] < 1.}$$

Remarque 3

Introducing assumption **A2** intuitively ensures "not to wait too long until the walk reaches n " : according to theorem 16, the waiting time T_n depends linearly of n asymptotically. Moreover, as we showed it in the proof of lemma 12, assumption **A2** combined with the Jensen's inequality yields :

$$\forall \nu \in \mathcal{P}, \mathbb{E}^\nu [\log(\omega_0)] > -\log(2).$$

6.5 The parametric setting

Coming back closer to the articles [FC14], [FLM13], [FLG14] and [CFL14], we can now restrict to parametric setting : $\mathcal{P} = \{\nu_\theta / \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^d$.

As often in parametric setting, we are now going to assume the *identifiability* of the parameter :

$$\boxed{\mathbf{A3} : \forall (\theta, \theta') \in \Theta^2, \theta = \theta' \Leftrightarrow \nu_\theta = \nu_{\theta'}.$$

Remarque 4

This assumptions of *identifiability* is especially true if we make the assumption that the mapping $\theta \in \Theta \mapsto \nu_\theta \in \mathcal{P}$ is one-to-one.

Instead of working with general probability distribution ν , we can now directly work with parameter θ just seen as a special case of the somewhat general former framework and consider the *maximum likelihood estimator* $\widehat{\theta}_n$ of θ as the M-estimator defined by :

$$\widehat{\theta}_n \in \underset{\theta \in \Theta}{\text{Argmax}}(l_n(\theta)).$$

6.5.1 The particular case of parametric ballistic RWRE

In order to ensure the condition of compactness of the general setting is met, we will assume the compactness of the parameter set and the continuity of the parametric mapping :

$$\boxed{\mathbf{A4} : \Theta \text{ is compact and } \theta \in \Theta \mapsto \nu_\theta \in \mathcal{P} \text{ is a continuous mapping.}}$$

And so, in the parametric setting, provided **A3** and **A4** are verified (what is the case from now on), we are in the same situation as the one we have been interested in so far. The result proven under (slightly adapted) assumptions **A1** or **A2** are naturally true and we'll see how it's possible to go even further with some additional regularity assumptions.

6.5.1.1 Consistency of $\widehat{\theta}_n$ for parametric ballistic RWRE

In order to get that $\widehat{\theta}_n$ is a consistent estimator of θ^* , we have to meet assumption :

$$\boxed{\mathbf{A5} : \inf_{\theta \in \Theta} (\{\mathbb{E}^\theta [\log(1 - \omega_0)]\}) > -\infty.}$$

Lemme 14

Under assumption **A1**, **A3**, **A4** and **A5**, if $\theta^* \in \Theta$ is such that $\mathbb{E}^{\theta^*} [\rho_0] < 1$, then, θ^* is the unique maximizer of the criterion function l^+ and it even holds for every $\epsilon > 0$:

$$\sup (\{l^+(\theta) / \|\theta - \theta^*\| \geq \epsilon\}) < l^+(\theta^*).$$

Démo 29

On one hand, $\theta^* \in \Theta$ being such that $\mathbb{E}^{\theta^*} [\rho_0] < 1$ and **A1**, **A3**, **A4** being verified, lemma 13 gives that θ^* is the unique maximizer of the criterion function l^+ .

On the other hand, as we showed in the proof of lemma 12, it holds under assumptions **A1** and **A2** for any $(m, n) \in \mathbb{N}^2$:

$$\phi_\nu(m, n) \geq (m + 1) \mathbb{E}^\nu [\log(\omega_0)] + n \mathbb{E}^\nu [\log(1 - \omega_0)]$$

and

$$\inf_{\theta \in \Theta} (\{\mathbb{E}^\theta [\log(\omega_0)]\}) > -\log(2),$$

so that it exists $A > 0$ such that for any $(m, n) \in \mathbb{N}^2$:

$$\begin{aligned} |\phi_\nu(m, n)| &\leq (m + 1) \log(2) + n \sup_{\theta \in \Theta} (\{\mathbb{E}^\theta [\log(1 - \omega_0)]\}) \\ &\leq (\log(2) + \mathbb{E}^\nu [\log(1 - \omega_0)]) (1 + m + n) = A(1 + m + n). \end{aligned}$$

Hence, it holds $|\phi_\nu(\widetilde{Z}_0, \widetilde{Z}_1)| \leq A(1 + \widetilde{Z}_0 + \widetilde{Z}_1)$. Since assumption **A2** combined with the very definition of $(\widetilde{Z}_n)_{n \in \mathbb{N}}$ and since π_{θ^*} ensures that :

$$\mathbb{E}^{\theta^*} [\widetilde{Z}_0] = \mathbb{E}^{\theta^*} [\widetilde{Z}_1] = \sum_{k=0}^{+\infty} k \pi_{\theta^*}(k) = \sum_{n=1}^{+\infty} (\mathbb{E}^{\nu_{\theta^*}} [\rho_0])^n < +\infty,$$

the continuity of the parameter mapping assumed in **A4** applied with Lebesgue's theorem of continuity under integral sign gives that $\theta \in \Theta \mapsto l^+(\theta)$ is continuous.

Finally, continuity of l^+ and the fact that θ^* is the unique maximizer of the criterion function l^+ yields : $\sup (\{l^+(\theta) / \|\theta - \theta^*\| \geq \epsilon\}) < l^+(\theta^*)$. ■

For proving consistency of $\widehat{\theta}_n$, we recall Wald's consistency theorem :

Théorème 25 (Wald)

If Θ is a compact set, if $\theta \in \Theta \mapsto m_\theta(y)$ is continuous for almost every y and $\forall \theta \in \Theta, \exists \delta_\theta > 0$ such that $\mathbb{E}[\sup (\{m_{\theta'}(Y_1) / \|\theta - \theta'\| \leq \delta_\theta\})] < +\infty$, then, for $(Y_n)_{n \in \mathbb{N}^*}$ i.i.d. random variables, defining $M_n(\theta) = \frac{1}{n} \sum_{i=1}^n m_\theta(Y_i)$ and $M(\theta) = \mathbb{E}[m_\theta(Y_1)]$, any estimator $\widehat{\theta}_n$ such that $M_n(\widehat{\theta}_n) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}_{\theta^*}} M(\theta^*)$ where θ^* is in the set $\Theta_M = \{\theta \in \Theta / M(\theta) = \sup (\{M(\theta') / \theta \in \Theta\})\}$ is such that : $d(\widehat{\theta}_n, \Theta_M) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}_{\theta^*}} 0$

Consistency of $\widehat{\theta}_n$ is then entailed by Wald's consistency theorem applied to functions $m_\theta = \phi_\theta$ since θ^* is the unique maximizer of l .

6.5.1.2 Asymptotic normality of parametric ballistic RWRE

It's also possible to prove that $\widehat{\theta}_n$ is asymptotically normal but it is much more involved and again further regularity assumption have to be introduced.

First of all, the fact that $\theta \mapsto \phi_\theta(m, n)$ is continuous is strengthened in :

A6 : $\forall (m, n) \in \mathbb{N}^2, \theta \in \Theta \mapsto \phi_\theta(m, n)$ is twice continuously differentiable.

Then, we make the two following integrability assumptions :

A7 : $\forall \theta \in \Theta, \exists q \in]1, +\infty[/ \tilde{\pi}_\theta \left(\|\dot{\phi}_\theta\|_2^{2q} \right) < +\infty$
and
 $\forall n \in \mathbb{N}, \sum_{m \in \mathbb{N}} \dot{Q}_\theta(n, m) = 0$

as well as

A8 : $\forall \theta \in \Theta, \exists \mathcal{V}_\theta$ neighborhood of θ such that :
 $\tilde{\pi}_\theta \left(\sup \left(\{ \|\dot{\phi}_{\theta'}\|_2^2 / \theta' \in \mathcal{V}_\theta \} \right) \right) < +\infty$
and
 $\tilde{\pi}_\theta \left(\sup \left(\{ \|\ddot{\phi}_{\theta'}\|_\infty / \theta' \in \mathcal{V}_\theta \} \right) \right) < +\infty$

Defining the Fisher information matrix as $\Sigma_\theta = \tilde{\pi}_\theta (\dot{\phi}_\theta \dot{\phi}_\theta^t) = -\tilde{\pi}_\theta (\ddot{\phi}_\theta)$, we assume finally :

A9 : $\forall \theta \in \Theta, \Sigma_\theta$ is nonsingular.

Proposition 1

Under assumptions **A1**, **A2**, **A3**, **A4**, **A5**, **A6**, **A7** and **A8**, the M-estimator $\widehat{\theta}_n$ is such that :

$$\widehat{\Sigma}_n = -\frac{1}{n} \sum_{k=0}^{n-1} \ddot{\phi}_{\widehat{\theta}_n}(L_{k+1}^n, L_k^n) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^*} \Sigma_{\theta^*}.$$

Démo 30

On one hand, simply reusing that for any $n \in \mathbb{N}$, $(L_{n-k}^n)_{k \in \{0, \dots, n\}}$ and $(Z_k)_{k \in \{0, \dots, n\}}$ have the same distribution under annealed law \mathbf{P}_0^* in the definition of $\widehat{\Sigma}_n$, it comes that $\widehat{\Sigma}_n$ and $-\frac{1}{n} \sum_{k=0}^{n-1} \ddot{\phi}_{\theta}(Z_k, Z_{k+1})$ have the same distribution under \mathbf{P}_0^* . Thus we can replace (L_{k+1}^n, L_k^n) by (Z_k, Z_{k+1}) in the proof

On the other hand, the claimed "uniform weak convergence" mainly relies on the compactness of the set Θ and the continuity of the application $(x, y, \theta) \mapsto \phi_{\theta}(x, y)$.

Indeed, using that when $(Y_i)_{i \in I}$ is a family of real random variables, it holds $\sup_{i \in I} (\mathbb{E}[Y_i]) \leq \mathbb{E} \left[\sup_{i \in I} (Y_i) \right]$, we get :

$$\begin{aligned} & \sup_{\theta \in \Theta} \left(\mathbb{E}^{\tilde{\theta}} \left[\sup_{\|\theta' - \theta\| < \delta} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) - \inf_{\|\theta' - \theta\| < \delta} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right] \right) \\ & \leq \mathbb{E}^{\tilde{\theta}} \left[\sup_{\theta \in \Theta} \left(\sup_{\|\theta' - \theta\| < \delta} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) - \inf_{\|\theta' - \theta\| < \delta} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right) \right] \end{aligned}$$

Combining this inequality with the continuity of $\theta \mapsto \phi_{\theta}(Z_k, Z_{k+1})$ for any $(x, y) \in \mathbb{N}^2$ and applying the dominated convergence theorem, we get that for any $\delta > 0$ and for any $\tilde{\theta} \in \Theta$:

$$\begin{aligned} & \overline{\lim}_{\delta \rightarrow 0} \left(\sup_{\theta \in \Theta} \left(\mathbb{E}^{\tilde{\theta}} \left[\sup_{\|\theta' - \theta\| < \delta} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) - \inf_{\|\theta' - \theta\| < \delta} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right] \right) \right) \\ & \leq \overline{\lim}_{\delta \rightarrow 0} \left(\mathbb{E}^{\tilde{\theta}} \left[\sup_{\theta \in \Theta} \left(\sup_{\|\theta' - \theta\| < \delta} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) - \inf_{\|\theta' - \theta\| < \delta} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right) \right] \right) \\ & = \lim_{\delta \rightarrow 0} \left(\mathbb{E}^{\tilde{\theta}} \left[\sup_{\theta \in \Theta} \left(\sup_{\|\theta' - \theta\| < \delta} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) - \inf_{\|\theta' - \theta\| < \delta} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right) \right] \right) = 0 \end{aligned}$$

which ensures that it exists $\delta_0 > 0$ such that :

$$\mathbb{E}^{\tilde{\theta}} \left[\sup_{\theta \in \Theta} \left(\sup_{\|\theta' - \theta\| < \delta_0} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) - \inf_{\|\theta' - \theta\| < \delta_0} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right) \right] \leq \frac{\epsilon}{4}.$$

Assumption **A9** ensures the existence of a neighbourhood \mathcal{V}_{θ} for any $\theta \in \Theta$ such that $\tilde{\pi}_{\theta}(\sup(\{\|\ddot{\phi}_{\theta'}\|_{\infty} / \theta' \in \mathcal{V}_{\theta}\})) < +\infty$.

For any $\theta \in \Theta$, it thus exists a positive δ_{θ} such that :

$$B_{\delta_{\theta}}(\theta) = \{\theta' / \|\theta' - \theta\| < \delta_{\theta}\} \subset \mathcal{V}_{\theta} \quad \text{and} \quad \tilde{\pi}_{\theta}(\sup(\{\|\ddot{\phi}_{\theta'}\|_{\infty} / \theta' \in B_{\delta_{\theta}}(\theta)\})) < +\infty.$$

Defining $\tilde{\delta}_\theta = \min(\delta_\theta, \delta_0)$ and noticing that $\bigcup_{\theta \in \Theta} B_{\tilde{\delta}_\theta}(\theta)$ is an open cover of the compact set Θ , we know it exists $N \in \mathbb{N}^*$ and $(\theta_1, \dots, \theta_N)$ such that :

$$\Theta \subset \bigcup_{i=1}^N B_{\tilde{\delta}_{\theta_i}}(\theta_i).$$

Thanks to several upper bounding, we also can get that for any $\delta > 0$:

$$\begin{aligned} & \sup_{\|\theta' - \theta\| < \delta} \left(\left| \frac{1}{n} \sum_{k=1}^n \ddot{\phi}_{\theta'}(Z_k, Z_{k+1}) - \mathbb{E}^* [\ddot{\phi}_{\theta'}(Z_0, Z_1)] \right| \right) \\ & \leq 2 \left| \frac{1}{n} \sum_{k=1}^n \sup_{\|\theta' - \theta\| < \delta} (\ddot{\phi}_{\theta'}(Z_k, Z_{k+1})) - \mathbb{E}^* \left[\sup_{\|\theta' - \theta\| \leq \delta} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right] \right| \\ & + 2 \left| \frac{1}{n} \sum_{k=1}^n \inf_{\|\theta' - \theta\| < \delta} (\ddot{\phi}_{\theta'}(Z_k, Z_{k+1})) - \mathbb{E}^* \left[\inf_{\|\theta' - \theta\| \leq \delta} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right] \right| \\ & + 2 \mathbb{E}^{\tilde{\theta}} \left[\sup_{\|\theta' - \theta\| < \delta} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) - \inf_{\|\theta' - \theta\| < \delta} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right]. \end{aligned}$$

Putting together all previous remarks, it first comes using a union bound :

$$\begin{aligned} & \mathbb{P}^* \left(\sup_{\|\theta' - \theta\| < \min(\{\tilde{\delta}_{\theta_i} / 1 \leq i \leq N\})} \left(\left| \frac{1}{n} \sum_{k=1}^n \ddot{\phi}_{\theta'}(Z_k, Z_{k+1}) - \mathbb{E}^* [\ddot{\phi}_{\theta'}(Z_0, Z_1)] \right| \right) \geq \epsilon \right) \\ & \leq \mathbb{P}^* \left(\max_{1 \leq i \leq N} \left(\sup_{\|\theta' - \theta_i\| < \tilde{\delta}_{\theta_i}} \left(\left| \frac{1}{n} \sum_{k=1}^n \ddot{\phi}_{\theta'}(Z_k, Z_{k+1}) - \mathbb{E}^* [\ddot{\phi}_{\theta'}(Z_0, Z_1)] \right| \right) \right) \geq \epsilon \right) \\ & \leq \sum_{1 \leq i \leq N} \mathbb{P}^* \left(\sup_{\|\theta' - \theta_i\| < \tilde{\delta}_{\theta_i}} \left(\left| \frac{1}{n} \sum_{k=1}^n \ddot{\phi}_{\theta'}(Z_k, Z_{k+1}) - \mathbb{E}^* [\ddot{\phi}_{\theta'}(Z_0, Z_1)] \right| \right) \geq \epsilon \right) \end{aligned}$$

Assumption **A8** ensures that $\tilde{\pi}_\theta(\sup(\{ \|\ddot{\phi}_{\theta'}\|_\infty / \theta' \in \mathcal{V}_\theta \})) < +\infty$ which implies that $\forall i \in \{1 \leq i \leq N\}$:

$$\tilde{\pi}_{\theta^*} \left(\sup_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right) = \mathbb{E}^* \left[\sup_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right] < +\infty$$

and

$$\tilde{\pi}_{\theta^*} \left(\inf_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right) = \mathbb{E}^* \left[\inf_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right] < +\infty.$$

Since $(Z_k, Z_{k+1})_{k \in \mathbb{N}}$ is a positive recurrent Markov chain admitting $\tilde{\pi}_{\theta^*}$ as unique invariant probability distribution and thanks to the two finiteness results implied by assumption **A8**, the ergodic theorem ensures that :

$$\forall \theta \in \Theta, \frac{1}{n} \sum_{k=0}^{n-1} \sup_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_k, Z_{k+1})) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}^* \text{-a.s.}} \mathbb{E}^* \left[\sup_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right]$$

and

$$\forall \theta \in \Theta, \frac{1}{n} \sum_{k=0}^{n-1} \inf_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_k, Z_{k+1})) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^* \text{-a.s.}} \mathbb{E}^* \left[\inf_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right].$$

It finally comes :

$$\begin{aligned} & \mathbb{P}^* \left(\sup_{\|\theta' - \theta\| < \min(\{\tilde{\delta}_{\theta_i}/1 \leq i \leq N\})} \left(\left| \frac{1}{n} \sum_{k=1}^n \ddot{\phi}_{\theta'}(Z_k, Z_{k+1}) - \mathbb{E}^* [\ddot{\phi}_{\theta'}(Z_0, Z_1)] \right| \right) \geq \epsilon \right) \\ & \leq \sum_{1 \leq i \leq N} \mathbb{P}^* \left(2 \left| \frac{1}{n} \sum_{k=1}^n \sup_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_k, Z_{k+1})) - \mathbb{E}^* \left[\sup_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right] \right| \right. \\ & \quad \left. + 2 \left| \frac{1}{n} \sum_{k=1}^n \inf_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_k, Z_{k+1})) - \mathbb{E}^* \left[\inf_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right] \right| \right. \\ & \quad \left. + 2 \mathbb{E}^{\tilde{\theta}} \left[\sup_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) - \inf_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right] \geq \epsilon \right) \\ & \leq \sum_{1 \leq i \leq N} \mathbb{P}^* \left(\left| \frac{1}{n} \sum_{k=1}^n \sup_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_k, Z_{k+1})) - \mathbb{E}^* \left[\sup_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right] \right| \right. \\ & \quad \left. + \left| \frac{1}{n} \sum_{k=1}^n \inf_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_k, Z_{k+1})) - \mathbb{E}^* \left[\inf_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right] \right| \geq \frac{\epsilon}{4} \right) \\ & \leq \sum_{1 \leq i \leq N} \mathbb{P}^* \left(\left| \frac{1}{n} \sum_{k=1}^n \sup_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_k, Z_{k+1})) - \mathbb{E}^* \left[\sup_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right] \right| \geq \frac{\epsilon}{8} \right) \\ & \quad + \sum_{1 \leq i \leq N} \mathbb{P}^* \left(\left| \frac{1}{n} \sum_{k=1}^n \inf_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_k, Z_{k+1})) - \mathbb{E}^* \left[\inf_{\|\theta' - \theta\| \leq \tilde{\delta}_{\theta_i}} (\ddot{\phi}_{\theta'}(Z_0, Z_1)) \right] \right| \geq \frac{\epsilon}{8} \right) \end{aligned}$$

Each of the N term tending to zero when $n \rightarrow +\infty$, we thus proved that :

$$\mathbb{P}^* \left(\sup_{\|\theta' - \theta^*\| < \min(\{\tilde{\delta}_{\theta_i}/1 \leq i \leq N\})} \left(\left| \frac{1}{n} \sum_{k=1}^n \ddot{\phi}_{\theta^*}(Z_k, Z_{k+1}) - \mathbb{E}^* [\ddot{\phi}_{\theta^*}(Z_0, Z_1)] \right| \right) \geq \epsilon \right) \xrightarrow[n \rightarrow +\infty]{} 0$$

$$\text{that is : } \sup_{\|\theta' - \theta^*\| < \min(\{\tilde{\delta}_{\theta_i}/1 \leq i \leq N\})} \left(\left| \frac{1}{n} \sum_{k=1}^n \ddot{\phi}_{\theta^*}(Z_k, Z_{k+1}) - \tilde{\pi}_{\theta^*}(\ddot{\phi}_{\theta^*}(Z_0, Z_1)) \right| \right) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^* \text{-a.s.}} 0$$

Thanks to the equality in distribution mentioned at the beginning and the fact that $\tilde{\theta}_n$ is weakly consistent, this yields :

$$\hat{\Sigma}_n = -\frac{1}{n} \sum_{k=0}^{n-1} \ddot{\phi}_{\tilde{\theta}_n}(L_{k+1}^n, L_k^n) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^* \text{-a.s.}} -\tilde{\pi}_{\theta^*}(\ddot{\phi}_{\theta^*}) = \Sigma_{\theta^*}.$$

■

Tanks to fundamental equality in distribution between $(L_k^n)_{0 \leq k \leq n}$ and $(Z_k)_{0 \leq k \leq n}$, it comes that :

$$\frac{\dot{i}_n(\theta^*)}{\sqrt{n}} \sim \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \dot{\phi}_{\theta^*}(Z_k, Z_{k+1}).$$

Expressing $\dot{\phi}_{\theta}(x, y)$ as $\frac{\dot{Q}_{\theta}(x, y)}{Q_{\theta}(x, y)}$ and using the second assertion of assumption **A3**, we get for all x in \mathbb{N} that :

$$\mathbb{E}^* [\dot{\phi}_{\theta^*}(Z_k, Z_{k+1}) | Z_k] = \sum_{y=0}^{+\infty} \frac{\dot{Q}_{\theta^*}(Z_k, y)}{Q_{\theta^*}(Z_k, y)} Q_{\theta^*}(Z_k, y) = \partial_{\theta} \left(\sum_{y=0}^{+\infty} Q_{\theta}(Z_k, y) \right)_{\theta=\theta^*} = 0.$$

We now remind of the central limit theorem given in the book [HH80] :

Théorème 26

If $(S_{n,i})_{\substack{n \in \mathbb{N}^* \\ 1 \leq i \leq k_n}}$ adapted to the filtration $(\mathcal{F}_{n,i})_{\substack{n \in \mathbb{N}^* \\ 1 \leq i \leq k_n}}$ is a square-integrable centered martingale array with differences $Y_{n,i} = S_{n,i} - S_{n,i-1}$ for $1 \leq i \leq k_n$ and if η^2 is an a.s. finite random variable such that :

- $\max_{1 \leq i \leq k_n} (|Y_{n,i}|) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}} 0$,
- $\sum_{i=1}^{k_n} Y_{n,i}^2 \xrightarrow[n \rightarrow +\infty]{\mathbf{P}} \eta^2$,
- The sequence $\left(\mathbb{E} \left[\max_{1 \leq i \leq k_n} (Y_{n,i}^2) \right] \right)_{n \in \mathbb{N}^*}$ is bounded in n ,
- for all n in \mathbb{N}^* , for $1 \leq i \leq k_n$, $\mathcal{F}_{n,i} \subset \mathcal{F}_{n+1,i}$,

then, $S_{n,k_n} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} Z$ where the random variable Z has characteristic function $\mathbb{E} \left[-\frac{1}{2} \eta^2 t^2 \right]$.

Théorème 27

Under assumptions **A1**, **A2**, **A3**, **A4**, **A5** and **A6**, the score vector sequence $\left(\frac{\dot{i}_n(\theta^*)}{\sqrt{n}} \right)_{n \in \mathbb{N}^*}$ is asymptotically normal :

$$\frac{\dot{i}_n(\theta^*)}{\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \Sigma_{\theta^*}).$$

Since we saw that the variables $\dot{\phi}_{\theta^*}(Z_k, Z_{k+1})$ were centered, defining for all $1 \leq k \leq n$:

$$Y_{n,k} = \frac{1}{\sqrt{n}} \dot{\phi}_{\theta^*}(Z_k, Z_{k+1}) \quad \text{and} \quad S_{n,k} = \sum_{j=1}^k Y_{n,j} \quad \text{and} \quad \mathcal{F}_{n,k} = \mathcal{F}_k = \sigma(\{Z_j / 1 \leq j \leq k\}).$$

and using then the first assertion of assumption **A6**, the sequence $(S_{n,k})_{\substack{n \in \mathbb{N}^* \\ 1 \leq k \leq n}}$ is a centered and square-integrable martingale array with differences $Y_{n,k}$. Since it's direct that σ -fields $(\mathcal{F}_{n,k})_{\substack{n \in \mathbb{N}^* \\ 1 \leq k \leq n}}$ are nested, to get the asymptotic normality of the score vector sequence stated in theorem 27, it's sufficient to prove that the three other conditions of theorem 26 are met by the sequence $(Y_{n,k})$.

Démo 31

First of all, since $\sum_{k=1}^n Y_{n,k}^2 = \sum_{k=1}^n \frac{1}{\sqrt{n}} \dot{\phi}_{\theta^*}(Z_k, Z_{k+1})^2 = \frac{1}{n} \sum_{k=1}^n \dot{\phi}_{\theta^*}(Z_k, Z_{k+1})^2$, the ergodic theorem applied to the ergodic Markov chain $(Z_k, Z_{k+1})_{k \in \mathbb{N}}$ gives that :

$$\sum_{k=1}^n Y_{n,k}^2 \xrightarrow[n \rightarrow +\infty]{\mathbf{P}\text{-a.s.}} \Sigma_{\theta^*}.$$

Then, applying Markov inequality to the random variable $\max_{1 \leq i \leq n} (|Y_{n,k}|)$, it comes for all $\epsilon > 0$:

$$\begin{aligned} \mathbb{P}^* \left(\max_{1 \leq k \leq n} (|Y_{n,k}|) \geq \epsilon \right) &= \mathbb{P}^* \left(\max_{1 \leq k \leq n} \left(\|\dot{\phi}_{\theta^*}(Z_k, Z_{k+1})\|^{2q} \right) \geq \epsilon \sqrt{n} \right) \\ &\leq \frac{1}{n^q \epsilon^{2q}} \mathbb{E}^* \left[\max_{1 \leq k \leq n} \left(\|\dot{\phi}_{\theta^*}(Z_k, Z_{k+1})\|^{2q} \right) \right] \\ &\leq \frac{1}{n^q \epsilon^{2q}} \sum_{k=1}^n \mathbb{E}^* \left[\|\dot{\phi}_{\theta^*}(Z_k, Z_{k+1})\|^{2q} \right] \\ &= \frac{1}{n^{q-1} \epsilon^{2q}} \frac{1}{n} \sum_{k=1}^n \mathbb{E}^* \left[\|\dot{\phi}_{\theta^*}(Z_k, Z_{k+1})\|^{2q} \right] \end{aligned}$$

But, using again the ergodicity of the Markov chain $(Z_k, Z_{k+1})_{k \in \mathbb{N}}$ and the fact that $\tilde{\pi}_{\theta} \left(\|\dot{\phi}_{\theta}\|_2^{2q} \right) < +\infty$ ensured by assumption **A7**, we know that $\frac{1}{n} \sum_{k=1}^n \mathbb{E}^* \left[\|\dot{\phi}_{\theta^*}(Z_k, Z_{k+1})\|^{2q} \right] \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^*\text{-a.s.}} \mathbb{E}^* \left[\|\dot{\phi}_{\theta^*}(\tilde{Z}_0, \tilde{Z}_1)\|^{2q} \right] < +\infty$ and since $q - 1 > 0$ implies that $\frac{1}{n^{q-1} \epsilon^{2q}} \xrightarrow[n \rightarrow +\infty]{} 0$, it comes at the end :

$$\frac{1}{n^{q-1} \epsilon^{2q}} \frac{1}{n} \sum_{k=1}^n \mathbb{E}^* \left[\|\dot{\phi}_{\theta^*}(Z_k, Z_{k+1})\|^{2q} \right] \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^*} 0$$

so that $\mathbb{P}^* \left(\max_{1 \leq k \leq n} (|Y_{n,k}|) \geq \epsilon \right) \xrightarrow[n \rightarrow +\infty]{} 0$ that is $\max_{1 \leq k \leq n} (|Y_{n,k}|) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^*} 0$.

Finally, the bound $\mathbb{E}^* \left[\max_{1 \leq k \leq n} \left(\|Y_{n,k} Y_{n,k}^t\|_{\infty} \right) \right] \leq \sum_{k=1}^n \mathbb{E}^* [\|Y_{n,k}\|^2]$ ensure that the sequence $\left(\mathbb{E} \left[\max_{1 \leq k \leq n} \left(\|Y_{n,k} Y_{n,k}^t\|_{\infty} \right) \right] \right)_{n \in \mathbb{N}^*}$ is bounded in n . Indeed, the ergodicity of the Markov chain $(Z_k, Z_{k+1})_{k \in \mathbb{N}}$ and again the fact that $\tilde{\pi}_{\theta} \left(\|\dot{\phi}_{\theta}\|_2^{2q} \right) < +\infty$ with $q > 1$ implies that :

$$\sum_{k=1}^n \mathbb{E}^* [\|Y_{n,k}\|^2] = \frac{1}{n} \sum_{k=1}^n \mathbb{E}^* \left[\|\dot{\phi}_{\theta^*}(Z_k, Z_{k+1})\|^2 \right] \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^*\text{-a.s., } \mathbb{L}^1} \mathbb{E}^* \left[\|\dot{\phi}_{\theta^*}(\tilde{Z}_0, \tilde{Z}_1)\|^2 \right] < +\infty$$

so that in particular the sequence $\left(\sum_{k=1}^n \mathbb{E}^* [\|Y_{n,k}\|^2] \right)_{n \in \mathbb{N}}$ is bounded. \blacksquare

Théorème 28

Under assumptions **A1**, **A2**, **A3**, **A4**, **A5**, **A6**, **A7** and **A8**, the M -estimator $\widehat{\theta}_n$ is asymptotically normal :

$$\sqrt{n} \left(\widehat{\theta}_n - \theta^* \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \Sigma_{\theta^*}^{-1}).$$

6.5.2 The case of parametric sub-ballistic RWRE

One major problem when dealing with consistency of parametric sub-ballistic RWRE is that the foreseen approach doesn't work directly and the whole estimation procedure has to be modified. We still assume that assumption **A1**, **A3** and **A4** are satisfied but assumption **A2** and **A5** are respectively replaced by following assumptions :

$$\boxed{\mathbf{A2bis} : \forall \nu \in \mathcal{P}, \mathbb{E}^\nu [\rho_0] \geq 1.}$$

and

$$\boxed{\mathbf{A5bis} : \forall \theta \in \Theta, \widetilde{\pi}_\theta(\sup\{|\phi_{\theta'} - \phi_{\theta_0}| / \theta' \in \Theta\}) < +\infty.}$$

Similarly to the reminder on RWRE, we can define κ_θ as the unique solution to $\mathbb{E}^\theta [\rho_0^\kappa]$ if it exists and as $+\infty$ otherwise.

Lemme 15

If assumption **A1** and **A2bis** are met, it holds for any $\alpha > 0$ the dichotomy :

$$\alpha < \kappa_0 \Rightarrow \sum_{k=0}^{+\infty} k^\alpha \pi(k) < +\infty \quad \text{or} \quad \alpha \geq \kappa_0 \Rightarrow \sum_{k=0}^{+\infty} k^\alpha \pi(k) = +\infty.$$

Démo 32

Let first consider the real-valued random variable Λ defined by : $1 - S = e^{-\Lambda}$. It then comes by Fubini's theorem and definition of π_θ :

$$\sum_{k=0}^{+\infty} k^\alpha \pi_\theta(k) = \sum_{k=0}^{+\infty} k^\alpha \mathbb{E}^\theta [S(1-S)^k] = \mathbb{E}^\theta \left[\sum_{k=0}^{+\infty} k^\alpha S(1-S)^k \right] = \mathbb{E}^\theta \left[S \sum_{k=0}^{+\infty} k^\alpha e^{-k\Lambda} \right].$$

Since the functions $x \in \mathbb{R}^+ \mapsto x^\alpha$ and $x \in \mathbb{R}^+ \mapsto e^{-\lambda x}$ are respectively increasing and decreasing for any $\lambda > 0$ and any $\alpha > 0$, for all k in \mathbb{N} , it comes first :

$$k^\alpha e^{-\lambda(k+1)} \leq x^\alpha e^{-\lambda x} \leq (k+1)^\alpha e^{-\lambda(k+1)} e^{-\lambda},$$

an then integrating on the interval $[k, k+1]$, it comes then :

$$k^\alpha e^{-\lambda(k+1)} \leq \int_k^{k+1} x^\alpha e^{-\lambda x} dx \leq (k+1)^\alpha e^{-\lambda(k+1)} e^{-\lambda},$$

so that summing those two inequalities for k in \mathbb{N} , we get :

$$\begin{aligned} \sum_{k=0}^{+\infty} k^\alpha e^{-\lambda(k+1)} &\leq \sum_{k=0}^{+\infty} \int_k^{k+1} x^\alpha e^{-\lambda x} dx \leq \sum_{k=0}^{+\infty} (k+1)^\alpha e^{-\lambda(k+1)} e^{-\lambda} \\ \Leftrightarrow e^\lambda \sum_{k=0}^{+\infty} k^\alpha e^{-\lambda k} &\leq \int_0^{+\infty} x^\alpha e^{-\lambda x} dx \leq e^{-\lambda} \sum_{k=1}^{+\infty} k^\alpha e^{-\lambda k} = \sum_{k=0}^{+\infty} k^\alpha e^{-\lambda k}. \end{aligned}$$

Since for any $(\lambda, \alpha) \in \mathbb{R}^{*+2}$, it holds $\int_0^{+\infty} x^\alpha e^{-\lambda x} dx = \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}}$, the former inequalities applied to $\lambda = \Lambda$ yields : $(1-S) \frac{\Gamma(\alpha+1)}{\Lambda^{\alpha+1}} \leq \sum_{k=0}^{+\infty} k^\alpha e^{-\Lambda k} \leq \frac{1}{1-S} \frac{\Gamma(\alpha+1)}{\Lambda^{\alpha+1}}$.

Then we notice that :

$$S \sum_{k=0}^{+\infty} k^\alpha e^{-\Lambda k} \mathbb{1}_{\{S > \frac{1}{2}\}} \leq \sum_{k=0}^{+\infty} k^\alpha \frac{1}{2^k} = C_\alpha < +\infty$$

and using the convexity of the function $x \in \mathbb{R} \mapsto e^x$, we notice also that : $\Lambda > S$ so that with the upper bound on $\sum_{k=0}^{+\infty} k^\alpha e^{-\Lambda k}$, we get :

$$\begin{aligned} \mathbb{E}^\theta \left[S \sum_{k=0}^{+\infty} k^\alpha e^{-k\Lambda} \right] &= \mathbb{E}^\theta \left[S \sum_{k=1}^{+\infty} k^\alpha e^{-k\Lambda} \right] \leq C_\alpha + \mathbb{E}^\theta \left[S \frac{1}{1-S} \mathbb{1}_{\{S > \frac{1}{2}\}} \frac{\Gamma(\alpha+1)}{\Lambda^{\alpha+1}} \right] \\ &\leq C_\alpha + \mathbb{E}^\theta \left[2 \mathbb{1}_{\{S > \frac{1}{2}\}} \frac{\Gamma(\alpha+1)}{S^\alpha} \right] = C_\alpha + 2\Gamma(\alpha+1) \mathbb{E}^\theta [S^{-\alpha}]. \end{aligned}$$

Using then that the graph of the convex function $x \in]0, 1[\mapsto |\log(1-x)|$ is under its cord between 0 and $\frac{1}{2}$, we also notice that : $\frac{\mathbb{1}_{\{S < \frac{1}{2}\}}}{2 \log(2)} \leq \frac{\mathbb{1}_{\{S < \frac{1}{2}\}}}{\Lambda}$ so that with the lower bound on $\sum_{k=0}^{+\infty} k^\alpha e^{-\Lambda k}$, we get :

$$\begin{aligned} \mathbb{E}^\theta \left[S \sum_{k=0}^{+\infty} k^\alpha e^{-k\Lambda} \right] &= \mathbb{E}^\theta \left[S(1-S) \frac{\Gamma(\alpha+1)}{\Lambda^{\alpha+1}} \right] \geq \mathbb{E}^\theta \left[(1-S) \mathbb{1}_{\{S < \frac{1}{2}\}} S \Gamma(\alpha+1) \left(\frac{\mathbb{1}_{\{S < \frac{1}{2}\}}}{\Lambda} \right)^{\alpha+1} \right] \\ &\geq \frac{\Gamma(\alpha+1)}{2} \mathbb{E}^\theta \left[S \left(\frac{\mathbb{1}_{\{S < \frac{1}{2}\}}}{2 \log(2) S} \right)^{\alpha+1} \right] = \frac{\Gamma(\alpha+1)}{2(2 \log(2))^{\alpha+1}} \mathbb{E}^\theta [S^{-\alpha} \mathbb{1}_{\{S < \frac{1}{2}\}}] \end{aligned}$$

but $\mathbb{E}^\theta [S^{-\alpha}] = \mathbb{E}^\theta [S^{-\alpha} \mathbb{1}_{\{S < \frac{1}{2}\}}] + \mathbb{E}^\theta [S^{-\alpha} \mathbb{1}_{\{S \geq \frac{1}{2}\}}] \leq \mathbb{E}^\theta [S^{-\alpha} \mathbb{1}_{\{S < \frac{1}{2}\}}] + 2^\alpha$ and thus $\mathbb{E}^\theta [S^{-\alpha} \mathbb{1}_{\{S < \frac{1}{2}\}}] \geq \mathbb{E}^\theta [S^{-\alpha}] - 2^\alpha$ so that it finally comes :

$$\mathbb{E}^\theta \left[S \sum_{k=0}^{+\infty} k^\alpha e^{-k\Lambda} \right] \geq \frac{\Gamma(\alpha+1)}{2(2 \log(2))^{\alpha+1}} (\mathbb{E}^\theta [S^{-\alpha}] - 2^\alpha)$$

Thanks to Fubini's theorem, we get that :

$$\mathbb{E}^\theta [S^{-\alpha}] = \int_0^{+\infty} \mathbb{P}^\theta (S^{-\alpha} > t) dt = \int_0^{+\infty} \mathbb{P}^\theta \left(S < \frac{1}{t^{\frac{1}{\alpha}}} \right) dt.$$

Remembering that there exist $c_\theta \in \mathbb{R}^+$ such that $\mathbb{P}^\theta \left(S < \frac{1}{t} \right) \underset{t \rightarrow +\infty}{\sim} \frac{c_\theta}{t^{k_\theta}}$, we get $\mathbb{P}^\theta \left(S < \frac{1}{t^{\frac{1}{\alpha}}} \right) \underset{t \rightarrow +\infty}{\sim} \frac{c_\theta}{t^{\frac{k_\theta}{\alpha}}}$, so that $\mathbb{E}^\theta [S^{-\alpha}] < +\infty \iff 1 < \frac{k_\theta}{\alpha} \iff \alpha < k_\theta$

and it follows from the upper and lower bounds on $\mathbb{E}^\theta \left[S \sum_{k=0}^{+\infty} k^\alpha e^{-k\Lambda} \right]$ that :

$$\alpha < \kappa_0 \Rightarrow \sum_{k=0}^{+\infty} k^\alpha \pi(k) < +\infty \quad \text{or} \quad \alpha \geq \kappa_0 \Rightarrow \sum_{k=0}^{+\infty} k^\alpha \pi(k) = +\infty. \quad \blacksquare$$

As have already been mentioned, the foreseen approach on ballistic random walk can't be applied directly in the sub-ballistic case but it is possible to adapt it by a slight modification of the criterion function l_n^+ . Letting $\theta_0 \in \Theta$ be constant, we define the criterion function $l_n^{+,sb}$ by :

$$l_n^{+,sb}(\theta) = \sum_{k=0}^{n-1} (\phi_\theta(L_{k+1}^n, L_k^n) - \phi_{\theta_0}(L_{k+1}^n, L_k^n)).$$

Similarly we entitle *maximum likelihood estimator* of θ any M-estimator $\widehat{\theta}_n^{sb}$ defined by :

$$\widehat{\theta}_n^{sb} \in \underset{\theta \in \Theta}{\text{Argmax}}(l_n^{+,sb}(\theta)).$$

Consistency of parametric sub-ballistic RWRE

Lemme 16

If assumptions **A1**, **A2bis**, **A3**, **A4**, **A5bis**, **A6**, **A7**, **A8** and **A9** are met, then, defining $l^{+,sb}(\theta) = \widetilde{\pi}_{\theta^*}(\phi_\theta - \phi_{\theta_0})$ we have the following uniform convergence :

$$\sup_{\theta \in \Theta} \left(\left| \frac{1}{n} (l_n^{+,sb}(\theta) - l^{+,sb}(\theta)) \right| \right) \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^* \text{-a.s.}} 0.$$

Démo 33

The demonstration is nearly the same than the one of the consistency of the Fisher information matrix in the ballistic case. \blacksquare

Lemme 17

Under assumption **A1**, **A3**, **A4** and **A5bis**, if $\theta^* \in \Theta$ is such that $\mathbb{E}^{\theta^*}[\rho_0] \geq 1$, then, θ^* is the unique maximizer of the criterion function $l^{+,sb}$ and it even holds for every $\epsilon > 0$:

$$\sup(\{l^{+,sb}(\theta) / \|\theta - \theta^*\| \geq \epsilon\}) < l^{+,sb}(\theta^*).$$

Démo 34

For any $\theta \in \Theta$, since assumption **A5** ensures that $l^{+,sb}(\theta) \in]-\infty, 0]$, writing $l^{+,sb}(\theta) - l^{+,sb}(\theta^*) = \widetilde{\pi}_{\theta^*}(\phi_\theta - \phi_{\theta^*})$ makes sense and using the fact that $(\widetilde{Z}_k, \widetilde{Z}_{k+1})_{k \in \mathbb{N}}$ is a Markov chain with transition kernel $\widetilde{Q}^* = \widetilde{Q}^{\theta^*}$ given by $\widetilde{Q}^\theta((m, n), (m', n')) = Q^\theta(m', n') \mathbb{1}_{\{n=m'\}}$ and initial distribution $\widetilde{\pi}^* = \widetilde{\pi}_{\theta^*}$, we get :

$$\begin{aligned} l^{+,sb}(\theta) - l^{+,sb}(\theta^*) &= \mathbf{E}^* [\phi_\theta(\widetilde{Z}_0, \widetilde{Z}_1)] - \mathbf{E}^* [\phi_{\theta^*}(\widetilde{Z}_0, \widetilde{Z}_1)] = \mathbf{E}^* [\phi_\theta(\widetilde{Z}_0, \widetilde{Z}_1) - \phi_{\theta^*}(\widetilde{Z}_0, \widetilde{Z}_1)] \\ &= \sum_{m=0}^{+\infty} \sum_{n=0}^{+\infty} [\phi_\theta(m, n) - \phi_{\theta^*}(m, n)] Q^*(m, n) \pi^*(m) \\ &= \sum_{m=0}^{+\infty} \sum_{n=0}^{+\infty} \left[\log \left(\frac{Q^\theta(m, n)}{\binom{m+n}{m}} \right) - \log \left(\frac{Q^*(m, n)}{\binom{m+n}{m}} \right) \right] Q^*(m, n) \pi^*(m) \\ &= \sum_{m=0}^{+\infty} \pi^*(m) \sum_{n=0}^{+\infty} \log \left(\frac{Q^\theta(m, n)}{Q^*(m, n)} \right) Q^*(m, n). \end{aligned}$$

For every $m \in \mathbb{N}$, Jensen's inequality applied with the probability measure $Q^*(m, \cdot)$ on \mathbb{N} yields :

$$\begin{aligned} \sum_{n=0}^{+\infty} \log \left(\frac{Q^\nu(m, n)}{Q^*(m, n)} \right) Q^*(m, n) &\leq \log \left(\sum_{n=0}^{+\infty} \frac{Q^\nu(m, n)}{Q^*(m, n)} Q^*(m, n) \right) \\ &= \log \left(\sum_{n=0}^{+\infty} Q^\nu(m, n) \right) = \log(Q^\nu(m, \mathbb{N})) = \log(1) = 0 \end{aligned}$$

with equality, if and only if, $n \in \mathbb{N} \mapsto \frac{Q^\nu(m, n)}{Q^*(m, n)}$ is $Q^*(m, \cdot)$ -a.s. constant and equals 1, if and only if, $n \in \mathbb{N} \mapsto \frac{Q^\nu(m, n)}{Q^*(m, n)}$ is constant and equals 1 (since for all

n in \mathbb{N} , $Q^*(m, n) > 0$), if and only if, $\forall n \in \mathbb{N}$, $Q^\nu(m, n) = Q^*(m, n)$, if and only if, $Q^\nu(m, \cdot) = Q^*(m, \cdot)$.

It thus comes $l^+(\nu) - l^+(\nu^*) \leq 0$ with equality, if and only if, for all $m \in \mathbb{N}$, $Q^\nu(m, \cdot) = Q^*(m, \cdot)$. Hence ν^* is a maximizer of the criterion function l^+ and moreover $l^+(\nu) = l^+(\nu^*)$, if and only if, for all $m \in \mathbb{N}$, $Q^\nu(m, \cdot) = Q^*(m, \cdot)$.

In particular, if $l^+(\nu) = l^+(\nu^*)$, then, for all $m \in \mathbb{N}$, $Q^\nu(m, 0) = Q^*(m, 0)$, that is, for all $m \in \mathbb{N}$, $\int_{]0,1[} t^{m+1} \nu(dt) = \phi_\nu(m, 0) = \phi_{\nu^*}(m, 0) = \int_{]0,1[} t^{m+1} \nu^*(dt)$, that is, the probability distributions ν and ν^* have the same moments and since their supports are bounded, it implies that they are the same : $\nu = \nu^*$. Hence, ν^* is the unique maximizer of the criterion function l^+ .

At the end, the fact that $\forall \epsilon > 0$, $\sup(\{l^{+,sb}(\theta) / \|\theta - \theta^*\| \geq \epsilon\}) < l^{+,sb}(\theta^*)$ comes from the continuity of the application $\theta \mapsto l^{+,sb}(\theta)$ which itself results from Lebesgue's theorem of continuity under integral sign applied to the continuity assumption **A4** and the dominating condition **A5bis**. ■

Théorème 29

If assumption **A1**, **A2bis**, **A3**, **A4**, **A5bis** are met, then for all ν in \mathcal{P} , it comes :

$$\widehat{\theta}_n^{sb} \xrightarrow[n \rightarrow +\infty]{\mathbf{P}^* \text{-a.s.}} \theta^*.$$

Démo 35

The consistency of $\widehat{\theta}_n^{sb}$ is then entailed by Wald's consistency theorem 25 applied to functions $m_\theta = \phi_\theta - \phi_{\theta_0}$ since θ^* is the unique maximizer of $l^{+,sb}$. ■

6.5.3 Examples of parametric settings

In order to illustrate what we have been dealing with so far from a more applied point of view, we are now going to consider two particular settings for the law of the environment :

- an environment with two possible known states
- an environment with Beta distribution

Environment with finite and known support

We assume here that $\nu = \nu_\theta = \nu_p = p\delta_{e_1} + (1-p)\delta_{e_2}$ where $e_1 \leq e_2$ in $]0, 1[$ are fixed and p in $]0, 1[$ is unknown. In this setting, we directly compute the useful quantities :

$$\begin{aligned} \mathbb{E}^\theta[\omega_0] &= pe_1 + (1-p)e_2, \\ \mathbb{E}^\theta[\rho_0] &= p \frac{1-e_1}{e_1} + (1-p) \frac{1-e_2}{e_2}, \\ \mathbb{E}^\theta[\log(\rho_0)] &= p \log\left(\frac{1-e_1}{e_1}\right) + (1-p) \log\left(\frac{1-e_2}{e_2}\right), \\ Q^p(x, y) &= e^{-\phi(x, y)} = \log(pe_1^{x+1}(1-e_1)^y + (1-p)e_2^{x+1}(1-e_2)^y), \end{aligned}$$

$$l_n^+(p) = \sum_{k=0}^{n-1} \log \left(p e_1^{L_{k+1}^n + 1} (1 - e_1)^{L_k^n} + (1 - p) e_2^{L_{k+1}^n + 1} (1 - e_2)^{L_k^n} \right),$$

Hence, it comes the conditions :

$$\mathbb{E}^\theta [\log(\rho_0)] < 0 \iff p < \frac{\log \left(\frac{e_2}{1 - e_2} \right)}{\log \left(\frac{e_2}{e_1} \frac{1 - e_1}{1 - e_2} \right)} \quad \text{and} \quad \mathbb{E}^\theta [\log(\rho_0)] = 0 \iff p = \frac{\log \left(\frac{e_2}{1 - e_2} \right)}{\log \left(\frac{e_2}{e_1} \frac{1 - e_1}{1 - e_2} \right)}$$

as well as

$$\mathbb{E}^\theta [\rho_0] < 1 \iff p < \frac{a_1(2a_2 - 1)}{a_2 - a_1} \quad \text{and} \quad \mathbb{E}^\theta [\rho_0] \geq 1 \iff p \geq \frac{a_1(2a_2 - 1)}{a_2 - a_1}$$

From the form of $l_n^+(p)$, it's straightforward that the function \log being concave the function $p \in]0, 1[\mapsto l_n(p)$ is also concave so that the M-estimator $\widehat{p}_n \in \underset{p \in]0, 1[}{\text{Argmax}}(l_n(p))$ is not only well-defined in this case but also uniquely determined.

Environment with beta distribution

We assume here that $\nu = \nu_{(a,b)} = \nu_\theta$ is the Beta-distribution with parameters a and b such that $a > b + 1$:

$$\nu(ds) = \frac{1}{\beta(a,b)} s^{a-1} (1-s)^{b-1} ds = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} s^{a-1} (1-s)^{b-1} ds$$

where we have set :

$$\beta(a,b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt \quad \text{and} \quad \Gamma(x) = x \int_0^1 t^{x-1} e^{-x} dt.$$

In this setting, remembering that $\Gamma(x+1) = x\Gamma(x) = x \int_0^1 t^{x-1} e^{-x} dt$ and $\beta(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$, we can also compute directly the useful quantities :

$$\mathbb{E}^\theta [\omega_0] = \frac{\beta(a+1,b)}{\beta(a,b)},$$

$$\mathbb{E}^\theta [\rho_0] = \frac{\beta(a-1,b+1)}{\beta(a,b)} = \frac{\frac{\Gamma(a-1)\Gamma(b+1)}{\Gamma(a+b)}}{\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}} = \frac{\Gamma(a-1)}{\Gamma(a)} \frac{\Gamma(b+1)}{\Gamma(b)} = \frac{b}{a-1},$$

$$\mathbb{E}^\theta [\log(\rho_0)] = \int_0^1 \log \left(\frac{t}{1-t} \right) t^{a-1} (1-t)^{b-1} dt,$$

$$Q^p(x,y) = e^{-\phi(x,y)} = \frac{\beta(x+a+1,y+b)}{\beta(a,b)},$$

$$l_n^+(\theta) = \sum_{k=0}^{n-1} \log \left(\frac{\beta(L_{k+1}^n + a + 1, L_k^n + b)}{\beta(a,b)} \right) = -n \log(\beta(a,b)) + \sum_{k=0}^{n-1} \log(\beta(L_{k+1}^n + a + 1, L_k^n + b)).$$

Hence, it comes directly $\mathbb{E}^\theta [\rho_0] < 1 \iff b + 1 < a$ but as the exact computation of the integral giving $\mathbb{E}^\theta [\log(\rho_0)]$ is intractable, it is not possible to give a necessary condition ensuring its positivity. Nevertheless, observing that $\mathbb{E}^\theta [\log(\rho_0)] \leq \log(\mathbb{E}^\theta [\rho_0])$ thanks to concavity of the function \log and Jensen's inequality, we derive $b + 1 < a \Rightarrow \mathbb{E}^\theta [\log(\rho_0)] < 0$.

6.5.4 Numerical simulations and effective estimation

The two examples of transient environment introduced in previous section have lead to numerical simulations which enables to verify some of the reminded probabilistic results on RWRE and above all we will try to illustrate the statistical properties of the estimator $\hat{\theta}_n$ and also to pick out different computational problems when dealing practically with the estimation problem.

6.5.4.1 Numerical simulations

We first simulated a 1000-sample of trajectories of RWRE in transient ballistic regime until they reach the integer 10000 for each of the two former parameter settings. In the case of environment with finite and known support (we name it "Bernoulli" in the simulations) we used $a_1 = 0.4$ and $a_2 = 0.7$ and $p = 0.3$ as value for the known and unknown parameters and in the case of environment with Beta distribution we took $\alpha = 5$ and $\beta = 1$ to simulate the 1000 trajectory.

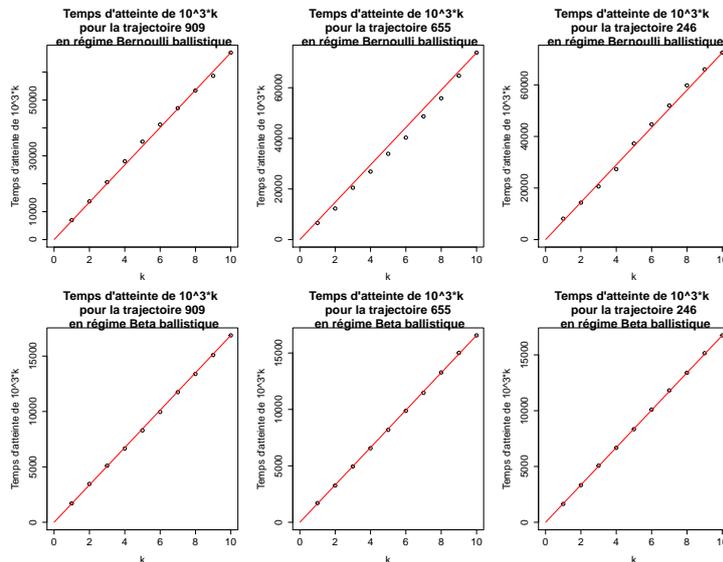


FIGURE 6.1 – Waiting time T_n as a function of $n \in \{k 10^3/1 \leq k \leq 10\}$.

Once those trajectory were simulated, it was then straightforward to know the value of T_n that is when each of them reached for the first time the integer $n \in \{k 10^3/1 \leq k \leq 10\}$. This first allows to effectively notice the "asymptotic linearity" of T_n in the ballistic regime : this can be seen in figure 6.1. It was secondly the occasion to stress the difference of simulation time experienced between the two different environment : the simulations of the trajectory for environment Beta were shorter than those of the trajectory for environment Bernoulli (see figure 6.2).

Furthermore having a look at the density (see figure 6.3) of the two different environment can give an idea of why the RWRE with environment Beta get faster than this with environment Bernoulli : *the probability for the environment to be big* is higher with Beta environment than with Bernoulli environment.

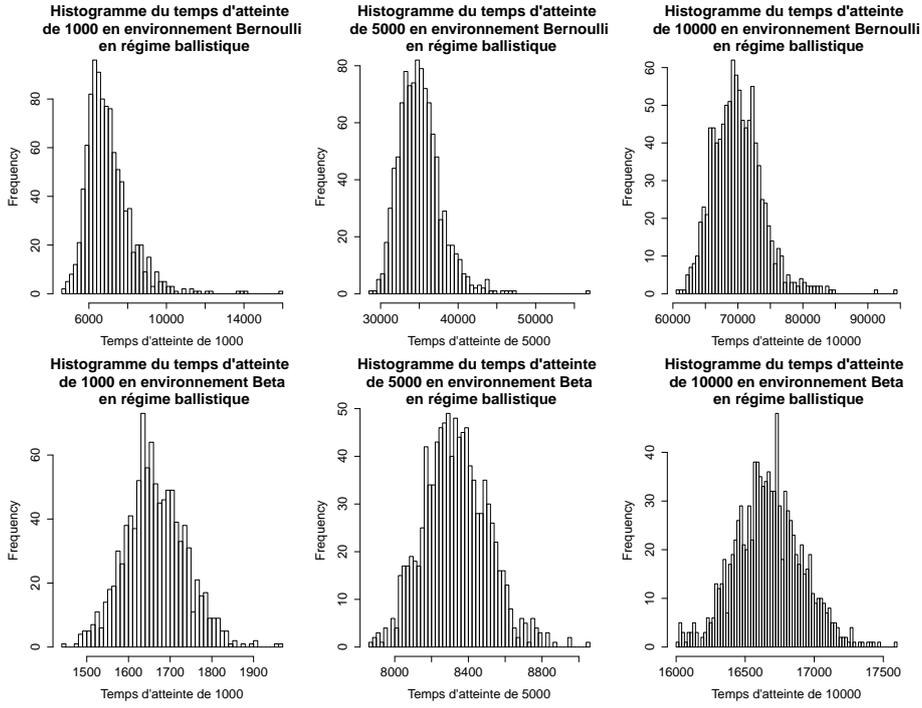


FIGURE 6.2 – Histogram of T_n for $n \in \{k 10^3 / k \in \{1, 5, 10\}\}$.

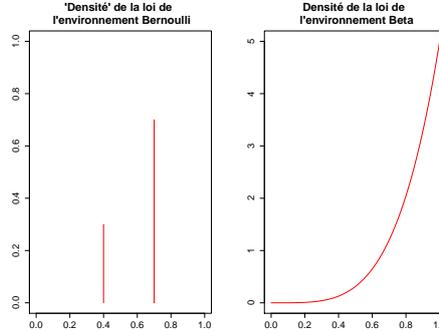


FIGURE 6.3 – Density of the law of the two different environments.

We then only simulated a 100-sample of trajectories of RWRE in transient sub-ballistic regime until they reach the integer 1000 and only for the Bernoulli setting using $a_1 = 0.4$ and $a_2 = 0.7$ and $p = 0.458$ as value for the parameters and in the case of environment.

This simulation in sub-ballistic regime having been much longer, it was an obvious task to compare it with the simulation of the RWRE made for the Bernoulli environment in ballistic regime by looking if the time T_n still depends linearly asymptotically on n and if the distribution of T_n presented much changes.

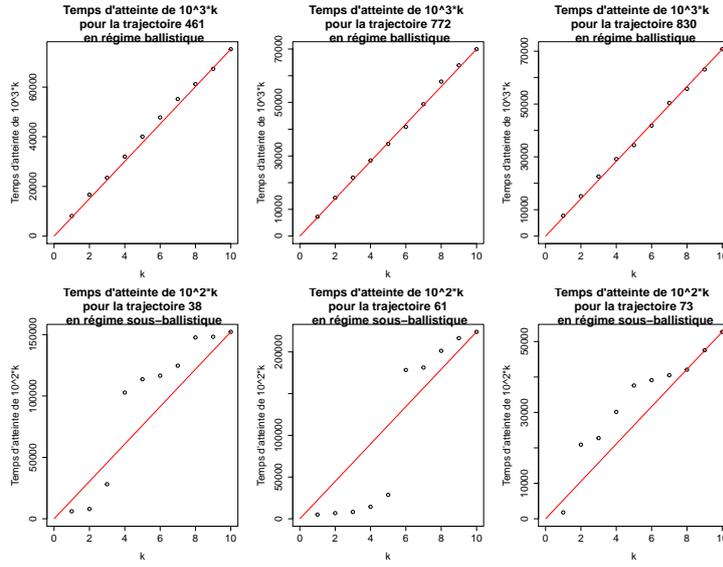


FIGURE 6.4 – Waiting time T_n as a function of $n \in \{k 10^3/1 \leq k \leq 10\}$ in ballistic regime and of $n \in \{k 10^2/1 \leq k \leq 10\}$ in sub-ballistic regime for randomly chosen trajectory.

Looking at the three lower windows of figure 6.4, it is clear that there is no more linear asymptotic dependency of T_n and it also enables to understand why some trajectory were much longer to simulate than others in this ballistic regime.

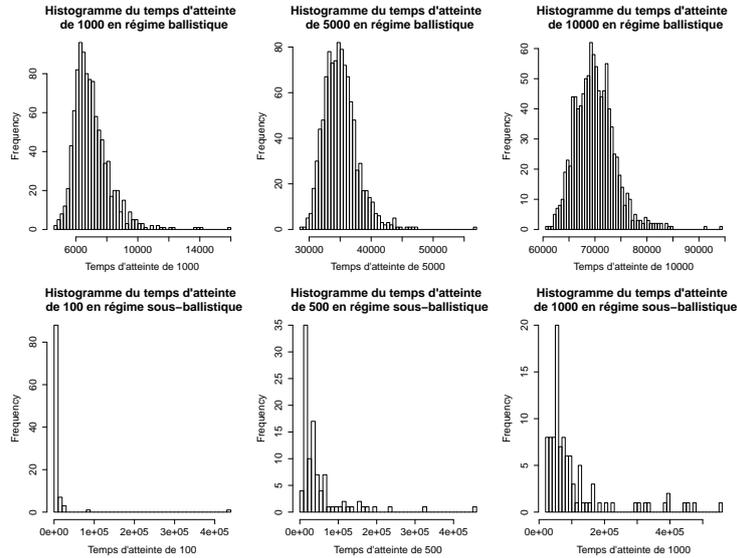


FIGURE 6.5 – Histogram of T_n for $n \in \{k 10^3/k \in \{1, 5, 10\}\}$ in ballistic regime and for $n \in \{k 10^2/k \in \{1, 5, 10\}\}$ in sub-ballistic regime.

Figure 6.5 shows that the distribution of T_n in ballistic regime is very different : T_n takes much bigger values (even for small values of n as on figure 6.5) and its dispersion is also bigger (whatever the value of n is).

6.5.4.2 Effective estimation

We restricted to the effective computation of the estimator $\hat{\theta}_n$ (written p_n in Bernoulli environment and (a_n, b_n) in Beta environment) only in the case of the ballistic regime from the 1000-samples simulated beforehand.

So we computed the value of p_n and (a_n, b_n) for $n \in \{k 10^3/1 \leq k \leq 10\}$ for each of the 1000 sample. By the way, an interesting thing to note is that the computation of the estimator in the case of the Beta environment was longer than this in the case of the Bernoulli environment contrary to the trend noticed in trajectory simulation : Is that due to the optimization procedure, to that it was now a multivariate estimator or to something else ?

On figures 6.6, 6.7 and 6.8, we illustrate that the estimators p_n , α_n and β_n are consistent as foreseen by the theory seeing that they become less biased and with a smaller dispersion when n increases.

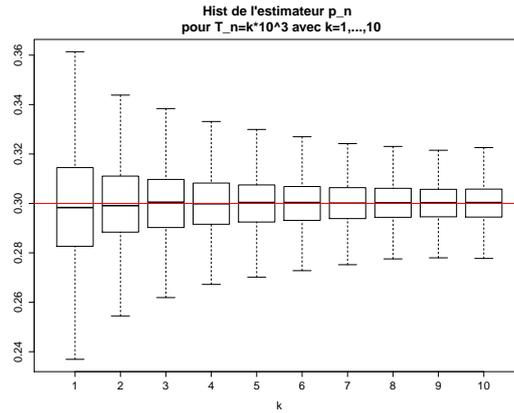


FIGURE 6.6 – Histogram of p_n for $n \in \{k 10^3/1 \leq k \leq 10\}$ in ballistic regime.

Finally, we illustrate on figure 6.9 the asymptotic normality of the estimator p_n foreseen by the theorem 28.

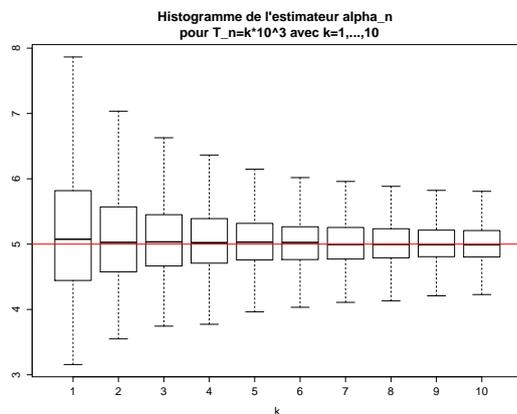


FIGURE 6.7 – Histogram of α_n for $n \in \{k 10^3/1 \leq k \leq 10\}$ in ballistic regime.

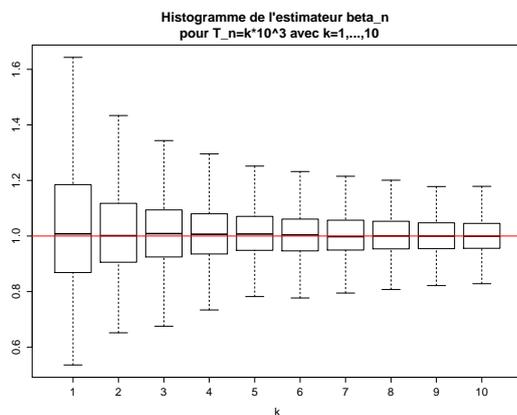


FIGURE 6.8 – Histogram of β_n for $n \in \{k 10^3/1 \leq k \leq 10\}$ in ballistic regime.

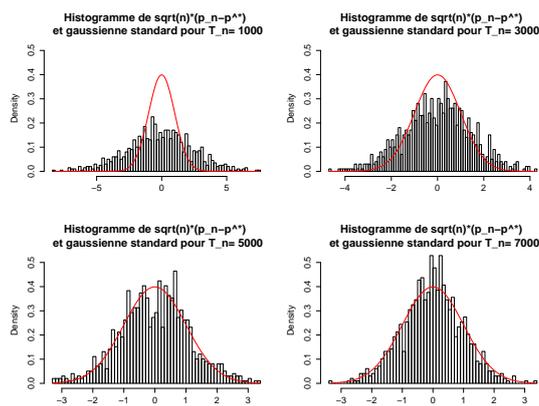


FIGURE 6.9 – Histogram of β_n for $n \in \{k 10^3/1 \leq k \leq 10\}$ in ballistic regime.

6.6 Conclusion

After having explained what is a random walk in random environment on \mathbb{Z} , the present work first reminded some of the most important probability results so far proven but it mostly retook the statistical results obtained in the papers [FC14], [FLM13], [FLG14] and [CFL14] and introduce the problem of the estimation of the environment of a RWRE in a little more general and unifying framework. We finally cope with the problem of numerical simulations and then effective parameter estimation.

The non-parametric framework we introduce allows to deal the parametric case as a special one. Nevertheless, our extension is not very far from the parametric setting and in order to get usual asymptotic results it's often necessary to restrict oneself to much stronger assumptions than the parametric ones. One major thing, we didn't reach is to find a unique approach to ballistic and sub-ballistic case. An other case, we didn't worked on is the problem of estimation in recurrent regime which had already been considered in [CFL14] but only in the situation of a parametric environment with finite support.

The numerical simulation work we achieved allowed us to stress two tricky points of the computational aspects of RWRE. On one hand, when considering simulation of the trajectory of RWRE in parametric settings, we observed that the simulation step can be very time consuming depending on the RWRE is ballistic or not. On the other hand, once the trajectory has been simulated, the optimization proceeded in order to compute the "maximum-likelihood" estimator is also time consuming depending on the environment.

This work was so an opportunity to learn what's the state of the art in the problem of statistical estimation of the environment of a RWRE but also to discover what are the most apparent obstacles as well in its theoretical study as in its computational approach.

Bibliography

- [CFL14] 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. *ArXiv e-prints*, April 2014.
- [FC14] O.Loukianov D.Loukianova F.Comets, M.Falconnet and C.Matias. Maximum likelihood estimator consistency for ballistic random walk in a parametric random environment. *Stochastic Processes and their Applications*, 124(1) :166–188, 2014.
- [FLG14] M. Falconnet, D. Loukianova, and A. Gloter. Maximum likelihood estimation in the context of a sub-ballistic random walk in a parametric random environment. *ArXiv e-prints*, May 2014.
- [FLM13] 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. *ArXiv e-prints*, February 2013.
- [HH80] Philip Hall and Christopher Charles Heyde. *Martingale limit theory and its application*. Probability and mathematical statistics : a series of monographs and textbooks. Academic Press, Boston, San Diego, New York, 1980.
- [Kes75] Kozlov M. V. Spitzer F. Kesten, H. A limit law for random walk in a random environment. *Compositio Mathematica*, 30(2) :145–168, 1975.
- [Sol75] Fred Solomon. Random walks in a random environment. *Ann. Probab.*, 3(1) :1–31, 02 1975.
- [Tav04] Tavaré, Simon and Zeitouni, Ofer and Picard, Jean. *Lectures on probability theory and statistics : École d’été de probabilités de Saint-Flour XXXI - 2001*, École d’été de probabilités de Saint-Flour, 31, 2001, Saint-Flour, Cantal, 2004. Springer. Conférences tenues à l’école d’été de théorie des probabilités de Saint-Flour, du 8 au 25 juillet 2001.