

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université A.MIRA-BEJAIA



Faculté des Sciences Exactes
Département de Mathématiques
Laboratoire de Mathématiques Appliquées (LMA)

THÈSE

EN VUE DE L'OBTENTION DU DIPLOME DE
DOCTORAT

Domaine : Mathématiques et Informatique Filière : Mathématiques
Spécialité : Probabilités et Statistiques

Présentée par
HARFOUCHE Zineb

Thème

**Méthodes d'estimation et d'approximation pour l'analyse
des modèles de risque**

Soutenue le : 13/07/2022

Devant le Jury composé de :

Nom et Prénom

Grade

Mme. BELAIDE Karima

Professeur

Univ. de Bejaia

Président

Mme. BARECHE Aicha

Professeur

Univ. de Bejaia

Rapporteur

Mr. HAMAZ Abdelghani

MCA

Univ. de Tizi Ouzou

Examineur

Mr. ZOUGAB Nabil

Professeur

Univ. de Bejaia

Examineur

Année Universitaire : 2021/2022

** Remerciements **

Je tiens particulièrement et très sincèrement à remercier Prof. **BARECHE Aicha** d'avoir accepté de diriger cette thèse. Sa rigueur, sa disponibilité et sa patience ont été indispensables pour la réalisation de ce travail. Qu'elle trouve à travers ces quelques mots l'expression de ma profonde et totale gratitude et reconnaissance.

Mes vifs remerciements vont également au Prof. **TIMERIDJINE Karima**, professeur à l'Université de Bejaia, de m'avoir fait l'honneur d'accepter de présider le jury de cette thèse.

Je remercie aussi, Prof. **ZOUGAB Nabil**, professeur à l'Université de Bejaia, et Dr. **HAMAZ Abdelghani**, Maître de conférences (A) à l'Université de Tizi Ouzou, d'avoir accepté d'examiner ce travail.

Je présente tous mes respects et mes sincères remerciements à mes inestimables parents pour leurs contributions et leur soutien énorme.

J'adresse mes plus sincères remerciements à mon fiancé Amine, qui m'a toujours soutenu et encouragé au cours de la réalisation de cette thèse.

Je tiens à remercier mes collègues de notre Laboratoire "LMA", d'avoir passé ensemble des moments de travail agréables.

※ *Dédicaces* ※

Je dédie ce modeste travail à

Mon cher père : ELHACEN pour le soutien que vous m'avez donné

Pendant toutes ces années

Ma chère mère : merci pour l'amour et la tendresse que vous m'avez

Donnés pendant toutes ces années

Mon cher fiancé AMINE : qui a partagé avec moi le chemin du savoir et qui m'a toujours

Soutenu dans mes moments difficiles, merci pour ta présence dans ma vie.

Mon cher frère CHAOUKI et mon oncle LOUNES qui ont

Toujours été là pour moi

*Ma chère sœur NOURIA et son mari CHERIF et son fils ADEM, ses deux filles
DIHIA et MALAK*

Ma chère sœur CHABHA et son fiancé ABDELLAH

Mes oncles et toute ma famille surtout HARFOUCHE & MOUHOUB

Mon oncle Fahim et sa famille, ma cousine Karima et sa famille

Mon grand père et ma grand mère

Mes petits cousins LIDIA & TINHINEN & INES & CHIHEB & RAHIMA

& CHANAISE & CHAHINAISE & MARBOUHA & TORKIA

*Ma chère HANANE, KATIA, WISSEM, MARBOUHA qui sont toujours avec moi
A tous ceux qui me connaissent*

M^{elle} HARFOUCHE Zineb

Table des matières

Table des matières	i
Table des figures	vi
Liste des algorithmes	vii
Listes des tableaux	ix
Notations et symboles	x
Liste des contributions	xi
Introduction générale	1

Partie Approximation d'un modèle de risque classique par la méthode de stabilité forte

Introduction	9
1 Théorie de risque et modèle de risque classique	10
1.1 Introduction	11
1.2 La théorie de risque	11
1.3 Modèle de risque général	12
1.4 Quelques modèles de risque	13
1.4.1 Modèle individuel	13
1.4.2 Modèle collectif	13
1.4.3 Modèle de risque multi-branches	14
1.5 Modèle de risque Poisson composé (classique)	14
1.6 Probabilité de ruine	15
1.7 Conclusion	18

2	Stabilité des modèle de risque en assurance	19
2.1	Introduction	20
2.2	Concept général de stabilité	20
2.3	Stabilité forte	23
2.3.1	Préliminaires et notations	24
2.4	Stabilité forte d'un modèle de risque classique univarié	27
2.4.1	Stabilité forte d'une chaîne de Markov	28
2.4.2	Stabilité forte d'un processus régénératif	30
2.5	Conclusion	31

Partie Estimation de la densité de probabilité

	Introduction	33
3	Estimation non-paramétrique	34
3.1	Introduction	35
3.2	Estimation par la méthode du noyau	37
3.2.1	Noyau	37
3.2.2	Estimateurs à noyaux continus classiques (symétriques)	38
3.2.3	Estimateurs à noyaux continus asymétriques	42
3.2.4	Choix du paramètre de lissage h_n	49
3.3	Conclusion	53
4	Estimation semi-paramétrique de la densité de probabilité	54
4.1	Introduction	55
4.2	Distribution de Champernowne	55
4.3	Distribution de Champernowne généralisée	56
4.3.1	Propriétés de la distribution de Champernowne généralisée	56
4.3.2	Estimation des paramètres de la distribution de Champernowne généralisée	57
4.4	Estimateur semi-paramétrique basé sur la distribution de Champernowne généralisée et l'estimateur à noyau Beta (KMCE)	58
4.5	Conclusion	60

Partie Application de la méthode semi-paramétrique à l'analyse de la stabilité forte d'un modèle de risque classique

Introduction	62
5 Application numérique basée sur la simulation	63
5.1 Introduction	64
5.2 Version 1 : Stabilité d'une chaîne de Markov	64
5.3 Version 2 : Stabilité d'un processus de Lindley	79
5.4 Conclusion	91
6 Application numérique basée sur des données réelles	92
6.1 Introduction	93
6.2 Présentation des données (sinistres)	93
6.3 Analyse exploratoire des données	93
6.4 Application des méthodes d'estimation et d'approximation sur les sinistres	94
6.5 Conclusion	99
Conclusion générale	100
Bibliographie	102
A Annexe : Données réelles	109

Table des figures

1	Comment décrire les richesses d'une compagnie d'assurance?	2
1.1	Processus des réserves et ruine.	16
1.2	Trajectoire de $X(t)$	16
2.1	Le modèle $(\mathcal{F}, \mathbb{A}^*, \mathbb{B}^*)$	20
3.1	Plusieurs histogrammes avec différents nombres d'intervalles, tous calculés à partir du même échantillon de taille 1000 provenant d'une loi normale de moyenne 0 et de variance 1.	36
3.2	Histogramme calculé à partir d'un échantillon provenant d'une loi normale de moyenne 0 et de variance 1 dont les points centraux des barres sont reliés par une droite.	36
3.3	Exemples de noyaux symétriques.	38
3.4	L'influence du paramètre de lissage h_n sur la qualité de l'estimation.	50
4.1	Effet du paramètre c selon les valeurs de α ($c = 0$ pour graphes en pointillé, $c = 2$ trait plein).	57
5.1	Densités des quatre distributions étudiées.	65
5.2	Les estimateurs finaux de KMCE et la densité réelle pour la distribution de Weibull.	67
5.3	Erreur Γ_1 en fonction de ϵ avec la distribution de Weibull. Estimateur β_1 (bleu), Estimateur Mac- β_1 (turquoise), Estimateur Mic- β_1 (orange), Estimateur β_2 (vert), Estimateur Mac- β_2 (violet), Estimateur Mic- β_2 (marron).	67
5.4	Erreur Γ_1 en fonction de n avec la distribution de Weibull.	68
5.5	Les estimateurs KMCE finaux et la densité réelle pour la distribution lognormale.	68
5.6	Erreur Γ_1 en fonction de ϵ avec une distribution lognormale. Estimateur β_1 (bleu), Estimateur Mac- β_1 (turquoise), Estimateur Mic- β_1 (orange), Estimateur β_2 (vert), Estimateur Mac- β_2 (violet), Estimateur Mic- β_2 (marron).	69
5.7	Erreur Γ_1 en fonction de n avec une distribution lognormale.	69
5.8	Les estimateurs KMCE finaux et la densité réelle pour la distribution mélange Pareto-lognormale 30%70%.	70

5.9	Erreur Γ_1 en fonction de ϵ avec la distribution mélange Pareto-lognormale 30%70%. Estimateur β_1 (bleu), Estimateur Mac- β_1 (turquois), Estimateur Mic- β_1 (orange), Estimateur β_2 (vert), Estimateur Mac- β_2 (violet), Estimateur Mic- β_2 (marron). . . .	70
5.10	Erreur Γ_1 en fonction de n avec avec une distribution mélange Pareto-lognormale 30%70%.	71
5.11	Les estimateurs KMCE finaux et la densité réelle pour la distribution mélange Pareto-lognormale 70%30%.	71
5.12	Erreur Γ_1 en fonction de ϵ avec une distribution mélange Pareto-lognormale 70%30%. Estimateur β_1 (bleu), estimateur Mac- β_1 (turquois), estimateur Mic- β_1 (orange), estimateur β_2 (vert), estimateur Mac- β_2 (violet), estimateur Mic- β_2 (marron). . . .	72
5.13	Erreur Γ_1 en fonction de n avec une distribution mélange Pareto-lognormale 70%30% distribution.	72
5.14	L'estimateur KMCE final ($\hat{f}_{Mic-\beta_2}$), l'estimateur BS et la densité réelle pour les distributions de Weibull et lognormale.	73
5.15	Estimateur KMCE final ($\hat{f}_{Mic-\beta_2}$), estimateur BS et densité réelle pour les distri- butions mélange Pareto-lognormale 70%30% et mélange Pareto-lognormale 30%70%. . . .	73
5.16	Les estimateurs finaux de KMCE et la densité réelle pour la distribution de Weibull.	80
5.17	Erreur Γ_2 (gauche) (Γ_1 (droite)) en fonction de ϵ avec la distribution de Weibull. Estimateur β_1 (bleu), estimateur Mac- β_1 (turquois), estimateur Mic- β_1 (orange), estimateur β_2 (vert), estimateur Mac- β_2 (violet), estimateur Mic- β_2 (marron). . . .	80
5.18	Erreur Γ_2 (haut) (Γ_1 (bas)) en fonction de n avec la distribution de Weibull.	81
5.19	Les estimateurs KMCE finaux et la densité réelle pour la distribution lognormale. . . .	81
5.20	Erreur Γ_2 (gauche) (Γ_1 (droite)) en fonction de ϵ avec la distribution lognormale. Estimateur de β_1 (bleu), estimateur de Mac- β_1 (turquois), estimateur de Mic- β_1 (orange), estimateur de β_2 (vert), estimateur de Mac- β_2 (violet), estimateur de Mic- β_2 (marron).	82
5.21	Erreur Γ_2 (haut) (Γ_1 (bas)) en fonction de n avec la distribution lognormale.	82
5.22	Les estimateurs KMCE finaux et la densité réelle pour la distribution mélange Pareto-lognormale 30%70%.	83
5.23	Erreur Γ_2 (gauche) (Γ_1 (droite)) en fonction de ϵ avec la distribution mélange Pare- to-lognormale 30%70%. Estimateur β_1 (bleu), estimateur Mac- β_1 (turquois), estima- teur Mic- β_1 (orange), estimateur β_2 (vert), estimateur Mac- β_2 (violet), estimateur Mic- β_2 (marron).	83
5.24	Erreur Γ_2 (haut) (Γ_1 (bas)) en fonction de n avec la distribution mélange Pareto- lognormale 30%70%.	84
5.25	Les estimateurs KMCE finaux et la densité réelle pour la distribution mélange Pareto-lognormale 70%30%.	84

5.26	Erreur Γ_2 (gauche) (Γ_1 (droite)) en fonction de ϵ avec la distribution mélange Pareto-lognormale 70%30%. Estimateur β_1 (bleu), estimateur Mac- β_1 (turquoise), estimateur Mic- β_1 (orange), estimateur β_2 (vert), estimateur Mac- β_2 (violet), estimateur Mic- β_2 (marron).	85
5.27	Erreur Γ_2 (haut) (Γ_1 (bas)) en fonction de n avec la distribution mélange Pareto-lognormale 70%30%.	85
6.1	Représentation graphique de la fonction de distribution de probabilité avec estimation du noyau (a) et boxplot (b) pour les données de sinistres.	94
6.2	Les estimateurs finaux KMCE, BS, GAM, RIG et la densité réelle pour les données de sinistres.	95
6.3	Borne de stabilité Γ (approche régénérative (b) et approche par chaîne de Markov (a)) en fonction de ϵ avec les données réelles. Estimateur β_1 (bleu), estimateur Mac- β_1 (turquoise), estimateur Mic- β_1 (orange), estimateur β_2 (vert), estimateur Mac- β_2 (violet), estimateur Mic- β_2 (marron).	95
6.4	L'estimateur KMCE final ($\hat{f}_{Mic-\beta_2}$), l'estimateur BS et la densité pour les données de sinistres.	96

Liste des algorithmes

1	Algorithme 1: Stabilité d'une chaîne de Markov	65
2	Algorithme 2: Stabilité d'un processus de Lindley	79

Liste des tableaux

3.1	Exemples des noyaux symétriques.	38
5.1	Estimation du paramètre $\theta = (\alpha, M, c)$ de la loi de Champernowne.	74
5.2	Borne de stabilité Γ_1 via l'estimateur semi-paramétrique à noyau Beta 1 et ses versions normalisées.	74
5.3	Borne de stabilité Γ_1 via l'estimateur semi-paramétrique à noyau Beta 2 et ses versions normalisées.	75
5.4	L'erreur quadratique moyenne (MSE) des estimateurs étudiés.	75
5.5	Une étude comparative de la borne Γ_1 entre les méthodes non paramétriques et semi-paramétriques.	76
5.6	Estimation du paramètre $\theta = (\alpha, M, c)$ de la loi de Champernowne.	86
5.7	Borne de stabilité Γ_2 via l'estimateur semi-paramétrique à noyau Beta 1 et ses versions normalisées.	86
5.8	Borne de stabilité Γ_2 via l'estimateur semi-paramétrique à noyau Beta 2 et ses versions normalisées.	87
5.9	L'erreur quadratique moyenne (MSE) avec les estimateurs étudiés.	87
5.10	Une étude comparative des intervalles de stabilité entre l'approche régénérative et l'approche par chaîne de Markov.	88
5.11	Une étude comparative de la borne Γ_1 entre les méthodes non paramétriques et semi-paramétriques.	89
5.12	Une étude comparative de la borne Γ_2 entre les méthodes non paramétriques et semi-paramétriques.	89
6.1	Statistiques descriptives pour les données de sinistres d'une compagnie d'assurance australienne.	93
6.2	Une étude comparative de la borne de stabilité Γ entre l'approche régénérative (RP) et l'approche par chaîne de Markov (MC) pour les données de sinistres : cas non-paramétrique.	97

6.3	Une étude comparative de la borne de stabilité Γ entre l'approche régénérative (RP) et l'approche par chaîne de Markov (MC) pour les données de sinistres : cas semi-paramétrique.	97
6.4	Une étude comparative de la borne de stabilité Γ en fonction de la perturbation ϵ entre l'approche régénérative et l'approche par chaîne de Markov et entre l'approche semi-paramétrique et non paramétrique pour les données de sinistres.	98

Notations et symboles

$X(t)$: processus de réserve.

$\Psi(u)$: probabilité de ruine.

θ : chargement de sécurité.

\mathbb{R} : ensemble des nombres réels.

X : variable aléatoire .

$\mathbb{E}(X)$: espérance de la variable aléatoire X .

$\text{Var}(X)$: variance de la variable aléatoire X .

f : densité de probabilité.

F : fonction de répartition associée à la densité f .

v.a. : variable aléatoire.

i.i.d. : indépendantes et identiquement distribuées.

I : fonction indicatrice d'un ensemble.

Liste des contributions

Dans le cadre de cette thèse, nous avons réalisé les contributions scientifiques suivantes :

Articles en rapport avec la thèse

1. Z. Harfouche and A. Bareche. Semi-parametric approach for approximating the ruin probability of classical risk models with large claims. *Communications in Statistics – Simulation and Computation*, DOI :10.1080/03610918.2021.1992636, 2021.
2. Z. Harfouche and A. Bareche. Semi-parametric approach for approximating the ruin probability of classical risk models using a Lindley process, (A soumettre).

Communications

1. Z. Harfouche and A. Bareche. Approximation de la probabilité de ruine de modèles de risque : Approche semi-paramétrique. *Conférence internationale sur les Mathématiques Financières : Outils et Applications*. 28-29 Octobre, 2019, Bejaia.
2. Z. Harfouche and A. Bareche. Combination of the strong stability method and semi-parametric approach for the evaluation of ruin probability of a classical risk model. *2nd National Seminar of Mathematics*, 02 June 2021, Mentouri University Constantine 1, Algeria.

Introduction générale

L'assurance fait aujourd'hui totalement partie de notre cadre de vie quotidien. Souscrire un contrat d'assurance est devenu un acte naturel chez la plupart des personnes désirant se prémunir des pertes financières entraînées par la réalisation casuelle d'un événement entraînant des conséquences fâcheuses (incendie, vol, accident, maladie, etc.). Cependant, bien que l'assurance soit désormais un concept familier pour un très grand nombre d'individus, bien souvent, peu en ont une idée « claire et distincte », dans la mesure où ils ignorent la plupart des mécanismes qui entrent en jeu dans la réalisation d'une opération d'assurance. L'assurance donc est une opération par laquelle l'assureur offre à un assuré une protection en cas de réalisation d'un événement incertain prédéfini moyennant le paiement d'une prime par cet assuré. En contrepartie, la compagnie d'assurance s'engage à payer une partie ou la totalité de la perte encourue à la suite d'un sinistre résultant d'accidents, d'incendies, d'inondations, de catastrophes naturelles, etc. Ce sont ces deux engagements (paiement de la prime contre paiement, le cas échéant, de la prestation) qui constituent le contrat d'assurance.

Les sciences actuarielles sont une discipline relativement récente bien que les bases de l'assurance soient quand à elles très anciennes. Les principales théories en actuariat ont été développées en grande partie dans le dernier siècle. De plus, avec l'amélioration de l'informatique, entre autres, on remarque une progression rapide dans un grand nombre de domaines en actuariat. Un de ces domaines est la théorie de la ruine où la complexité mathématique est souvent un facteur limitant les chercheurs. Cependant, en utilisant des ordinateurs pour des approximations numériques ou encore des simulations, du progrès est encore fait continuellement dans ce domaine.

Dans la littérature actuarielle, l'objectif de la théorie de la ruine est de modéliser la richesse d'une compagnie d'assurance par un processus stochastique et d'évaluer la probabilité de ruine. Cette dernière est l'une des caractéristiques fondamentales des modèles de risque (qui sont des modèles dynamiques en général à temps continu et qui décrivent l'évolution d'un stock avec des

entrées (input) et des sorties (output) et sont donc de ce point de vue très proches des modèles de files d'attente). Le concept de la théorie de la ruine se comprend comme étant la survenance d'un scénario défavorable pouvant conduire à l'impossibilité pour la société d'assurance de faire face à certains de ses engagements, aussi bien envers ses assurés que ses actionnaires, voire à devoir cesser son activité pour cause d'insolvabilité. Le premier but de la théorie de la ruine a donc logiquement été de modéliser l'évolution de la richesse de la compagnie par un processus stochastique, d'évaluer la probabilité de la ruine, soit la probabilité que le scénario traduisant un échec se réalise, et bien entendu d'estimer le niveau de réserve initiale pour prendre cette probabilité de ruine suffisamment faible.

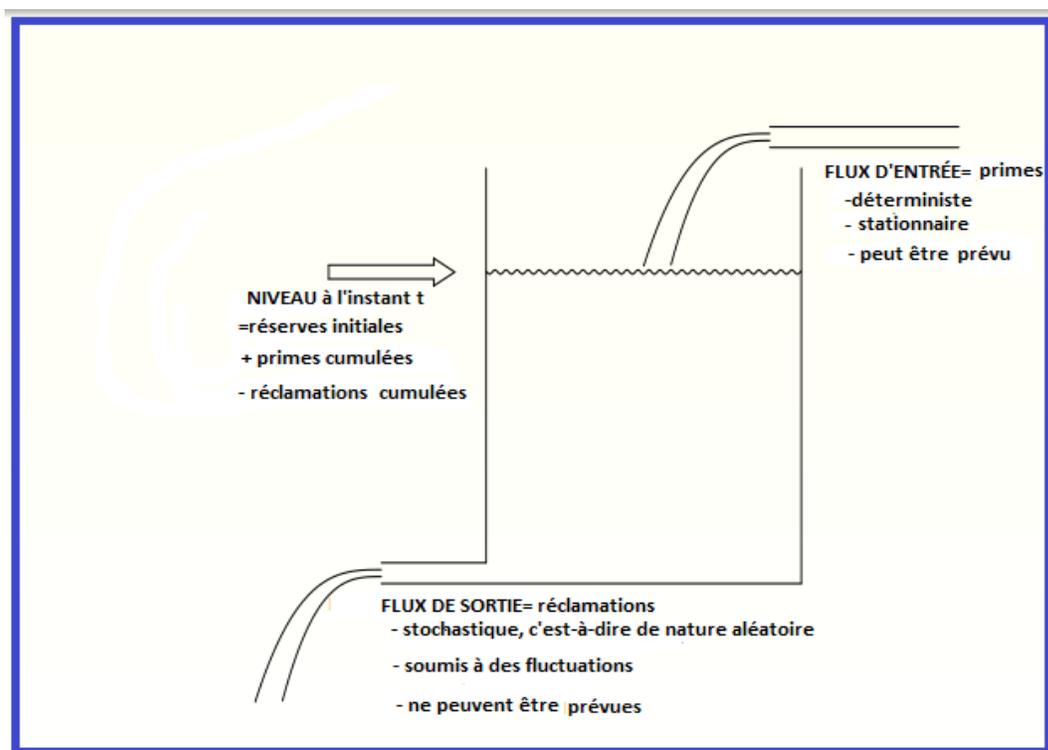


FIG. 1: Comment décrire les richesses d'une compagnie d'assurance ?

Dans le cadre des assurances, que nous prendrons comme cadre de référence dans cette thèse, plusieurs auteurs se sont penchés sur le calcul de la probabilité de ruine (voir [6], [13]). Cependant, il n'existe pas de formule explicite pour son évaluation. De plus, les paramètres régissant ces modèles de risque sont souvent inconnus et ne peuvent être estimés qu'avec incertitude. Par conséquent, on ne peut donner que quelques bornes pour leurs valeurs. Dans de telles situations, la question de la stabilité devient cruciale. En mathématiques des assurances, on peut considérer un modèle stochastique comme une approximation de l'activité d'assurance réelle. Les éléments

stochastiques considérés représentent une idéalisation des phénomènes d'assurance réels. Le problème consiste donc à établir des conditions pour lesquelles notre modèle "idéal" peut être utilisé comme approximation du modèle réel. Le praticien doit se rendre compte de l'exactitude des recommandations, dérivées des investigations sur le modèle idéal [12]. L'utilisation d'approximations implique que nous étudions des modèles "idéaux" qui sont plutôt simples mais proches, dans un certain sens, du modèle réel exprimé sous forme de modèle perturbé.

L'aspect général du concept de stabilité dans les modèles de risque peut être énoncé de la manière suivante : si les éléments d'entrée des modèles idéal et réel sont "proches" les uns des autres, alors comment estimer l'écart entre les sorties correspondantes ? En d'autres termes, la théorie de la stabilité en général, dans le cas des modèles de risque, donne l'image suivante : Si nous avons comme caractéristiques d'entrée la fonction de distribution des réclamations d'un modèle de risque et comme caractéristiques de sortie sa probabilité de ruine, alors, en utilisant la définition d'une certaine norme (la norme pondérée qui sera définie dans le chapitre 2, partie I), la stabilité signifie que la convergence en \mathcal{L}^1 des caractéristiques d'entrée implique la convergence faible des caractéristiques de sortie.

En général, la méthode de stabilité des modèles de risque est basée sur les trois étapes suivantes. Premièrement, nous identifions la probabilité de ruine Ψ_a associée au modèle de risque régi par un vecteur de paramètres a , avec une distribution stationnaire d'un processus aléatoire spécifique, noté par $\{V_n\}_{n \geq 0}$, que l'on appelle un processus inverse (ou dual). Une telle identification est bien connue et a été étudiée dans de nombreux travaux récents (voir par exemple, [13] ; [36] ; [50] ; [78, 79]). La construction du processus inverse est typiquement algébrique (utilise des arguments de type trajectoire) et ne repose pas sur les propriétés probabilistes du processus de risque. Par conséquent, dans certains cas, $\{V_n\}_{n \geq 0}$ ne représente pas une chaîne de Markov. Typiquement, cela peut être résolu en la complétant avec des coordonnées supplémentaires provenant d'une certaine séquence stationnaire $\{Y_n\}_{n \geq 0}$, ce qui fait du processus élargi $\{W_n = (V_n - Y_n)\}_{n \geq 0}$ une chaîne de Markov et éventuellement aussi un processus régénératif (deuxième étape). Ceci permet d'utiliser les résultats bien développés de la stabilité. Ainsi, la dernière étape consiste à appliquer l'aspect quantitatif de l'approche de stabilité, en donnant les déviations possibles des distributions stationnaires des deux processus de Markov.

Après l'introduction du problème de stabilité dans les mathématiques des assurances par Beirlant et Rachev [12], Kalashnikov [50] a étudié l'estimation de la probabilité de ruine dans les

modèles de risque univariés, en utilisant la méthode de stabilité forte (voir [56]), la notion de processus inverse et la technique des variables supplémentaires. Par la suite, plusieurs auteurs ont étendu cette approche à différents types de processus de risque. Kalashnikov et al. [36] ont proposé une approche générale, basée sur trois étapes, permettant d'obtenir des bornes de stabilité (continuité) explicites pour les probabilités de ruine. Rusaityte [78] a appliqué cette approche, basée sur les bornes de stabilité des chaînes de Markov générales, à un risque particulier. Dans ce modèle, le capital dépassant un niveau donné (important) est investi dans un portefeuille dirigé par un processus de Lévy, et lorsque le capital tombe en dessous d'un certain niveau donné, la société est obligée d'emprunter le montant manquant à un intérêt constant. Dans [79], le même auteur, inspirée par les travaux de Kalashnikov [50], a utilisé l'approche régénérative, pour étudier la stabilité de la probabilité de ruine dans un modèle de risque modulé de Markov des investissements basés sur le processus de Lévy. Dans le cadre bidimensionnel, Benouaret et Aissani [13] ont montré l'applicabilité de la méthode de stabilité forte au modèle de risque classique bivarié avec sinistres indépendants. Bareche et Cherfaoui [11] ont étudié, sur la base d'exemples numériques obtenus par simulation, l'impact des réclamations larges sur l'approximation de la probabilité de ruine dans un modèle de risque classique et univarié en utilisant la méthode de stabilité forte et les processus régénératifs. Touazi et al. [88] considèrent l'analyse de la stabilité forte dans les modèles de risque univariés en utilisant l'estimation non paramétrique de la densité (méthode du noyau) pour la distribution inconnue des montants d réclamations (voir également [94]).

Comme mentionné ci-dessus, l'évaluation de la probabilité de ruine n'est pas une tâche facile et reste un problème délicat. La méthode de stabilité forte peut être un outil utile pour fournir des bornes sensibles (qualitatives et quantitatives) pour cette mesure de risque pour de nombreux modèles de risque (voir [11, 13, 36, 50, 78, 79, 88]). Cependant, ces bornes impliquent souvent une inférence statistique sur la distribution des montants de réclamations (également appelée fonction de perte). En effet, cette fonction dans l'actuariat et la gestion des risques financiers est fondamentale et d'une utilité ultime. Elle décrit la distribution de probabilité du paiement à l'assuré. Dans la plupart des situations, les pertes sont faibles, et les pertes extrêmes se produisent rarement. Mais le nombre et la taille des pertes extrêmes peuvent avoir une influence substantielle sur le bénéfice de l'entreprise. Traditionnellement, les praticiens utilisent souvent des spécifications paramétriques (Lognormale, Weibull, Pareto ou un mélange de distributions Lognormale et Pareto) pour modéliser les distributions de pertes par un seul modèle paramétrique, ou décident

d'analyser séparément les grandes et les petites pertes car aucun modèle paramétrique classique ne convient à toutes les tailles de sinistres. Cette approche soulève d'importants défis : choisir le modèle paramétrique approprié, identifier la meilleure façon d'estimer les paramètres et identifier le seuil entre les grandes et les petites pertes. Le lissage non paramétrique ne nécessite pas la spécification d'un modèle paramétrique. Il permet une inférence valide dans une classe de structures beaucoup plus large que celles imposées par les modèles paramétriques. Malheureusement, cette robustesse a un prix. Le taux de convergence des estimateurs non paramétriques est plus lent que le taux paramétrique, et le biais inclus par la procédure de lissage peut être substantiel même pour des tailles d'échantillon modérées. Les pertes étant des variables positives, l'estimateur standard à noyau proposé par Rosenblatt [75] et Parzen [71] présente un biais au bord. Ce dernier est le résultat de l'allocation de poids par le noyau symétrique fixe en dehors du support de la distribution, lorsque le lissage près du bord est effectué. Par conséquent, le mode proche du bord, typique des distributions de perte, est souvent sous estimé. En outre, les méthodes de noyau standard donnent des estimations fluctuantes dans la queue de la distribution parce que l'atténuation du biais au bord conduit à favoriser une petite largeur de fenêtre, ce qui empêche de regrouper suffisamment de données. Cela empêche une mesure précise de la queue des distributions de perte, ce qui est d'une importance primordiale pour obtenir des mesures de risque appropriées lors de la conception d'un système de gestion de risques efficace. Une alternative intéressante aux méthodes paramétriques et non paramétriques est l'approche semi-paramétrique qui convient aux distributions à queue lourde. Elle a l'avantage de donner une fonction de densité qui peut bien s'adapter à toute la gamme de données, sans qu'il soit nécessaire de laisser de côté toute information sur l'échantillon (petites, moyennes et grandes pertes) (voir [10, 15]). L'estimateur semi-paramétrique est obtenu en transformant l'ensemble de données avec une distribution paramétrique, puis en estimant la densité de l'ensemble de données transformées à l'aide de l'estimateur classique de densité à noyau (voir [71, 75, 85]).

De plus, nous sommes souvent confrontés, en finance et en assurance, à des sinistres importants décrits par des distributions à queue lourde (Pareto, lognormale, Weibull, ...). Ce cas a été étudié dans [11], où les auteurs utilisent des exemples numériques basés sur simulation pour monter l'impact des réclamations larges sur l'approximation de la probabilité de ruine dans un modèle de risque classique univarié en utilisant la méthode de stabilité forte et les processus régénératifs (voir [50]).

Notons que les estimateurs semi-paramétriques ont été introduits dans la littérature statistique pour démontrer que l'estimateur classique à noyau de la densité pouvait être amélioré en transformant l'ensemble de données avec certaines transformations spécifiques et en réduisant l'asymétrie des données (voir [15, 21, 29, 42, 44, 91, 93]).

Contrairement à la procédure paramétrique utilisée dans Bareche et Cherfaoui [11] et à la méthode non paramétrique du noyau adoptée dans [88], **l'objectif** de cette thèse est de proposer une approche semi-paramétrique, suivant [21], pour estimer la fonction de perte dans le but d'évaluer la probabilité de ruine d'un modèle de risque classique avec réclamations larges, en utilisant la méthode de stabilité forte. Cette technique semi-paramétrique vise à améliorer la qualité de l'approximation de la probabilité de ruine, en prenant en compte lors de l'estimation de la fonction de perte inconnue à la fois la correction du biais au bord et les problèmes de queue en capturant implicitement les données en queue. La procédure consiste à estimer la fonction de perte inconnue en deux étapes. Tout d'abord, une distribution de départ paramétrique (distribution de Champernowne généralisée (voir [21])) est utilisée pour estimer les données initiales. Nous appliquons ensuite à l'échantillon résultant de la première étape l'estimateur à noyau asymétrique Beta (voir [27]). Cet estimateur est approprié pour ce type de données transformées définies sur $[0, 1]$ car il permet la correction des effets du biais au bord. Pour éviter le problème d'inconsistance de cet estimateur au bord, nous utilisons ses deux versions normalisées : locale (mico-Beta) et globale (macro-Beta) (voir [38]).

Cette thèse comporte une introduction générale, trois parties, une conclusion générale et une bibliographie. Chaque partie comprend deux chapitres. Le premier chapitre est consacré aux outils mathématiques de modèles de risque. Nous introduisons le modèle de risque classique, et nous présentons également une revue de littérature sur la théorie de ruine. Dans le deuxième chapitre, nous présentons les bases de la méthode de stabilité forte et nous donnons un aperçu de son application dans l'analyse des modèles de risque classique et dans l'évaluation de la probabilité de ruine. Le premier chapitre de la deuxième partie aborde la méthode d'estimation non-paramétrique par la méthode du noyau. Le second présente en détail l'approche semi-paramétrique, via la combinaison de la distribution paramétrique de Champernowne généralisée et de l'estimateur non-paramétrique à noyau Beta. Dans le premier chapitre de la troisième partie, nous réalisons quelques études de simulation ainsi qu'une étude comparative pour évaluer la performance et l'efficacité de l'approche semi-paramétrique proposée, et aussi pour montrer son importance dans l'amélioration de la qua-

lité de la borne de stabilité sur la probabilité de ruine dans un modèle de risque classique avec réclamations larges, par rapport à l'approche non paramétrique. Une autre étude de simulation sert à comparer entre l'approche de stabilité forte basée sur les chaînes de Markov et celle basée sur l'approche régénérative pour évaluer la performance de la borne de stabilité. Dans le deuxième chapitre, nous considérons une étude sur des données réelles, afin de concrétiser les résultats obtenus par simulation.

Partie

Approximation d'un modèle de risque classique par la méthode de stabilité forte

Introduction

Le risque est inhérent à la plupart des activités humaines, en particulier dans les compagnies d'assurance. Il semble délicat de mesurer ces différents risques. L'assurance est, par définition, un système qui permet de prémunir un individu, une société ou une entreprise contre les conséquences financières et économiques liées à la survenance d'un risque (événement aléatoire) particulier.

Cette partie est composée de deux chapitres. Le premier chapitre, est consacré aux outils mathématiques de modèles de risque. Nous introduisons donc le modèle de risque classique, et nous présentons aussi une revue de littérature sur la théorie de ruine. Dans le deuxième chapitre, nous présentons les outils essentiels de la théorie de stabilité forte et nous discutons son application au modèle de risque classique. Ces éléments seront utilisés dans les chapitres suivants.

Théorie de risque et modèle de risque classique

Sommaire

1.1	Introduction	11
1.2	La théorie de risque	11
1.3	Modèle de risque général	12
1.4	Quelques modèles de risque	13
1.4.1	Modèle individuel	13
1.4.2	Modèle collectif	13
1.4.3	Modèle de risque multi-branches	14
1.5	Modèle de risque Poisson composé (classique)	14
1.6	Probabilité de ruine	15
1.7	Conclusion	18

1.1 Introduction

L'assurance consiste en une opération de transfert de risque d'un assuré vers un assureur qui, en contrepartie d'une prime, s'engage à supporter les coûts financiers de sinistres éventuels. Afin de garantir le respect de leurs engagements auprès des assurés, les assureurs et réassureurs ont recours à une modélisation actuarielle des risques. Des modèles mathématiques sont sans cesse développés pour répondre à une meilleure compréhension des risques et de leur évolution. Des ouvrages de référence dans le domaine de la théorie du risque sont Klugman et al. [58], Kaas et al. [47], Denuit et Charpentier [33].

Le modèle classique de la théorie de ruine (modèle de Cramér-Lundberg) représente l'évolution des réserves d'une compagnie d'assurance par un processus de Poisson-composé (voir Dufresne [34]). Il est connu comme le fondement théorique de la théorie de risque et a fait l'objet de plusieurs études, et de nombreux résultats sur la probabilité de ruine existent pour ce modèle. Par la suite, plusieurs modèles de risque ont vu le jour et différentes méthodes pour l'évaluation de leurs probabilités de ruine ont été développées dans la littérature, nous citons entre autres : le modèle individuel [33], le modèle collectif [33] et le modèle multidimensionnel [87] comprenant le cas particulier du modèle de risque bidimensionnel [13, 25].

Dans le présent chapitre, nous allons donner les différents outils mathématiques usuels dans la théorie de risque. Nous présentons quelques modèles de risque classique. Nous mettons l'accent particulièrement sur le modèle de Cramér-Lundberg (Poisson composé). Ce modèle, est caractérisé par le fait qu'on peut utiliser les sinistres via un processus de Poisson. Nous nous intéressons à l'évaluation de sa probabilité de ruine. Le calcul de cette mesure de risque se fera d'une façon exacte ou approximative.

1.2 La théorie de risque

La théorie du risque, également appelée théorie de la ruine, est une branche des science actuarielles, qui étudie la vulnérabilité d'un assureur à l'insolvabilité basé sur la modélisation mathématique de l'excédent de l'assureur et l'étude de l'évolution des richesses d'une compagnie d'assurance. Elle sert à l'étude des problématiques (à court terme et à long terme) d'un portefeuille d'assurance non vie. Elle regroupe entre autres la théorie de la ruine et la réassurance.

La théorie de la ruine a pris naissance en Suède au début du 20ème siècle dans les travaux

de Lundberg [65]. Elle suppose l'indépendance entre les différentes variables aléatoires et autres paramètres intervenant dans la modélisation actuarielle.

Le but de la théorie du risque est l'analyse mathématique des fluctuations aléatoires dans les opérations d'assurance. Une bonne partie des travaux sur la théorie du risque est divisée en deux grandes classes : l'étude du problème de la ruine et l'étude de certaines fonctions de répartition nécessaires pour analyser les fluctuations aléatoires d'une opération d'assurance. Le problème de la ruine suscite un grand intérêt, par exemple dans la question bien actuelle des marges de solvabilité. Pour tout modèle de risque d'un surplus financier, la première quantité d'intérêt est la probabilité de ruine.

La théorie du risque peut être classée en deux branches, à savoir la théorie classique ou individuelle du risque et la théorie collective du risque. Selon la théorie classique du risque, la compagnie d'assurance s'intéresse aux contrats individuels assurés par la compagnie. Ainsi, le gain net ou la perte nette pour la compagnie sera la somme de toutes ces variables aléatoires correspondant aux contrats individuels, et un modèle de probabilité approximatif lié au gain ou à la perte de la compagnie peut alors être déterminé dans des conditions spécifiques pour étudier le progrès de la compagnie d'assurance d'un point de vue probabiliste. Dans la théorie collective du risque, aucune attention particulière n'est accordée aux contrats individuels et l'activité de risque de l'entreprise d'assurance concernée est considérée comme un tout.

1.3 Modèle de risque général

Les modèles de risque sont des modèles dynamiques en général à temps continu qui décrivent l'évolution d'un stock avec des entrées (input) et des sorties (output). Ils représentent la réserve d'une compagnie d'assurance et peuvent être décrits en fonction des deux processus suivants :

- $S = \{S(t), t \geq 0\}$: représente la richesse entre les primes et les réclamations ; c'est-à-dire l'activité purement assurance.
- $R = \{R(t), t \geq 0\}$ est le processus des intérêts : il s'agit des intérêt avec un taux déterministe ou stochastique, reçus par la compagnie dans le cas d'un investissement d'une partie de sa réserve.

Pour une entrée $a = (S, R)$ donnée, le modèle de risque est alors défini par son processus des

réserves $X = \{X(t), t \geq 0\}$ tel que

$$X(t) = e^{R(t)} \left(u + \int_0^t e^{-R(x)} dS(x) \right), \quad t \geq 0 \quad (1.1)$$

où u est le surplus initial, i.e : $X(0) = u$, avec la supposition que l'intégrale dans la formule (1.1) est bien définie. C'est une supposition qui reste vraie dans le cas où S et R sont des processus de Lévy indépendants. L'équation précédente représente donc la forme générale du processus des réserves.

La déduction de tous les modèles de risque est basée sur les différents comportements du processus X (voir Embrecht and Schmidli [35]).

1.4 Quelques modèles de risque

1.4.1 Modèle individuel

Le modèle individuel modélise la charge totale générée par les sinistres individuels (voir Denuit et Charpentier [33]). La charge totale pour un portefeuille comprenant n contrats est définie par la formule :

$$S^{Ind} = \sum_{i=1}^N X_i, \quad (1.2)$$

où, X_i est une v.a. positive qui indique le montant total des sinistres subit par l'assuré i sur la période d'observation.

Dans ce type de modèles, les variables X_i sont supposées indépendentes mais pas forcément d'une même loi de probabilité.

1.4.2 Modèle collectif

Le modèle collectif modélise la charge totale subite par un portefeuille vue, non pas contrat par contrat, mais suivant un nombre total de sinistres, tout assuré confondu, (voir [33]). Le montant global des pertes est :

$$S^{Coll} = \sum_{i=1}^N X_i, \quad (1.3)$$

où, N est une v.a. pour le nombre de pertes, X_i représente le montant de la $i^{\text{ème}}$ perte.

Les variables aléatoires X_i sont supposées de même loi et indépendantes. De plus la variable N est indépendante des X_i .

Remarque 1.4.1. On parle de modèle collectif car on associe la même loi de probabilité pour les pertes. Par conséquent, le modèle individuel avec les pertes de même loi de probabilité est un cas particulier du modèle collectif quand N est une constante : $P(N = n) = 1$.

1.4.3 Modèle de risque multi-branches

Les modèles de risque multidimensionnels ont été introduits et étudiés dans la littérature (voir Sundt [87]). Ces modèles sont composés de p branches d'activité, donc p sous réserves. Le nombre d'évènements de réclamation jusqu'au temps t est modélisé généralement par un processus de Poisson $\{N(t), t \geq 0\}$, sachant qu'un seul évènement peut produire jusqu'à p types de réclamations.

Notons pour tout $j = 1, \dots, p$,

- u_j : la réserve initiale de la $j^{\text{ème}}$ branche,
- c_j : le taux de prime constant de la $j^{\text{ème}}$ branche,
- Z_i^j : le montant de la réclamation de type j produite par le $i^{\text{ème}}$ évènement.

Alors, le modèle de risque multi-branches est décrit par le processus suivant :

$$X(t) = \begin{pmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_p(t) \end{pmatrix} = \begin{pmatrix} u_1 + c_1 t - \sum_{i=1}^{N(t)} Z_i^1 \\ u_2 + c_2 t - \sum_{i=1}^{N(t)} Z_i^2 \\ \vdots \\ u_p + c_p t - \sum_{i=1}^{N(t)} Z_i^p \end{pmatrix}, \quad t \geq 0.$$

Un cas particulier du modèle multidimensionnel est le modèle de risque bidimensionnel (bivarié) qui contient deux branches d'activité (voir Chan et al. [25], Benouaret et Aïssani [13]). Il est défini par :

$$X(t) = \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} u_1 + c_1 t - \sum_{i=1}^{N(t)} Z_i^1 \\ u_2 + c_2 t - \sum_{i=1}^{N(t)} Z_i^2 \end{pmatrix}, \quad t \geq 0.$$

1.5 Modèle de risque Poisson composé (classique)

Le modèle de risque Poisson composé (classique) est un modèle utilisé pour décrire le mécanisme d'arrivée des sinistres et de leurs montants. Dans ce modèle, on suppose que les temps d'arrivés et les montants des sinistres sont des variables aléatoires indépendantes.

Le modèle de risque classique est un cas particulier du modèle collectif qui considère le portefeuille comme un tout, au lieu de considérer les contrats individuellement. Le risque global associé au portefeuille est caractérisé par le nombre de sinistres ainsi que les montants de leurs réclamations. Une description plus détaillée du modèle collectif est présentée dans [58], [70] et [19].

Le résultat d'une compagnie d'assurance à la fin de chaque exercice dépend de la réalisation de nombreuses activités. Nous allons considérer ici l'activité provenant du côté "purent assurance".

On suppose que la compagnie d'assurance reçoit des cotisations de ses assurés, appelées primes, de façon déterministe et continue, à raison de c unités de compte par unité de temps. Elle dispose d'une réserve initiale u pour absorber un éventuel excès de sinistralité et doit indemniser ses assurés pour les sinistres qui la concernent.

Le processus stochastique $\{Z(t), t \geq 0\}$ avec $Z(t) = \sum_{i=1}^{N(t)} Z_i$ représente le montant cumulé des sinistres, avec la convention $Z(t) = 0$ si $N(t) = 0$ et Z_i est une variable aléatoire non négative qui modélise le montant du $i^{\text{ème}}$ sinistre.

Le nombre de réclamations (ou de sinistres) $N(t)$ survenues jusqu'au temps $t \geq 0$ est représenté par un processus de Poisson de paramètre $\lambda > 0$. $\{Z_i\}_{i \in \mathbb{N}^*}$ est une suite de variables aléatoires indépendantes et identiquement distribuées et indépendante du processus de Poisson $\{N(t), t \geq 0\}$, de fonction de répartition F et de moyenne μ supposée finie. Par conséquent, le modèle de risque classique (Poisson composé) est entièrement décrit par le processus suivant :

$$X(t) = u + ct - Z(t), \quad t \geq 0, \quad (1.4)$$

où

$$Z(t) = \sum_{i=1}^{N(t)} Z_i.$$

$X(t)$ représente la réserve de la compagnie d'assurance à l'instant t et le processus $\{S(t) = ct - Z(t), t \geq 0\}$ est dit processus de risque. De plus,

$$\mathbb{E}[ct - Z(t)] = ct - \mathbb{E}[N(t)]\mu = ct - \lambda t\mu = (c - \lambda\mu)t,$$

est le risque moyen sur l'intervalle $(0, t]$ avec $\mu = \mathbb{E}[Z(t)]$.

1.6 Probabilité de ruine

La probabilité de ruine de l'activité de l'assurance ou son équivalent, la « value at risk » en provenance du monde bancaire, est le point central de la théorie du risque. Ces deux concepts sont

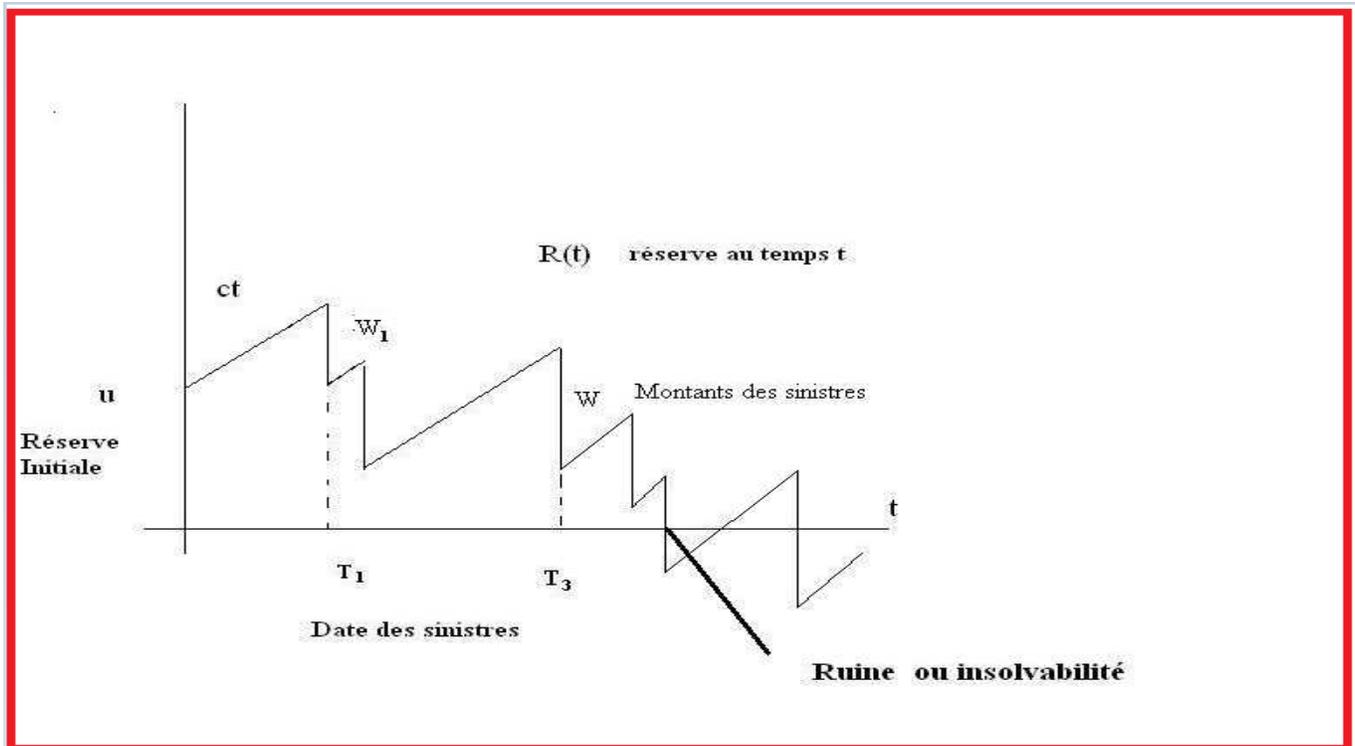


FIG. 1.1: Processus des réserves et ruine.

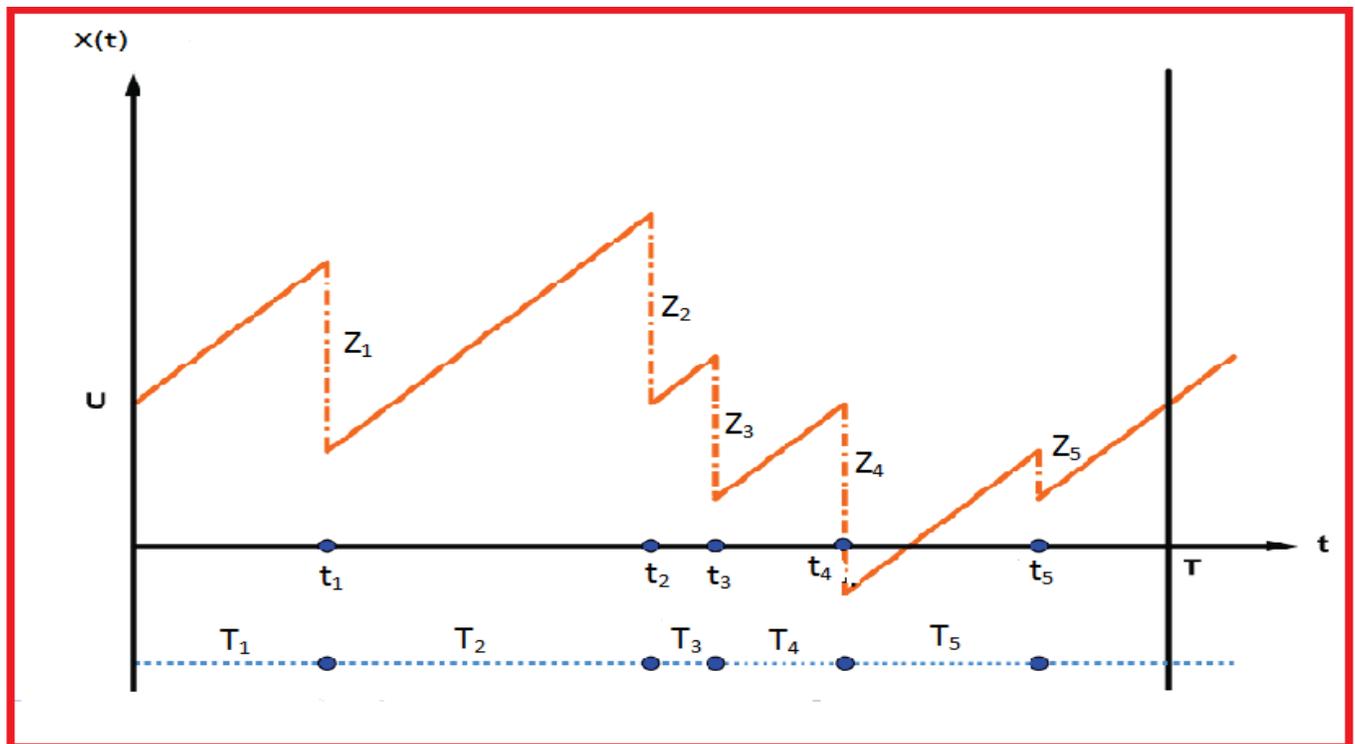


FIG. 1.2: Trajectoire de $X(t)$.

fortement liés, puisque si une compagnie, disposant d'un montant de ressources égal à $Var(X, \alpha)$, assure un unique risque X , sa probabilité de ruine est égale à $1 - \alpha$.

Définition 1.6.1. (Le temps de ruine).

$$\tau(u) = \inf \{t : X(t) < 0 \mid X(0) = u\}. \quad (1.5)$$

est le premier instant où le processus de réserve devient négatif ou de manière équivalente où le processus de surplus excède le niveau u . La probabilité de ruine peut ainsi être définie à travers le moment de la ruine.

Définition 1.6.2. Nous appelons probabilité de ruine en temps fini t , la fonction donnée par

$$\Psi(u, t) = P(\exists s \in [0, t] \mid X(s) < 0), \quad \forall u \geq 0.$$

— En temps infini, elle est définie comme suit

$$\Psi(u, \infty) = P(\exists s \geq 0 \mid X(s) < 0) = \Psi(u), \quad \forall u \geq 0.$$

On peut aussi considérer les probabilités de non ruine en temps fini et en temps infini qui sont définies respectivement par

$$\varphi(u, t) = 1 - \Psi(u, t),$$

et

$$\varphi(u) = 1 - \Psi(u).$$

Chargement de sécurité

La notion de chargement de sécurité est étroitement liée à celle de tarification : un principe de prime contient un chargement de sécurité s'il conduit à exiger une prime supérieure à celle qui est exigée si la mutualisation des risques est parfaite.

On appelle chargement ou coefficient de sécurité, la quantité définie par :

$$\theta = c - \lambda\mu. \quad (1.6)$$

La constante θ définit le chargement de sécurité de la compagnie. Il mesure la rentabilité de la compagnie. Le coefficient $\lambda\mu$ est interprété comme le montant moyen des sinistres par unité de temps. Il paraît prudent que l'assureur fixe un taux de prime c supérieur à $\lambda\mu$ pour que, en moyenne, les primes reçues soient supérieures aux indemnités payées par la compagnie d'assurance.

1.7 Conclusion

Dans ce chapitre, nous avons rappelé les outils mathématiques de modèles des risque, nous avons introduit aussi le modèle de risque classique, et nous avons présenté une revue de littérature sur la théorie de ruine. Ces différents concepts et notions seront utiles dans les applications du prochain chapitre et dans le dernier chapitre (aspect numérique : simulation et application aux données réelles).

Stabilité des modèle de risque en assurance

Sommaire

2.1	Introduction	20
2.2	Concept général de stabilité	20
2.3	Stabilité forte	23
2.3.1	Préliminaires et notations	24
2.4	Stabilité forte d'un modèle de risque classique univarié	27
2.4.1	Stabilité forte d'une chaîne de Markov	28
2.4.2	Stabilité forte d'un processus régénératif	30
2.5	Conclusion	31

2.1 Introduction

Nous donnons dans ce chapitre un aperçu sur la méthode de stabilité forte comprenant son concept général, quelques notations particulières et son application aux modèles de risque classique.

2.2 Concept général de stabilité

Plusieurs auteurs (voir Kalashnikov [49], [52]) donnent un aperçu général des principaux résultats du concept de stabilité des modèles stochastiques, que nous présentons brièvement par la suite :

Soit le modèle décrit par l'application suivante :

$$\mathcal{F} : (\mathbb{A}, \mu) \rightarrow (\mathbb{B}, \nu),$$

avec (\mathbb{A}, μ) représente l'espace des paramètres $a \in \mathbb{A}$, et (\mathbb{B}, ν) représente l'ensemble de ses caractéristiques $b \in \mathbb{B}$. L'association des deux métriques μ et ν aux espaces \mathbb{A} et \mathbb{B} respectivement, nous permettra d'obtenir des estimations quantitatives de la stabilité.

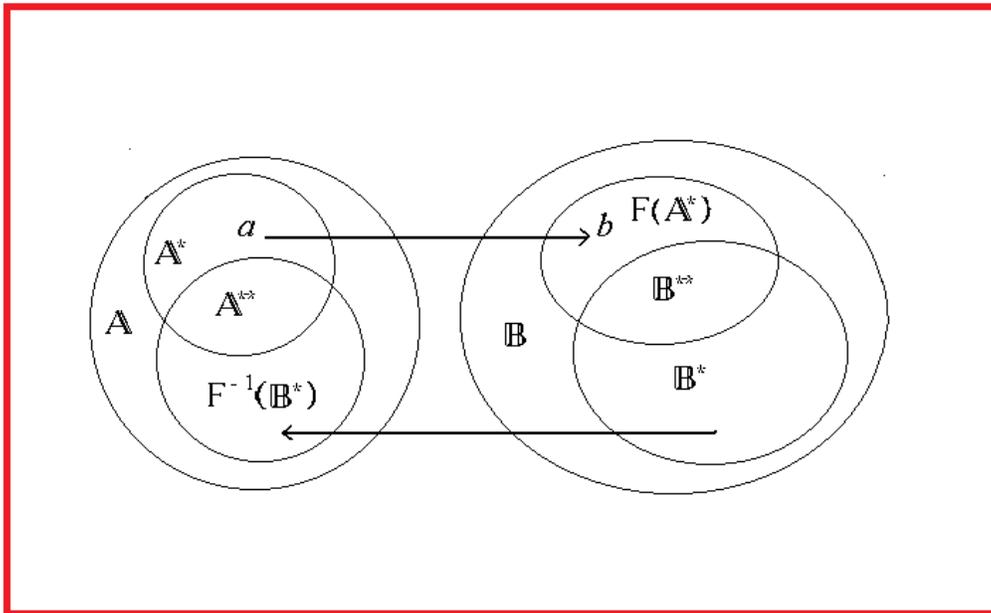


FIG. 2.1: Le modèle $(\mathcal{F}, \mathbb{A}^*, \mathbb{B}^*)$.

Supposons que le modèle limite l'entrée a au sous-ensemble $\mathbb{A}^* \subset \mathbb{A}$ et la sortie b à $\mathbb{B}^* \subset \mathbb{B}$.

\mathbb{B} . On désigne ce modèle par un triplet $(\mathcal{F}, \mathbb{A}^*, \mathbb{B}^*)$. Soit

$$\begin{aligned}\mathcal{F}(\mathbb{A}^*) &= \{b : b = \mathcal{F}(a), \quad a \in \mathbb{A}^*\}, \\ \mathcal{F}^{-1}(\mathbb{B}^*) &= \{a : \mathcal{F}(a) \in \mathbb{B}^*, \quad a \in \mathbb{A}\}, \\ \mathbb{A}^{**} &= \mathbb{A}^* \cap \mathcal{F}^{-1}(\mathbb{B}^*), \\ \mathbb{B}^{**} &= \mathbb{B}^* \cap \mathcal{F}(\mathbb{A}^*) = \mathcal{F}(\mathbb{A}^{**}).\end{aligned}$$

Sur le modèle $(\mathcal{F}, \mathbb{A}^*, \mathbb{B}^*)$, on peut citer deux problèmes importants :

- Le premier, dit problème direct, consiste à décrire l'ensemble \mathbb{B}^{**} sous les restrictions imposées. Le cas particulier le plus important de ce problème se réfère à $\mathbb{B}^* = \mathbb{B}$ où aucune restriction a priori sur les sorties n'est imposée. Si, par exemple, nous considérons un processus de risque classique satisfaisant la condition de Cramér (ensemble \mathbb{A}^*) et nous mettons $\mathbb{B}^* = \mathbb{B}$, alors $\mathbb{B}^{**} = \mathcal{F}(\mathbb{A}^*)$ consiste en des probabilités de ruine qui décroissent asymptotiquement de façon exponentielle.
- Le deuxième, dit problème inverse, consiste à décrire l'ensemble \mathbb{A}^{**} . Le problème inverse traite de la restauration des valeurs des entrées à condition que les restrictions \mathbb{A}^* et \mathbb{B}^* soient connues. Une telle configuration est typique des problèmes de caractérisation (voir [51] et [81]). Pour en revenir au modèle de risque, supposons que \mathbb{B}^* consiste en des probabilités de ruine qui décroissent asymptotiquement de façon exponentielle et supposons que l'ensemble \mathbb{A}^* ne distingue que les modèles de risque classiques. La solution au problème inverse peut alors être énoncée comme suit : \mathbb{A}^{**} est constitué de tous les modèles satisfaisant la condition de Cramér.

Signalons que pour les deux problèmes précédents (direct et inverse), il est clair que $a \in \mathbb{A}^*$ et $b \in \mathbb{B}^*$ où \mathbb{A}^* et \mathbb{B}^* sont donnés. Par contre, dans la théorie de stabilité, nous nous intéressons aux modèles perturbés. Autrement dit, les relations $a \in \mathbb{A}^*$ et $b \in \mathbb{B}^*$ ne sont pas nécessairement vérifiées. L'analyse de stabilité permet de répondre à la question suivante : à quelle distance peut être $b = \mathcal{F}(a)$ de \mathbb{B}^{**} si a est proche de \mathbb{A}^* et b proche de \mathbb{B}^* . Définissons alors

$$\kappa(a) = \mu(a, \mathbb{A}^*) + \nu(\mathcal{F}(a), \mathbb{B}^*), \quad (2.1)$$

comme une mesure de la perturbation.

En particulier, si $\mathbb{B}^* = \mathbb{B}$, on a $\kappa(a) = \mu(a, \mathbb{A}^*)$.

Évidemment,

$$\{a \in \mathbb{A}^*\} \cap \{\mathcal{F}(a) \in \mathbb{B}^*\} \iff \kappa(a) = 0.$$

Considerons les fonctions non-négatives :

$$\phi_i = [0, \infty] \rightarrow [0, \infty], \quad i = 1, 2$$

qui sont continues en 0 et $\phi_i(0) = 0$, $i = 1, 2$ (notez qu'elles peuvent prendre des valeurs infinies, en général).

Nous passons maintenant aux définitions de la stabilité. Les définitions 2.2.1 et 2.2.2 traitent de la correspondance "directe" \mathcal{F} lorsque nous étudions si une "petite" perturbation du modèle implique une petite déviation des sorties correspondantes.

Définition 2.2.1. Le modèle $(\mathcal{F}, \mathbb{A}^*, \mathbb{B}^*)$ est d-stable si pour toute séquence $a_n \in \mathbb{A}$ telle que $\kappa(a_n) \rightarrow 0$ quand $n \rightarrow \infty$, la relation $\nu(\mathcal{F}(a_n), \mathbb{B}^{**}) \rightarrow 0$ est vérifiée.

S'il existe deux fonctions ϕ_i , $i = 1, 2$ telles que

$$\phi_1(\kappa(a)) \leq \nu(\mathcal{F}(a_n), \mathbb{B}^{**}) \leq \phi_2(\kappa(a)) \quad \forall a \in \mathbb{A}, \quad (2.2)$$

alors les inégalités données par (2.2) sont appelées des bornes de d-stabilité (supérieure et inférieure respectivement).

Définition 2.2.2. Le modèle $(\mathcal{F}, \mathbb{A}^*, \mathbb{B}^*)$ est d-stable au point $a \in \mathbb{A}^{**}$ si pour toute séquence $a_n \in \mathbb{A}$ telle que $\theta(a_n, a) = \mu(a_n, a) + \nu(\mathcal{F}(a_n), \mathbb{B}^*) \rightarrow 0$ quand $n \rightarrow \infty$, la relation $\nu(\mathcal{F}(a_n), \mathcal{F}(a)) \rightarrow 0$ est vérifiée.

S'il existe deux fonctions ϕ_i : $i = 1, 2$ (elles peuvent dépendre de $a \in \mathbb{A}^{**}$) telles que

$$\phi_1(\theta(a', a)) \leq \nu(\mathcal{F}(a'), \mathcal{F}(a)) \leq \phi_2(\theta(a', a)) \quad \forall a' \in \mathbb{A}, \quad (2.3)$$

alors les inégalités données par (2.3) sont dites des bornes de d-stabilité au point a .

Définition 2.2.3. Le modèle $(\mathcal{F}, \mathbb{A}^*, \mathbb{B}^*)$ est d-stable en tout point $a \in \mathbb{A}^{**}$ s'il existe deux fonctions ϕ_1 et ϕ_2 indépendantes de a telles que :

$$\phi_1(\theta(a', a)) \leq \nu(\mathcal{F}(a'), \mathcal{F}(a)) \leq \phi_2(\theta(a', a)), \quad (2.4)$$

pour tout $a \in \mathbb{A}^{**}$ et $a' \in \mathbb{A}^*$. Le modèle est dit uniformément stable et les inégalités données par (2.4) sont dites des bornes de d-stabilité uniforme.

Des exemples illustrant l'application des définitions ci-dessus de stabilité aux modèles de risque sont présentés dans Kalashnikov [50].

L'aspect général du concept de stabilité dans les modèles de risque peut être énoncé comme suit :

Supposons que le paramètre directeur a d'un modèle de risque prenne ses valeurs dans l'espace métrique (\mathbb{A}, μ) . Par exemple, notre modèle de risque est régi par le taux de prime c , l'intensité de survenance des sinistres λ et la fonction de distribution de la taille des sinistres F avec la taille moyenne des sinistres μ . Par conséquent, $a = (\lambda, \mu, c)$, et donc $\mathbb{A}_+ \times \mathbb{R}_+ \times \mathbb{R}_+$. Pour un paramètre a fixe, nous considérons la probabilité de ruine $\Psi_a(x)$ en fonction du capital initial x et supposons que l'espace $\Psi = \{\Psi_a(x), a \in \mathbb{A}\}$ est doté d'une métrique ν . La probabilité de ruine $\Psi_a(x)$ est dite stable au point a si pour toute séquence $\{a'\} \subset \mathbb{A}$ qui converge vers a , on a la convergence des probabilités de ruine correspondantes, soit

$$(\mu(a, a') \rightarrow 0) \implies (\nu(\Psi_a, \Psi_{a'}) \rightarrow 0).$$

De plus, si on peut trouver une fonction continue f en 0 ($f(0) = 0$) et un sous-ensemble $\mathbb{A}_d \subset \mathbb{A}$ tel que :

$$\nu(\Psi_a, \Psi_{a'}) \leq f(\mu(a, a')), \quad a, a' \in \mathbb{A}_d, \quad (2.5)$$

alors l'inégalité (2.5) est appelée une borne de stabilité pour la probabilité de ruine. En général, la fonction f dépend des paramètres a, a' .

2.3 Stabilité forte

Le concept de stabilité a été largement et longtemps utilisé dans diverses sciences. Cependant, il n'existe pas une définition universelle pour la stabilité, qui a été souvent adaptée aux besoins spécifiques d'une science particulière ou d'un problème particulier. Ainsi, la stabilité est peut être l'un des termes scientifiques ayant le plus de significations possibles. Le sens large de la stabilité peut se comprendre comme l'aptitude d'un système à maintenir son fonctionnement sans changement de sa structure interne malgré les perturbations externes. Le concept de stabilité a été défini sous forme mathématique par Liapunov [63] dans sa thèse intitulée "The general problem of the stability of motion". Il y décrit le comportement de l'équilibre du système solaire. Plusieurs notions de stabilité et d'instabilité existent dans la théorie des modèles stochastiques, comme c'est le cas dans les systèmes déterministes classiques. C'est le cas, par exemple au niveau des chaînes de Markov. Pour une revue sur les autres méthodes de stabilité, le lecteur peut se référer

à [16, 41, 73, 76, 86]. Dans la section suivante, nous présentons un aperçu de l'application de l'approche de stabilité forte aux modèles stochastiques, en général.

2.3.1 Préliminaires et notations

Soit $m\mathcal{E}$ l'espace des mesures finies sur l'espace probabilisable (E, \mathcal{E}) , et $f\mathcal{E}$ l'espace des fonctions mesurables bornées sur E . Nous associons à chaque noyau de transition P la correspondance linéaire suivante :

$$\begin{aligned}\mu P(A) &= \int_E \mu(dx) P(x, A), & \forall A \in \mathcal{E}, \\ Pf(x) &= \int_E P(x, dy) f(y), & \forall x \in E.\end{aligned}$$

Introduire sur $m\mathcal{E}$ la classe des normes de la forme :

$$\|\mu\|_v = \int_E v(x) |\mu|(dx), \quad (2.6)$$

où v est une fonction mesurable arbitraire (pas nécessairement finie) bornée inférieurement par 0 et croissante (typiquement vers ∞), et $|\mu|$ est la variation de la mesure μ .

La norme $\|\cdot\|_v$ est appelée la métrique de variation totale pondérée.

Cette norme induit dans l'espace $f\mathcal{E}$ la norme suivante :

$$\|f\|_v = \sup_{x \in E} \{|\mu f|, |\mu f|_v \leq 1\} = \sup_{x \in E} [v(x)]^{-1} |f(x)|, \quad \forall f \in f\mathcal{E}.$$

Considérons également \mathcal{B} , l'espace des opérateurs linéaires avec la norme :

$$\|P\|_v = \sup_{x \in E} \left([v(x)]^{-1} \int_E v(y) |P(x, dy)| \right). \quad (2.7)$$

Pour plus de détails sur la méthode de stabilité forte, le lecteur est renvoyé à Kartashov [9, 56]. Toutefois, rappelons la définition de base suivante, qui donne l'aspect qualitatif de la méthode de stabilité forte.

Définition 2.3.1. (Kartashov [56]) Une chaîne de Markov X avec noyau de transition P et de distribution stationnaire π est dite fortement v -stable par rapport à la norme $\|\cdot\|_v$ (où $\|\alpha\|_v$ est définie dans l'expression (2.6) pour toute mesure α), si $\|P\|_v < \infty$ et chaque noyau stochastique Q dans un certain voisinage $\{Q : \|Q - P\|_v < \varepsilon\}$ pour $\varepsilon > 0$ admet une unique distribution stationnaire $\mu = \mu(Q)$ telle que $\|\pi\mu\|_v \rightarrow 0$ lorsque $\|Q - P\|_v \rightarrow 0$.

Définition 2.3.2. (Kartashov [56]) On dit qu'une chaîne de Markov X avec un noyau de transition P est uniformément ergodique par rapport à une certaine norme $\|\cdot\|$ si elle admet une mesure invariante unique π et :

$$\lim_{t \rightarrow +\infty} \|P^t - \Pi\| = 0,$$

où Π est le projecteur stationnaire du noyau de transition P , défini par $\Pi = I \circ \pi$, et \circ désigne la convolution entre une mesure et une fonction et I est la fonction identité.

D'autres définitions légèrement différentes de l'ergodicité uniforme des chaînes de Markov peuvent être trouvées dans [67] et [68].

Théorème 2.3.1. (Aïssani et Kartashov [9], Kartashov [54]) Une chaîne de Markov X est uniformément ergodique par rapport à la norme $\|\cdot\|$ si et seulement si l'opérateur $(I - P + \Pi)$ est inversible et borné

$$\|(I - P + \Pi)^{-1}\| < \infty, \quad (2.8)$$

où I est l'opérateur identité. De plus, la relation (2.8) entraîne $\|\Pi\| < \infty$.

Définition 2.3.3. Une chaîne de Markov X est dite fortement stable par rapport à la norme $\|\cdot\|$ si

1. $\|P\| < \infty$.
2. Chaque noyau de transition Q dans un certain voisinage $\{Q : \|Q - P\| < \varepsilon\}$, admet une mesure invariante unique $\nu = \nu(Q)$.
3. Il existe une constante $C = C(P)$, telle que $\|\nu - \pi\| \leq C \|P - Q\|$.

Théorème 2.3.2. (Aïssani et Kartashov [9], Kartashov [54]) Une chaîne de Markov X de noyau de transition P tel que $\|P\| < \infty$ est fortement stable par rapport à la mesure $\|\cdot\|$ si et seulement si elle est uniformément ergodique par rapport à la même norme.

Pour cela, quel que soit le noyau Q , de mesure invariante ν , on a

$$\|\nu - \pi\| = \mathcal{O}(\|Q - P\|)$$

et

$$\sup_t \|Q^{(t)} - P^{(t)}\| = \mathcal{O}(\|Q - P\|),$$

quand $\|Q - P\| \rightarrow 0$.

Théorème 2.3.3. (Aïssani et Kartashov [9], Aïssani [7]) Soit X une chaîne uniformément ergodique. Alors chaque noyau stochastique Q dans un certain voisinage $\{Q : \|Q - P\| < \varepsilon\}$ correspond à une chaîne de Markov uniformément ergodique et fortement stable par rapport à la même norme $\|\cdot\|$.

Cela veut dire que la propriété d'ergodicité uniforme, par rapport à une norme donnée, se préserve pour de petites perturbations du noyau de transition par rapport à la même norme.

Le théorème suivant donne l'équivalence entre l'ergodicité uniforme et la stabilité forte :

Théorème 2.3.4. (Aïssani and Kartashov [9]) Une chaîne de Markov X avec un noyau de transition P ($\|P\|_v < \infty$) est dite fortement v -stable par rapport à la norme pondérée $\|\cdot\|_v$ si et seulement si elle est uniformément ergodique par rapport à la même norme $\|\cdot\|_v$.

Théorème 2.3.5. (Aïssani and Kartashov [9]) Une chaîne de Markov X avec un noyau de transition P et une mesure invariante π , est fortement v -stable par rapport à la norme $\|\cdot\|_v$, si et seulement si il existe une mesure α et une fonction mesurable non négative h sur E telles que $\pi h > 0$, $\alpha I = 1$, $\alpha h > 0$ et :

1. $\|P\|_v < \infty$,
2. L'opérateur $T = P - h \circ \alpha$ est non négatif,
3. Il existe $\rho < 1$ tel que $Tv(x) \leq \rho v(x)$, pour $x \in E$.

Dans la suite, nous avons besoin du résultat suivant :

Théorème 2.3.6. (Kartashov [56]) Soit v une fonction poids et supposons qu'une chaîne de Markov avec le noyau de transition P , satisfaisant $\|P\|_v < \infty$, possède une distribution stationnaire unique π . Supposons également qu'il existe une fonction non négative h et une mesure de probabilité α telles que P peut être écrit comme suit :

$$P(x, \cdot) = T(x, \cdot) + h(x) \alpha(\cdot), \quad (2.9)$$

où

$$\|\pi\|_h > 0, \quad \|\alpha\|_h > 0, \quad (2.10)$$

et

$$\|T\|_v \leq \rho \leq 1. \quad (2.11)$$

Alors, toute chaîne de Markov avec noyau de transition P' satisfaisant l'inégalité :

$$\Delta = \|P - P'\|_v < \Delta_0 \equiv \frac{(1 - \rho)^2}{1 - \rho + \rho \|\alpha\|_v}, \quad (2.12)$$

a une distribution stationnaire unique π' et de plus

$$\|\pi - \pi'\|_v \leq \frac{\Delta \|\alpha\|_v}{(1 - \rho)(\Delta_0 - \Delta)}. \quad (2.13)$$

2.4 Stabilité forte d'un modèle de risque classique univarié

C'est l'académicien Kalashnikov dans son article "The stability concept for stochastic risk models" [50], qui a initié l'étude de stabilité forte dans les modèles de risque.

Processus inverse "reversed process"

Le processus dual (processus inverse) du processus de réserve $X(t)$ défini dans l'équation (1.4) peut être simplement écrit, pour tout $n \geq 0$ comme suit (voir [36] ; [50]) :

$$V_{n+1} = (V_n - cU + Z_{n+1})_+, \quad V_0 = 0, \quad (2.14)$$

où U est une variable aléatoire suivant une distribution exponentielle de paramètre λ .

Le processus $V = \{V_n\}_{n \geq 0}$ est défini sur l'espace $(E, \mathcal{E}) = (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$, avec $\mathcal{B}(\mathbb{R}_+)$ la σ -algèbre générée par les boréliens de \mathbb{R}_+ .

Par conséquent, sur la base des propriétés des processus duaux (voir [8], [36], [50]), il est possible d'écrire les probabilités de ruine en temps fini $(\Psi_a)_n(x)$ et en temps infini $\Psi_a(x)$ comme suit :

$$(\Psi_a)_n(x) = P(V_n > x), \quad n \geq 0, \quad (2.15)$$

$$\Psi_a(x) = \lim_{n \rightarrow +\infty} (\Psi_a)_n(x) = \lim_{n \rightarrow +\infty} P(V_n > x). \quad (2.16)$$

Ainsi, à partir des relations (2.15) et (2.16), la caractéristique fondamentale du modèle peut être exprimée par l'ensemble des fonctions cumulatives suivantes :

$$G_n(x) = 1 - P(V_n > x), \quad n \geq 0,$$

$$G(x) = \lim_{n \rightarrow +\infty} G_n(x) = 1 - \lim_{n \rightarrow +\infty} P(V_n > x).$$

Par ailleurs, selon la relation (2.16), la probabilité de ruine est liée à la distribution stationnaire du processus dual $V = \{V_n\}_{n \geq 0}$, qui forme une chaîne de Markov homogène.

Soit $a' = (\lambda', \mu', c')$ le vecteur de paramètres régissant un autre modèle de risque univarié défini comme précédemment, sa probabilité de ruine et son processus inverse étant respectivement $\Psi_{a'}(u)$ et $\{V'_n\}_{n \geq 0}$. Soit également π et π' les distributions stationnaires des processus $\{V_n\}_{n \geq 0}$ et $\{V'_n\}_{n \geq 0}$, respectivement. Alors, d'après l'équation (2.16), nous avons la relation suivante :

$$\|\Psi_a - \Psi_{a'}\|_v = \|\pi - \pi'\|_v. \quad (2.17)$$

Nous présenterons, dans les sections suivantes, deux résultats importants pour le modèle de risque que nous considérons, basés sur la stabilité forte d'une chaîne de Markov et la stabilité forte d'un processus de Lindley.

2.4.1 Stabilité forte d'une chaîne de Markov

Le théorème suivant détermine la stabilité forte d'un modèle de risque classique univarié en utilisant la théorie générale de chaînes de Markov. Il donne également les estimations des écarts entre les opérateurs de transition et les probabilités de ruine à l'état stationnaire.

Théorème 2.4.1. (*Enikeeva et al. [36], Kalashnikov [50]*) *Considérons un modèle de risque classique univarié gouverné par le vecteur de paramètres a . Alors, il existe $\varepsilon > 0$ tel que le processus inverse $\{V_n\}_{n \geq 0}$ (chaîne de Markov) associé à ce modèle est fortement stable par rapport à la fonction poids $v(x) = e^{\varepsilon x}$ ($\varepsilon > 0$), $x \in \mathbb{R}_+$. De plus, si $\mu(a, a') < (1 - \rho(\varepsilon))^2$, alors nous obtenons la marge entre les opérateurs de transition P et P' des chaînes de Markov $\{V_n\}_{n \geq 0}$ et $\{V'_n\}_{n \geq 0}$:*

$$\|P - P'\|_v \leq 2\mathbb{E}e^{\varepsilon z} \left| \ln \frac{\lambda c'}{\lambda' c} \right| + \|F - F'\|_v, \quad (2.18)$$

où,

$$\begin{aligned} \mu(a, a') &= 2\mathbb{E}e^{\varepsilon z} \left| \ln \frac{\lambda c'}{\lambda' c} \right| + \|F - F'\|_v, \\ \rho(\varepsilon) &= \mathbb{E}(\exp\{\varepsilon(Z - cU)\}), \\ \|F - F'\|_v &= \int_0^\infty v(u) |d(F - F')|(u) = \int_0^\infty v(u) |f - f'| (u) du. \end{aligned}$$

De plus, nous avons l'écart entre les probabilités de ruine :

$$\|\Psi_a, \Psi_{a'}\|_v \leq \frac{\mu(a, a')}{(1 - \rho(\varepsilon))(1 - \rho(\varepsilon))^2 - \mu(a, a')} = \Gamma_1. \quad (2.19)$$

Démonstration. La preuve de ce théorème est basée sur les résultats du Théorème 2.3.6.

Supposons que, pour un point donné a , la condition suivante soit vérifiée : il existe un $\varepsilon > 0$ tel que :

$$\mathbb{E}(\exp\{\varepsilon(Z - cU)\}) = \rho(\varepsilon) < 1. \quad (2.20)$$

Notons que l'équation (2.20) donne la condition de profit net.

Nous allons d'abord vérifier les conditions (2.9), (2.10) et (2.11) du Théorème 2.3.6. La décomposition (2.9) du noyau de transition P est basée sur la représentation suivante :

$$P(x, A) = \mathbb{P}(x + Z_n - cU \in A, x + Z_n - cU > 0) + \mathbb{P}(x + Z_n - cU \leq 0) \delta_0(A),$$

où $\delta_0(A)$ est la mesure de probabilité concentrée en 0. Ainsi le noyau P de l'équation (2.9) est défini par la fonction $h(x) = \mathbb{P}(x + Z_n - cU \leq 0)$ et la mesure de probabilité $\alpha(\cdot) = \delta_0(\cdot)$.

Prenons la fonction poids $v(x) = e^{\varepsilon x}$. En utilisant l'expression (2.7) et la condition (2.20), on a :

$$\begin{aligned} \|T\|_v &= \sup_{x \geq 0} e^{-\varepsilon x} \int_0^{+\infty} e^{\varepsilon y} T(x, dy) \\ &= \sup_{x \geq 0} e^{-\varepsilon x} \mathbb{E}(e^{\varepsilon(x+Z_n-cU)}) I(Z_n - cU + x > 0) \\ &\leq \mathbb{E}(\exp\{\varepsilon(Z - cU)\}) = \rho(\varepsilon) < 1, \end{aligned}$$

où I désigne la fonction indicatrice. Ainsi, les conditions du Théorème 2.3.6 sont satisfaites.

La distance $\|P - P'\|_v$ peut être estimée comme suit (voir [36]) :

$$\|P - P'\|_v \leq 2\mathbb{E}e^{\varepsilon z} \left| \ln \frac{\lambda c'}{\lambda' c} \right| + \|F - F'\|_v := \mu(a, a'). \quad (2.21)$$

Notons que la "petitesse" de la distance $\mu(a, a')$ exige que F et F' soient proches en termes de la distance $\|F - F'\|_v$.

Puisque $\|\alpha\|_v = \|\delta_0\|_v = 1$ alors $\Delta_0 = (1 - \rho(\varepsilon))^2$. En remplaçant ces expressions dans l'inégalité (2.13) et en utilisant la relation (2.17), on obtient :

$$\|\Psi_a, \Psi_{a'}\|_v \leq \frac{\mu(a, a')}{(1 - \rho(\varepsilon))(1 - \rho(\varepsilon))^2 - \mu(a, a')} = \Gamma_1,$$

qui est une borne supérieure de stabilité au point a . □

2.4.2 Stabilité forte d'un processus régénératif

2.4.2.1 Processus régénératif

Un processus stochastique $X = \{X(t) : t > 0\}$ est appelé processus régénératif s'il existe une variable aléatoire $R_1 > 0$ telle que :

- $\{X(t + R_1) : t > 0\}$ est indépendant de $\{X(t) : t < R_1\}$,
- $\{X(t + R_1) : t > 0\}$ est stochastiquement équivalent à $\{X(t) : t > 0\}$,

où R_1 appelée époque de régénération (ou temps de régénération) et on dit que X se régénère ou se réinitialise à ce point.

Ce type de processus est d'un grand intérêt dans plusieurs modèles stochastiques, comme les systèmes de files d'attente. En 2000, Kalashnikov [50] a proposé la théorie des processus régénératifs comme solution pour l'application de la méthode de stabilité dans le modèle de risque avec des réclamations larges.

2.4.2.2 Stabilité d'un processus de Lindley

Pour pouvoir prouver le dernier théorème de cette section (Théorème 2.4.4), nous énonçons les deux résultats intermédiaires suivants (Théorème 2.4.2 et 2.4.3) :

Théorème 2.4.2. (Kartashov [55]) *Supposons que $\mathbb{E}(\xi) < 0$ et qu'il existe $\theta > 0$ tel que $\mathbb{E}[\exp(\theta\xi)] < \infty$. Alors, la chaîne de Markov V est apériodique et uniformément ergodique par rapport $\|\cdot\|_v$ pour tout ε tel que $\rho(\varepsilon) = \mathbb{E}[\exp(\theta\xi)] < 1$.*

Théorème 2.4.3. (Bareche et Cherfaoui [11]) *Supposons que les conditions du Théorème 2.4.2 soient vérifiées. Alors, pour tout ε tel que $\mathbb{E}[\exp(\theta\xi)] < 1$ et $\gamma(\varepsilon) = \|P - P'\|_v < 1 - \rho(\varepsilon)$, on a l'inégalité suivante :*

$$\|\pi - \pi'\|_v \leq \frac{1}{(1 - \rho(\varepsilon) - \gamma(\varepsilon))^2} \left(\frac{2 - \rho(\varepsilon)}{1 - \rho(\varepsilon)} \right) \|P - P'\|_v. \quad (2.22)$$

La preuve du Théorème 2.4.3 est basée sur les résultats du Théorème 2.3.6.

Le théorème suivant donne de nouvelles estimations quantitatives de la probabilité de ruine du modèle défini en (1.4) basé sur la stabilité forte d'un processus de Lindley (voir [36], [50]).

Théorème 2.4.4. (Bareche et Cherfaoui [11]) *Soit $\varepsilon > 0$ tel que $\mathbb{E}(\exp\{\varepsilon(Z - cU)\}) < 1$. Si*

$$\mu(a, a') < 1 - \rho(\varepsilon), \quad (2.23)$$

alors on obtient la marge entre les probabilités de ruine des chaînes de Markov $\{V_n\}_{n \geq 0}$ et $\{V'_n\}_{n \geq 0}$:

$$\|\Psi_a - \Psi_{a'}\|_v \leq \frac{\mu(a, a')(2 - \rho(\varepsilon))}{(1 - \rho(\varepsilon))(1 - \rho(\varepsilon) - \mu(a, a'))^2} = \Gamma_2. \quad (2.24)$$

Démonstration. Considérons les deux distributions stationnaires $\pi = (\pi(\Gamma), \Gamma \in \mathcal{B}(\mathbb{R}_+))$ et $\pi' = (\pi'(\Gamma), \Gamma \in \mathcal{B}(\mathbb{R}_+))$ des chaînes de Markov V et V' , respectivement. Ainsi, pour tout $x \in \mathbb{R}_+$, on a :

$$\begin{aligned} \pi([0, x]) &= \mathbb{P}(V_\infty \leq x) \\ &= 1 - \mathbb{P}(V_\infty > x) \\ &= 1 - \lim_{n \rightarrow +\infty} \mathbb{P}(V_n > x) \\ &= 1 - \Psi_a(x), \end{aligned}$$

et de même $\pi'([0, x]) = 1 - \Psi_{a'}(x)$. Par conséquent, on trouve que $\|\pi - \pi'\|_v = \|\Psi_a - \Psi_{a'}\|_v$. En substituant l'estimation (2.18) dans l'inégalité (2.22), le résultat est immédiatement obtenu. \square

Remarque 2.4.1. Sans perte de généralité, nous relaxons certaines conditions en prenant $\lambda' = \lambda$, $c' = c$. On a alors, $\mu(a, a') = \|F - F'\|_v = \int_0^\infty v(u) |f - f'| (u) du$. La perturbation peut concerner la taille moyenne de réclamations (i.e. $\mu' = \mu + \epsilon$) ou la distribution des montants de réclamations F elle-même.

Remarque 2.4.2. Il n'est pas facile de comparer, en général, la qualité des deux estimations (2.19) et (2.24). Cependant, il est clair que le domaine de validité de l'estimation (2.24) définie par (2.23) est beaucoup plus large que celui donné dans le Theorème 2.4.1 et lié à l'expression (2.19). C'est pourquoi nous proposons, dans la dernière partie, d'effectuer une telle comparaison numériquement.

2.5 Conclusion

Dans ce chapitre, nous avons donné les fondements de la méthode de stabilité forte et un aperçu de son application à l'analyse des modèles de risque et à l'évaluation de la probabilité de ruine. Cette méthode d'approximation sera combinée avec les concepts d'estimation de la densité donnés dans le prochain chapitre pour être utilisés ensemble dans la dernière partie numérique (simulation et application aux données réelles).

Partie

Estimation de la densité de probabilité

Introduction

L'estimation d'une densité de probabilité sous-jacente à un ensemble fini d'observations est un problème fondamental en statistique qui a fait l'objet d'une vaste littérature. On retrouve cette problématique dans de nombreux domaines des sciences et techniques tels que le traitement du signal et des images, la mécanique, la robotique, etc.

On trouve deux types d'approches d'estimation de la densité de probabilité : estimation paramétrique et non paramétrique. L'inconvénient principal de l'approche paramétrique est qu'elle nécessite la connaissance au préalable de la loi du phénomène étudié. L'approche non paramétrique estime la densité à partir de l'information disponible. On dit souvent que dans cette approche les données parlent d'elles mêmes. L'avantage principal de l'estimation non paramétrique de la densité de probabilité est de ne pas nécessiter d'hypothèses a priori sur l'appartenance de cette densité à une famille de lois connues.

Nous proposons dans cette partie la méthode semi-paramétrique qui est une combinaison entre la méthode paramétrique (nous proposons d'utiliser une distribution de Champernowne généralisée) et un estimateur non-paramétrique à noyau asymétrique.

Dans le premier chapitre, une brève description des estimateurs à noyau standard sera faite à la section [3.2.2](#). Certains problèmes des noyaux standards seront discutés et le noyau Beta et ses versions seront alors proposés à la section [3.2.3.5](#). Les propriétés asymptotiques ainsi que le choix de la fenêtre optimale pour un estimateur à noyau Beta seront également traités à la section [3.2.3.5](#).

Dans le deuxième chapitre, nous allons présenter la méthode d'estimation semi-paramétrique proposée, qui sera utilisée ultérieurement, dans la troisième partie, pour estimer la densité des réclamations (fonction de perte).

Estimation non-paramétrique

Sommaire

3.1	Introduction	35
3.2	Estimation par la méthode du noyau	37
3.2.1	Noyau	37
3.2.2	Estimateurs à noyaux continus classiques (symétriques)	38
3.2.3	Estimateurs à noyaux continus asymétriques	42
3.2.4	Choix du paramètre de lissage h_n	49
3.3	Conclusion	53

3.1 Introduction

La théorie de l'estimation est l'une des préoccupations majeures des statisticiens. L'estimation de la densité, par exemple est un sujet qui a donné lieu à un grand nombre de travaux à cause de son application dans divers domaines, comme l'analyse de la régression des séries chronologiques et la théorie de fiabilité.

L'approche non paramétrique prend son sens lorsqu'on ne possède aucune information précise sur la forme et la classe de la vraie densité. Dans cette approche, ce sont les observations qui vont nous permettre de déterminer l'estimation de la densité f . Cette approche sera utilisée puisque l'on ne peut supposer aucune forme de fonction prédéfinie pour f .

Il existe plusieurs méthodes d'estimation d'une densité. Parmi les méthodes non paramétriques, nous citons la méthode des histogrammes (voir Abou-Jouadé [1] et Geffroy [37]), la méthode d'interpolation par les fonctions Splines (voir Wahba [89]), la méthode des séries orthogonales (voir Concov [30], Saadi et Adjabi [80]), etc. Mais la plus utilisée est la méthode du noyau dont nous présentons quelques aspects. Elle a été introduite par Rosenblatt [75] et développée par Parzen [71].

La méthode la plus simple serait d'utiliser un histogramme comme estimation de la densité. Toutefois, un premier problème vient du fait que l'histogramme nous donne une densité qui n'est pas continue partout, alors qu'ici la variable modélisée est continue. On pourrait augmenter le nombre de barres dans l'histogramme, mais alors il y a de moins en moins d'observations dans chaque intervalle. Ce qui rend l'estimation très instable, comme observée à la figure 3.1.

Une autre solution intuitive, pour transformer l'histogramme en une fonction continue, est de relier les points centraux des barres de l'histogramme, comme à la figure 3.2. Ici encore, il pourrait être tentant d'augmenter le nombre de barres de l'histogramme pour réduire l'espace entre les points centraux, mais le problème d'instabilité reste le même. De plus, il faudrait alors s'assurer que la fonction ainsi obtenue intègre à 1 pour que ce soit une densité

Une solution simple et élégante, pour utiliser l'idée de l'histogramme comme estimation de la fonction de densité, est ce que l'on pourrait appeler l'histogramme mobile. L'idée de l'histogramme mobile est que pour estimer la densité au point x , on construit une barre de l'histogramme ayant comme point central le point x et une largeur de $2h_n$. La valeur de la densité estimée évaluée au

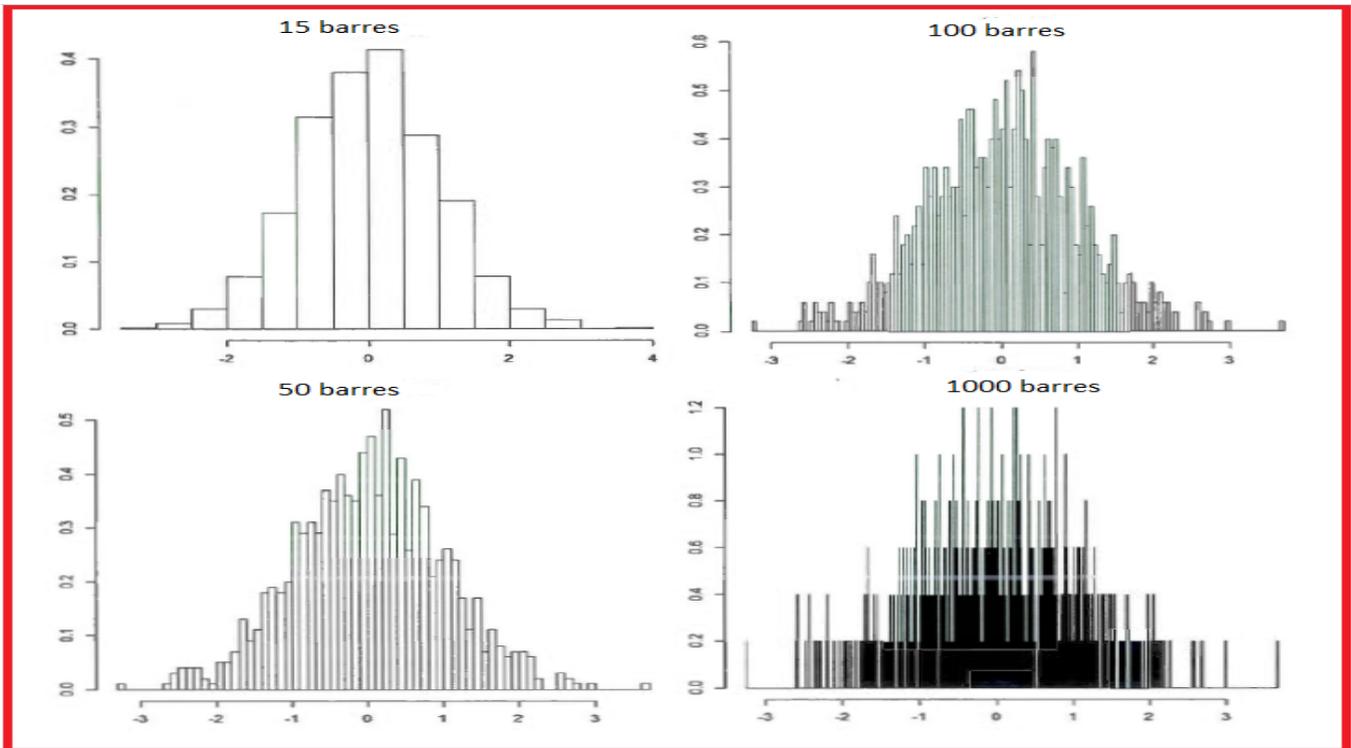


FIG. 3.1: Plusieurs histogrammes avec différents nombres d'intervalles, tous calculés à partir du même échantillon de taille 1000 provenant d'une loi normale de moyenne 0 et de variance 1.

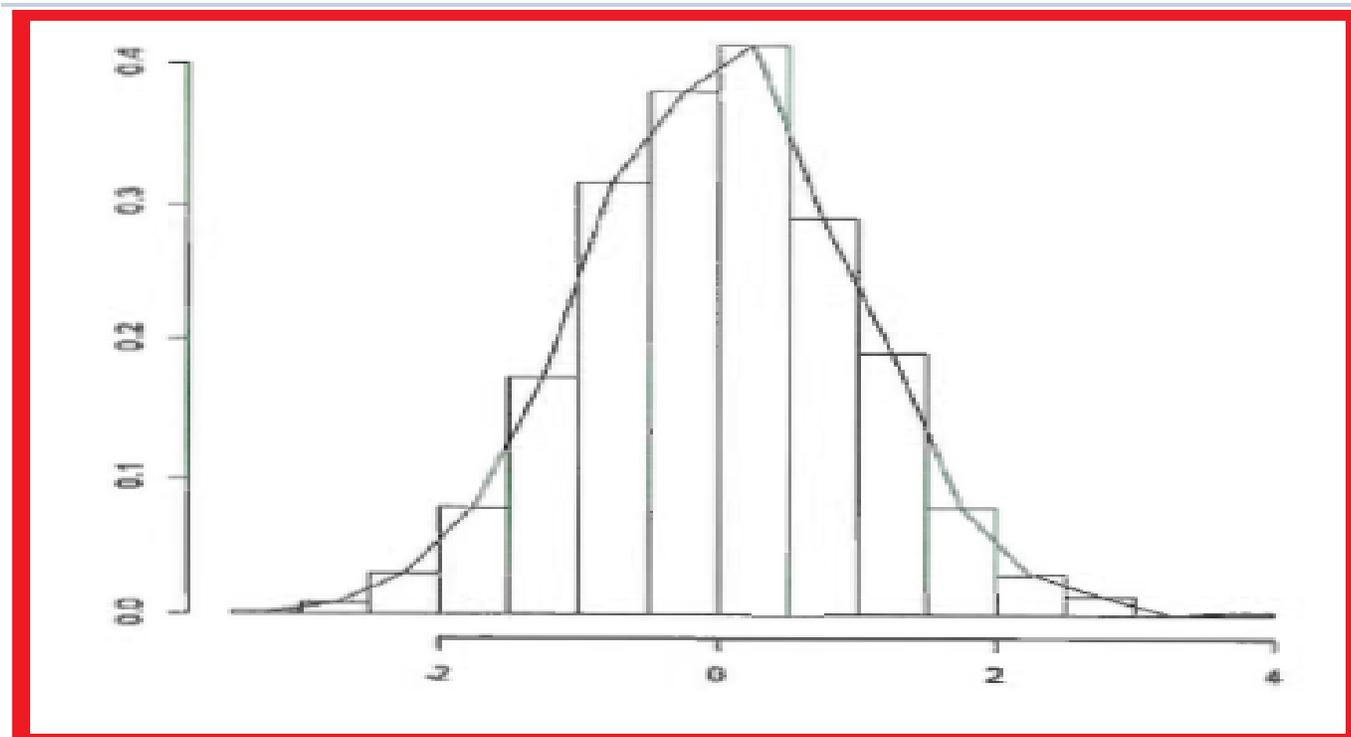


FIG. 3.2: Histogramme calculé à partir d'un échantillon provenant d'une loi normale de moyenne 0 et de variance 1 dont les points centraux des barres sont reliés par une droite.

point x serait donc :

$$\hat{f}(x) = (nh_n)^{-1} \sum_{i=1}^n I_{x-h_n < x_i < x+h_n}, \quad (3.1)$$

où I est la fonction indicatrice et les x_i sont les éléments d'un échantillon et $i = 1, \dots, n$. L'équation (3.1) correspond justement à ce qui est appelé un estimateur à noyau.

Dans le présent chapitre, nous allons présenter la méthode non-paramétrique à noyau symétrique et asymétrique pour l'estimation d'une densité.

3.2 Estimation par la méthode du noyau

Le concept de noyau a d'abord été introduit par Rosenblatt [75], mais c'est Cacoullos [22] qui a été le premier à utiliser le terme "noyau" pour désigner la fonction que l'on utilise dans les méthodes non paramétriques. En hydrologie, c'est Yakowitz et Feluch [92] qui ont introduit indépendamment la méthode du noyau lors d'une conférence de l'AGU (American Geophysical Union) à l'automne 1983.

Dans la méthode du noyau, une fonction K est associée à chaque observation de l'échantillon. La seule véritable restriction concernant le noyau K est que son intégration sur tout le domaine de définition de x doit être égale à un. On rencontre parfois d'autres restrictions théoriques qui sont appliquées à K , comme la symétrie ou la positivité sur tout le domaine de définition du noyau (voir Adamowski [2]). Toutefois, ces restrictions sont surtout introduites afin de simplifier les développements théoriques.

La méthode du noyau est aujourd'hui la méthode d'estimation non paramétrique la plus classique, connue et utilisée pour une fonction de densité inconnue. Dans l'estimation de la densité par la méthode du noyau, la forme de la densité estimée est déterminée par les données et, en principe, étant donné un ensemble de données suffisamment grand, la technique est capable d'estimer une densité arbitraire f de manière assez précise. Cette méthode ne fait aucune hypothèse concernant la forme fonctionnelle globale de la densité sous-jacente (voir Silverman [85] pour une bonne introduction au sujet).

3.2.1 Noyau

Définition 3.2.1. Soit $K : \mathbb{R} \rightarrow \mathbb{R}$, on dit que K est un noyau si et seulement si :

$$\int K(u) du = 1.$$

Noyaux	Support	Densité	Efficacité
Epanechnikov	$[-1, 1]$	$\frac{3}{4}(1-x^2)$	1
Cosinus	$[-1, 1]$	$\frac{\pi}{4} \cos\left(\frac{\pi}{2}x\right)$	0.999
Biweight	$[-1, 1]$	$\frac{15}{16}(1-x^2)^2$	0.994
Triweight	$[-1, 1]$	$\frac{35}{32}(1-x^2)^3$	0.987
Triangulaire	$[-1, 1]$	$1 - x $	0.986
Gaussien	\mathbb{R}	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$	0.946
Uniforme	$[-1, 1]$	$\frac{1}{2}$	0.930
Double Epanechnikov	$[-1, 1]$	$3 x (1- x)$	0.816
Double Exponentielle	\mathbb{R}	$\frac{1}{2} \exp\left\{-\frac{1}{2} x \right\}$	0.759

Tableau 3.1: Exemples des noyaux symétriques.

- K est dit positif si $K(u) \geq 0 \quad \forall u$.
- K est dit symétrique si $K(u) = K(-u) \quad \forall u$.

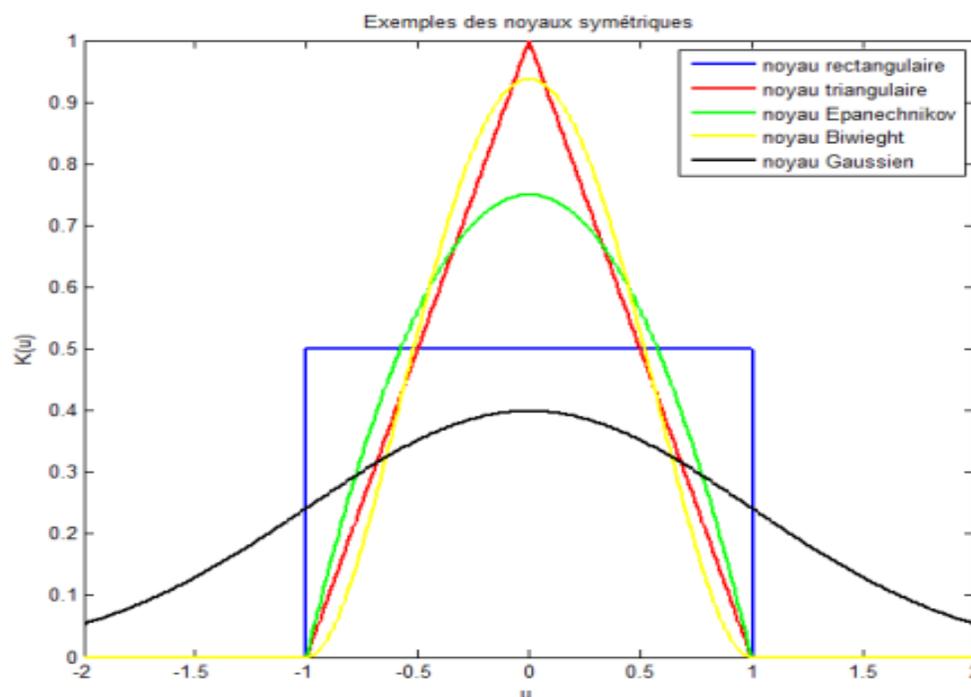


FIG. 3.3: Exemples de noyaux symétriques.

3.2.2 Estimateurs à noyaux continus classiques (symétriques)

L'utilisation des modèles paramétriques dans l'estimation statistique est souvent faite pour des raisons de simplicité seulement. Les modèles statistiques qui expliquent plus profondément les

données sont plus complexes. Les inconnues de ces modèles sont, en général, des fonctions possédant certaines propriétés de régularité. Le problème statistique, dans ce contexte, consiste à estimer, à partir des observations, une fonction inconnue appartenant à une certaine classe fonctionnelle : problème dit de l'estimation non-paramétrique.

Nous présentons, dans cette section, la méthode d'estimation de densité par noyaux continus (classiques). Pour simplifier, on suppose que les X_i sont à valeurs réelles et que f est la densité par rapport à la mesure de Lebesgue sur \mathbb{R} .

La méthode du noyau a été introduite par Rosenblatt [75] dans le cadre de la densité et par Nadaraya [69] dans le cadre de la régression. Elle représente une approche qui fait référence à une classe de techniques générales d'estimation non-paramétrique de fonctions.

La méthode du noyau permet de trouver une structure dans l'ensemble des données, sans l'imposition d'un modèle spécifique. Dans ce cas, seules les observations peuvent suggérer un modèle.

Nous donnons ci-après les définitions d'un noyau continu (classique) et de l'estimateur basé sur ce dernier, respectivement.

Définition 3.2.2. Une fonction K de support \mathfrak{N} est dite noyau si elle est une densité de probabilité symétrique (i.e. $K(-u) = K(u)$), de moyenne μ_K nulle ($\mu_K := \int_{\mathfrak{N}} uK(u) du = 0$), de variance σ_K^2 finie ($\sigma_K^2 := \int_{\mathfrak{N}} u^2 K(u) du < +\infty$) et de carré intégrable ($\int_{\mathfrak{N}} K^2(u) du < +\infty$).

Précisons ici qu'en tant que densité de probabilité, le noyau K est positif et de masse totale égale à 1 (i.e. pour tout élément u de \mathfrak{N} , $K(u) \geq 0$ et $\int_{\mathfrak{N}} K(u) du = 1$).

Définition 3.2.3. Soit $h_n > 0$ la fenêtre de lissage et K la fonction noyau vérifiant la Définition 3.2.2. L'estimateur à noyau continu (classique) de f est défini en un point $x \in \mathbb{R}$ par :

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right). \quad (3.2)$$

De ce fait, on peut considérer l'estimation par noyau en un point donné $x \in \mathbb{R}$ comme la moyenne des valeurs des fonctions noyaux centrés sur chaque observation X_i de l'échantillon.

Dans l'expression (3.2), le paramètre de lissage h_n a une influence assez évidente. Il est largement admis (et partiellement démontré) que le choix de h_n est d'une importance capitale dans le processus. Plus il tend vers 0, la fonction noyau s'apparente à la fonction de Dirac, plus on a d'informations (sur lissage); tandis que, plus h_n devient grand, plus on manque d'informations (sous lissage). Ainsi, l'obtention d'une estimation plus proche de la réalité implique donc un choix optimal de h_n .

Après la présentation de l'estimateur à noyau symétrique, nous donnons quelques unes de ses propriétés.

3.2.2.1 Propriétés de l'estimateur à noyau symétrique

Juste après l'introduction de l'estimateur à noyau de la densité par Rosenblatt [75], Parzen [71] à étudié ses propriétés fondamentales. Depuis, cet estimateur est devenu un objet classique étudié par les statisticiens. Les expressions de l'espérance, du biais et de la variance de l'estimateur à noyau sont données respectivement par :

Esperance de l'estimateur

L'espérance mathématique de l'estimateur à noyau est définie par :

$$\mathbb{E} \left[\hat{f}_n(x) \right] = \mathbb{E} \left[\frac{1}{nh_n} \sum_{i=0}^n K \left(\frac{x - X_i}{h_n} \right) \right] = f(x) + \frac{h_n^2}{2} f''(x) \mu_2(K) + o(h_n^2), \quad (3.3)$$

où $\mu_2 = \int_{\mathbb{R}} u^2 K(u) du$.

Biais de l'estimateur

Le biais de l'estimateur \hat{f}_n est :

$$\text{Biais} \left[\hat{f}_n(x) \right] = f(x) + \frac{h_n^2}{2} f''(x) \mu_2(K) + o(h_n^2) - f(x) = \frac{h_n^2}{2} f''(x) \mu_2(K) + o(h_n^2).$$

Variance de l'estimateur

Soit x fixé dans \mathbb{R} . La variance de l'estimateur \hat{f}_n est définie par :

$$\text{Var} \left[\hat{f}_n(x) \right] = \text{var} \left[\frac{1}{nh_n} \sum_{i=0}^n K \left(\frac{x - X_i}{h_n} \right) \right] = \frac{f(x)}{n} R(K) + o\left(\frac{1}{nh_n}\right),$$

où $R(K) = \int_{\mathbb{R}} K^2(y) dy$.

3.2.2.2 Critères d'erreur

Lorsqu'on définit l'estimateur à noyau, il y a un certain nombre de critères qui permettent d'évaluer la similarité de l'estimateur \hat{f}_n par rapport à la vraie densité f à estimer. Parmi les nombreux critères proposés dans la littérature nous citons :

Erreur Quadratique Moyenne

L'Erreur Quadratique Moyenne ponctuelle en anglais "pointwise Mean Squared Error" (MSE) est définie par :

$$MSE(x) = \mathbb{E} \left[\left\{ \hat{f}_n(x) - f(x) \right\}^2 \right], \quad x \in \mathbb{R}. \quad (3.4)$$

Il est intéressant de développer cette expression pour y faire apparaître la variance et le biais ponctuels de l'estimateur $\hat{f}_n(x)$ notés $Var \left\{ \hat{f}_n(x) \right\}$ et $Biais \left\{ \hat{f}_n(x) \right\}$ définis respectivement en (3.5) et (3.6) pour tout point $x \in \mathbb{R}$ par :

$$Var \left\{ \hat{f}_n(x) \right\} = \mathbb{E} \left(\left[\hat{f}_n(x) - \mathbb{E} \left\{ \hat{f}_n(x) \right\} \right]^2 \right), \quad (3.5)$$

$$Biais \left\{ \hat{f}_n(x) \right\} = \mathbb{E} \left\{ \hat{f}_n(x) \right\} - f(x). \quad (3.6)$$

Ceci conduit à une nouvelle expression du $MSE(x)$ comme suit

$$\begin{aligned} MSE(x) &= \mathbb{E} \left[\left\{ \hat{f}_n(x) \right\}^2 \right] - 2\mathbb{E} \left\{ \hat{f}_n(x) \right\} f(x) + [\mathbb{E} \left\{ f(x) \right\}]^2, \\ &= \mathbb{E} \left[\left\{ \hat{f}_n(x) \right\}^2 \right] - \left[\mathbb{E} \left\{ \hat{f}_n(x) \right\} \right]^2 + \left[\mathbb{E} \left\{ \hat{f}_n(x) \right\} \right]^2 - 2\mathbb{E} \left\{ \hat{f}_n(x) \right\} f(x) + f^2(x), \\ &= Var \left\{ \hat{f}_n(x) \right\} + Biais^2 \left\{ \hat{f}_n(x) \right\}. \end{aligned}$$

Erreur Quadratique Moyenne Intégrée (MISE)

Propriété 3.2.1. *L'Erreur Quadratique Moyenne Intégrée ("Mean Integrated Squared Error") est la mesure théorique commune la plus utilisée pour évaluer l'erreur entre la fonction f et \hat{f}_n . Il est également convenable d'évaluer l'erreur globale sur le support \mathbb{R} de cet estimateur.*

$$MISE(\hat{f}_n) = \int_{\mathbb{R}} MSE(\hat{f}_n) dx = \frac{1}{nh_n} R(K) \frac{h_n^4}{4} R(f''(x)) \mu_2^2(K) + o\left(h_n^4 + \frac{1}{nh_n}\right),$$

$$\text{où } R(K) = \int_{\mathbb{R}} K^2(y) dy \text{ pour une fonction } K \text{ carrée intégrable, } \mu_2 = \int_{\mathbb{R}} u^2 K(u) du.$$

Pour plus de détails voir Silverman [85].

MISE asymptotique

La définition 3.2.2 et l'expression (3.3), nous permettent de trouver une expression asymptotique pour le $MISE$, définie par

$$AMISE(\hat{f}_n) = \frac{1}{nh_n} R(K) + \frac{h_n^4}{4} R(f''(x)) \mu_2^2(K),$$

$$\text{où } R(K) = \int_{\mathbb{R}} K^2(y) dy \text{ pour une fonction } K \text{ de carré intégrable, } \mu_2 = \int_{\mathbb{R}} u^2 K(u) du.$$

3.2.2.3 Choix du noyau

Pour mesurer l'efficacité d'un noyau symétrique on peut calculer le rapport de AMISE des 2 noyaux

$$eff(K_1, K_2) = \frac{AMISE(K_1, n, h_n)}{AMISE(K_2, n, h_n)} < 1.$$

Le choix du noyaux n'influe pas trop dans le cas de noyaux symétriques.

Plusieurs résultats sont connus dans la littérature lors de l'estimation d'une fonction de densité définie sur \mathbb{R} (voir [75, 85]) pour le choix du noyau. Lorsque cette dernière est définie sur un support borné, utiliser un noyau symétrique n'est pas approprié pour l'estimer puisque les effets du biais de bord sont présents. Ce problème est le résultat de l'allocation de poids par le noyau symétrique fixe en dehors du support de la distribution, lorsque le lissage près du bord est effectué. Pour résoudre ce problème, plusieurs méthodes ont été introduites récemment (voir [18], [27], [83]). Schuster [83] suggère de créer l'image miroir des données de l'autre côté du bord, puis appliquer l'estimateur (3.2) pour l'ensemble des données initiales et leur image miroir .

Une autre idée simple pour éviter les effets de bord dans l'estimation non paramétrique de la densité consiste en l'utilisation d'un noyau flexible, qui n'attribue jamais un poids hors du support de la fonction de densité et qui corrige automatiquement et implicitement les effets de bord. Une catégorie de noyaux flexibles consiste en des noyaux asymétriques (voir [18], [27], [28], [82], [66]).

3.2.3 Estimateurs à noyaux continus asymétriques

Si le support de f est non borné (par exemple $[0, +\infty[$), les noyaux symétriques donnent une estimation de plus en plus biaisée lorsque x est proche de 0. D'où le problème de biais au bord. Plusieurs auteurs ont proposé d'utiliser des noyaux asymétriques.

Dans cette section, nous présentons l'estimateur à noyau associé continu asymétrique dans le cas univarié. Cet estimateur est approprié pour estimer des densités à support compact ou borné d'un coté. Nous allons traiter quatres noyaux différents : Beta, Gamma, Gaussien inverse réciproque (RIG) et Noyau Birnbaum-Saunders (BS). Pour de récentes références nous pouvons consulter Chen [27, 28], Kokonendji et Kiessé [59] et Scaillet [82]. Nous montrons les propriétés élémentaires telles que Biais, Variance et MISE. Ensuite nous déterminons les fenêtres optimales pour chaque noyau associé considéré et l'erreur en fonction de ces valeurs.

Soit X_1, \dots, X_n un échantillon de variables aléatoires i.i.d. de densité de probabilité continue f à support $\mathfrak{N} = [a, b]$, avec $a \in \mathbb{R}$ et $b \in \bar{\mathbb{R}}$ (\mathfrak{N} est par exemple le support $[0, 1]$ ou $[0, +\infty[$).

De manière générale, l'estimateur à noyau continu est de la forme suivante :

$$\begin{aligned}\hat{f}_n(x) &= \frac{1}{n} \sum_{i=1}^n K_{x,h_n}(X_i) \\ &= \hat{f}_{n,h_n,K}(x),\end{aligned}\tag{3.7}$$

où x est fixé dans \mathfrak{N} , K_{x,h_n} est la fonction "noyau associé" et h_n est un réel strictement positif appelé paramètre de lissage.

Dans le cas où K_{x,h_n} est associé à un noyau continu symétrique, il vérifie :

$$K_{x,h_n}(\cdot) = \frac{1}{h_n} K\left(\frac{x - \cdot}{h_n}\right).$$

Dans le cas purement asymétrique, K_{x,h_n} est un noyau variable en fonction de la cible x (point d'estimation). Il change de forme chaque fois que x varie dans \mathfrak{N} .

Définition 3.2.4. Soit x dans \mathfrak{N} et $h_n > 0$. Nous appelons "noyau associé continu" K_{x,h_n} toute densité de probabilité d'une variable aléatoire \mathcal{X}_{x,h_n} sur le support \mathfrak{N}_{x,h_n} tels que :

$$\mathfrak{N}_{x,h_n} \cap \mathfrak{N} \neq \emptyset\tag{3.8}$$

$$\cup_x \mathfrak{N}_{x,h_n} \supseteq \mathfrak{N}\tag{3.9}$$

$$\mathbb{E}(\mathcal{X}_{x,h_n}) \sim x \quad \text{quand } h_n \rightarrow 0\tag{3.10}$$

$$\text{Var}(\mathcal{X}_{x,h_n}) < \infty\tag{3.11}$$

$$\text{Var}(\mathcal{X}_{x,h_n}) \rightarrow 0 \quad \text{quand } h_n \rightarrow 0.\tag{3.12}$$

Commentaire : La relation (3.8) traduit le fait que l'intersection entre le support des observations et le support du noyau associé continu asymétrique doit contenir au moins un élément. Pour h_n fixé, quand x parcourt \mathfrak{N} , le support \mathfrak{N}_{x,h_n} change, l'expression (3.9) suppose que \mathfrak{N} doit être toujours contenu dans la réunion des \mathfrak{N}_{x,h_n} . La condition (3.10) permet d'assurer la convergence ponctuelle de l'estimateur ; elle met en évidence que le noyau K_{x,h_n} est un noyau variable ou adaptatif à la cible x . Enfin la relation (3.12) assure la convergence de la variance de la variable aléatoire du noyau associé et va nous servir dans les calculs à suivre.

Nous allons montrer la technique de construction des noyaux associés asymétriques à partir de n'importe quel type de noyau.

3.2.3.1 Propriétés asymptotiques

Proposition 3.2.1. *Le Biais et la Variance de l'estimateur défini dans la formule (3.7) sont donnés par*

$$\begin{aligned} \text{Biais} [\hat{f}_{h_n}(x)] &= f[\mathbb{E}(\mathcal{X}_{x,h_n})] + \frac{1}{2} \text{Var}(\mathcal{X}_{x,h_n}) f''(x) f(x) + o(h_n^2). \\ \text{Var} [\hat{f}_{h_n}(x)] &= \frac{1}{n} \int_{\mathfrak{X}} K_{x,h_n}^2(u) f(u) du - \frac{1}{n} \left\{ \text{Biais} [\hat{f}(x)] + f(x) \right\}^2. \end{aligned}$$

Tel que, le Biais de \hat{f}_{h_n} peut s'écrire comme

$$\begin{aligned} \text{Biais} [\hat{f}_{h_n}(x)] &= \mathbb{E} [\hat{f}_{h_n}(x)] - f(x) \\ &= \mathbb{E} [K_{x,h_n}(X)] - f(x) \\ &= \int_{\mathfrak{X}} K_{x,h_n}(u) f(u) du - f(x) \\ &= \mathbb{E} [f(\mathcal{X}_{x,h_n})] - f(x). \end{aligned}$$

Le développement limite de Taylor de $f(\mathcal{X}_{x,h_n})$ au point moyen $\mathbb{E}(\mathcal{X}_{x,h_n}) = m_{x,h_n}$ est :

$$f(\mathcal{X}_{x,h_n}) = f(m_{x,h_n}) + (\mathcal{X}_{x,h_n} - m_{x,h_n}) f'(m_{x,h_n}) + \frac{1}{2} (\mathcal{X}_{x,h_n} - m_{x,h_n})^2 f''(m_{x,h_n}).$$

En développant l'espérance $\mathbb{E}[f(\mathcal{X}_{x,h_n})]$, on obtient

$$\begin{aligned} \text{Biais} [\hat{f}_{h_n}(x)] &= f(m_{x,h_n}) \frac{1}{2} \mathbb{E}[(\mathcal{X}_{x,h_n} - m_{x,h_n})^2] f''(x) - f(x) + o(h_n^2) \\ &= f[\mathbb{E}(\mathcal{X}_{x,h_n})] + \frac{1}{2} \text{Var}(\mathcal{X}_{x,h_n}) f''(x) - f(x) + o(h_n^2). \end{aligned}$$

D'un autre côté, la Variance peut s'exprimer comme

$$\begin{aligned} \text{Var} [\hat{f}_{h_n}(x)] &= \left[\frac{1}{n} \sum_{i=1}^n K_{x,h_n}(X_i) \right] \\ &= \frac{1}{n} \text{Var} [K_{x,h_n}(X)] \\ &= \frac{1}{n} \left\{ \mathbb{E} [K_{x,h_n}^2(X)] - \mathbb{E}^2 [K_{x,h_n}(X)] \right\} \\ &= \frac{1}{n} \int_{\mathfrak{X}} K_{x,h_n}^2(u) f(u) du - \frac{1}{n} \left\{ \text{Biais} [\hat{f}(x)] + f(x) \right\}^2. \end{aligned}$$

$$\text{Var} [\hat{f}_{h_n}(x)] \rightarrow 0 \text{ quand } \int_{\mathfrak{X}} K_{x,h_n}^2(u) f(u) du \rightarrow 0.$$

Les expressions de l'erreur quadratique moyenne ($MSE [\hat{f}_{h_n}(x)]$) et l'erreur quadratique moyenne intégrée ($MISE [\hat{f}_{h_n}(x)]$) sont données par la proposition suivante.

Proposition 3.2.2.

$$\begin{aligned}
MSE \left[\hat{f}_{h_n}(x) \right] &= \text{Biais}^2 \left[\hat{f}_{h_n}(x) \right] + \text{Var} \left[\hat{f}_{h_n}(x) \right] \\
&= \left\{ f \left[\mathbb{E}(\mathcal{X}_{x,h_n}) \right] + \frac{1}{2} \text{Var}(\mathcal{X}_{x,h_n}) f''(x) - f(x) \right\}^2 \\
&\quad + \frac{1}{2} \int_{\mathfrak{N}} K_{x,h_n}^2(u) f(u) du - \frac{1}{n} \left\{ \text{Biais} \left[\hat{f}(x) \right] + f(x) \right\}^2.
\end{aligned}$$

$$\begin{aligned}
MISE \left[\hat{f}_{h_n}(x) \right] &= \int_{\mathfrak{N}} MSE \left[\hat{f}_{h_n}(x) \right] \\
&= \int_{\mathfrak{N}} \left\{ f \left[\mathbb{E}(\mathcal{X}_{x,h_n}) \right] + \frac{1}{2} \text{Var}(\mathcal{X}_{x,h_n}) f''(x) - f(x) \right\}^2 dx \\
&\quad + \frac{1}{2} \int_{\mathfrak{N}} \left\{ \int_{\mathfrak{N}} K_{x,h_n}^2(u) f(u) du - \frac{1}{n} \left\{ \text{Biais} \left[\hat{f}(x) \right] + f(x) \right\}^2 \right\} dx.
\end{aligned}$$

3.2.3.2 Convergences de l'estimateur à noyau asymétrique

Théorème 3.2.1. (cf. Kokonendji et al. [60]).

Soit f une densité continue sur son support \mathfrak{N} , et \hat{f}_{h_n} son estimateur à noyau associé. On suppose qu'il existe un nombre réel $r_0 = r_0(K) > 0$ dépendant du type de noyau K tel que, pour tout $x \in \mathfrak{N}$,

$$h_n^{r_0} \int_{\mathfrak{N}_{\cap, \simeq}} |K_{x,h_n}(u)| du \leq c_0(x),$$

avec, $c_0(x)$ constante positive et bornée sur tout compact $I \subseteq \mathfrak{N}$ contenant x . Pour tout sous ensemble compact I de \mathfrak{N} , on a :

- ★ Si $\lim_{n \rightarrow \infty} nh_n^{2r_0} = +\infty$ alors $\sup_{x \in I} \left| \hat{f}_{h_n}(x) - f(x) \right| \xrightarrow{P} 0$ quand $n \rightarrow \infty$.
- ★ Si $\lim_{n \rightarrow \infty} \frac{nh_n^{2r_0}}{\log n} = +\infty$ alors $\sup_{x \in I} \left| \hat{f}_{h_n}(x) - f(x) \right| \xrightarrow{P,S} 0$ quand $n \rightarrow \infty$.

Proposition 3.2.3. (Zougab et al. [96])

Si $\frac{1}{n} \int_{\mathbb{R}_+} K_{x,h_n}^2(u) f(u) du \rightarrow 0$ et $h_n \rightarrow 0$ quand $n \rightarrow \infty$, alors

$$\lim_{n \rightarrow \infty} MSE \left[\hat{f}_{h_n}(x) \right] = 0 \quad \forall x > 0.$$

Proposition 3.2.4. (Zougab et al. [96])

Si $\int_{\mathbb{R}_+} \left[\int_{\mathbb{R}_+} K_{x,h_n}^2(u) f(u) du \right] dx \rightarrow 0$ et $h_n \rightarrow 0$ quand $n \rightarrow \infty$, alors

$$\lim_{n \rightarrow \infty} MISE \left[\hat{f}_{h_n} \right] = 0.$$

3.2.3.3 Quelques estimateurs à noyau asymétrique de la densité

Nous présentons dans cette section les noyaux asymétriques utiles dans notre travail, tel que le noyau Beta, Gamma (GAM), réciproque-inverse gaussien (RIG) et Birnbaum-Saunders (BS).

3.2.3.4 Construction des noyaux associés continus par la méthode mode-dispersion

Principe de construction

Etant donné un type de noyau $K_{\theta(a,b)}$ sur $\aleph_{\theta(a,b)}$, uni-modal de mode $M(a,b)$ et d'un paramètre de dispersion $D(a,b)$, la méthode Mode-Dispersion permet la construction d'une fonction $K_{\theta(x,h_n)}$, dépendant de x et h_n en résolvant en a et b le système :

$$\begin{cases} M(a,b) = x \\ D(a,b) = h_n. \end{cases} \quad (3.13)$$

Les solutions $a = a(x, h_n)$ et $b = b(x, h_n)$ du système (3.13) permettent d'exprimer $\theta(a,b)$ en fonction de x et h_n . On obtient alors $\theta(a,b) = \theta(a(x, h_n), b(x, h_n))$ que nous désignons simplement par $\theta(x, h_n)$, en remplaçant $\theta(a,b)$ par $\theta(x, h_n)$ puis a et b par leurs valeurs $a(x, h_n)$ et $b(x, h_n)$ dans les expressions des moyenne et variance du type de noyau $K_{\theta(a,b)}$.

Nous montrons dans la proposition suivante que la fonction noyau $K_{\theta(x,h_n)}$ issue de la méthode mode-dispersion est bien un noyau associé.

Proposition 3.2.5. *Soit \aleph le support de la densité f , à estimer. Pour tout $x \in \aleph$ et $h_n > 0$, la fonction noyau $K_{\theta(x,h_n)}$ construite par la méthode mode-dispersion, de support $S_{\theta(x,h_n)} = S_{\theta(a(x,h_n), b(x,h_n))}$, vérifie :*

$$x \in S_{\theta(x,h_n)},$$

$$\mathbb{E}(Z_{\theta(x,h_n)} - x) = A_{\theta}(x, h_n),$$

$$\text{Var}(Z_{\theta(x,h_n)}) = B_{\theta}(x, h_n),$$

où $Z_{\theta(x,h_n)}$ est une variable aléatoire de densité $K_{\theta(x,h_n)}$ et $A_{\theta}(x, h_n)$ et $B_{\theta}(x, h_n)$ tendent vers 0 quand h_n tend vers 0.

3.2.3.5 Noyau Beta

Le noyau Beta a été proposé par Chen [27, 28], pour estimer la densité à support compact $[0, 1]$ et pour régler ainsi le problème du biais aux bornes.

La première classe prend la forme suivante

$$K_{\beta_1} \left(\frac{x}{h_n}, \frac{(1-x)}{h_n} + 1 \right) (u) = \frac{u^{h_n-1} (1-u)^{x-1}}{B \left(\frac{x}{h_n}, \frac{(1-x)}{h_n} \right)}, \quad u \in [0, 1], \quad (3.14)$$

où $B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt$, ($\alpha > 0$ et $\beta > 0$) est la fonction Beta et h_n est le paramètre de lissage satisfaisant les conditions $h_n \rightarrow 0$ et $nh_n \rightarrow \infty$ quand $n \rightarrow \infty$. L'estimateur à noyau Beta défini dans (3.14) est donné par :

$$\hat{f}_{h_n}(x) = \frac{1}{n} \sum_{i=1}^n K_{\beta_1} \left(\frac{x}{h_n}, \frac{(1-x)}{h_n} \right) (X_i). \quad (3.15)$$

La forme de cet estimateur présente deux avantages par rapport aux noyaux classiques. D'abord, le paramètre forme du noyau varie naturellement en fonction du point d'estimation de la densité, ce qui permet à l'estimateur à noyau Beta d'être souple et adaptatif. L'autre avantage est que le support du noyau Beta épouse celui de la densité à estimer permettant ainsi de concentrer la masse sur les points du bord du support. Pour réduire encore le biais du bord et pour des raisons de convergence uniforme sur tout l'intervalle $[0, 1]$ de l'estimateur défini ci-haut, Chen [27] a proposé une autre version du noyau Beta, qu'on appellera par la suite Beta modifié ou Beta2, qui accentue encore plus le lien entre le paramètre forme du noyau, et le point d'estimation de la densité. Ce noyau est donné par l'équation (3.17) afin de réduire le biais de $\hat{f}_{h_n}(x)$. L'estimateur à noyau Beta modifié ou Beta2 est

$$\hat{f}_{h_n}(x) = \frac{1}{n} \sum_{i=1}^n K_{\beta_2}(X_i), \quad (3.16)$$

où K_{β_2} est le noyau Beta borné défini comme suit

$$K_{\beta_2}(u) = \begin{cases} K_{\beta_1} \left(\frac{x}{h_n}, \frac{1-x}{h_n} \right) (u) & \text{si } x \in [2h_n, 1 - 2h_n]; \\ K_{\beta_1} \left(\rho_{h_n}(x), \frac{1-x}{h_n} \right) (u) & \text{si } x \in [0, 2h_n]; \\ K_{\beta_1} \left(\frac{x}{h_n}, \rho_{h_n}(1-x) \right) (u) & \text{si } x \in (1 - 2h_n, 1]; \end{cases} \quad (3.17)$$

où $\rho_{h_n} = 2h_n^2 + 2.5 - \sqrt{4h_n^4 + 6h_n^2 + 2.25 - x^2 - \frac{x}{h_n}}$.

Notons que pour chaque h_n fixé, $\rho_{h_n}(x)$ est une fonction monotone croissante pour $x \in [0, 2h_n]$. Le noyau Beta a deux avantages, premièrement il peut parfaitement estimer les densités à support compact et deuxièmement il possède une forme flexible qui change le lissage dans le sens naturel quand on s'éloigne des bornes. Par conséquent, le noyau Beta élimine le biais aux bornes et fournit une réduction de la variance.

$K_{\beta_1}(\alpha, \beta)(0) = K_{\beta_1}(\alpha, \beta)(1) = 0$. Donc s'il y a une masse ponctuelle aux points 0 ou 1, l'estimateur devient inconsistant, c'est-à-dire

$$\begin{aligned}\hat{f}_{h_n}(x) &= \frac{1}{n} \sum_{i=1}^n K_{\beta_1} \left(1 + \frac{x}{h_n}, 1 + \frac{1-x}{h_n} \right) (X_i), \quad x \in [0, 1] \\ &\approx (1 - \mathbb{P}(x=0) - \mathbb{P}(x=1)) \cdot f_0(x), \quad x \in [0, 1];\end{aligned}$$

Pour éviter le problème d'inconsistance de cet estimateur, Gouriéroux et Montfort [38] ont proposé d'utiliser des versions normalisées de la distribution Beta et ont obtenu l'estimateur suivant

$$\hat{f}_{h_n}^1(x) = \frac{\hat{f}_{h_n}(x)}{\int_0^1 \hat{f}_{h_n}(t) dt}, \quad \text{pour tout } x \in [0, 1]. \quad (3.18)$$

Il est appelé estimateur macro-Beta (Mac- β) car la correction est effectuée de manière globale. Les mêmes auteurs proposent également

$$\hat{f}_{h_n}^2(x) = \frac{1}{n} \sum_{i=1}^n \frac{K_{\beta}(x, h_n)(X_i)}{\int_0^1 K_{\beta}(t, h_n)(X_i) dt}, \quad \text{pour tout } x \in [0, 1]. \quad (3.19)$$

Il est appelé estimateur micro-Beta (Mic- β) car la correction est effectuée localement. Le noyau Beta K_{β} est soit de type 1, K_{β_1} , soit de type 2, K_{β_2} .

3.2.3.6 Noyau Gamma (GAM)

Ce noyau associé a été proposé par Chen [28] dans le but d'estimer des densités définies sur $[0, +\infty[$. Il est défini comme suit

$$K_{GAM} \left(\frac{x}{h_n} + 1, h_n \right) (u) = \frac{u^{\frac{x}{h_n}} \exp\left(\frac{-u}{h_n}\right)}{h_n^{\frac{x}{h_n} + 1} \Gamma\left(\frac{x}{h_n} + 1\right)}, \quad u > 0, \quad (3.20)$$

où $\Gamma(\alpha) = \int_0^{\infty} \exp(-t) t^{\alpha-1} dt$, $\alpha > 0$ est la fonction Gamma et h_n est le paramètre de lissage satisfaisant les conditions $h_n \rightarrow 0$ et $nh_n \rightarrow \infty$ quand $n \rightarrow \infty$.

L'estimateur à noyau Gamma (GAM) est donné par

$$\hat{f}_{GAM}(x) = \frac{1}{n} \sum_{i=1}^n K_{\frac{x}{h_n} + 1, h_n}(X_i). \quad (3.21)$$

3.2.3.7 Noyau réciproque-inverse gaussien (RIG)

Le noyau réciproque-inverse gaussien (RIG) a été introduit par Scaillet [82] pour estimer des densités à support $]0, +\infty[$. Il est donné par

$$K_{RIG}\left(\frac{1}{x-h_n}, \frac{1}{h_n}\right)(u) = \frac{1}{\sqrt{2\pi h_n u}} \exp\left(\frac{-(x-h_n)}{2h_n} \left(\frac{u}{x-h_n} - 2 + \frac{x-h_n}{u}\right)\right), \quad u > 0. \quad (3.22)$$

L'estimateur à noyau réciproque-inverse gaussien (RIG) est donné par

$$\hat{f}_{RIG}(x) = \frac{1}{n} \sum_{i=1}^n K_{RIG}\left(\frac{1}{x-h_n}, \frac{1}{h_n}\right)(X_i). \quad (3.23)$$

3.2.3.8 Noyau Birnbaum-Saunders (BS)

Le noyau BS a été introduit par Jin et Kawczak [43] qui est basé sur les distributions BS. Il est donné par

$$K_{BS}\left(\frac{1}{h_n^2}, x\right)(u) = \frac{1}{2\sqrt{2\pi h_n}} \left(\sqrt{\frac{1}{xu}} + \sqrt{\frac{x}{u^3}}\right) e^{-\frac{1}{2h_n} \left(\frac{u}{x} - 2 + \frac{x}{u}\right)}, \quad u > 0 \quad (3.24)$$

L'estimateur à noyau Birnbaum-Saunders (BS) est donné par

$$\hat{f}_{BS}(x) = \frac{1}{n} \sum_{i=1}^n K_{BS}\left(\frac{1}{h_n^2}, x\right)(X_i). \quad (3.25)$$

3.2.4 Choix du paramètre de lissage h_n

Le paramètre de lissage h_n est le second élément de la méthode d'estimation à noyau. Ce paramètre est indispensable pour la convergence de l'estimateur à noyau et donc pour l'efficacité du lissage et la qualité de l'estimation (voir figure 3.4).

En pratique, le choix de ce paramètre est crucial dans l'estimation de densité par noyau, qui contrôle le lissage de l'estimateur à noyau (3.2). Plusieurs méthodes pour sélectionner la fenêtre h_n ont été proposées dans la littérature et quelques études comparatives entre elles ont été effectuées.

Il y a essentiellement deux approches pour trouver une largeur de fenêtre optimale. La première consiste à trouver le paramètre qui minimise l'erreur quadratique moyenne de $\hat{f}_n(x)$, c'est-à-dire

$$\arg \min \left[\mathbb{E} \left(\hat{f}_n(x) - f(x) \right)^2 \right].$$

On obtient donc un paramètre de lissage optimal, qui varie en fonction du x où l'on veut estimer la fonction de densité f .

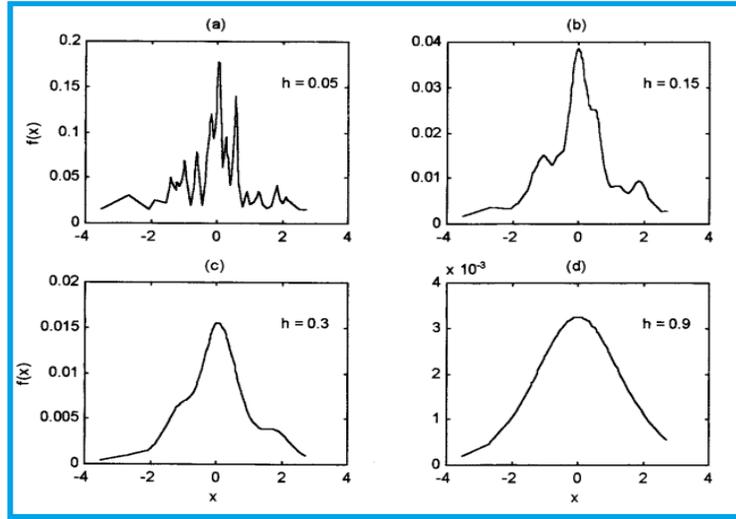


FIG. 3.4: L'influence du paramètre de lissage h_n sur la qualité de l'estimation.

La seconde approche nous donne un paramètre de lissage optimal global, qui ne dépend pas de x . Pour ce faire, on cherche le h_n qui minimise l'erreur quadratique moyenne intégrée (MISE), c'est-à-dire

$$h_{n_{optimal}} = \arg \min [MISE(n, h_n)] = \arg \min \left[\int_{\mathbb{R}} \mathbb{E} \left\{ \hat{f}(x) - f(x) \right\}^2 dx \right].$$

On suppose que la densité à estimer f et le noyau K sont des fonctions de carré intégrable, de sorte que le MISE est fini.

$$MISE(n, h_n) = \frac{h^4}{4} R(f'') (\mu_2(K))^2 + \frac{1}{nh_n} R(K) + o\left(\frac{1}{nh_n} + h^4\right).$$

L'approximation asymptotique du MISE est donné par

$$AMISE(n, h_n) = \frac{h^4}{4} R(f'') (\mu_2(K))^2 + \frac{1}{nh_n} R(K).$$

On dérive ce AMISE par rapport à h_n et on égale à 0, on obtient

$$\frac{\partial}{\partial h_n} AMISE(n, h_n) = h_n^3 R(f'') (\mu_2(K))^2 - \frac{R(K)}{nh_n^2}.$$

$$\text{Si } \frac{\partial}{\partial h_n} AMISE(n, h_n) = 0$$

$$\implies h_{n_{optimal}} = \left[\frac{R(K)}{R(f'') (\mu_2(K))^2} \right]^{\frac{1}{5}} n^{-\frac{1}{5}}.$$

Par ailleurs, plusieurs auteurs ont regroupé les méthodes de sélection du paramètre de lissage en deux classes : la première classe est basée sur des techniques de validation croisée (voir Bowman

[17] et Rudemo [77]) et la deuxième classe est basée sur les méthodes de réinjection (plug-in) (voir Silverman [85]). Une autre méthode utilisant l'approche bayésienne (voir Zougab et al. [96]), est appliquée lorsque la taille de l'échantillon est modérée. Pour des synthèses sur le choix du paramètre de lissage, le lecteur est renvoyé à Jones et al. [45] et à Loader [64].

On présente ci-après les principales méthodes des deux classes citées ci-dessus.

3.2.4.1 Première classe (validation croisée)

La majorité des méthodes de cette classe ont été proposées avant 1990. Dans ce qui suit, on donne une présentation de quelques méthodes les plus connues.

1. La méthode de validation croisée par moindres carrés (*UCV*)

Bowman [17] et Rudemo [77] ont donné l'idée de cette méthode. En utilisant la formule développée de l'erreur quadratique intégrée $ISE(h_n)$, on choisit le paramètre de lissage h_n qui minimise cette erreur.

$$\begin{aligned} ISE(h_n) &= \int_{\mathbb{R}} (f(x) - f_{h_n}(x))^2 dx \\ &= \int_{\mathbb{R}} f_{h_n}^2(x) dx - 2 \int_{\mathbb{R}} f_{h_n}(x) f(x) dx + \int_{\mathbb{R}} f^2(x) dx. \end{aligned} \quad (3.26)$$

On remarque que le troisième terme de la formule (3.26) ne dépend pas de h_n , donc on peut choisir le h_n de façon à ce qu'il minimise le critère de la validation croisée défini par :

$$\begin{aligned} UCV(h_n) &= ISE(h_n) - \int_{\mathbb{R}} f^2(x) dx \\ &= \int_{\mathbb{R}} f_{h_n}^2(x) dx - 2 \int_{\mathbb{R}} f_{h_n}(x) f(x) dx. \end{aligned} \quad (3.27)$$

On doit donc trouver un estimateur pour $\int_{\mathbb{R}} f_{h_n}(x) f(x) dx$. Remarquons que :

$$\int_{\mathbb{R}} f_{h_n}(x) f(x) dx = \mathbb{E}(f_{h_n}(x)). \quad (3.28)$$

Son estimateur empirique est alors

$$\frac{1}{n} \sum_{\substack{j \neq 1 \\ i \neq j}}^n f_{h_n,i}(x), \quad (3.29)$$

avec $f_{h_n, i}(x) = \frac{1}{n} \sum_{\substack{j \neq 1 \\ i \neq j}}^n K\left(\frac{x-x_j}{h_n}\right)$ est l'estimateur de la densité construit à partir de l'ensemble des points sauf le point x_i (la méthode de leave one out). En remplaçant cet estimateur dans (3.27), pour un noyau K , le critère de la validation croisée par les moindres carrés est donné par :

$$UCV(h_n) = \int_0^\infty \left(\frac{1}{n} \sum_{i=1}^n K_{x, h_n}(X_i) \right)^2 dx - \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j \neq 1 \\ i \neq j}}^n K_{x, h_n}(X_i). \quad (3.30)$$

2. **Validation croisée biaisée (BCV)** Le critère de validation croisée biaisée a été introduit par Scott et Terrel [84] pour remédier aux problèmes de la méthode "validation croisée non biaisée". Il s'agit d'introduire un biais dans UCV afin de réduire sa variance. Le paramètre de lissage basé sur la méthode BCV est la valeur de h_n qui minimise un estimateur de l' $AMISE$ défini dans (??). Il est clair que afin d'estimer l' $AMISE$, il suffit d'estimer $R(f'')$. Un estimateur naturel de ce dernier terme est donnée par $R(\hat{f}'')$.

Finalement, Scott et Terrel [84] ont proposé la forme de l'estimateur de $AMISE$ à minimiser qui se résume comme suit :

Proposition 3.2.6. (Scott et Terrel [84]) Soit X_1, X_2, \dots, X_n un n -échantillon issu d'une variable aléatoire X de fonction de densité f . Pour un noyau K , on obtient :

$$BCV(h_n) = \frac{R(K)}{nh_n} + \frac{h_n^4}{4n^2} \mu_2^2(K) \sum_i \sum_{j, j \neq i} K_{h_n}^{(2)} K_{h_n}^{(2)}(X_i - X_j).$$

3.2.4.2 Deuxième classe réinjection (plug-in)

Beaucoup de méthodes ont été proposées dans cette classe (voir la monographie de Silverman [85] et celle de Wand et Jones [90]). Nous donnons ici l'exemple de la règle du pouce.

La règle du pouce (Rule of thumb) L'idée de cette méthode revient à Deheuvels [32] avant d'être publiée par Silverman [85]. Le choix du paramètre de lissage par cette méthode consiste à remplacer la partie inconnue $R(f'')$ dans l'expression de l'estimateur optimal h_{AMISE} , donné par l'équation (3.31), par une distribution classique afin d'obtenir un estimateur pour h_n .

$$h_{AMISE} = \left[\frac{R(K)}{nR(f'')\sigma_2^2(K)} \right]. \quad (3.31)$$

Si on choisit f comme étant la distribution normale de moyenne 0 et de variance σ^2 , on a alors :

$$\begin{aligned} R(f'') &= \int (f''(x)) dx \\ &= \frac{3}{8\sqrt{\pi}} \sigma^{-\frac{1}{5}}. \end{aligned} \quad (3.32)$$

De plus, si on utilise un noyau gaussien, alors la valeur de h_{AMISE} que l'on note dans ce cas par h_{rot} et en substituant la valeur obtenue dans (3.32), on aura :

$$\begin{aligned} h_{rot} &= (4\pi)^{\frac{-1}{10}} \left(\frac{3}{8} \pi^{\frac{-1}{2}} \sigma \right) n^{\frac{-1}{5}} \\ &= \left(\frac{4}{3} \right)^{\frac{1}{5}} \sigma n^{\frac{-1}{5}} \\ &= 1.06 \sigma n^{\frac{-1}{5}}. \end{aligned} \quad (3.33)$$

Il suffit donc d'estimer σ à partir des données et de le remplacer dans la formule (3.33) de h_{rot} .

3.3 Conclusion

Dans ce chapitre, nous avons présenté le concept de l'estimation de densité où nous avons essayé de mettre l'accent sur la notion d'estimation non-paramétrique et plus particulièrement sur l'utilisation de la méthode du noyau. Cette approche est relativement simple d'utilisation puisqu'il ne s'agit que de sélectionner la fonction noyau et de calculer le paramètre de lissage qui détermine le degré d'influence des observations pour l'estimation. La méthode non-paramétrique décrite dans ce chapitre sera appliquée à des données transformées par une distribution paramétrique, qui sera décrite dans le prochain chapitre.

Estimation semi-paramétrique de la densité de probabilité

Sommaire

4.1	Introduction	55
4.2	Distribution de Champernowne	55
4.3	Distribution de Champernowne généralisée	56
4.3.1	Propriétés de la distribution de Champernowne généralisée	56
4.3.2	Estimation des paramètres de la distribution de Champernowne généralisée	57
4.4	Estimateur semi-paramétrique basé sur la distribution de Champernowne généralisée et l'estimateur à noyau Beta (KMCE)	58
4.5	Conclusion	60

4.1 Introduction

Nous allons présenter, dans ce chapitre, la méthode semi-paramétrique pour estimer la fonction de densité qui se fait en deux étapes. Tout d'abord, une distribution de départ paramétrique (distribution de Champernowne généralisée) est utilisée pour transformer les données initiales, puis l'estimation du noyau Beta est appliquée à l'échantillon résultant.

4.2 Distribution de Champernowne

Cette distribution est un peu moins bien connue par les actuaires et fut principalement utilisée pour modéliser la distribution des revenus dans les sociétés (pour plus de détails, voir Kleiber et Kotz [57]). Par contre, dernièrement la distribution de Champernowne est de plus en plus souvent utilisée comme estimation préliminaire dans les estimateurs semi-paramétriques. De plus, la queue de la distribution de Champernowne converge vers une loi de Pareto ce qui fait d'elle un choix intéressant pour modéliser des distributions à queues lourdes.

La distribution originale de Champernowne a pour densité (voir Buch-Larsen et al [21])

$$t(x) = \frac{c}{x \left(\left(\frac{1}{2}\right) \left(\frac{x}{M}\right)^{-\alpha} + \lambda + \left(\frac{1}{2}\right) \left(\frac{x}{M}\right)^{\alpha} \right)}; x > 0,$$

où c est une constante de normalisation et α, λ et M sont les paramètres de la distribution. Cette distribution a été mentionnée pour la première fois en 1936 par D.G. Champernowne lors du congrès "the Oxford Meeting of the Econometric Society" comme outil de modélisation de la distribution des revenus des ménages en Grande Bretagne (voir [23, 24]).

Lorsque $\lambda = 1$ et $c = \frac{1}{2}\alpha$, on obtient ce qu'on appelle dans la littérature la vraie densité de Champernowne :

$$t(x) = \frac{\alpha M^{\alpha} x^{\alpha-1}}{(x^{\alpha} + M^{\alpha})^2},$$

qui a pour fonction de répartition la fonction :

$$T(x) = \frac{x^{\alpha}}{x^{\alpha} + M^{\alpha}}.$$

On remarque que la densité de Champernowne se comporte comme une loi de Pareto lorsque x est assez grand et elle se comporte comme la distribution log-normale pour des petites valeurs de x lorsque $\alpha > 1$. Mais lorsque $\alpha < 1$, cette densité n'est pas bornée au voisinage de zéro. Lorsqu'on fait de l'estimation non-paramétrique à noyau basée sur des données transformées avec

cette distribution, des simulations ont montré que son comportement au voisinage de zéro crée un biais dit de bord (voir Buch-Larsen et al. [21]). Pour remédier à ce problème, Buch-Larsen et al. [21] ont généralisé cette distribution en introduisant un nouveau paramètre permettant à la nouvelle distribution d'être bornée au voisinage de zéro pour toutes les valeurs de α .

4.3 Distribution de Champernowne généralisée

La distribution de Champernowne généralisée (également appelée distribution de Champernowne modifiée) convient pour modéliser les sinistres d'assurance (c'est-à-dire des variables positives), en raison de sa grande flexibilité qui permet d'obtenir des densités qui s'adaptent bien à toute la gamme de données. Elle est particulièrement adaptée à la modélisation des données de sinistres car elle a une forme similaire près du bord à celle du modèle lognormal, qui est considéré comme fournissant un bon ajustement pour les pertes de petite taille. Cependant, contrairement au modèle lognormal à queue moyenne, la distribution de Champernowne généralisée converge dans la queue vers la distribution de Pareto, qui est à queue lourde, et peut donc également capturer la caractéristique typique à queue lourde présentée par les données de perte empiriques. L'étape de transformation peut être considérée comme un type de procédure de stabilisation de la variance (voir [21, 39]).

Définition 4.3.1. Pour tout $x > 0$: La fonction de répartition de la distribution de Champernowne généralisée est donnée par

$$T_{\alpha, M, c}(x) = \frac{(x+c)^\alpha - c^\alpha}{(x+c)^\alpha + (M+c)^\alpha - 2c^\alpha} \quad \text{avec } \alpha > 0, c > 0 \text{ et } M > 0. \quad (4.1)$$

Sa densité de probabilité est donnée par :

$$t_{\alpha, M, c}(x) = \frac{\alpha (x+c)^{\alpha-1} ((M+c)^\alpha - c^\alpha)}{((x+c)^\alpha + (M+c)^\alpha - 2c^\alpha)^2}. \quad (4.2)$$

4.3.1 Propriétés de la distribution de Champernowne généralisée

- ★ la distribution de Champernowne généralisée converge vers la distribution de Pareto lorsque x converge vers l'infini :

$$t_{\alpha, M, c}(x) \sim \frac{\alpha \left(((M+c)^\alpha - c^\alpha)^{\frac{1}{\alpha}} \right)^\alpha}{((x+c)^\alpha + (M+c)^\alpha - 2c^\alpha)^2}.$$

- ★ lorsque $\alpha = 1$, le paramètre c n'a aucun effet.

★ lorsque $\alpha \neq 1$, le choix du paramètre c affecte la nouvelle densité de trois manières :

- ◆ **Effet sur la queue** : quand $\alpha < 1$, la distribution a une queue fine et elle a une queue lourde lorsque $\alpha > 1$ (voir Figure 4.1 (d), (e) et (f)).
- ◆ **Effet au voisinage de 0** lorsque $c > 0$, la densité est bornée pour tout $\alpha > 0$ (voir Figure 4.1 (d), (e) et (f)).

$$0 < f_{\alpha, M, c}(0) = \frac{\alpha c^{\alpha-1}}{(M+c)^\alpha - c^\alpha} < \infty.$$

- ◆ **Effet sur le mode de la distribution** : lorsque $\alpha > 1$, la densité a en général un mode et le paramètre c le décale vers la gauche (voir Figure 4.1 (f)).

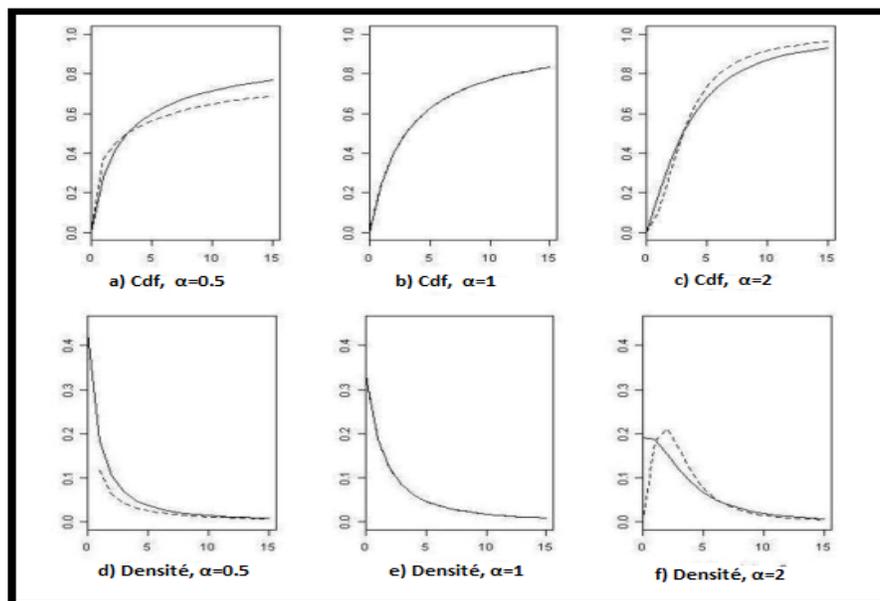


FIG. 4.1: Effet du paramètre c selon les valeurs de α ($c = 0$ pour graphes en pointillé, $c = 2$ trait plein).

4.3.2 Estimation des paramètres de la distribution de Champernowne généralisée

Il existe deux méthodes pour l'estimation des paramètres de la distribution de Champernowne généralisée. Pour les deux méthodes, nous utiliserons le fait que $T_{\alpha, M, c}(M) = 0.5$ et nous estimerons M par la médiane empirique (voir par exemple Lehmann [62]). La différence entre les deux méthodes est sur l'estimation du couple de paramètres (α, c) .

4.3.2.1 Méthode du maximum de vraisemblance

On estime le couple (α, c) par maximum de vraisemblance. Cette méthode est la plus courante dans la littérature (voir Charpentier et Oulidi [26] ou Buch-Larsen et al. [21] entre autres).

On donne ici la log-vraisemblance :

$$l(\alpha, M, c) = n(\log(\alpha) + \log((M+c)^\alpha - c^\alpha)) + (\alpha - 1) \sum_{i=1}^n \log(Y_i + c) - 2 \sum_{i=1}^n \log((Y_i + c)^\alpha + (M+c)^\alpha - 2c^\alpha).$$

Pour M fixé, la log-vraisemblance est concave et admet donc un maximum (voir [21]).

4.3.2.2 Méthode des percentiles

Pour la deuxième approche, on estime α de manière à ce que le 95^{ème} percentile de la distribution estimée coïncide avec le 95^{ème} percentile empirique. On estime ensuite c pour que la moyenne de la distribution empirique soit le plus près possible de la moyenne empirique. Dans certains cas, il serait peut-être plus approprié de choisir d'autres critères, mais ici, comme ces distributions sont utilisées pour la tarification en assurance, il est important de bien modéliser la queue, d'où le choix d'utiliser un quantile élevé, et que la moyenne soit bonne. Cette méthode est introduite dans Buch-Larsen [20].

4.4 Estimateur semi-paramétrique basé sur la distribution de Champernowne généralisée et l'estimateur à noyau Beta (KMCE)

L'estimateur semi-paramétrique de la densité est basé sur la distribution de Champernowne généralisée et l'estimateur à noyau Beta.

L'idée consiste à transformer l'échantillon initial $\{X_1, \dots, X_n\}$ en un nouvel échantillon $\{Y_1, \dots, Y_n\}$ où $Y_i = T_{\hat{\alpha}, \hat{M}, \hat{c}}(X_i), i = 1, \dots, n$. On doit estimer les paramètres α, M, c de cette distribution de telle sorte que la distribution de Champernowne ainsi estimée soit proche de la distribution de l'échantillon original $\{X_1, \dots, X_n\}$. Ainsi, l'échantillon transformé $\{Y_1, \dots, Y_n\}$ aura une distribution proche de la distribution uniforme sur l'intervalle $[0, 1]$.

Ceci peut être justifié par le théorème suivant :

Théorème 4.4.1. (Charpentier et Oulidi [26]) *Pour une variable aléatoire de fonction de répartition F , on note G son inverse :*

$$G(p) = \inf \{x \in \mathbb{R} \mid F(x) \geq p\}.$$

Si U désigne une variable aléatoire réelle uniforme sur $[0, 1]$, alors $X = G(U)$ a pour fonction de répartition F .

Comme les estimateurs à noyau standard supposent que le domaine de la distribution estimée est $(-\infty, +\infty)$ et qu'ici le domaine est $[0, 1]$, on doit tenir compte du problème de biais aux bornes. Pour cela nous proposons d'utiliser un noyau Beta donné dans la sous-section 3.2.3.5 (avec ses différentes versions) pour estimer la densité des données transformées Y_i .

L'estimateur résultant est défini comme suit :

$$\hat{f}_{trans}(y) = \frac{1}{nh_n} \sum_{i=1}^n K_{h_n}(y - Y_i). \quad (4.3)$$

Une fois que la densité des données transformées est estimée, nous trouvons la densité estimée pour les données originales :

$$\hat{f}(x) = \frac{\hat{f}_{trans}(T_{\hat{\alpha}, \hat{M}, \hat{c}}(x))}{\left| (T_{\hat{\alpha}, \hat{M}, \hat{c}}^{-1})'(T_{\hat{\alpha}, \hat{M}, \hat{c}}(x)) \right|}. \quad (4.4)$$

L'expression finale de l'estimateur KMCE donne le résultat suivant :

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n K_{h_n} \left(T_{\hat{\alpha}, \hat{M}, \hat{c}}(x) - T_{\hat{\alpha}, \hat{M}, \hat{c}}(X_i) \right) T'_{\hat{\alpha}, \hat{M}, \hat{c}}(x). \quad (4.5)$$

En résumé, la méthode proposée est la suivante :

1. Calculer les paramètres de la distribution de Champernowne généralisée à l'aide des données originales (les X_i).
2. Poser $Y_i = \hat{T}_{\hat{\alpha}', \hat{M}', \hat{c}'}(X_i)$, où $\hat{T}_{\hat{\alpha}', \hat{M}', \hat{c}'}$ est la fonction de répartition d'une distribution de Champernowne généralisée, avec les paramètres estimés à l'étape précédente.
3. Estimer la densité des données transformées à l'aide de l'estimateur à noyau Beta (et ses versions).
4. Finalement, on retrouve la densité pour les données d'origine.

Quelques propriétés asymptotiques de l'estimateur KMCE

Certaines propriétés asymptotiques de l'estimateur KMCE défini dans l'équation (4.5) peuvent être trouvées dans Buch-Larsen et al. [21].

Le biais et la variance de $\hat{f}(x)$ sont donnés par

$$\mathbb{E} \left[\hat{f}(x) \right] - f(x) = \frac{1}{2} \mu_2(K) h_n^2 \left(\left(\frac{f(x)}{T'(x)} \right)' \frac{1}{T'(x)} \right)' + o(h_n^2),$$

$$\mathbb{V} \left[\hat{f}(x) \right] = \frac{1}{nh_n} R(K) T'(x) f(x) + o\left(\frac{1}{h_n^2}\right),$$

quand $n \rightarrow \infty$, où $\mu_2(K) = \int u^2 K(u) du$ et $R(K) = \int K^2(u) du$.

L'estimateur KMCE suit une distribution normale asymptotiquement :

$$\sqrt{nh_n} \left(\hat{f}(x) - \mathbb{E} \left[\hat{f}(x) \right] \right) \sim \mathbf{N} \left(0, \frac{1}{nh_n} R(K) T'(x) f(x) \right).$$

4.5 Conclusion

Dans ce chapitre, nous avons présenté le concept de l'estimation de la densité en utilisant une approche semi-paramétrique, via la combinaison de la distribution paramétrique de Champernowne généralisée et de l'estimateur non paramétrique à noyau Beta. Cette technique servira dans les applications numériques qui seront considérées dans la prochaine partie pour estimer l'erreur d'approximation entre le modèle idéal et le modèle perturbé en évaluant les bornes de stabilité données par l'équation (2.19) et (2.24) respectivement.

Partie

**Application de la méthode
semi-paramétrique à l'analyse de la
stabilité forte d'un modèle de risque
classique**

Introduction

Cette partie est composée de deux chapitres. Nous allons réaliser dans le premier chapitre des études de simulation ainsi qu'une étude comparative pour évaluer la performance et l'efficacité de l'approche semi-paramétrique proposée, et pour montrer son importance pour l'amélioration de la qualité de la borne de stabilité sur la probabilité de ruine dans un modèle de risque classique avec des réclamations larges, par rapport à l'approche non-paramétrique. Une autre comparaison entre les deux bornes de stabilité (la borne de chaîne de Markov et la borne de l'approche régénérative) sera effectuée. Le deuxième chapitre sera consacré à une étude de données réelles pour confirmer les résultats obtenus dans le premier chapitre.

Application numérique basée sur la simulation

Sommaire

5.1 Introduction	64
5.2 Version 1 : Stabilité d'une chaîne de Markov	64
5.3 Version 2 : Stabilité d'un processus de Lindley	79
5.4 Conclusion	91

5.1 Introduction

Nous appliquons l'approche semi-paramétrique définie dans le Chapitre 2 de la deuxième partie, afin de juger de la qualité des bornes données dans la formule (2.19) du Théorème 2.4.1 et (2.24) du Théorème 2.4.4 respectivement. Nous utilisons les algorithmes 1 et 2 (qui seront définis ultérieurement dans les sections 5.2 et 5.3 respectivement) et nous prenons les mêmes distributions que celles utilisées dans Buch-Larsen et al. [21] : Weibull, lognormale et mélange Paretolognormale. Dans ce travail, nous considérons la perturbation de la taille moyenne des réclamations $\mu' = \mu + \epsilon$. Pour les estimateurs impliquant le noyau Beta asymétrique, nous suivons [74] et prenons

$$h_{rot} = \hat{\sigma} n^{-\frac{2}{5}}, \quad (5.1)$$

où $\hat{\sigma}$ est l'écart type empirique.

5.2 Version 1 : Stabilité d'une chaîne de Markov

Distributions simulées

Nous prenons les mêmes distributions que celles utilisées dans Buch-Larsen et al. [21] (voir aussi Bolancé et al. [15]), à savoir :

1. Distribution de Weibull (Wbl) :

$$f_{Wbl}(x) = \gamma x^{\gamma-1} e^{-x^\gamma/\beta}; x > 0. \quad \text{avec } (\gamma, \beta) = (3, 2).$$

2. Distribution lognormale (LogN) :

$$f_{LogN}(x) = \left(\frac{2}{x * \sigma \sqrt{2\pi}} \right) e^{-1/2((\log(x)-\mu)/\sigma)^2}; x > 0. \quad \text{avec } (\mu, \sigma^2) = (0, 0.5).$$

3. Mélange de distributions pareto/lognormale (Mixt) :

$$f_{Mixt}(x) = \alpha f_{LogN}(x) \mathbb{I}_{\{x>0\}} + (1 - \alpha) \left(\frac{\rho \lambda^\rho}{(x - c)^{\rho+1}} \right) \mathbb{I}_{\{x>c\}}, \quad \text{avec}$$

$$(\alpha, \mu, \sigma, \lambda, \rho, c) = \begin{cases} (0.7; 0; 0.5; 1; 1.5; 1), & \text{i.e. 30\% - 70\% (Mixt1)} \\ (0.3; 0; 0.5; 1; 1.5; 1), & \text{i.e. 70\% - 30\% (Mixt2)} \end{cases}$$

Les six estimateurs KMCE suivants sont utilisés pour estimer les différentes densités définies précédemment.

Algorithm 1 Algorithme 1: Stabilité d'une chaîne de Markov

1. Introduire les paramètres λ, μ, c du modèle idéal, et λ', μ', c' du modèle perturbé (réel).
2. Vérifier la positivité des charges de sécurité relatives θ et θ' définies par : $\theta = \frac{c-\lambda\mu}{\lambda\mu}$ et $\theta' = \frac{c'-\lambda'\mu'}{\lambda'\mu'}$. Si oui, ("la ruine des modèles n'est pas certaine") passer à l'étape 3. Sinon retournez à l'étape 1.
3. Initialiser ε ($\varepsilon > 0$) de sorte que $0 < \rho(\varepsilon) < 1$ et Γ_1 soit minimal.
4. Calculer $\mu(a, a') = 2\mathbb{E}e^{\varepsilon Z} \left| \ln \frac{\lambda'\mu'}{\lambda\mu} \right| + \int_0^{\infty} v(u) - f'(u) du$, et tester : $\mu(a, a') < (1 - \rho(\varepsilon))^2$.
Si oui, ("nous pouvons déduire l'inégalité de stabilité forte") passer à l'étape 5; sinon augmenter ε avec un pas p , puis retourner à l'étape 4.
5. Calculer la borne de l'écart $\|\Psi_a - \Psi_{a'}\|_v$:

$$\|\Psi_a - \Psi_{a'}\|_v \leq \frac{\mu(a, a')}{(1 - \rho(\varepsilon)) ((1 - \rho(\varepsilon))^2 - \mu(a, a'))} = \Gamma_1.$$

Puisque l'étape 4 implique le calcul d'une quantité dépendant des distributions de pertes inconnues f et f' , alors elle peut être développée, en utilisant la technique semi-paramétrique, comme suit :

- 4.a Générer un échantillon $X = X_1, X_2, \dots, X_n$ de taille n , à partir d'une distribution parmi les quatre distributions (qui seront définies dans la section suivante).
- 4.b Estimer les paramètres de la distribution de Champernowne généralisée : M (médiane empirique), et α et c en utilisant la technique de maximum de vraisemblance.
- 4.c Transformer les X_i en Y_i de sorte que $Y_i = T_{\hat{\alpha}, \hat{M}, \hat{c}}(X_i)$.
- 4.d Estimer la fonction de densité des données transformées Y_i selon l'équation (4.3).
- 4.e Estimer la fonction de densité des données d'origine X_i selon l'équation (4.5).

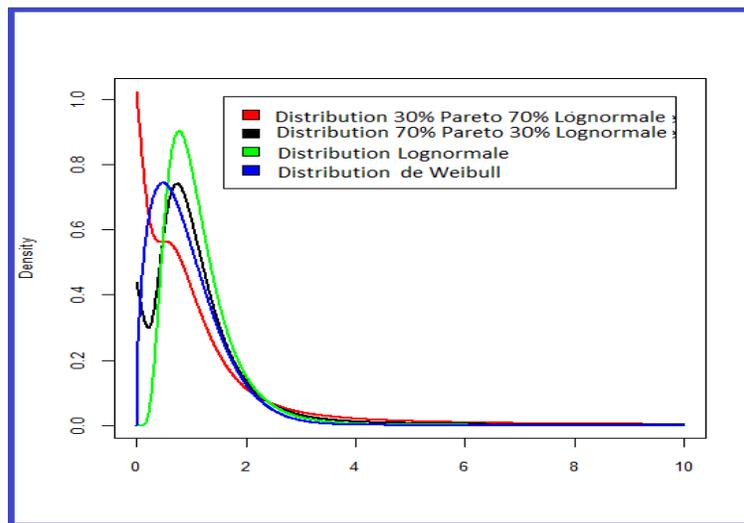


FIG. 5.1: Densités des quatre distributions étudiées.

1. $\hat{f}_{\beta 1}$ est l'estimateur, avec le noyau Beta1 défini en (3.14).
2. $\hat{f}_{Mac-\beta 1}$ est l'estimateur macro-Beta donné dans (3.18), avec le noyau Beta1.
3. $\hat{f}_{Mic-\beta 1}$ est l'estimateur micro-Beta donné dans (3.19), avec le noyau Beta1.
4. $\hat{f}_{\beta 2}$ est l'estimateur avec le noyau modifié Beta2, défini en (3.17).
5. $\hat{f}_{Mac-\beta 2}$ est l'estimateur macro-Beta donné dans (3.18), avec le noyau Beta2.
6. $\hat{f}_{Mic-\beta 2}$ est l'estimateur micro-Beta donné dans (3.19), avec le noyau Beta2.

Pour analyser l'effet du choix de l'estimateur nous fixons les paramètres comme suit :

- ★ La taille de l'échantillon $n = \{50, 100, 200, 500\}$.
- ★ Le nombre de réplifications $m = 2000$ (pour appliquer le principe de Monte Carlo).
- ★ Le paramètre $\lambda = 20$.
- ★ Le taux de prime $c = 100$.

Les études de simulation sont réalisées sous l'environnement R .

Résultats graphiques et numériques

Les résultats des simulations sont présentés dans les figures et tableaux suivants, tels que

:

- Les estimateurs KMCE finaux et la densité réelle sont présentés sur les figures 5.2, 5.5, 5.8 et 5.11 pour toutes les lois données précédemment.
- La variation de la borne de stabilité Γ_1 par rapport à la perturbation ϵ est présentée sur les figures 5.3, 5.6, 5.9 et 5.12 pour toutes les lois données précédemment.
- La variation de la borne de stabilité Γ_1 en fonction de la taille de l'échantillon n est présentée sur les figures 5.4, 5.7, 5.10 et 5.13 pour toutes les lois données précédemment.
- La comparaison entre les estimateurs Mic- $\beta 2$ et BS est donnée dans les figures 5.14 et 5.15.
- Les paramètres estimés par la méthode du maximum de vraisemblance, $\hat{\theta} = (\hat{\alpha}, \hat{M}, \hat{c})$, de la distribution de Champernowne généralisée sont présentés dans le Tableau 5.1.
- La borne de stabilité Γ_1 , calculé à l'étape 5 de l'algorithme précédent, est présentée dans les tableaux 5.2 et 5.3, dans les colonnes 4, 6 et 8.
- La borne théorique $\Gamma_{1_{théorique}}$ est présentée dans les tableaux 5.2 et 5.3, dans la colonne 9.
- Le paramètre de lissage optimal, au sens de la méthode de la règle du pouce, est présenté dans les tableaux 5.2 et 5.3, dans les colonnes 3, 5 et 7.

- L'erreur quadratique moyenne (MSE) entre la densité réelle et les estimateurs KMCE est présentée dans le Tableau 5.4. Le critère MSE est donné dans l'expression (3.4)
- La comparaison entre les méthodes non paramétriques et semi-paramétriques est présentée dans le Tableau 5.10.

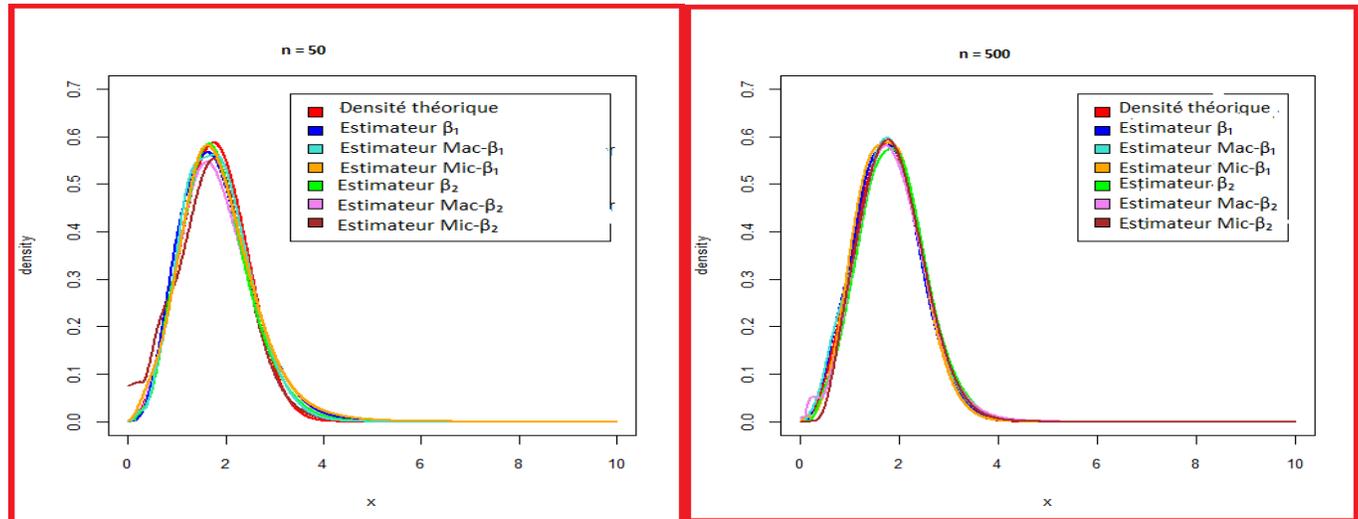


FIG. 5.2: Les estimateurs finaux de KMCE et la densité réelle pour la distribution de Weibull.

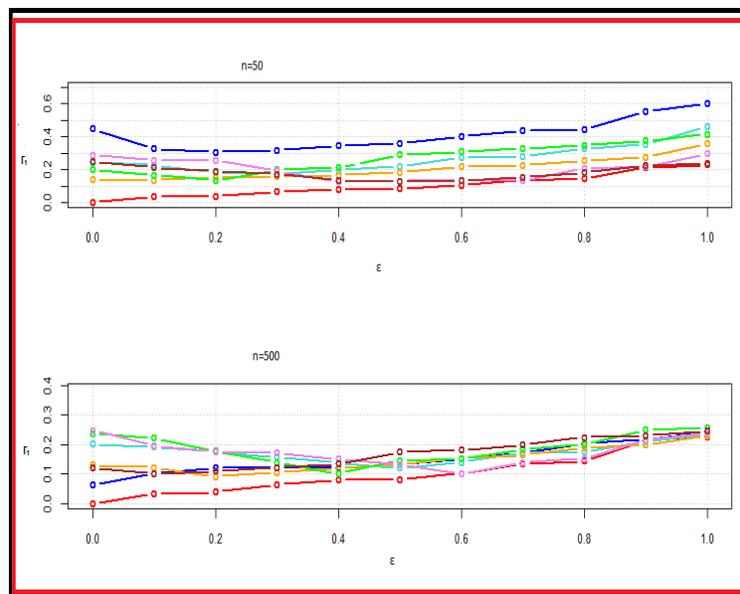


FIG. 5.3: Erreur Γ_1 en fonction de ϵ avec la distribution de Weibull. Estimateur β_1 (bleu), Estimateur Mac- β_1 (turquoise), Estimateur Mic- β_1 (orange), Estimateur β_2 (vert), Estimateur Mac- β_2 (violet), Estimateur Mic- β_2 (marron).

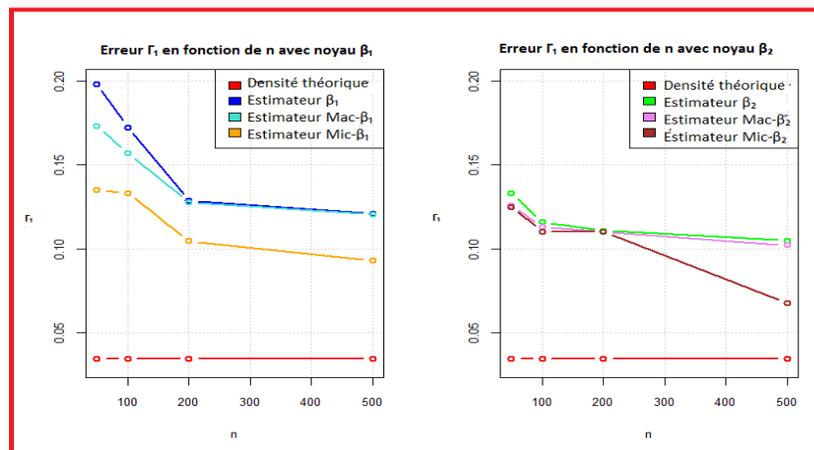


FIG. 5.4: Erreur Γ_1 en fonction de n avec la distribution de Weibull.

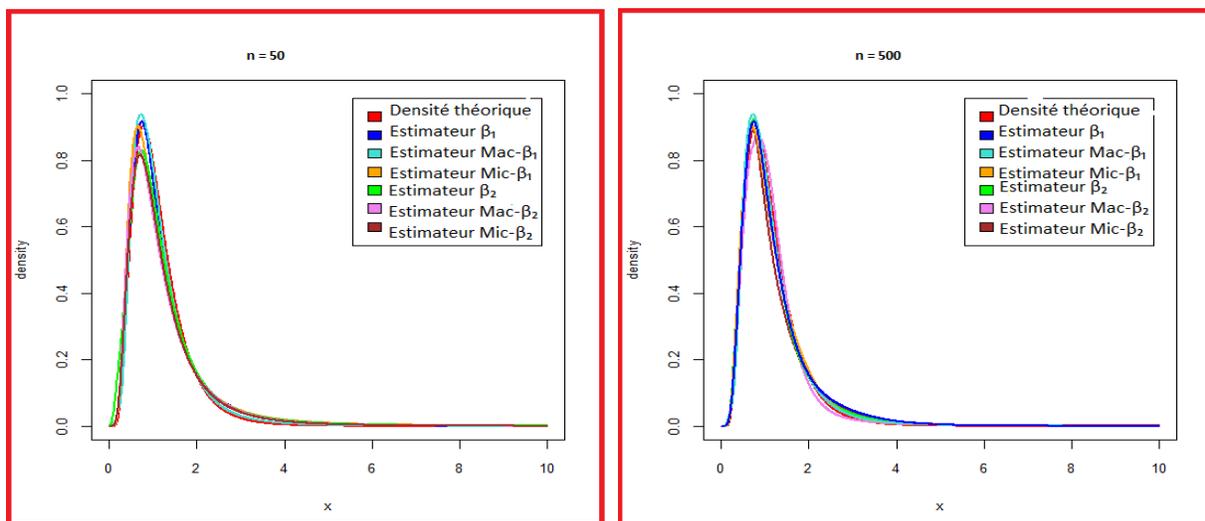


FIG. 5.5: Les estimateurs KMCE finaux et la densité réelle pour la distribution lognormale.

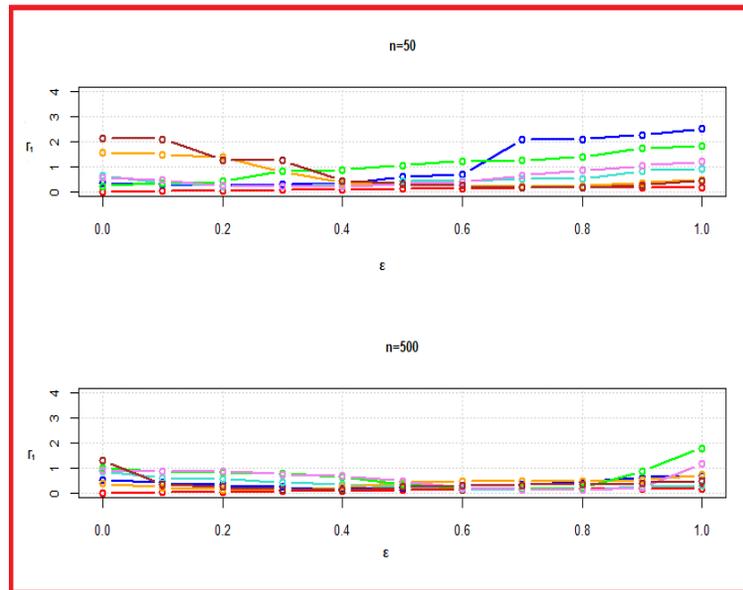


FIG. 5.6: Erreur Γ_1 en fonction de ϵ avec une distribution lognormale. Estimateur β_1 (bleu), Estimateur Mac- β_1 (turquoise), Estimateur Mic- β_1 (orange), Estimateur β_2 (vert), Estimateur Mac- β_2 (violet), Estimateur Mic- β_2 (marron).

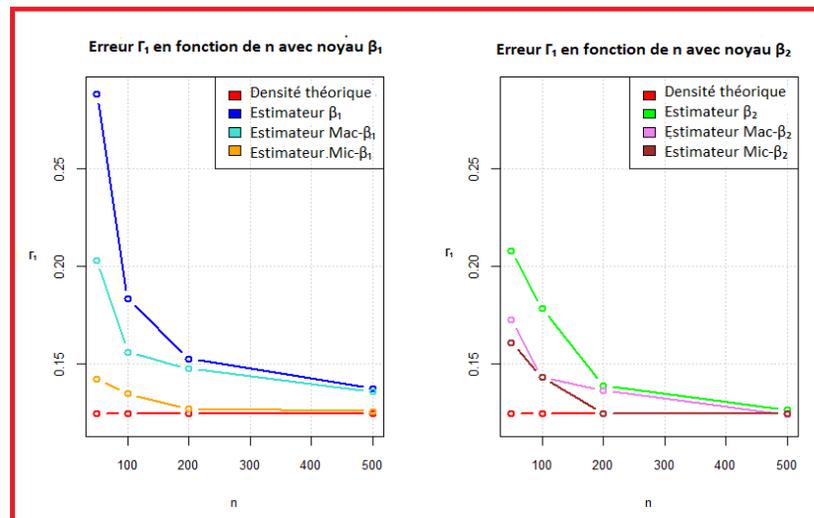


FIG. 5.7: Erreur Γ_1 en fonction de n avec une distribution lognormale.

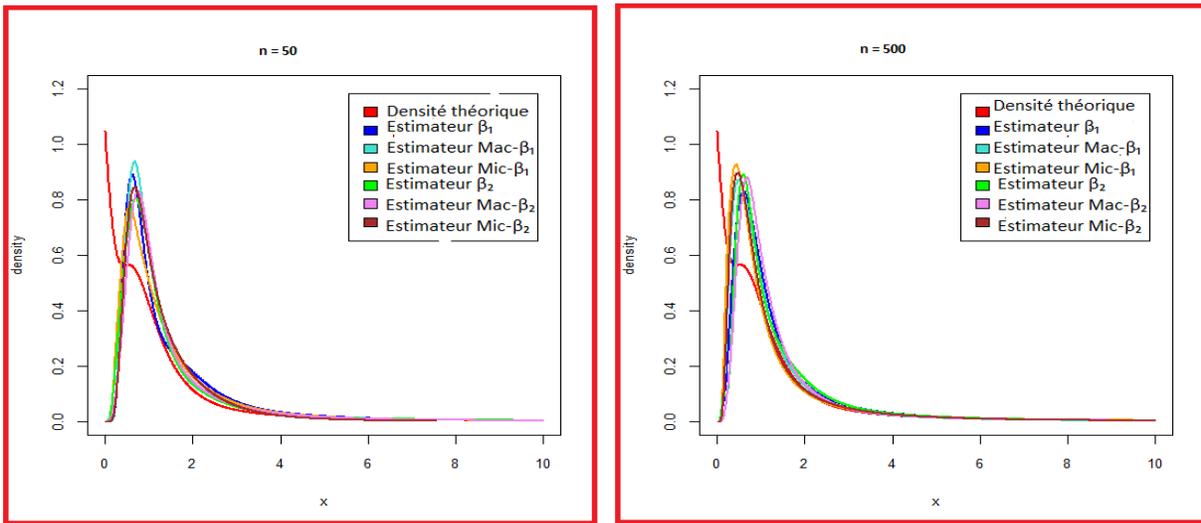


FIG. 5.8: Les estimateurs KMCE finaux et la densité réelle pour la distribution mélange Pareto-lognormale 30%70%.

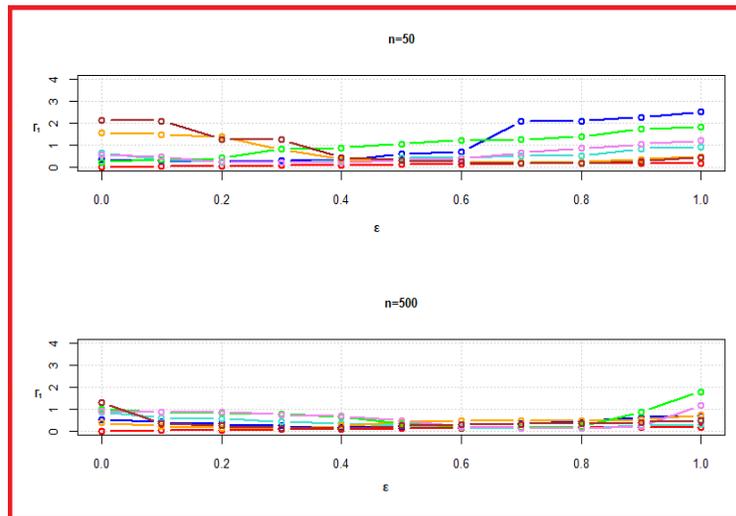


FIG. 5.9: Erreur Γ_1 en fonction de ϵ avec la distribution mélange Pareto-lognormale 30%70%. Estimateur β_1 (bleu), Estimateur Mac- β_1 (turquoise), Estimateur Mic- β_1 (orange), Estimateur β_2 (vert), Estimateur Mac- β_2 (violet), Estimateur Mic- β_2 (marron).

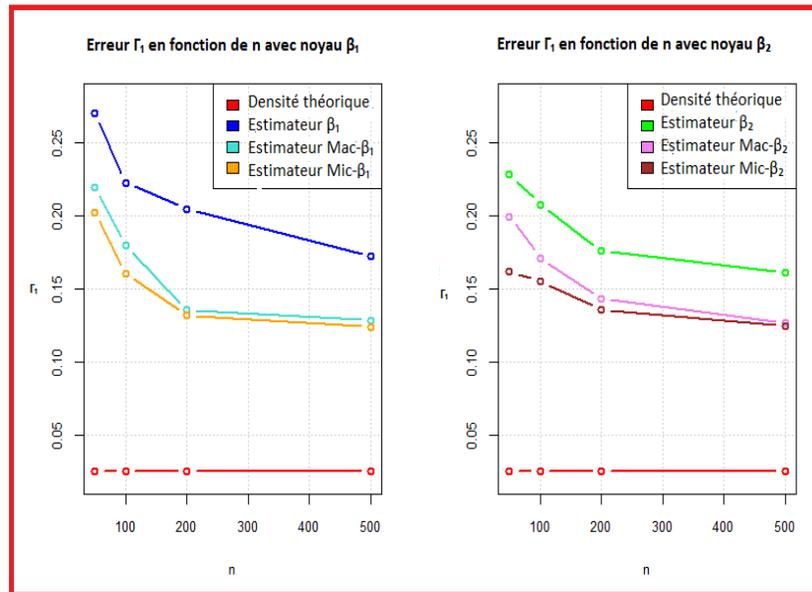


FIG. 5.10: Erreur Γ_1 en fonction de n avec une distribution mélange Pareto-lognormale 30%70%.

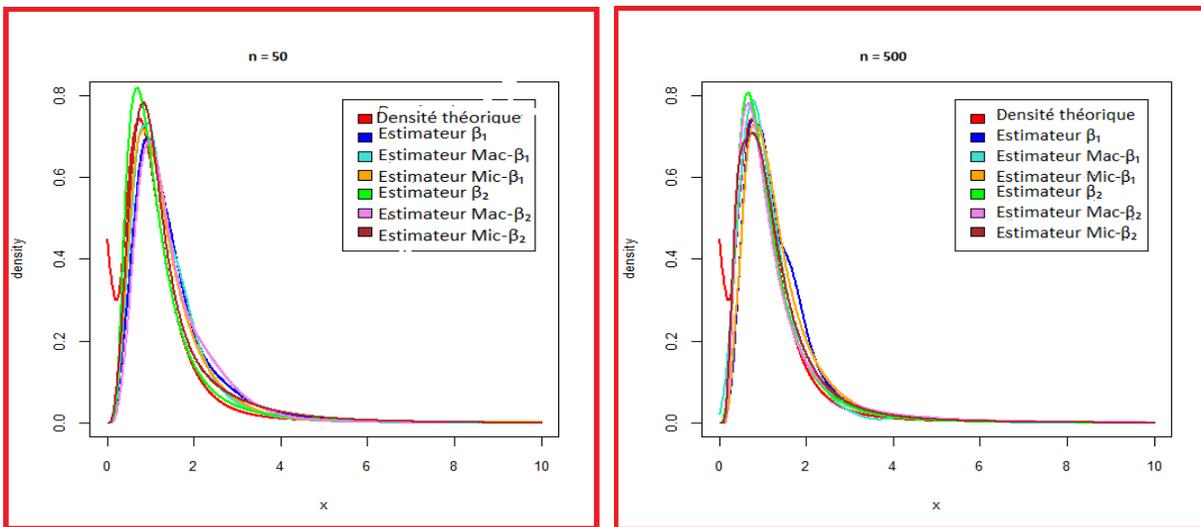


FIG. 5.11: Les estimateurs KMCE finaux et la densité réelle pour la distribution mélange Pareto-lognormale 70%30%.

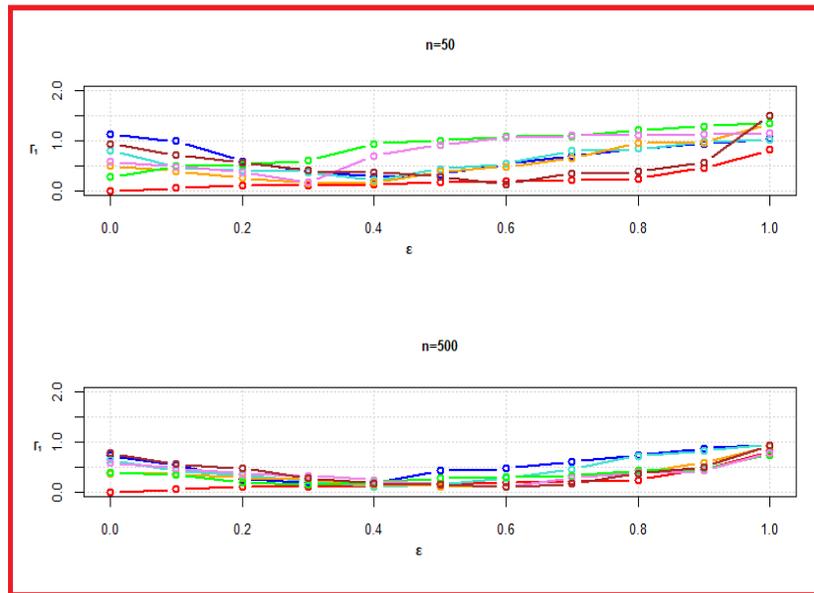


FIG. 5.12: Erreur Γ_1 en fonction de ϵ avec une distribution mélange Pareto-lognormale 70%30%. Estimateur β_1 (bleu), estimateur Mac- β_1 (turquoise), estimateur Mic- β_1 (orange), estimateur β_2 (vert), estimateur Mac- β_2 (violet), estimateur Mic- β_2 (marron).

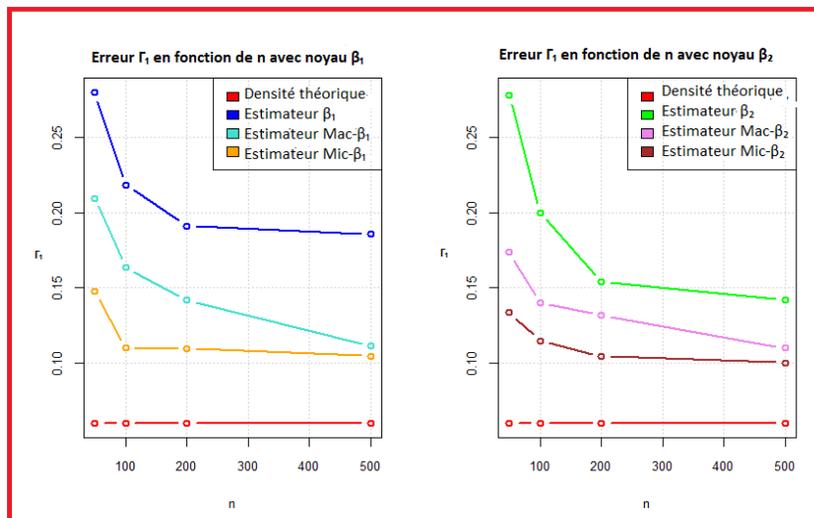


FIG. 5.13: Erreur Γ_1 en fonction de n avec une distribution mélange Pareto-lognormale 70%30% distribution.

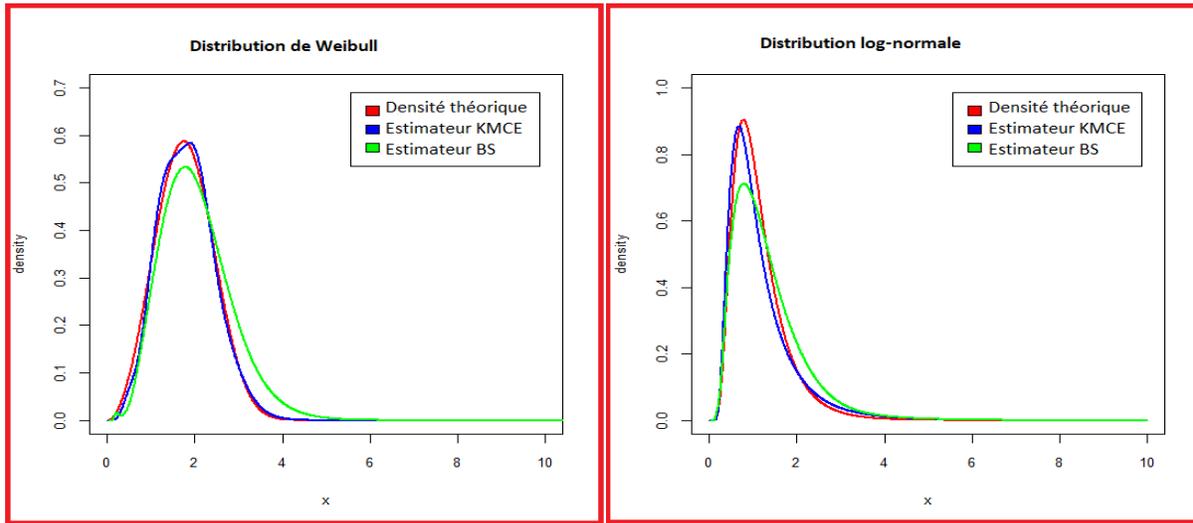


FIG. 5.14: L'estimateur KMCE final ($\hat{f}_{Mic-\beta 2}$), l'estimateur BS et la densité réelle pour les distributions de Weibull et lognormale.

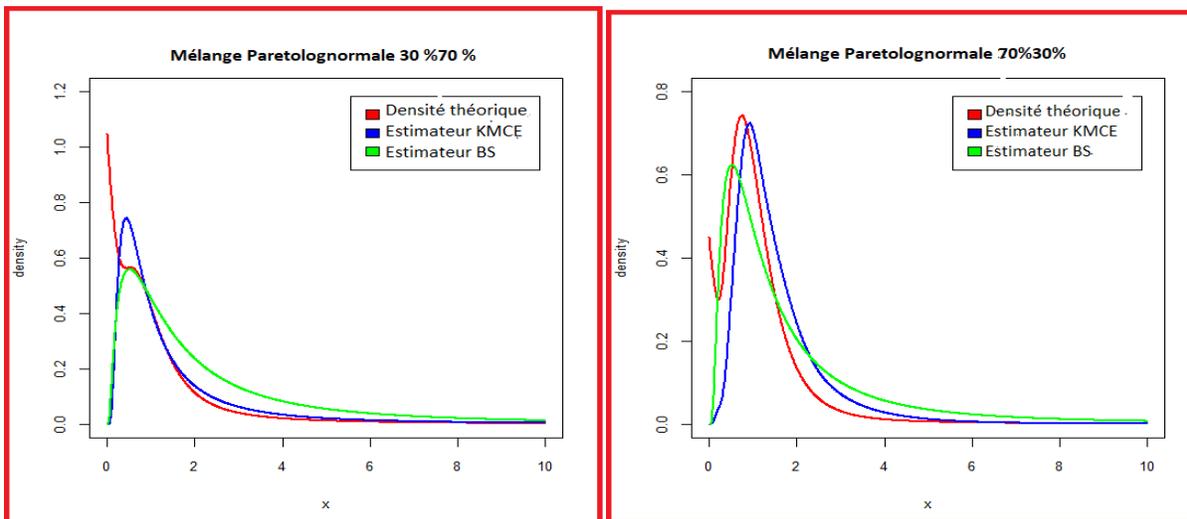


FIG. 5.15: Estimateur KMCE final ($\hat{f}_{Mic-\beta 2}$), estimateur BS et densité réelle pour les distributions mélange Paretolognormale 70%30% et mélange Paretolognormale 30%70%.

loi	n	$(\hat{\alpha}, \hat{M}, \hat{c})$
Wbl	50	(4.371378, 1.783584, 4.442755)
	100	(4.372817, 1.789351, 4.428579)
	200	(4.373976, 1.794501, 4.421431)
	500	(4.37412, 1.795682, 4.421121)
LogN	50	(0.002209702, 0.0005736382, 0.082373)
	100	(0.606, 0.7736681, 0.75201)
	200	(0.94187, 0.8661107, 0.6509)
	500	(0.9891943, 0.9445038, 0.58483)
Mixt1	50	(0.001226054, 0.00004408183, 0.3811156)
	100	(0.001611303, 0.000699596, 0.218442)
	200	(1.00380841, 0.04310538, 0.1916085)
	500	(1.0095565, 0.06009602, 0.02395939)
Mixt2	50	(2.250934, 1.183921, 3.667656)
	100	(2.464902, 1.273007, 3.264435)
	200	(2.4802959, 1.377527, 3.140869)
	500	(2.860573, 1.385846, 3.120029)

Tableau 5.1: Estimation du paramètre $\theta = (\alpha, M, c)$ de la loi de Champernowne.

loi	n	$\hat{f}_{\beta 1}$		$\hat{f}_{Mac-\beta 1}$		$\hat{f}_{Mic-\beta 1}$		Γ_1 théorique
		h_n	Γ_1	h_n	Γ_1	h_n	Γ_1	
Wbl	50	0.02930514	0.1982505	0.02898292	0.17306221	0.02680017	0.1349849	0.03400484
	100	0.0246552	0.172401	0.0235536	0.1570731	0.02087219	0.1331169	
	200	0.01632456	0.1287729	0.01871365	0.1277597	0.01623564	0.10453909	
	500	0.01159403	0.12069179	0.01179285	0.12043856	0.01179981	0.09281793	
LogN	50	0.03439691	0.2879552	0.03862293	0.2028934	0.03819431	0.1425789	0.1244944
	100	0.02438112	0.1832802	0.02857692	0.156224	0.02688863	0.13490687	
	200	0.02220641	0.1528544	0.01987875	0.1476125	0.01968361	0.1272244	
	500	0.01462523	0.137591	0.01446771	0.1360286	0.01422727	0.1259399	
Mixt1	50	0.03252294	0.2700395	0.03934995	0.2196325	0.01643765	0.2023622	0.02506683
	100	0.03031686	0.2228498	0.02818893	0.1798832	0.0115139	0.16036554	
	200	0.02316169	0.2046444	0.01952624	0.13537774	0.01070387	0.13199773	
	500	0.01395682	0.17203618	0.0134761	0.12841768	0.01032538	0.1238103	
Mixt2	50	0.03568085	0.27920345	0.03637747	0.2094644	0.03743662	0.1476961	0.06048445
	100	0.0257122	0.2181814	0.02764087	0.1636112	0.02661364	0.11037798	
	200	0.01797374	0.1908635	0.01914561	0.14185052	0.0192536	0.10991752	
	500	0.01312727	0.18555563	0.01395974	0.111715	0.01383699	0.104934546	

Tableau 5.2: Borne de stabilité Γ_1 via l'estimateur semi-paramétrique à noyau Beta 1 et ses versions normalisées.

loi	n	$\hat{f}_{\beta 2}$		$\hat{f}_{Mac-\beta 2}$		$\hat{f}_{Mic-\beta 2}$		Γ_1 théorique
		h_n	Γ	h_n	Γ_1	h_n	Γ_1	
Wbl	50	0.02943133	0.13311	0.02680017	0.1255507	0.02885693	0.12488363	0.03400484
	100	0.02375079	0.11590054	0.02529496	0.11278563	0.02186946	0.110136606	
	200	0.01799973	0.11067174	0.02020469	0.110107978	0.01712119	0.11001338	
	500	0.01192006	0.1048571	0.01171968	0.1020469	0.01176957	0.067412544	
LogN	50	0.03796038	0.2080006	0.03624853	0.17259018	0.03569399	0.16083142	0.1244944
	100	0.02973531	0.17835032	0.02827746	0.14322155	0.02671124	0.14313448	
	200	0.02198364	0.1390481	0.01996742	0.1364382	0.01994057	0.1249172	
	500	0.01036393	0.1266003	0.01497893	0.1252297	0.01461345	0.12449569	
Mixt1	50	0.03225347	0.228267	0.03815026	0.1992561	0.03941774	0.161668	0.02506683
	100	0.03108201	0.207791	0.03044834	0.1711076	0.03145446	0.1552654	
	200	0.02114798	0.17591976	0.02046885	0.14291887	0.01940284	0.135777379	
	500	0.01376534	0.1609048	0.01408615	0.126781	0.01317192	0.1244500223	
Mixt2	50	0.04427964	0.277258	0.03644457	0.1735601	0.03681481	0.133898	0.06048445
	100	0.0264847	0.1998723	0.02692092	0.140059	0.02776333	0.114973211	
	200	0.02090467	0.154211366	0.0208954	0.132006541	0.02006051	0.104785	
	500	0.01296031	0.142111	0.01349903	0.11014746	0.01346891	0.1004294	

Tableau 5.3: Borne de stabilité Γ_1 via l'estimateur semi-paramétrique à noyau Beta 2 et ses versions normalisées.

Estimateur	n	Wbl	LogN	Mixt1	Mixt2
$\hat{f}_{\beta 1}$	50	0.000042377238	0.0003210017	0.007843833	0.00275704
	100	0.00004038931	0.0002030713	0.006837752	0.002529511
	200	0.00003651917	0.0002011754	0.001217869	0.001142427
	500	0.00002819308	0.0001636075	0.001039574	0.001072434
$\hat{f}_{\beta 2}$	50	0.01205361	0.001370737	0.02914652	0.1746358
	100	0.005509124	0.0003927695	0.0263248	0.0013914
	200	0.0006955926	0.0002186766	0.01116965	0.001239643
	500	0.0001049646	0.0001823594	0.00118434346	0.0001108697
$\hat{f}_{Mac-\beta 1}$	50	0.0003630962	0.0001807518	0.02125305	0.013536
	100	0.0001124205	0.0001257097	0.0191764	0.001038345
	200	0.0001099355	0.00010303927	0.002777942	0.0010049876
	500	0.0001022499	0.0001325475	0.002178314	0.001002557
$\hat{f}_{Mac-\beta 2}$	50	0.01508669	0.02628705	0.02944201	0.02035151
	100	0.001002293	0.02048565	0.01542065	0.0100356
	200	0.0001698834	0.01910017	0.0103426	0.003754461
	500	0.0001001989	0.0162059	0.00125942	0.0001026326
$\hat{f}_{Mic-\beta 1}$	50	0.0002377894	0.0008708554	0.001749656	0.003858356
	100	0.0001680969	0.0008246252	0.001002485	0.0010084939
	200	0.000102663	0.0002475912	0.0003226237	0.0001055925
	500	0.00001404981	0.0001325475	0.00010733453	0.0001007698
$\hat{f}_{Mic-\beta 2}$	50	0.0004001963	0.006356391	0.0326984	0.010216871
	100	0.0001281099	0.0015762	0.02363471	0.00118815
	200	0.0001059428	0.001390225	0.0100745	0.001103037
	500	0.0001001163	0.00179502	0.001062365	0.0001738179

Tableau 5.4: L'erreur quadratique moyenne (MSE) des estimateurs étudiés.

loi	Γ_1 non-paramétrique				Γ_1 semi-paramétrique						Γ_1 théorique
	\hat{f}_{GAM}	\hat{f}_{BS}	\hat{f}_{RIG}	n	$\hat{f}_{\beta 1}$	$\hat{f}_{Mac-\beta 1}$	$\hat{f}_{Mic-\beta 1}$	$\hat{f}_{\beta 2}$	$\hat{f}_{Mac-\beta 2}$	$\hat{f}_{Mic-\beta 2}$	
	50 0.479156	0.4448456	0.5734506	50	0.1982505	0.1730622	0.1349849	0.13311	0.1255507	0.12488363	
Wbl	1000.4157367	0.349418	0.451785	1000	0.172401	0.1570731	0.1331169	0.11590054	0.11278563	0.110136606	0.03400484
	5000.3760803	0.3213912	0.39251936	5000	0.12069179	0.12043856	0.09281798	0.1048571	0.1020469	0.067412544	
	50 0.27889	0.25281	0.32311515	50	0.2879552	0.2028934	0.1425789	0.2080006	0.17259018	0.16083142	
LogN	1000.250341	0.238752	0.3096398	1000	0.1832802	0.156224	0.13490687	0.1783532	0.14322155	0.14313448	0.1244944
	5000.23818	0.1460287	0.1494301	5000	0.137591	0.1360286	0.1259399	0.1266003	0.1252297	0.12449569	
	50 13.75448	0.4219374	1.29846301	50	0.2700395	0.21963254	0.2023622	0.228267	0.1992561	0.161668	
Mixt1	1008.297866	0.3836091	1.118366	1000	0.2228498	0.1798832	0.16036554	0.207791	0.1711076	0.1552654	0.06048445
	5001.9972	0.322052840	0.7188342	5000	0.17203618	0.12841768	0.1238103	0.1609048	0.126781	0.1244500223	
	50 2.35146	0.4266026	0.6608563	50	0.27920345	0.2094644	0.1476961	0.277258	0.1735601	0.133898	
Mixt2	1002.018099	0.287889	0.4801103	1000	0.2181814	0.1636112	0.11037798	0.1998723	0.140059	0.114973211	0.02506683
	5001.058258	0.208721	0.321735	5000	0.18555563	0.111715	0.104934546	0.142111	0.11014746	0.1004294	

Tableau 5.5: Une étude comparative de la borne Γ_1 entre les méthodes non paramétriques et semi-paramétriques.

Discussion des résultats

- Pour toutes les distributions et les deux catégories d'estimation (non paramétrique et semi-paramétrique), la borne de stabilité Γ_1 est affectée de la même manière par la perturbation ϵ et augmente raisonnablement avec l'augmentation de la perturbation ϵ . Par exemple, à partir des graphiques présentés dans les figures "5.3, 5.6, 5.9 et 5.12" avec l'approche semi-paramétrique, nous remarquons que la borne de stabilité Γ_1 devient de plus en plus importante avec l'augmentation de la perturbation. La même remarque vaut pour la méthode non paramétrique (voir Tableau 5.10). Ceci est dû à l'importance de la distance entre les deux modèles (idéal et perturbé). On remarque également que la borne de stabilité Γ_1 est grande au début et diminue lorsque les valeurs de la perturbation ϵ sont au voisinage de la borne inférieure. Mais au voisinage de la borne supérieure, Γ_1 augmente rapidement. Ceci explique bien le fait que nous sommes dans la région critique.
- D'après les figures "5.4, 5.7, 5.10 et 5.13", nous remarquons que la borne de stabilité Γ_1 diminue lorsque la taille de l'échantillon augmente dans le cas de l'approche semi-paramétrique (voir également le tableau récapitulatif 5.10 pour l'approche non-paramétrique).
- Notons que dans la catégorie des estimations non paramétriques, la borne de stabilité Γ_1 obtenue à l'aide de l'estimateur à noyau BS est la meilleure comparée à celles obtenues par les estimateurs à noyau Gamma et RIG. Cela est dû aux propriétés statistiques de l'estimateur BS et à sa flexibilité qui permet d'éviter les effets de bord et de bien capturer les données dans la queue. D'autre part, pour la catégorie des estimateurs semi-paramétriques, la borne de stabilité Γ_1 obtenue en utilisant l'estimateur à noyau Mic- β_2 est plus proche de la borne de stabilité théorique Γ_1 par rapport aux autres estimateurs à noyau (Mac- β_2 , Mic- β_1 , Mac- β_1 , β_2 , β_1). Par conséquent, il s'agit de la meilleure estimation dans le contexte de notre approche semi-paramétrique pour notre étude. Nous remarquons également que l'estimateur à noyau Beta 2 nous fournit de meilleurs résultats en matière de borne de stabilité Γ_1 que l'estimateur à noyau Beta 1. Ce fait peut s'expliquer par les valeurs de h_n , qui régulent la qualité, la régularité et la cohérence de l'estimateur à noyau Beta 2 défini dans l'équation (3.17).
- Par ailleurs, les figures 5.14 et 5.15 montrent que les résultats obtenus à l'aide de l'approche semi-paramétrique sont meilleurs que ceux obtenus à l'aide de la méthode non-paramétrique. Comme nous pouvons le voir, l'estimateur KMCE trouvé par l'estimateur

Mic- β 2 (le meilleur parmi les estimateurs semi-paramétriques considérés) est plus proche de la vraie densité que l'estimateur donné en utilisant le noyau BS (le meilleur parmi les estimateurs non-paramétriques considérés). De plus, à partir du Tableau 5.10, nous remarquons que la borne de stabilité Γ_1 obtenue avec l'approche semi-paramétrique est meilleure que celle obtenue par la méthode non-paramétrique. Cela peut s'expliquer par le fait que la transformation de Champernowne permet de mieux utiliser et capturer toutes les données qui se trouvent dans la queue des distributions.

- On remarque d'après les tableaux " 5.2, 5.3, et 5.10 " que les bornes de stabilité Γ_1 obtenues pour la loi lognormale à l'aide de l'estimateur à noyau non paramétrique BS et de l'estimateur semi-paramétrique à noyau Mic- β 2 sont plus proches de la borne de stabilité théorique par rapport à celles obtenues par les autres lois et les autres estimateurs.
- Notons que pour la comparaison de la borne de stabilité Γ_1 entre la méthode non-paramétrique et l'approche semi-paramétrique, nous n'avons pas pris les mêmes noyaux, car les principes des deux techniques sont différents. En effet, dans le cas non-paramétrique, puisque nos données originales sont dans \mathbb{R}^+ , alors les estimateurs à noyaux BS, Gamma et RIG sont plus appropriées que l'estimateur à noyau Beta, qui est par conséquent plus approprié que les estimateurs à noyaux BS, Gamma et RIG pour les données transformées définies sur $[0, 1]$ en utilisant la transformation de Champernowne dans le cas semi-paramétrique.
- Le critère d'erreur quadratique moyenne (MSE) entre les estimateurs KMCE et la densité théorique, présenté dans le Tableau 5.4, confirme les résultats obtenus dans les graphiques 5.2, 5.5, 5.8 et 5.11 (c'est-à-dire que les densités estimées sont proches de la densité théorique). De plus, MSE diminue lorsque la taille de l'échantillon augmente. Ceci valide et sconsolide le choix de notre approche semi-paramétrique.
- On constate dans tous les cas, que contrairement aux bornes de stabilité Γ_1 obtenues à l'aide des estimateurs non paramétriques, celles obtenues à l'aide des estimateurs semi-paramétriques convergent rapidement vers la borne de stabilité théorique Γ_1 en fonction de la taille de l'échantillon n (voir Tableau 5.10), à savoir :

$$\Gamma_{1\text{semi-paramétrique}} \rightarrow \Gamma_{1\text{théorique}}, \text{ quand } n \rightarrow \infty.$$

Cela prouve que l'approche semi-paramétrique que nous proposons est satisfaisante puisqu'elle améliore la qualité de la borne de stabilité Γ_1 .

5.3 Version 2 : Stabilité d'un processus de Lindley

Algorithm 2 Algorithme 2: Stabilité d'un processus de Lindley

1. Introduire les paramètres λ, μ, c du modèle idéal, et λ', μ', c' du modèle perturbé (réel).
2. Vérifier la positivité des charges de sécurité relatives θ et θ' définies par : $\theta = \frac{c-\lambda\mu}{\lambda\mu}$ et $\theta' = \frac{c'-\lambda'\mu'}{\lambda'\mu'}$;
Si oui, ("la ruine des modèles n'est pas certaine") passer à l'étape 3 ;
Sinon retourner à l'étape 1.
3. Initialiser ε ($\varepsilon > 0$) de sorte que $0 < \rho(\varepsilon) < 1$ et Γ_2 soit minimale.
4. Calculer $\mu(a, a') = 2\mathbb{E}e^{\varepsilon Z} \left| \ln \frac{\lambda'\mu'}{\lambda\mu} \right| + \int_0^{\infty} v(u) - f'(u) du$, et tester : $\mu(a, a') < (1 - \rho(\varepsilon))$.
Si oui, ("nous pouvons déduire l'inégalité de stabilité forte") passer à l'étape 5; sinon augmenter ε avec le pas p , puis retourner à l'étape 4.
5. Calculer la borne de l'écart $\|\Psi_a - \Psi_{a'}\|_v$:

$$\|\Psi_a - \Psi_{a'}\|_v \leq \frac{\mu(a, a')(2 - \rho(\varepsilon))}{(1 - \rho(\varepsilon))(1 - \rho(\varepsilon) - \mu(a, a'))^2} = \Gamma_2.$$

L'étape 4 peut être développée en sous étapes (4.a – 4.e), similairement à l'Algorithme 1

Pour cette application, nous avons fixé les paramètres de simulation comme pour la première application tels que : la taille de l'échantillon $n = \{50, 100, 200, 500\}$, le nombre de répliques $m = 2000$ (pour appliquer le principe de Monte Carlo), le paramètre λ est 20 et le taux de prime c est égal à 100.

Résultats graphiques et numériques

Cette section est consacrée à la présentation des différents résultats numériques et graphiques obtenus lors de l'application de la méthode semi-paramétrique pour l'estimation de l'erreur d'approximation entre le modèle idéal et le modèle perturbé en évaluant l'erreur de stabilité basée sur le processus de Lindley, donnée par l'équation (2.24). Les résultats des simulations sont présentés dans les figures et les tableaux suivants :

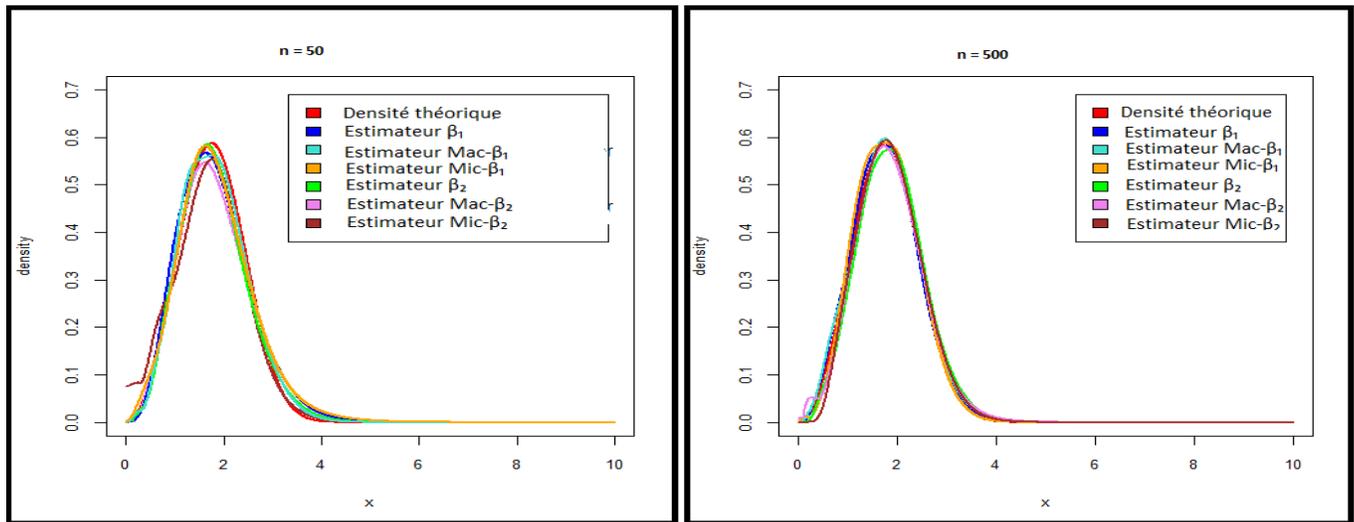


FIG. 5.16: Les estimateurs finaux de KMCE et la densité réelle pour la distribution de Weibull.

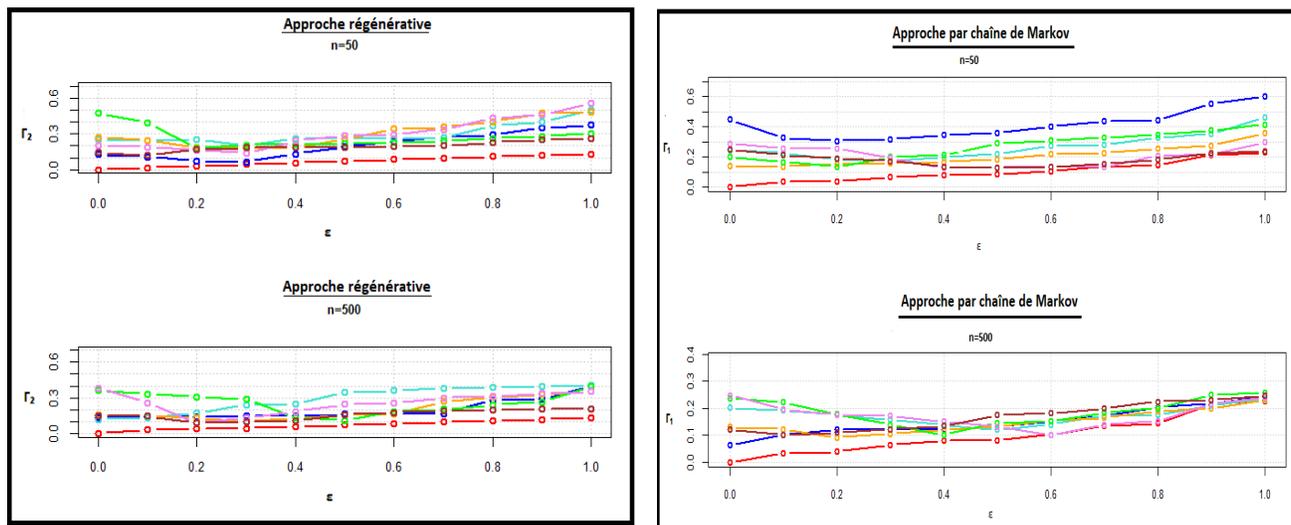


FIG. 5.17: Erreur Γ_2 (gauche) (Γ_1 (droite)) en fonction de ϵ avec la distribution de Weibull. Estimateur β_1 (bleu), estimateur Mac- β_1 (turquoise), estimateur Mic- β_1 (orange), estimateur β_2 (vert), estimateur Mac- β_2 (violet), estimateur Mic- β_2 (marron).

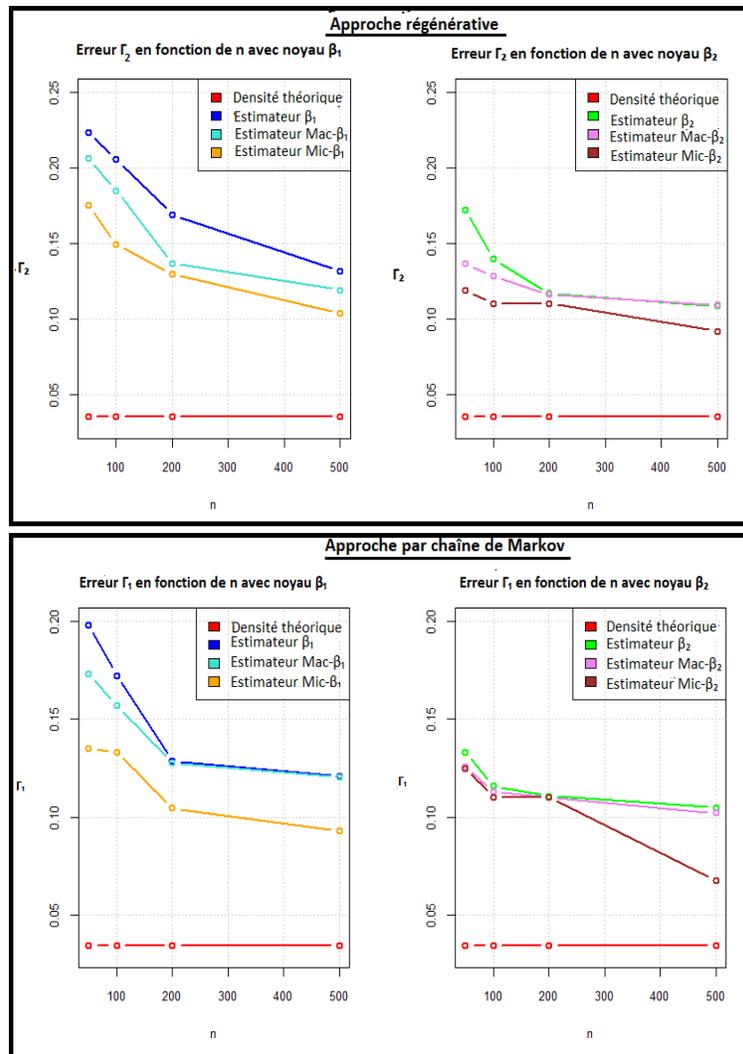


FIG. 5.18: Erreur Γ_2 (haut) (Γ_1 (bas)) en fonction de n avec la distribution de Weibull.

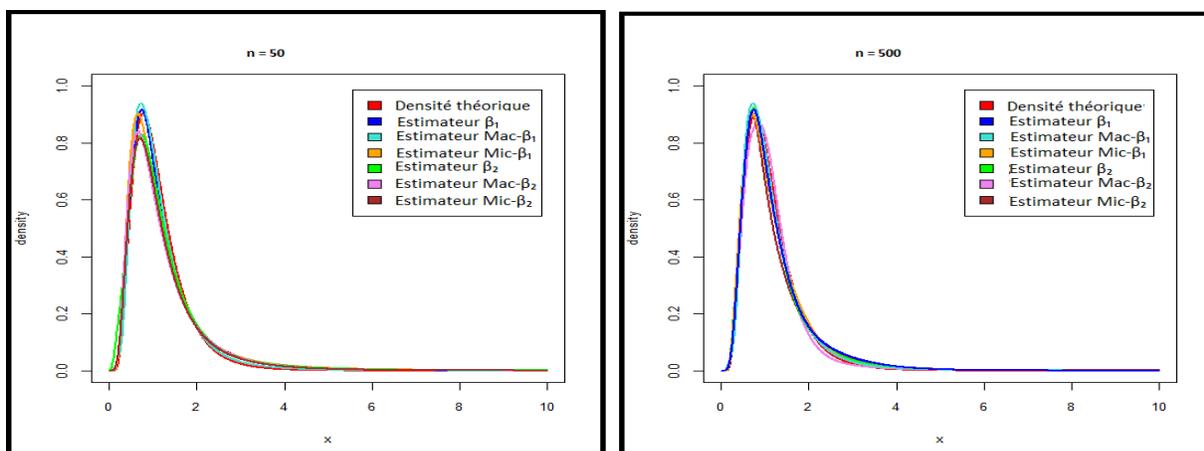


FIG. 5.19: Les estimateurs KMCE finaux et la densité réelle pour la distribution lognormale.

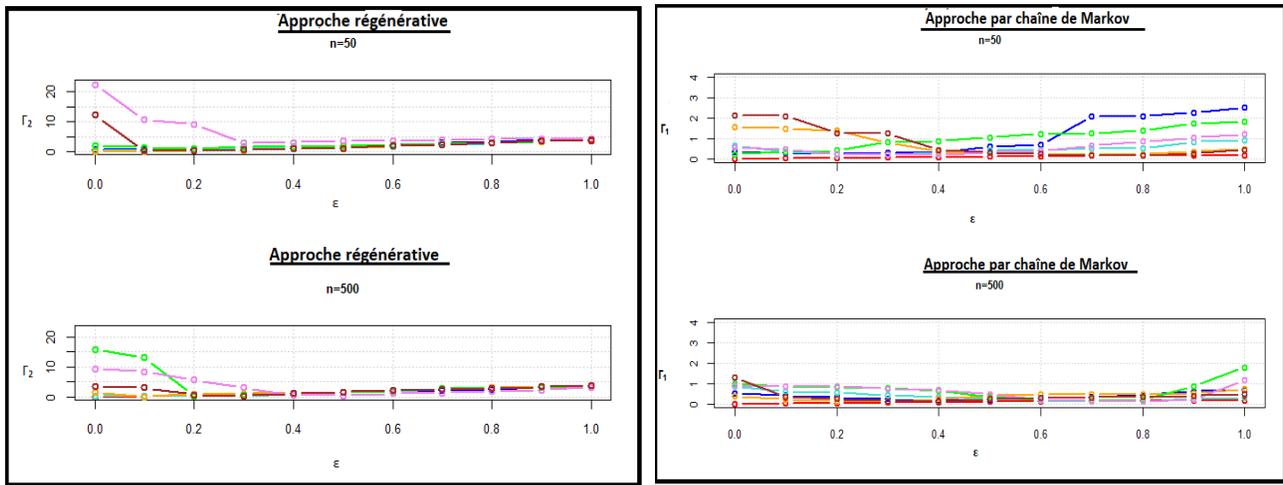


FIG. 5.20: Erreur Γ_2 (gauche) (Γ_1 (droite)) en fonction de ϵ avec la distribution lognormale. Estimateur de β_1 (bleu), estimateur de Mac- β_1 (turquoise), estimateur de Mic- β_1 (orange), estimateur de β_2 (vert), estimateur de Mac- β_2 (violet), estimateur de Mic- β_2 (marron).

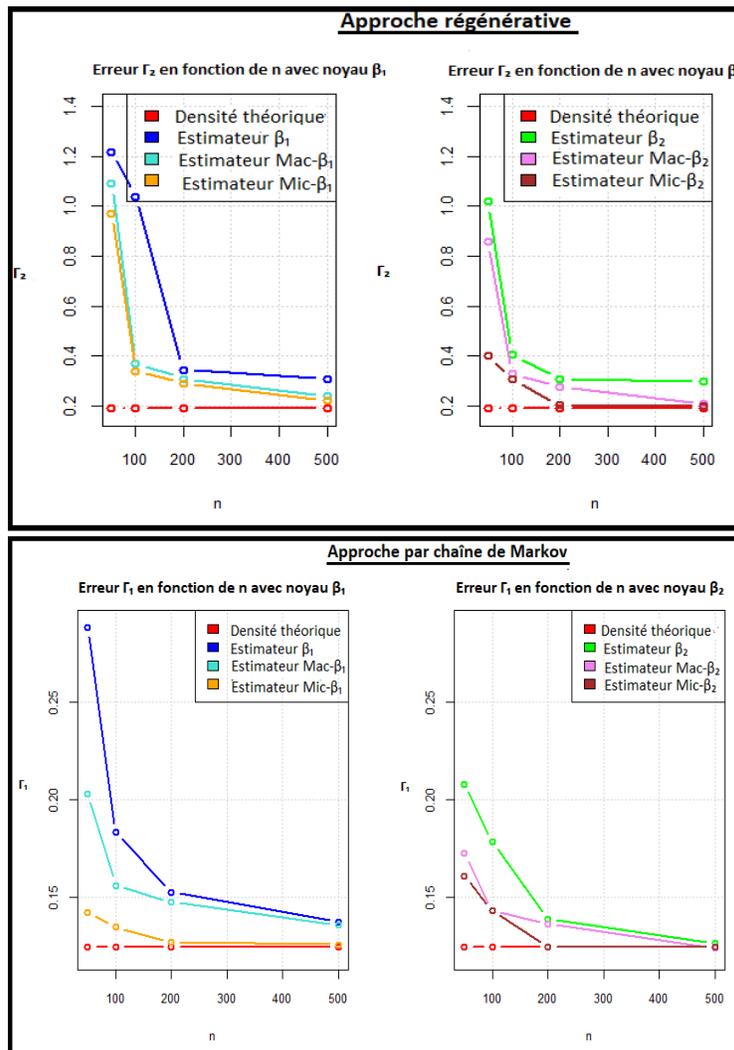


FIG. 5.21: Erreur Γ_2 (haut) (Γ_1 (bas)) en fonction de n avec la distribution lognormale.

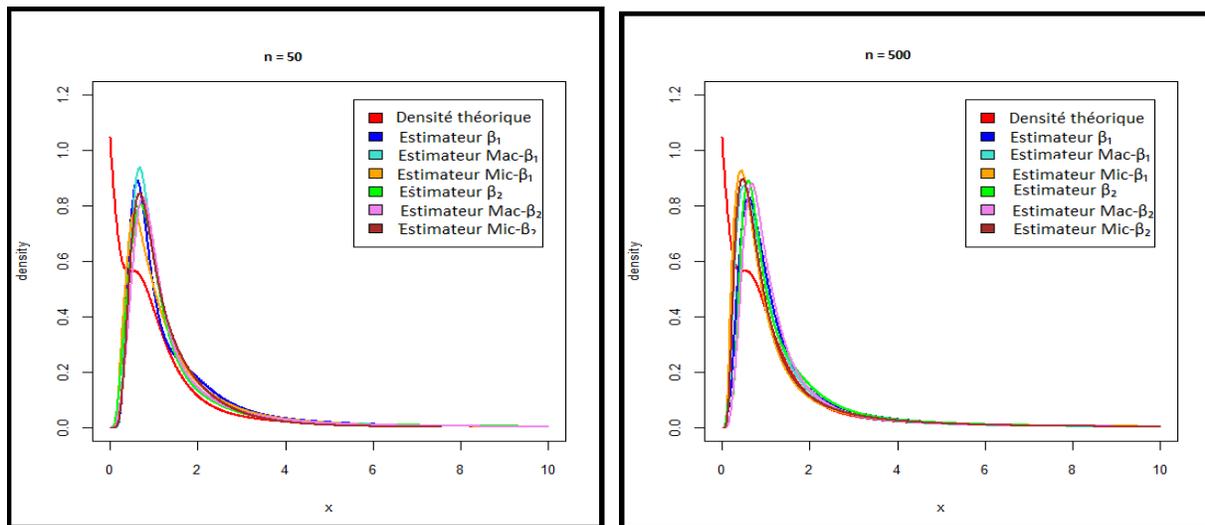


FIG. 5.22: Les estimateurs KMCE finaux et la densité réelle pour la distribution mélange Pareto-lognormale 30%70%.

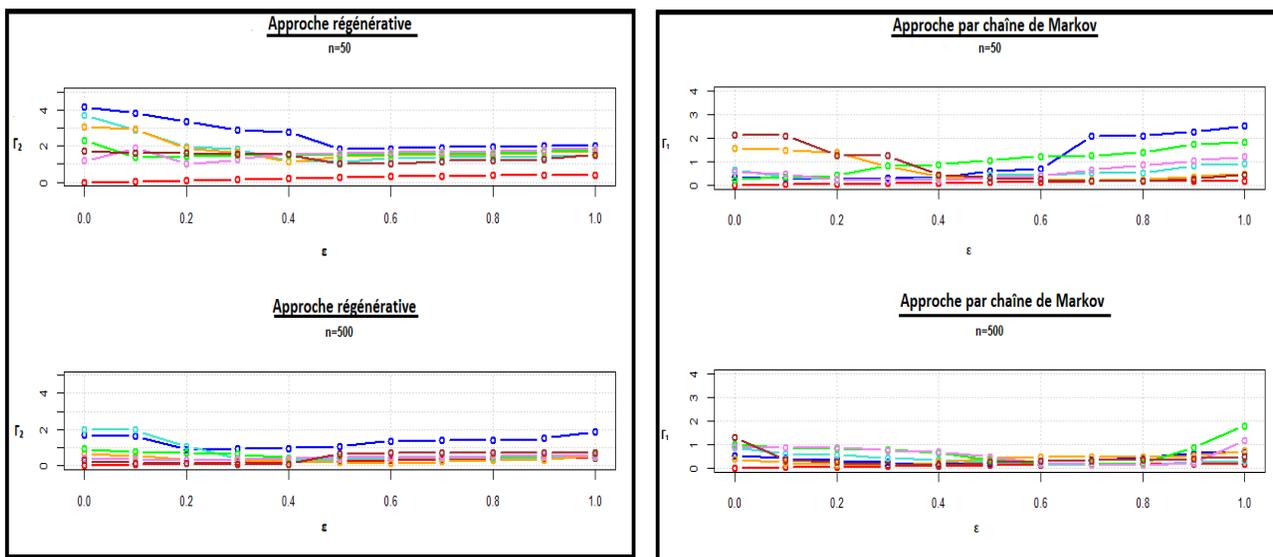


FIG. 5.23: Erreur Γ_2 (gauche) (Γ_1 (droite)) en fonction de ϵ avec la distribution mélange Pareto-lognormale 30%70%. Estimateur β_1 (bleu), estimateur Mac- β_1 (turquoise), estimateur Mic- β_1 (orange), estimateur β_2 (vert), estimateur Mac- β_2 (violet), estimateur Mic- β_2 (marron).

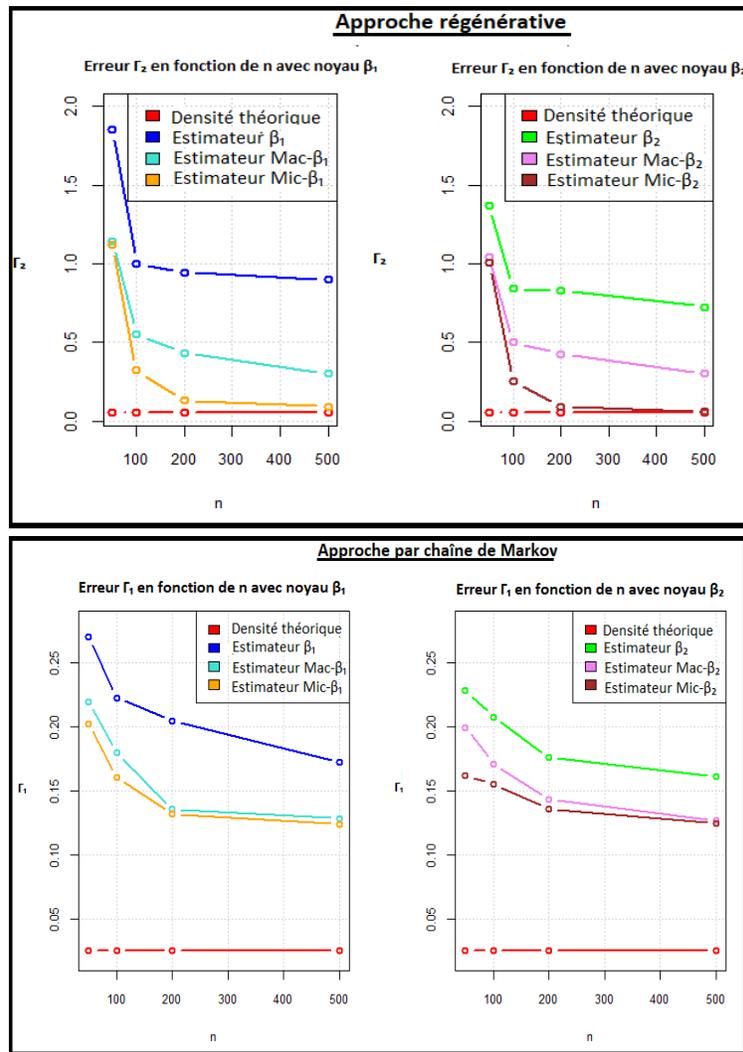


FIG. 5.24: Erreur Γ_2 (haut) (Γ_1 (bas)) en fonction de n avec la distribution mélange Pareto-lognormale 30%70%.

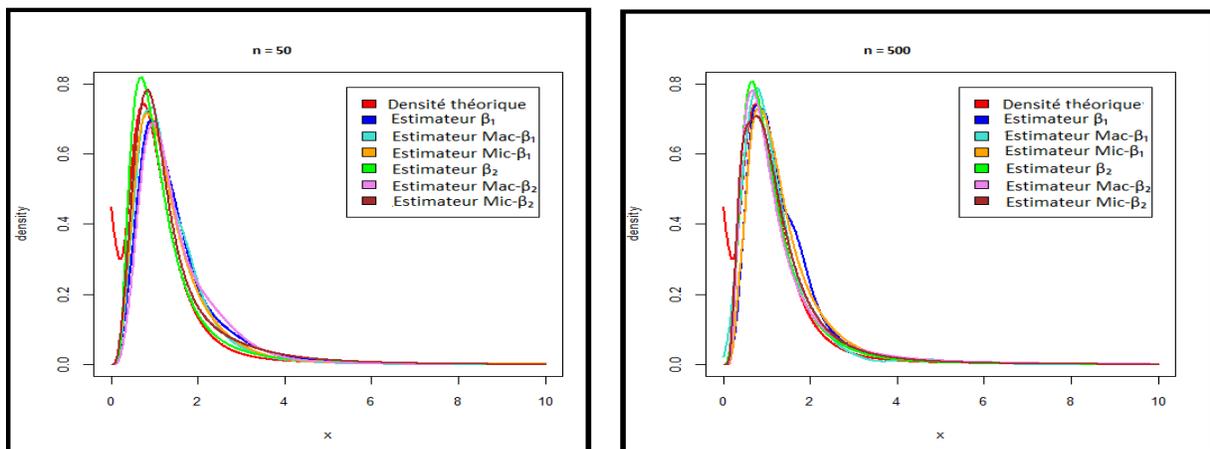


FIG. 5.25: Les estimateurs KMCE finaux et la densité réelle pour la distribution mélange Pareto-lognormale 70%30%.

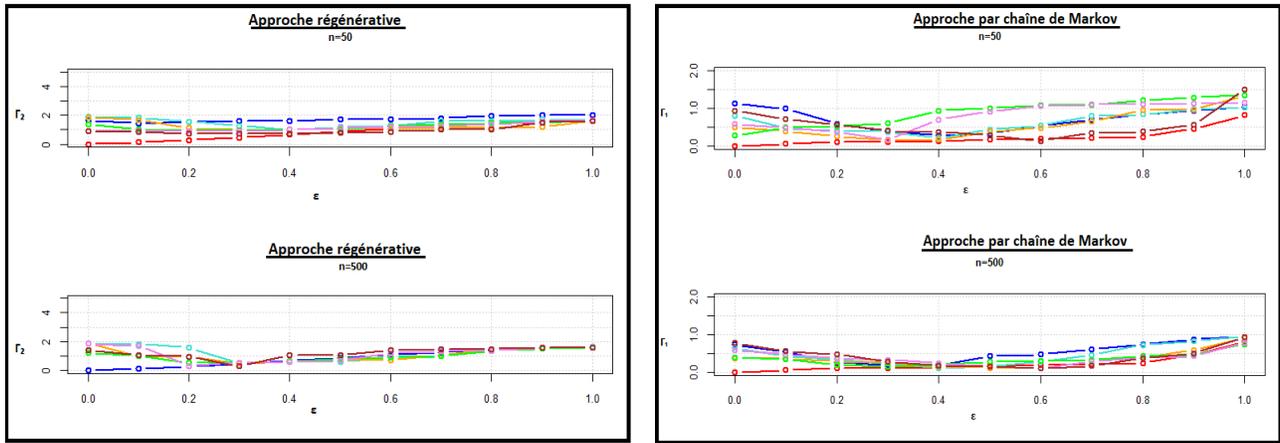


FIG. 5.26: Erreur Γ_2 (gauche) (Γ_1 (droite)) en fonction de ϵ avec la distribution mélange Pareto-lognormale 70%30%. Estimateur β_1 (bleu), estimateur Mac- β_1 (turquoise), estimateur Mic- β_1 (orange), estimateur β_2 (vert), estimateur Mac- β_2 (violet), estimateur Mic- β_2 (marron).

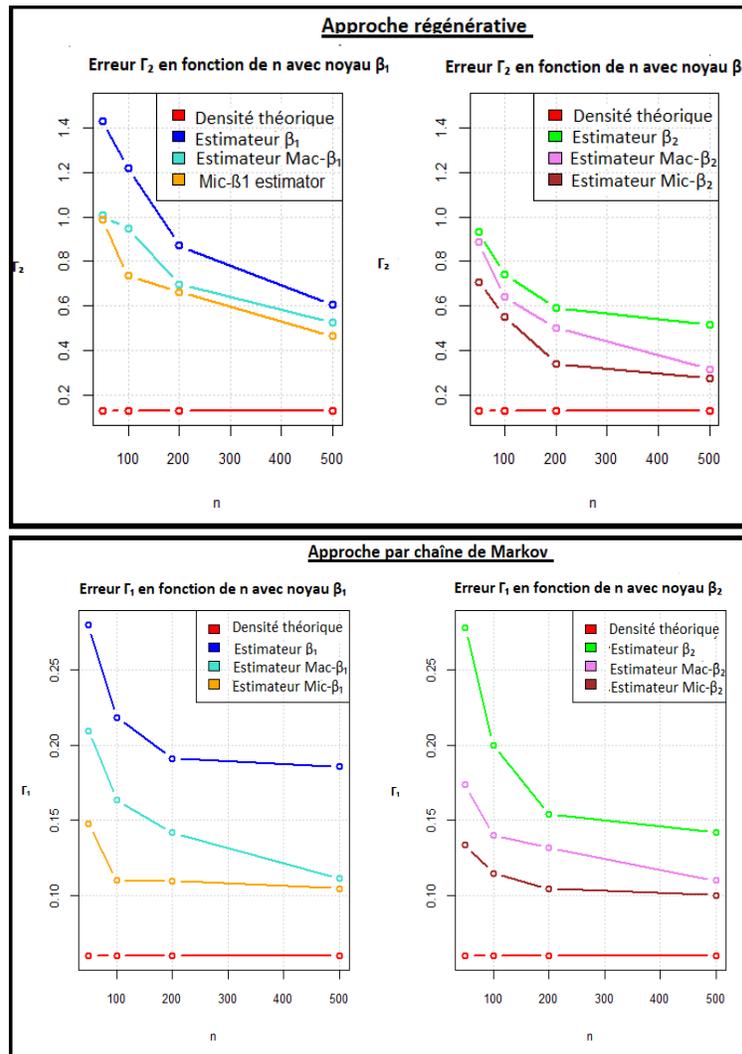


FIG. 5.27: Erreur Γ_2 (haut) (Γ_1 (bas)) en fonction de n avec la distribution mélange Pareto-lognormale 70%30%.

loi	n	$(\hat{\alpha}, \hat{M}, \hat{c})$
Wbl	50	(4.36828,1.77014,4.460991)
	100	(4.364112,1.771837,4.453584)
	200	(4.367716,1.770588,4.459425)
	500	(4.367499,1.770146,4.458352)
LogN	50	(0.04290641,0.01283803,0.04838946)
	100	(1.02545,1.000178,1.64254)
	200	(1.2545,1.106819,1.142072)
	500	(1.813884,1.105692,0.90422)
Mixt1	50	(0.04491615,0.04483989,0.0001191531)
	100	(0.07346703,0.04210168,0.060456)
	200	(1.885841,1.237261,1.776345)
	500	(1.630318,1.235095,0.5018638)
Mixt2	50	(0.07856115,0.02881375,0.03157385)
	100	(0.6284359,0.2947132,1.243975)
	200	(2.228614,1.272742,2.802394)
	500	(2.617114,1.172916,0.5495472)

Tableau 5.6: Estimation du paramètre $\theta = (\alpha, M, c)$ de la loi de Champernowne.

loi	n	$\hat{f}_{\beta 1}$		$\hat{f}_{Mac-\beta 1}$		$\hat{f}_{Mic-\beta 1}$		Γ_2 théorique
		h_n	Γ_2	h_n	Γ_2	h_n	Γ_2	
Wbl	50	0.03311009	0.2237045	0.030041129	0.20612252	0.03113344	0.17535606	0.03540733
	100	0.02291193	0.2054549	0.0228812	0.18505807	0.02546482	0.14973065	
	200	0.01749037	0.16925314	0.01848297	0.1367961	0.01732323	0.12996181	
	500	0.01191691	0.13190402	0.01198777	0.1193461	0.01248542	0.10395413	
LogN	50	0.02641758	1.21885	0.02600832	1.0896837	0.02357097	0.9718463	0.192847
	100	0.01900493	1.0385149	0.02076017	0.3707077	0.01945697	0.3398275	
	200	0.014377	0.344652	0.0133734	0.308795	0.01527376	0.28898286	
	500	0.1002556	0.3064222	0.010842309	0.2409755	0.01000516	0.2233451	
Mixt1	50	0.0335143	1.849355	0.0306601	1.139878	0.02392758	1.116968	0.05134723
	100	0.02176958	1.0020243	0.01921638	0.55309	0.0220048	0.322815	
	200	0.01178713	0.943303	0.01240313	0.432	0.01184386	0.12915	
	500	0.010745369	0.8974281	0.01068915	0.302	0.01025412	0.0911782	
Mixt2	50	0.036621134	1.431662	0.0334246	1.0069731	0.0397406	0.9864444	0.1281555
	100	0.02304693	1.217603	0.0273001	0.947537	0.01865918	0.735397	
	200	0.01976103	0.8711975	0.01253574	0.6969869	0.01457835	0.66349	
	500	0.01467545	0.6034079	0.01668522	0.5236578	0.01067268	0.4645264	

Tableau 5.7: Borne de stabilité Γ_2 via l'estimateur semi-paramétrique à noyau Beta 1 et ses versions normalisées.

loi	n	$\hat{f}_{\beta 2}$		$\hat{f}_{Mac-\beta 2}$		$\hat{f}_{Mic-\beta 2}$		Γ_2 théorique
		h_n	Γ_2	h_n	Γ_2	h_n	Γ_2	
Wbl	50	0.03457173	0.172281	0.03605102	0.136804	0.02853472	0.11919078	0.03540733
	100	0.02429661	0.1396848	0.02213863	0.12830851	0.02364514	0.11041772	
	200	0.01779244	0.11695428	0.01713474	0.11634619	0.01707127	0.11024113	
	500	0.01242709	0.1086956	0.01265158	0.0932046	0.011940261176957	0.09194366	
LogN	50	0.03630047	1.109533	0.034543	0.859464	0.03410212	0.417276	0.192847
	100	0.02561231	0.405837	0.02635094	0.3295542	0.02607184	0.3072363	
	200	0.01990918	0.3067672	0.01980111	0.2745395	0.02019111	0.2022039	
	500	0.01349426	0.296332	0.01382607	0.2095072	0.01339296	0.1996923	
Mixt1	50	0.03258238	1.369541	0.02329622	1.014397	0.02200726	1.0052472	0.05134723
	100	0.02145276	0.8415065	0.02339274	0.5000314	0.02268547	0.2518298	
	200	0.01083799	0.829198	0.01056232	0.426701	0.01134343	0.1917751	
	500	0.0100523177	0.725061	0.012645	0.3000332	0.016245829	0.0625424	
Mixt2	50	0.02306241	0.9306329	0.03339308	0.887353	0.02673394	0.707389	0.1281555
	100	0.01582547	0.7443261	0.01651221	0.638553	0.01914384	0.552469	
	200	0.01305006	0.5894007	0.01381022	0.5021367	0.01914384	0.340963	
	500	0.01863807	0.512991	0.01794989	0.3148249	0.01770174	0.273109	

Tableau 5.8: Borne de stabilité Γ_2 via l'estimateur semi-paramétrique à noyau Beta 2 et ses versions normalisées.

Estimateur	n	Wbl	LogN	Mixt1	Mixt2
$\hat{f}_{\beta 1}$	50	0.0007175608	0.000491315	0.008692001	0.004458542
	100	0.00003536721	0.0003120278	0.006150575	0.002523003
	200	0.00002503096	0.0001602303	0.00201052	0.001360483
	500	0.00001499054	0.0001331831	0.001197431	0.00120962
$\hat{f}_{\beta 2}$	50	0.05039347	0.003235766	0.0409701	0.02488146
	100	0.00427792	0.0001336867	0.02317183	0.00182045
	200	0.0002131846	0.0001906973	0.01659886	0.001465596
	500	0.000120557	0.0001884485	0.00134573	0.0001536603
$\hat{f}_{Mac-\beta 1}$	50	0.0004354851	0.0003064835	0.034376	0.02125433
	100	0.0002963337	0.0002692713	0.02957519	0.001877985
	200	0.0001293467	0.000173829	0.001892121	0.001556334
	500	0.0001117232	0.0001034084	0.00128569	0.001188483
$\hat{f}_{Mac-\beta 2}$	50	0.01381153	0.03442029	0.03138848	0.02641953
	100	0.001019789	0.02989985	0.02032918	0.016575
	200	0.0001243429	0.01552022	0.0150651	0.00149542
	500	0.00010132749	0.001709496	0.0011286	0.0001145577
$\hat{f}_{Mic-\beta 1}$	50	0.0002514937	0.0003604957	0.002986271	0.002844546
	100	0.000215793	0.0002851623	0.002260912	0.001321864
	200	0.0001336913	0.0001601206	0.000163269	0.0001231431
	500	0.00001022522	0.000102646	0.0001384901	0.0001079587
$\hat{f}_{Mic-\beta 2}$	50	0.0004883676	0.00326221	0.0209072	0.01532586
	100	0.0002844994	0.001452495	0.01935537	0.001913954
	200	0.0001758377	0.001209386	0.01154316	0.001050235
	500	0.00001455245	0.0000159448	0.0001181286	0.0000110315

Tableau 5.9: L'erreur quadratique moyenne (MSE) avec les estimateurs étudiés.

Perturbation (ϵ)	Approche régénératrice			Approche par chaîne de Markov				
	Intervalle de stabilité			Intervalle de stabilité				
	Wbl	LogN	Mixt1	Mixt2	Wbl	LogN	Mixt1	Mixt2
0.1	[1e - 04, 1.7341]	[0.0001, 1.8246]	[0.0001, 3.674]	[0.0001, 3.1339]	[0.001, 0.0856]	[0.0024, 0.2296]	[0.0026, 2.2157]	[0.0001, 0.638]
0.2	[1e - 04, 1.5282]	[0.0036, 1.8002]	[0.0001, 3.4476]	[0.0001, 2.1286]	[0.001, 0.0811]	[0.0035, 0.2154]	[0.0017, 2.0892]	[0.0001, 0.5868]
0.3	[1e - 04, 1.1063]	[0.004, 1.5047]	[0.0001, 2.744]	[0.0001, 2.0171]	[0.0009, 0.0561]	[0.0046, 0.2024]	[0.0013, 1.2557]	[0.0001, 0.0876]
0.4	[1e - 04, 0.7675]	[0.005, 1.4036]	[0.0001, 0.779]	[0.0001, 2.0156]	[0.0009, 0.0535]	[0.0054, 0.1955]	[0.0001, 0.8761]	[0.0005, 0.057]
0.5	[1e - 04, 0.7654]	[0.0056, 1.3581]	[0.0001, 0.5631]	[0.0001, 1.0383]	[0.0009, 0.0425]	[0.0057, 0.1355]	[0.0001, 0.7312]	[0.0011, 0.0402]
0.6	[1e - 04, 0.5222]	[0.0066, 1.2296]	[0.0001, 0.5017]	[0.0001, 1.0696]	[0.0009, 0.0246]	[0.0066, 0.0751]	[0.0008, 0.7057]	[0.0017, 0.0305]
0.7	[1e - 04, 0.5068]	[0.007, 1.1475]	[0.0001, 0.4236]	[0.0001, 1.0696]	[0.0009, 0.0203]	[0.0069, 0.0451]	[0.0009, 0.4518]	[0.0026, 0.026]
0.8	[1e - 04, 0.2882]	[0.0074, 0.6544]	[0.0001, 0.4236]	[0.0001, 0.8683]	[0.0009, 0.0145]	[0.0073, 0.0354]	[0.0014, 0.3933]	[0.0023, 0.0142]
0.9	[1e - 04, 0.2548]	[0.0076, 0.6258]	[0.0001, 0.3102]	[0.0001, 0.7822]	[0.0009, 0.0120]	[0.0077, 0.0314]	[0.0003, 0.3097]	[0.0032, 0.0065]
1	[1e - 04, 0.2491]	[0.0081, 0.3584]	[0.0001, 0.2281]	[0.0001, 0.3107]	[0.0009, 0.0061]	[0.0079, 0.0248]	[0.0032, 0.0041]	[0.0035, 0.0059]

Tableau 5.10: Une étude comparative des intervalles de stabilité entre l'approche régénératrice et l'approche par chaîne de Markov.

loi	Approche MC										Γ_1 théorique	
	Γ_1 non-parametric					Γ_1 semi-paramétrique						
	n	\hat{f}_{GAM}	\hat{f}_{BS}	\hat{f}_{RIG}	n	\hat{f}_{β_1}	$\hat{f}_{Mac-\beta_1}$	$\hat{f}_{Mic-\beta_1}$	\hat{f}_{β_2}	$\hat{f}_{Mac-\beta_2}$		$\hat{f}_{Mic-\beta_2}$
Wbl	50	0.479156	0.4448456	0.5734506	50	0.1982505	0.17306221	0.1349849	0.13311	0.1255507	0.12488363	0.03400484
	100	0.4157367	0.349418	0.451785	100	0.172401	0.1570731	0.1331169	0.11590054	0.11278563	0.110136606	
	500	0.3760803	0.3213912	0.39251936	500	0.12069179	0.12043856	0.09281798	0.1048571	0.1020469	0.067412544	
LogN	50	0.27889	0.25281	0.32311515	50	0.2879552	0.2028934	0.1425789	0.2080006	0.17259018	0.16083142	0.1244944
	100	0.250341	0.238752	0.3096398	100	0.1832802	0.156224	0.13490687	0.1783532	0.14322155	0.14313448	
	500	0.23818	0.1460287	0.1494301	500	0.137591	0.1360286	0.1259399	0.1266003	0.1252297	0.12449569	
Mixt1	50	13.75448	0.4219374	1.29846301	50	0.2700395	0.21963254	0.2023622	0.228267	0.1992561	0.161668	0.06048445
	100	8.297866	0.3836091	1.118366	100	0.2228498	0.1798832	0.16036554	0.207791	0.1711076	0.1552654	
	500	1.9972	0.322052840	0.7188342	500	0.17203618	0.12841768	0.1238103	0.1609048	0.126781	0.1244500223	
Mixt2	50	2.35146	0.4266026	0.6608563	50	0.27920345	0.2094644	0.1476961	0.277258	0.1735601	0.133898	0.02506683
	100	2.018099	0.287889	0.4801103	100	0.2181814	0.1636112	0.11037798	0.1998723	0.140059	0.114973211	
	500	1.058258	0.208721	0.321735	500	0.18555563	0.111715	0.104934546	0.142111	0.11014746	0.1004294	

Tableau 5.11: Une étude comparative de la borne Γ_1 entre les méthodes non paramétriques et semi-paramétriques.

Loi	Approche régénérative									Γ_2 théorique	
	Γ_2 non-paramétrique				Γ_2 semi-paramétrique						
	n	\hat{f}_{GAM}	\hat{f}_{BS}	\hat{f}_{RIG}	\hat{f}_{β_1}	$\hat{f}_{Mac-\beta_1}$	$\hat{f}_{Mic-\beta_1}$	\hat{f}_{β_2}	$\hat{f}_{Mac-\beta_2}$		$\hat{f}_{Mic-\beta_2}$
Wbl	50	1.212405	0.6132141	1.04281348	0.2237045	0.20612252	0.17535606	0.172281	0.136804	0.11919078	0.03540733
	100	1.1120291	0.5494258	0.9347412	0.2054549	0.18505807	0.14973065	0.1396848	0.12830851	0.11041772	
	500	0.83681303	0.4800267	0.636936	0.13190402	0.1193461	0.10395413	0.1086956	0.0932046	0.09194366	
LogN	50	3.0260807	2.4039704	3.380544	1.21885	1.0896837	0.9718463	1.109533	0.859464	0.417276	0.192847
	100	2.8575701	1.708962	2.570929	1.0385149	0.3707077	0.3398275	0.405837	0.3295542	0.3072363	
	500	2.4753365	1.158653	1.5821225	0.3064222	0.2409755	0.2233451	0.296332	0.2095072	0.1996923	
Mixt1	50	20.8551049	10.341662	15.11901	1.849355	1.139878	1.116968	1.369541	1.014397	1.0052472	0.05134723
	100	10.423328	1.992304	5.200471	1.0020243	0.55309	0.322815	0.8415065	0.5000314	0.2518298	
	500	3.826074	1.359131	3.572913	0.8974281	0.302	0.0911782	0.725061	0.3000332	0.0625424	
Mixt2	50	20.6216408	1.819389	1.435319	1.431662	1.0069731	0.9864444	0.9306329	0.887353	0.707389	0.1281555
	100	4.940783	0.9811951	1.04317138	1.217603	0.947537	0.735397	0.7443261	0.638553	0.552469	
	500	1.059317	0.775409	0.957619	0.6034079	0.5236578	0.4645264	0.512991	0.3148249	0.273109	

Tableau 5.12: Une étude comparative de la borne Γ_2 entre les méthodes non paramétriques et semi-paramétriques.

Discussion des résultats

Dans cette section, nous comparons les deux versions de stabilité obtenues par l'approche de chaîne de Markov Γ_1 (voir la section 5.2) et par approche régénérative Γ_2 (voir la section 5.3). Nous obtenons les résultats suivants :

- Nous notons dans les tableaux "5.7, 5.8 et 5.12" que la borne de stabilité de l'approche régénérative Γ_2 diminue avec l'augmentation de la taille de l'échantillon n pour l'approche

semi-paramétrique aussi bien que pour la méthode non-paramétrique. De plus, elle se rapproche de l'erreur théorique $\Gamma_{2_{théorique}}$.

- D'après les figures "5.16, 5.19, 5.22 et 5.25", nous remarquons que la densité de l'estimateur KMCE pour toutes les distributions est proche de la densité réelle. Ceci est bien confirmé par le critère d'erreur quadratique moyenne (MSE) présenté dans le Tableau 5.9 (le MSE diminue avec l'augmentation de la taille de l'échantillon).
- D'après les figures "5.18, 5.21, 5.24 et 5.27", nous remarquons que pour les deux versions de stabilité, les bornes de stabilité Γ_1 et Γ_2 diminuent avec l'augmentation de la taille de l'échantillon, mais nous remarquons que Γ_2 est plus proche de la borne de stabilité théorique que la borne Γ_1 .
- Nous remarquons également que les bornes de stabilité Γ_1 et Γ_2 augmentent rapidement au voisinage de la borne supérieure du domaine de variation epsilon (voir figures "5.17, 5.20, 5.23 et 5.26"), c'est-à-dire qu'au voisinage de sa borne supérieure, Γ_1 et Γ_2 prennent des valeurs plus grandes et s'éloignent de l'écart des probabilités de ruine Γ_1 et Γ_2 respectivement.
- D'après les résultats donnés par le Théorème 2.4.4, on remarque que le domaine de variation d'epsilon (domaine de stabilité) est plus large dans l'approche régénérative et trop petit pour l'approche de chaîne de Markov. Ceci est dû à la sensibilité de la borne de stabilité Γ_1 par rapport à la variation d'epsilon et à la condition $(\mu(a, a') < (1 - \rho(\epsilon))^2)$ donnée par le Théorème 2.4.1 qui délimite le domaine de variation.
- Nous remarquons à partir des tableaux 5.11 et 5.12 que : Dans les deux versions de stabilité et pour toutes les distributions, les intervalles de stabilité diminuent avec l'augmentation de la perturbation ϵ . Il est évident que le modèle de risque tend à être instable avec une grande perturbation.
- Le Théorème 2.4.4 fournit des intervalles de stabilité plus grands que ceux donnés par le Théorème 2.4.1. Cela signifie que nous avons plus de flexibilité pour le choix de la norme $v(\epsilon)$ dans l'approche régénérative (Théorème 2.4.4) que dans l'approche par chaîne de Markov (Théorème 2.4.1). En d'autres termes, les conditions du Théorème 2.4.1 sont plus strictes que celles du Théorème 2.4.4.
- Le comportement de la borne de stabilité Γ_2 est similaire à celui de la borne Γ_1 . Cependant, l'erreur minimale Γ_2 fournie par le second théorème (Théorème 2.4.4) est plus grande que

l'erreur minimale Γ_1 donnée par le premier théorème (Théorème 2.4.1).

- D'après les résultats trouvés par les deux approches théorèmes (2.4.1 et 2.4.4), on remarque que pratiquement il est préférable d'utiliser l'approche par chaîne de Markov (Théorème 2.4.1) afin d'avoir une approximation raisonnable, puisqu'il donne des bornes d'erreur plus petites. Mais si les conditions de stabilité du Théorème 2.4.1 ne sont pas évidentes (pas facile à vérifier), alors on peut utiliser l'approche régénérative (Théorème 2.4.4) qui donne aussi des résultats satisfaisants (marge d'erreur). Cependant, pour éviter de procéder par essais et erreurs et ne pas perdre de temps, il peut être préférable d'utiliser directement le Théorème 2.4.4.

5.4 Conclusion

Dans ce chapitre, nous avons comparé entre deux bornes de stabilité forte, l'une est calculée par la méthode de chaîne de Markov et l'autre par l'approche régénérative. Pour chaque borne, nous comparons entre deux méthode d'estimation de la densité, la méthode semi-paramétrique et la méthode non-paramétrique. Ce travail de simulation montre bien l'importance de la méthode semi-paramétrique pour améliorer la qualité de la borne de stabilité forte par rapport à la méthode non paramétrique. Les résultats obtenus montrent aussi que les deux bornes de stabilité convergent dans un même sens (c'est-à-dire que le comportement des bornes de stabilité Γ_1 et Γ_2 sont similaires). Cependant, une petite préférence est attribuée à la stabilité d'une chaîne de Markov par rapport à la borne de stabilité d'un processus de Lindley, parce que la première borne est toujours plus petite que la seconde, mais comme les conditions de stabilité de l'approche par chaîne de Markov sont strictes, il est préférable d'utiliser directement l'approche régénérative.

Application numérique basée sur des données réelles

Sommaire

6.1	Introduction	93
6.2	Présentation des données (sinistres)	93
6.3	Analyse exploratoire des données	93
6.4	Application des méthodes d'estimation et d'approximation sur les sinistres	94
6.5	Conclusion	99

6.1 Introduction

Dans ce chapitre, nous allons utiliser également un estimateur à noyau appliqué à une transformation de données réelles. L'idée est d'utiliser une estimation paramétrique préliminaire de la distribution inconnue f comme fonction de transformation, puis d'utiliser un estimateur à noyau pour corriger cette estimation préliminaire.

6.2 Présentation des données (sinistres)

Dans la présente section, nous présentons un ensemble de données réelles correspondant aux sinistres d'une compagnie d'assurance australienne. Ces sinistres sont dus à des accidents survenus entre janvier 1998 et juin 1999. Les variables considérées sont notées : 'amount' (montant d'argent payé par une compagnie d'assurance en dollars australiens) (voir [72] et annexe A), qui seront analysées par l'approche que nous proposons : méthode de stabilité (par chaîne de Markov et processus régénératif) et méthodes d'estimation (semi-paramétrique et non paramétrique). La méthode du maximum de vraisemblance a été appliquée aux 542 données de sinistres corporels pour estimer les paramètres de la distribution de Champernowne généralisée. Nous obtenons $\hat{\alpha} = 3.4151$, $\hat{M} = 6765.25$ et $\hat{c} = 15.49685$. Des comparaisons entre les différentes approches seront réalisées.

6.3 Analyse exploratoire des données

Nous présentons ci-après le tableau des statistiques descriptives de l'ensemble des données présentées, qui comprend, entre autres, l'écart-type (SD) et les coefficients de variation (CV), d'asymétrie (CS) et d'aplatissement (CK) :

Observations (n)	Médiane	Moyenne	SD	CV	CS	CK	Min	Max
542	6765	8996.406	8790.712	97.71	5.58	51.11	109.000	116600.720

Tableau 6.1: Statistiques descriptives pour les données de sinistres d'une compagnie d'assurance australienne.

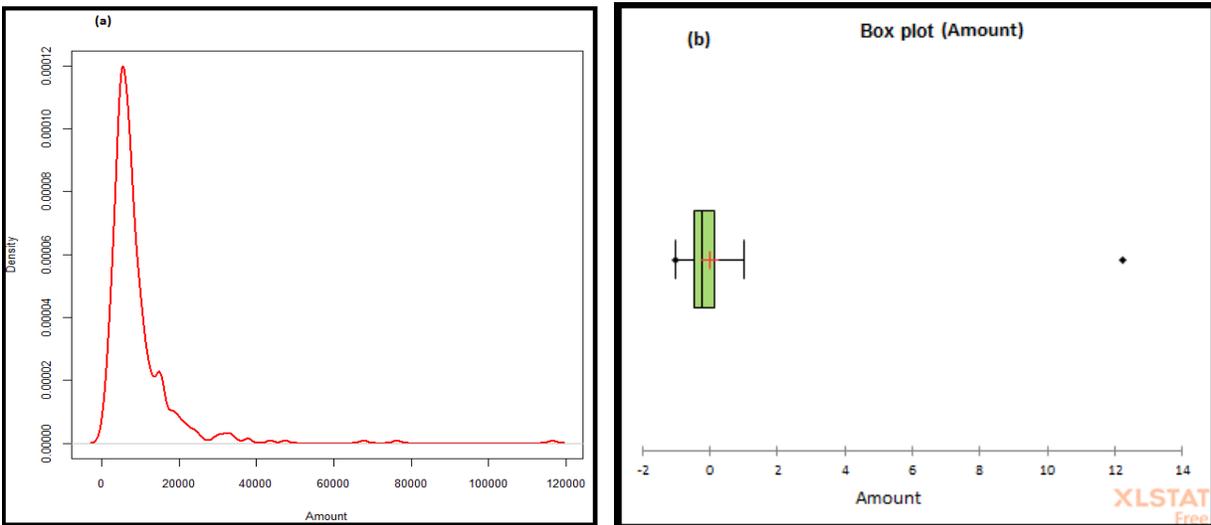


FIG. 6.1: Représentation graphique de la fonction de distribution de probabilité avec estimation du noyau (a) et boxplot (b) pour les données de sinistres.

6.4 Application des méthodes d'estimation et d'approximation sur les sinistres

Dans cette section, nous appliquons l'estimateur que nous avons (estimateur KMCE) aux montants des sinistres pour la couverture des dommages corporels, avec une taille d'échantillon $n = 542$, le paramètre $\lambda = 10$, le taux de prime $c = 100000$ et $\mu = 8996.406$. Le choix de la largeur de la fenêtre h_n sera fait avec la méthode de la règle du pouce définie dans la formule (5.1), on obtient $h_n = 0.02558697$.

L'environnement R est utilisé pour obtenir les résultats suivants. La distribution de Champnowne semble donc être un bon point de départ pour la modélisation des sinistres corporels.

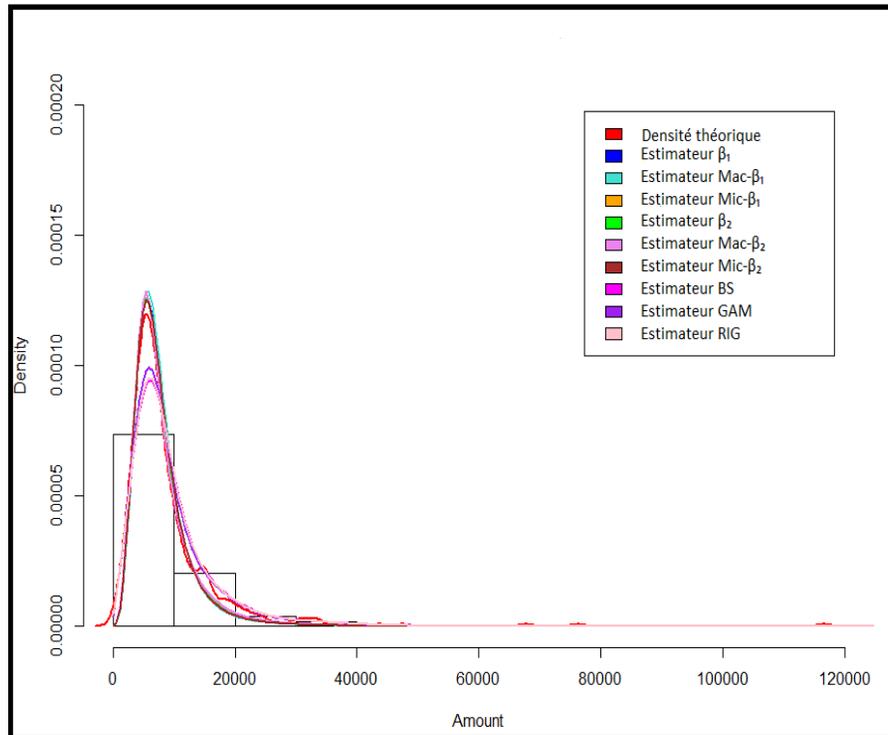


FIG. 6.2: Les estimateurs finaux KMCE, BS, GAM, RIG et la densité réelle pour les données de sinistres.

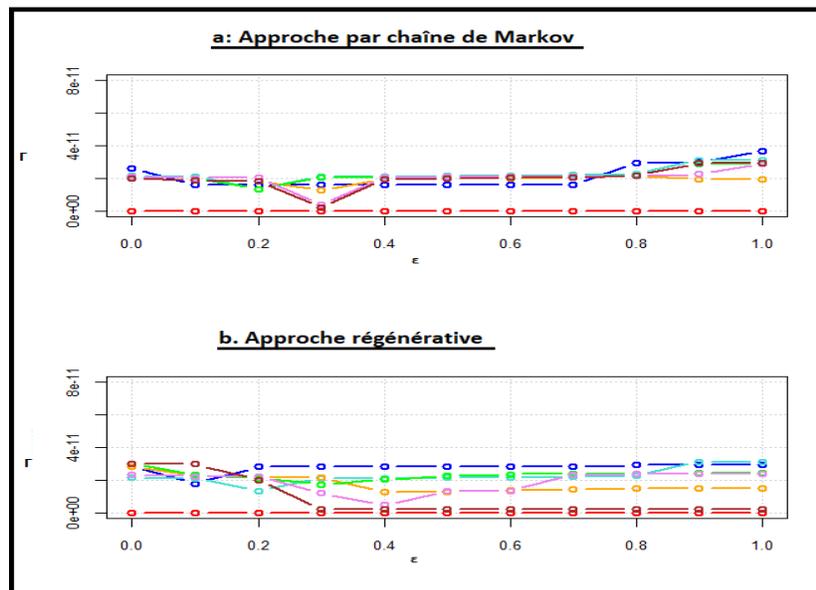


FIG. 6.3: Borne de stabilité Γ (approche régénérative (b) et approche par chaîne de Markov (a)) en fonction de ϵ avec les données réelles. Estimateur β_1 (bleu), estimateur Mac- β_1 (turquoise), estimateur Mic- β_1 (orange), estimateur β_2 (vert), estimateur Mac- β_2 (violet), estimateur Mic- β_2 (marron).

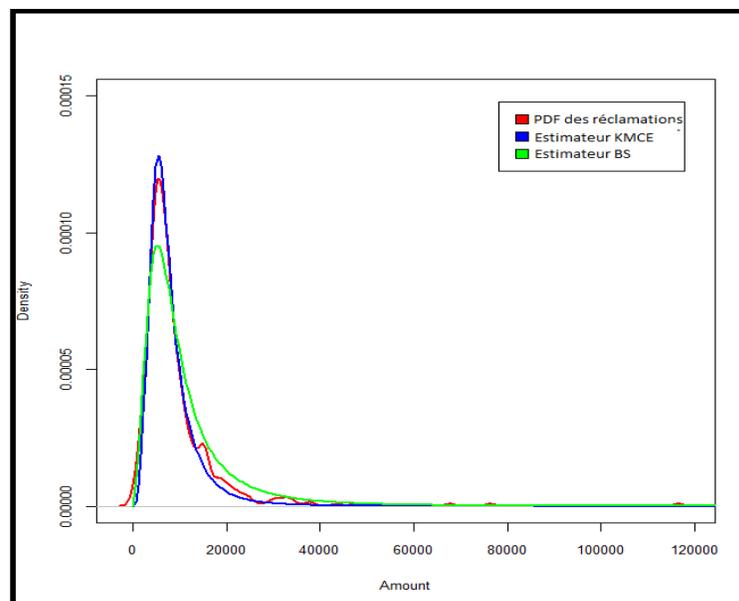


FIG. 6.4: L'estimateur KMCE final ($\hat{f}_{Mic-\beta 2}$), l'estimateur BS et la densité pour les données de sinistres.

	$\Gamma_{non-paramétrique}$			$\Gamma_{thorique}$
	\hat{f}_{GAM}	\hat{f}_{BS}	\hat{f}_{RIG}	
Approche régénérative	1.910783e - 5	3.140075e - 8	7.432278e - 6	5.293961e - 23
Approche chaîne de Markov	6.660041e - 7	2.204661e - 8	3.957162e - 6	5.059042e - 23

Tableau 6.2: Une étude comparative de la borne de stabilité Γ entre l'approche régénérative (RP) et l'approche par chaîne de Markov (MC) pour les données de sinistres : cas non-paramétrique.

	$\Gamma_{semi-paramétrique}$				$\Gamma_{thorique}$		
	\hat{f}_{β_1}	$\hat{f}_{Mac-\beta_1}$	$\hat{f}_{Mic-\beta_1}$	\hat{f}_{β_2}		$\hat{f}_{Mac-\beta_2}$	
Approche régénérative	7.795071e - 11	1.402177e - 11	1.290507e - 11	1.727212e - 11	4.640115e - 12	1.985772e - 12	5.293961e - 23
Approche chaîne de Markov	1.5813551e - 11	1.341769e - 11	1.267957e - 11	1.34605e - 11	4.067667e - 12	1.967957e - 12	5.059042e - 23

Tableau 6.3: Une étude comparative de la borne de stabilité Γ entre l'approche régénérative (RP) et l'approche par chaîne de Markov (MC) pour les données de sinistres : cas semi-paramétrique.

Perturbation (ϵ)	$\Gamma_{Non-paramtrique}$		$\Gamma_{Semi-paramtrique}$	
	RP	MC	RP	MC
0.0	$4.334189e - 08$	$2.204661e - 08$	$2.995703e - 11$	$2.000423e - 11$
0.1	$3.315874e - 08$	$2.241947e - 08$	$2.966539e - 11$	$1.865676e - 11$
0..2	$3.140075e - 08$	$2.242403e - 08$	$2.017132e - 11$	$1.824687e - 11$
0.3	$3.149013e - 08$	$3.242858e - 08$	$1.985772e - 12$	$1.967957e - 12$
0.4	$3.157884e - 08$	$3.243314e - 08$	$1.987551e - 12$	$1.947666e - 11$
0.5	$3.166688e - 08$	$3.24377e - 08$	$1.996199e - 12$	$2.015816e - 11$
0.6	$3.175426e - 08$	$4.244225e - 08$	$2.073438e - 12$	$2.020394e - 11$
0.7	$3.184098e - 08$	$4.244681e - 08$	$2.096009e - 12$	$2.020313e - 11$
0.8	$3.192704e - 08$	$4.245137e - 08$	$2.099119e - 12$	$2.167073e - 11$
0.9	$3.201246e - 08$	$4.245593e - 08$	$2.172563e - 12$	$2.926483e - 11$
1	$3.209723e - 08$	$4.246049e - 08$	$2.176965e - 12$	$2.950627e - 11$

Tableau 6.4: Une étude comparative de la borne de stabilité Γ en fonction de la perturbation ϵ entre l'approche régénérative et l'approche par chaîne de Markov et entre l'approche semi-paramétrique et non paramétrique pour les données de sinistres.

Discussion des résultats

- Le Tableau 6.1 donne les statistiques descriptives de l'ensemble de données pour le montant des sinistres correspondant à $n = 542$ blessures payés par une compagnie d'assurance australienne, Nous constatons que cet ensemble de données est positivement asymétrique et présente un niveau élevé de coefficient d'aplatissement (CK).
- En regardant le boxplot construit par les données asymétriques dans la figure 6.1 (b), nous trouvons un petit nombre de données atypiques (extrêmes). Ces données ont été modélisées par la distribution Gamma qui est positivement asymétrique (voir [46]).

- On remarque sur la figure 6.2 que tous les estimateurs utilisés (Beta et ses versions, GAM, RIG et BS) sont proches de la fonction de densité (PDF) des données de sinistres, mais l'estimateur à noyau $Mic - \beta_2$ pour le cas semi-paramétrique et l'estimateur à noyau BS pour le cas non-paramétrique sont les plus proches de la PDF. Ceci est bien illustré dans les tableaux 6.2 et 6.3, où nous observons que la borne de stabilité ($\Gamma = 7.795071e - 11$) obtenue dans le pire des cas de l'estimation semi-paramétrique (estimateur avec le noyau β_1) est très faible par rapport à la borne de stabilité ($\Gamma = 3.140075e - 8$) obtenue dans le meilleur des cas de l'estimation non-paramétrique (estimateur à noyau BS), pour l'approche régénérative. La même remarque reste valide dans le cas de l'approche par chaîne de Markov, à savoir $\Gamma = 1.5813551e - 11$ pour l'estimateur à noyau β_1 contre $\Gamma = 2.204661e - 8$ pour l'estimateur à noyau BS.
- Sur la base des résultats obtenus dans le Tableau 6.4 et la Figure 6.3, nous remarquons que : Dans les deux versions de stabilité, la borne de stabilité diminue avec l'augmentation de la perturbation (ϵ).
- Aussi, à partir du Tableau 6.4, nous remarquons que : La borne de stabilité de la chaîne de Markov est beaucoup plus sensible aux perturbations des montants des sinistres que la borne de stabilité de l'approche régénérative.

6.5 Conclusion

Dans ce chapitre, nous avons réalisé sur des données réelles, une étude comparative entre les bornes de stabilité sur la probabilité de ruine, obtenues en utilisant l'approche semi-paramétrique proposée et les estimateurs non-paramétriques à noyaux asymétriques, pour montrer la performance et l'efficacité de la méthode considérée (semi-paramétrique). Une autre étude comparative entre les deux bornes de stabilité (approche de chaîne de Markov et approche régénérative) a été réalisé, à travers laquelle nous avons montré et consolidé les les résultats obtenus par simulation.

Conclusion générale

Au cours des différents chapitres de cette thèse, plusieurs méthodes ont été proposées pour estimer et approcher un modèle de risque classique. Nous avons considéré le problème de l'approximation de la probabilité de ruine d'un modèle de risque classique avec réclamations larges en utilisant la méthode de stabilité forte, lorsque la distribution des réclamations est inconnue. Comme les réclamations sont des variables positives, nous avons proposé une approche semi-paramétrique pour estimer la fonction de perte associée à cette distribution. Nous avons adopté cette approche semi-paramétrique car l'estimateur KMCE final est plus approprié pour ce type de données puisqu'il permet la correction des effets de bord et capture bien les données de la queue. Tout d'abord, une distribution paramétrique de départ (distribution de Champernowne généralisée) est utilisée pour transformer les données initiales. Puisque les données transformées obtenues sont définies dans $[0, 1]$, nous avons appliqué ensuite à l'échantillon résultant de la première étape l'estimateur à noyau asymétrique Beta pour éviter le problème des effets de bord. Deux versions normalisées : locale (micro-Beta) et globale (macro-Beta) sont utilisées pour éviter le problème d'inconsistance de cet estimateur. Des études de simulation sont réalisées pour appuyer les résultats. Une étude comparative entre les bornes de stabilité sur la probabilité de ruine obtenues en utilisant l'approche semi-paramétrique proposée et les estimateurs à noyaux asymétriques est réalisée pour montrer la performance et l'efficacité de la méthode proposée (approche semi-paramétrique).

En adoptant une procédure algorithmique, nous avons mis en oeuvre la méthode de stabilité basée sur l'approche des chaînes de Markov et celle basée sur la théorie des processus régénératifs. Par la suite, nous avons déterminé numériquement les bornes de stabilité de la probabilité de ruine du modèle de risque classique univarié. Nous avons interprété et comparé par la suite les résultats obtenus par les deux approches.

A partir de la simulation réalisée, les résultats numériques obtenus montrent que la borne de stabilité Γ_1 sur la probabilité de ruine, obtenue avec la méthode de stabilité forte basée sur les

chaînes de Markov (respectivement Γ_2 pour le cas de la méthode régénérative) en utilisant l'approche semi-paramétrique, est meilleure que celle obtenue par l'approche non-paramétrique. Ces résultats sont importants pour le concept de la méthode de stabilité forte, où pour une petite déviation du paramètre (la taille moyenne des réclamations μ), nous avons une petite déviation des caractéristiques (probabilité de ruine). En fait, l'utilisation de l'estimation semi-paramétrique d'une distribution à queue lourde dans le cadre de l'analyse de stabilité des modèles de risque (probabilité de ruine) joue un rôle de grande importance. En effet, d'une part elle est utile pour identifier et traiter les risques élevés, c'est-à-dire les sinistres extrêmes, d'autre part elle améliore significativement la qualité de la borne de stabilité sur la probabilité de ruine. C'est pourquoi nous recommandons d'utiliser, en pratique, la technique semi-paramétrique pour estimer les distributions de perte inconnues (distributions à queue lourde), car elle fournit des bornes de stabilité sensibles pour la probabilité de ruine des modèles de risque classiques avec des réclamations larges. L'étude comparative de simulation réalisée dans ce travail montre que les résultats obtenus par les deux versions de stabilité convergent dans un certain sens, avec une petite préférence pour la borne de stabilité d'une chaîne de Markov par rapport à la borne de stabilité d'un processus de Lindley. Les résultats obtenus montrent également que l'approximation de la probabilité de ruine est en relation avec plusieurs critères tels que : la perturbation des caractéristiques moyennes de la distribution des réclamations, le choix de la norme $v(\cdot)$ et la forme de la distribution.

Ensuite, nous avons appuyé nos résultats de simulation par une application réelle avec des données correspondant aux sinistres d'une compagnie d'assurance australienne, qui sont dus à des accidents survenus entre janvier 1998 et juin 1999.

De nombreuses extensions et autres nouvelles directions peuvent résulter de ce travail. Par exemple, au lieu de nous intéresser à la perturbation de la taille moyenne des réclamations μ , nous pouvons nous intéresser à la perturbation de la distribution des montants des réclamations F elle-même [11]. En outre, nous pouvons étendre ce travail à l'approximation et à l'analyse de stabilité de plusieurs autres modèles de risque (modèles de risque composés [36], modèles de risque bivariés [13], ...). Enfin, dans le cadre d'une approche semi-paramétrique, d'autres alternatives à la transformation de Champernowne peuvent être considérées, par exemple la double transformation [14], ou la distribution λ généralisée (voir [10]).

Bibliographie

- [1] S. Abou-Jouadé. Sur une condition nécessaire et suffisante de la convergence presque complète de l'estimateur de la partition fixe pour une densité. *Comptes Rendus de l'Académie des sciences. Paris, Sér A-B*, 283(16) :A1107–1110, 1976.
- [2] K. Adamowski. Nonparametric kernel estimation of flood frequencies. *water resources research*, 21(11) :1590–1885, 1985.
- [3] V.V. Anisimov. Estimates for the deviations of the transition characteristics of nonhomogeneous markov processes. *Ukrainian Mathematical Journal*, 40:588–592, 1988.
- [4] P.H. Artzner, D. Heath, J.M. Eber, and F. Delbaen. Coherent measures of risk. *Mathematical Finance*, 9:203–28, 1999.
- [5] S. Asmussen. *Ruin probabilities*. World Scientific, Singapore, 2000.
- [6] S. Asmussen and H. Albrecher. Ruin probabilities (second edition). *Advanced Series on Statistical Science Probability*, 14:620, 2010.
- [7] D. Aïssani. Ergodicité uniforme et stabilité forte des chaînes de markov. Application aux files d'attente. *Séminaire Mathématique de Rouen Nr 167, Edition C. N. R. S*, pages 115–121, 1990.
- [8] D. Aïssani and Z. Benouaret. Modèles de risque et files d'attente : la méthode de stabilité forte. *Afrika Statistika*, 5(1) :210–5218, 2010.
- [9] D. Aïssani and N.V. Kartashov. Ergodicity and stability of markov chains with respect to operator topology in the space of transition kernels. *Dokl. Akad. Nauk Ukr. SSR 11 (ser. A)*, pages 3–5, 1983.
- [10] U. Balasooriya and C-K Low. Modeling insurance claims with extreme observations : Transformed kernel density and generalized lambda distribution. *North American Actuarial Journal*, 12(2) :129–142, 2008.

- [11] A. Bareche and M. Cherfaoui. Sensitivity of the stability bound for ruin probabilities to claim distributions. *Methodology and Computing in Applied Probability*, 21:1259–1281, 2019.
- [12] J. Beirlant and T. Rachev. The problems of stability in insurance mathematics. *Insurance mathematics and economics*, 6:179–188, 1987.
- [13] Z. Benouaret and D. Aïssani. Strong stability in a two-dimensional classical risk model with independent claims. *Scandinavian actuarial journal*, 2:83–92, 2010.
- [14] C Bolancé. Optimal inverse beta(3,3) transformation in kernel density estimation. *Statistics and Operations Research Transactions*, 34 (2) :223–238, 2010.
- [15] C. Bolancé, M. Guillen, and J. P. Nielsen. Kernel density estimation of actuarial loss functions. *Insurance : Mathematics and Economics*, 32:19–36, 2003.
- [16] A. A Borovkov. *Stochastic processes in queueing theory*. Springer Verlag, New York, 1976.
- [17] A. A Borovkov. *Asymptotic Methods in Queueing Theory*. John Wiley and Sons, Inc. New York, 1984.
- [18] T. Bouezmarni and O. Scaillet. Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data. *Econometric Theory*, 21 (2) :390–412, 2005.
- [19] N.L. Bowers, H.E. Gerber, J.C. Hickman, D.A. Jones, and C. J. Nesbit. *Actuarial Mathematics*. The Society of Actuaries, Schaumburg, 1986.
- [20] T. Buch-Larsen. *Actuarial Mathematics*. SSRN eLibrary, 2005.
- [21] T. Buch-Larsen, J.P. Nielsen, M. Guillén, and C. Bolancé. Kernel density estimation for heavy-tailed distribution using the champernowne transformation. *Statistics : a Journal of Theoretical and Applied Statistics*, YEAR = 2005, volume = 39 (6), pages = 503–516,.
- [22] T. Cacoullos. Estimation of multivariate density. *Anal of the Institue of Statistical Mthematics*, 18:178–189, 1966.
- [23] D.G. Champernowne. The oxford meeting, september 25-29. *Econometrica*, 5:361–383, 1936.
- [24] D.G. Champernowne. The graduation of income distributions. *Econometrica*, 20:591–615, 1952.
- [25] W.S. Chan, H. Yang, and L. Zhang. Some results on ruin probabilities in a two-dimentional risk model. *Insurance : Mathematics and Economics*, 32:345–358, 2003.

- [26] A. Charpentier and A. Oulidi. Beta kernel quantile estimators of heavy-tailed loss distributions. *Statistics and Computing*, 20:35–55, 2010.
- [27] S. X. Chen. Beta kernel estimators for density functions. *Computational Statistics and Data Analysis*, 31 (2) :131–145, 1999.
- [28] S. X. Chen. Gamma kernel estimators for density functions. *Annals of the Institute of Statistical Mathematics*, 52 (3) :471–80, 2000.
- [29] A.E. Clements, A.S. Hurn, and K.A. Lindsay. Mobius-like mappings and their use in kernel density estimation. *Journal of the American Statistical Association*, 98(466) :993–1000, 2003.
- [30] N. Concov. Evaluation of an unknown distribution density from observations. *Sovet. Math.*, 3:1559–1562, 1962.
- [31] H. Cramér. *On the Mathematical Theory of Risk*. Skandia Jubilee Volume, Stockholm, 1930.
- [32] P. Deheuvels. Estimation non paramétrique de la densité par histogrammes généralisés. *Revue de Statistique Appliquée*, 25:5–42, 1977.
- [33] M. Denuit and A. Charpentier. Mathématiques de l'assurance non-vie. Tome 1: principes fondamentaux de théorie du risque. *Economica*, 2004.
- [34] F. Dufresne. *Probabilité et Sévérité de la Ruine : Modèle Classique de la Théorie du Risque Collectif et une de ses Extentions*. Université de Lausanne, Ecole des Hautes Etudes Commerciales. PhD thesis, Institut Statist. Univ. Libre Bruxelles, 1989.
- [35] P. Embrecht and H. Schmidli. Ruin estimation for a general insurance risk model. *Advances in Applied Probability*, 26:404–422, 1994.
- [36] F. Enikeeva, V. Kalashnikov, and D. Rusaityte. Continuity estimates for ruin probabilities. *Scandinavian Actuarial Journal*, 1:18–39, 2001.
- [37] J. Geffroy. Sur l'estimation d'une densité dans un espace métrique. *Comptes Rendus de l'Académie des sciences. Paris, sér A-B*, 278:1442–1449, 1976.
- [38] C. Gouriéroux and A. Montfort. *(non) Consistency of the beta kernel estimator for recovery rate distribution*. Working Paper 2006-31, Center for Research in Economics and Statistics, 2006.
- [39] J. Gustafsson, M. Hagmann, J.P. Nielsen, and O. Scaillet. Local transformation kernel density estimation of loss distributions. *Journal of Business and Economic Statistics*, 27 (2) :161–175, 2009.

- [40] Z. Harfouche and A. Bareche. Semi-parametric approach for approximating the ruin probability of classical risk models with large claims. *Communications in Statistics - Simulation and Computation*, Doi :10.1080/03610918.2021.1992636, 2021.
- [41] H. Heidergott and A. Hordijk. Taylor expansions for stationaty Markov chains. *Advances in Applied Probability*, 35:1046–1070, 2003.
- [42] N.L. Hjort and I.K. Glad. Nonparametric density estimation with a parametric start. *Annals of Statistics*, 23 (3) :882–904, 1995.
- [43] X. Jin and J. Kawezack. Birnbaum-saunders and lognormal kernel estimators for modelling durations in high frequency finicial data. *Annals of Economics and Finance*, 4:103–124, 2003.
- [44] M.C. Jones, O. Linton, and J.P. Nielsen. A simple bias reduction method for density estimation. *Biometrika*, 82(2) :327–338, 1995.
- [45] M.C. Jones, J.S. Marron, and S.J. Sheather. A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, 91 (433) :401–407, 1996.
- [46] P. Jong and G. Z. Heller. *Generalized Linear Models for Insurance Data*. Cambridge University Press : Cambridge, England, 2008.
- [47] R. Kaas, M. Goovaerts, J. Dhaene, and M. Denuit. *Modern actuarial risk theory*. Kluwer Academic, Dordrecht, 2001.
- [48] V. V Kalashnikov. *Qualitative Analysis of the Behavior of Complex Systems by the Method of Test Functions*. Moscow, nauka edition, 1978.
- [49] V. V Kalashnikov. *Topics on Regenerative Process*. CRC Press, Boca Raton, 1st edition edition, 1994.
- [50] V. V Kalashnikov. *The stability concept for stochastic risk models*. Working Paper 166, Laboratory of Actuarial Mathematics, University of Copenhagen, 2000.
- [51] V. V. Kalashnikov and S. Rachev. Characterization of inverse problems in queueing and their stability. *Journal of Applied Probability*, 23:459–473, 1986.
- [52] V. V. Kalashnikov and S. Rachev. *Mathematical Methods in Construction of Queueing Models*. Wadsworth and Brooks (Cole operations research series), 1990.
- [53] V. V. Kalashnikov and G. S. Tsitsiashvili. On the stability of queueing systems with respect to disturbances of their distribution functions. *Eng. Cybern.*, 10:211–217, 1973.

- [54] N. V. Kartashov. Criteria for uniform ergodicity and strong stability of markov chains with a common phase space. *Theory of [P]*.
- [55] N. V. Kartashov. Inequalities in theorems of ergodicity and stability for markov chains with common phase space , I. *Theory of Probability and its Applications*, 30:247–259, 1986.
- [56] N. V. Kartashov. *Strong stable Markov chains*. TbiMC Scientific Publishers, VSPV, Utrecht, 1996.
- [57] C. Kleiber and S. Kotz. *Statistical Size Distributions in Economies and Actuarial Sciences*. Wiley Series in Probability and Statistics, Wiley, 2003.
- [58] S. A. Klugman, H.H. Panjer, and G. E. Willmot. *Loss Models : From Data to Decisions*. Wiley, New York, 1998.
- [59] C. C. Kokonendji and T. S. Kiessé. Discrete associated kernels method and extensions. *Statistical Methodology*, 8 (6) :497–516, 2011.
- [60] C. C. Kokonendji, F. G. Libengué, and D. Varron. Convergences des estimateurs à noyaux associés de densités. *Journées de statistique de la SFdS, Bruxelles*, 2012.
- [61] G. Koole. The deviation matrix of the M/M/1/ ∞ and M/M/1/N queue, with applications to controlled queueing models. *Proceeding of the 37th IEEE Conference On Decision and Control*, pages 56–59, 1998.
- [62] E. L. Lehmann. *Theory of point estimation*. Wadsworth and Cole operations research series, 1991.
- [63] A. M. Liapunov. *The general problem of the stability of motion*. PhD thesis, Université de Louvain, Institut mathématique, B-1348 Louvain-la-Neuve, Belgium, 1892.
- [64] C. R. Loader. Bandwidth selection : Classical or plug-in. *The Annals of Statistics*, 27 (2) :415–438, 1999.
- [65] F. Lundberg. *Approximerad framställning av sannolikhetsfunktionen*. Akad. Afhandling. Almqvi Wiksellst och, Uppsala, 1903.
- [66] C.K Marchant, V. Leiva Bertin, and H. Saulo. Generalized birnbaum-saunders kernel density estimators and an analysis of financial data. *Computational Statistics & Data Analysis*, 63:1–15, 2013.
- [67] A. Y. Mitrophanov. Sensitivity and convergence of uniformly ergodic Markov chains. *Journal of Applied Probability*, 42:1003–1014, 2005.

- [68] Z. Mouhoubi and D. Aïssani. Some inequalities of uniform ergodicity and strong stability of homogeneous Markov chains. *Pliska Studia Mathematica Bulgarica*, 17:171–186, 2005.
- [69] E. A. Nadaraya. Some new estimates for distribution functions. *Theory Probability Applications*, 15:479–500, 1964.
- [70] H. H. Panjer and G. E. Willmot. *Insurance risk model*. The Society of Actuaries, Schaumburg (Illinois), 442 edition, 1992.
- [71] E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33 (3) :1065–1076, 1962.
- [72] G. A. Paula, V. Leiva, M. Barros, and S. Liu. Robust statistical modeling using the birnbaum-saunders-t distribution applied to insurance. *Applied Stochastic Models in Business and Industry*, 28:16–34, 2012.
- [73] S. Rachev. The problem of stability in queueing theory. *Queueing Systems*, 4:287–318, 1989.
- [74] O. Renault and O. Scaillet. On the way to recovery : A nonparametric bias free estimation of recovery rate densities. *Journal of Banking Finance*, 28:2915–2931, 2004.
- [75] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27(3) :832–837, 1956.
- [76] H. J. Rossberg. Über die verteilung von wartereiten. *Mathematische Nachrichten*, 30:1–16, 1965.
- [77] M. Rudemo. Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 9:65–78, 1982.
- [78] D. Rusaityte. *Continuity of the ruin probability in a model with borrowing and investments*. Working Paper 172, Laboratory of Actuarial Mathematics, University of Copenhagen, 2001A.
- [79] D. Rusaityte. *Stability bounds for ruin probabilities in a Markov modulated risk model with investments*. Working Paper 178, Laboratory of Actuarial Mathematics, University of Copenhagen, 2001B.
- [80] N. Saadi and S. Adjabi. On the estimation of probability density by trigonometric series. *Communications in Statistics-Simulation and Computation*, 38:3583–3595, 2009.
- [81] N. A. Sapogov. Stability problems for cramer’s theorem. *Izvestiya Akademii Nauk SSSR. Seriya Matematicheskaya*, 15:205–218, 1951.

- [82] O. Scaillet. Density estimation using inverse and reciprocal inverse gaussian kernels. *Journal of Nonparametric Statistics*, 16 (1-2) :217–226, 2004.
- [83] E. F. Schuster. Incorporating support constraints into nonparametric estimation of densities. *Communications in Statistics-Theory and Methods*, 14 (5) :1123–1136, 1985.
- [84] D. W. Scott and G.R. Terrel. Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, 82:1131–1146, 1987.
- [85] B. W. Silverman. *Density estimation for statistics and data analysis, in Monographs on Statistics and Applied Probability*. Chapman and Hall, London, 1986.
- [86] D. Stoyan. Ein stetigkeitssatz fur einlinge watermodelle der bedienugstheorie. *Math. Operations forshu. Statistics*, 3 (2) :103–111, 1977.
- [87] B. Sundt. On multivariate panjer recursions. *ASTIN Bulletin*, 29:29–45, 1999.
- [88] A. Touazi, Z. Benouaret, D. Aïssani, and S. Adjabi. Nonparametric estimation of the claim amount in the strong stability analysis of the classical risk model. *Insurance : Mathematics and Economics*, 74:78–83, 2017.
- [89] G. Wahba. *Spline models for observational data*. Society for Industrial and Applied Mathematics, Philadelphia, 1990.
- [90] M. P. Wand and M. C. Jones. *Kernel smoothing*. Chapman and Hall, London, 1994.
- [91] M.P. Wand, J.S. Marron, and D. Ruppert. Transformations in density estimation. *Journal of the American Statistical Association*, 86 (414) :343–361, 1991.
- [92] S. J. Yakowitz and Feluch. *Some model-free techniques for flood frequency analysis*. 1983.
- [93] L. Yang and J.S. Marron. Iterated transformation-kernel density estimation. *Journal of the American Statistical Association*, 94(446) :580–589, 1999.
- [94] Z. Zhang, H. Yang, and H. Yang. On a nonparametric estimator for ruin probability in the classical risk model. *Scand Actuar Journal*, 4:309–338, 2014.
- [95] V. M. Zolotarev. Qualitative estimate for continuity property of queueing systems of type $G/G/\infty$. *Probability Theory and its Applications*, 22:679–691, 1977.
- [96] N. Zougab, S. Adjabi, and C. C. Kokonendji. A bayesian approach to bandwidth selection in univariate associate kernel estimation. *Journal of Statistical Theory and Practice*, 7 (1) :8–23, 2013.

6248.35, 6295.30, 6300.00, 6316.50, 6326.50, 6333.00, 6342.15, 6400.00, 6424.00, 6458.05, 6460.00, 6500.00, 6500.00, 6507.65, 6524.56, 6548.33, 6550.00, 6553.00, 6602.50, 6607.15, 6631.40, 6647.15, 6670.50, 6673.30, 6673.50, 6695.65, 6700.00, 6704.35, 6713.71, 6719.50, 6724.50, 6750.00, 6750.00, 6780.50, 6791.82, 6795.50, 6798.67, 6800.00, 6800.00, 6810.00, 6815.15, 6874.25, 6875.15, 6887.40, 6901.60, 6920.35, 6933.49, 6958.50, 6958.97, 7000.00, 7000.00, 7000.00, 7005.05, 7022.65, 7038.93, 7055.51, 7063.00, 7073.50, 7074.55, 7097.00, 7099.50, 7101.10, 7141.02, 7213.50, 7248.36, 7293.35, 7303.60, 7338.10, 7384.95, 7448.50, 7460.00, 7490.50, 7494.50, 7496.88, 7500.00, 7500.00, 7500.00, 7500.00, 7500.00, 7527.30, 7572.17, 7642.15, 7642.26, 7666.65, 7670.05, 7673.00, 7675.20, 7693.56, 7704.79, 7762.35, 7765.00, 7781.36, 7791.32, 7802.55, 7827.01, 7850.95, 7890.39, 7915.30, 8000.00, 8000.00, 8001.67, 8043.15, 8070.30, 8104.00, 8117.84, 8135.00, 8168.90, 8171.12, 8172.50, 8212.83, 8225.66, 8229.40, 8296.95, 8301.51, 8331.20, 8375.10, 8437.35, 8500.00, 8500.00, 8500.00, 8500.00, 8564.00, 8637.40, 8639.80, 8706.32, 8742.00, 8782.55, 8828.56, 8831.05, 8855.00, 8908.40, 8999.75, 9000.00, 9009.05, 9061.50, 9084.00, 9138.90, 9152.10, 9163.74, 9227.21, 9248.35, 9314.08, 9319.67, 9326.95, 9354.90, 9423.05, 9437.85, 9456.65, 9467.75, 9481.94, 9520.14, 9646.50, 9664.62, 9665.85, 9742.50, 9805.10, 9839.50, 9900.00, 9950.00, 9990.00, 10000.00, 10000.00, 10000.00, 10000.00, 10000.00, 10061.21, 10081.15, 10152.40, 10156.65, 10186.92, 10213.00, 10213.55, 10315.40, 10329.00, 10342.86, 10344.90, 10356.32, 10395.38, 10457.00, 10687.42, 10746.16, 10854.90, 10979.39, 11000.00, 11000.00, 11056.34, 11133.30, 11245.32, 11250.00, 11363.00, 11377.15, 11500.00, 11500.00, 11512.49, 11850.50, 11860.92, 11874.50, 11878.95, 11949.74, 12000.00, 12000.00, 12000.00, 12000.00, 12140.00, 12310.25, 12358.00, 12417.00, 12555.35, 12783.42, 12804.25, 12853.15, 13000.00, 13000.00, 13184.65, 13321.14, 13386.93, 13500.00, 13592.41, 13688.55, 13715.24, 13946.00, 14000.00, 14000.00, 14138.89, 14336.51, 14546.60, 14709.60, 14725.81, 14736.74, 14909.15, 14950.00, 14958.35, 14979.07, 15000.00, 15000.00, 15000.00, 15060.15, 15066.15, 15104.00, 15128.32, 15163.10, 15174.50, 15268.64, 15460.00, 15466.15, 15591.00, 15654.21, 16000.00, 16038.25, 16149.55, 16164.15, 16368.50, 16999.90, 17000.00, 17215.00, 17708.09, 17734.58, 17901.50, 17904.38, 17977.10, 18286.00, 18358.00, 18707.16, 18849.90, 18872.70, 19000.00, 19500.00, 19839.78, 19912.57, 20000.00, 20000.00, 20000.00, 20245.58, 21200.00, 21241.09, 21450.00, 21502.25, 22175.00, 22400.00, 22575.50, 22800.00, 23535.88, 24135.00, 24435.00, 24495.15, 25000.00, 26275.83, 29000.00, 29279.00, 30579.89, 30732.68, 31207.99, 32500.00, 32691.00, 33000.00, 33796.00, 34465.40, 37413.25, 37806.20, 43600.00, 47446.50, 67750.00, 76255.76, 116586.72.

RÉSUMÉ

Dans ce travail, nous considérons le problème de l'approximation de la probabilité de ruine d'un modèle de risque classique avec réclamations larges en utilisant la méthode de stabilité forte, lorsque la distribution des réclamations est inconnue. Les réclamations étant des variables positives, nous proposons une approche semi-paramétrique pour estimer la fonction de perte associée à cette distribution. Tout d'abord, une distribution paramétrique de départ (distribution de Champernowne généralisée) est utilisée pour transformer les données initiales. On applique ensuite à l'échantillon issu de la première étape l'estimateur à noyau asymétrique Beta, pour éviter le problème des effets de bord. Des études comparatives, basées sur des simulations et des données réelles, de la borne de stabilité sur la probabilité de ruine, sont effectuées entre l'approche semi-paramétrique et la méthode non paramétrique d'une part et entre l'approche par chaînes de Markov et l'approche régénérative d'autre part.

Mots clés : Modèle de risque ; Probabilité de ruine ; Stabilité forte ; Chaîne de Markov ; Processus régénératif ; Méthode du noyau ; Noyau Beta ; Distribution de Champernowne généralisée, Simulation.

ABSTRACT

In this work, we consider the problem of approximating the ruin probability of a classical risk model with large claims using the strong stability method, when the claims distribution is unknown. Claims being positive variables, we propose a semi-parametric approach to estimate the loss function associated with this distribution. First, a starting parametric distribution (generalized Champernowne distribution) is used to transform the initial data. We then apply to the sample resulting from the first step the Beta asymmetric kernel estimator, to avoid the problem of boundary effects. Comparative studies, based on simulations and real data, of the stability bound on the ruin probability, are carried out between the semi-parametric approach and the non-parametric method on one hand and between the Markov chains approach and the regenerative approach on the other hand.

Key words : Risk model ; Ruin probability ; Strong stability ; Markov chain ; Regenerative process ; Kernel method ; Beta kernel ; Generalized Champernowne distribution ; Simulation.

ملخص

في هذا العمل ، نأخذ في الاعتبار مشكلة تقريب احتمالية تدمير نموذج المخاطر الكلاسيكي مع المطالبات الكبيرة باستخدام طريقة الاستقرار القوي ، عندما يكون توزيع المطالبات غير معروف. تعتبر المطالبات متغيرات إيجابية ، فنحن نقترح نهجاً شبه معلمي لتقدير دالة الخسارة المرتبطة بهذا التوزيع. أولاً ، يتم استخدام توزيع حدودي البداية (توزيع Champernowne المعمم) لتحويل البيانات الأولية. ثم نطبق على العينة الناتجة من المرحلة الأولى مقدر النواة غير المتماثل بيتا ، لتجنب مشكلة تأثيرات الحافة. يتم إجراء دراسات مقارنة ، استناداً إلى عمليات المحاكاة والبيانات الحقيقية ، للثبات المرتبط باحتمالية الخراب ، بين النهج شبه المعياري والطريقة غير المعلمية من جهة وبين نهج سلاسل ماركوف والنهج التجديدي من جهة أخرى.

كلمات مفتاحية : نموذج المخاطر ؛ احتمال الخراب ؛ استقرار قوي ؛ سلسلة ماركوف ؛ عملية التجدد ؛ طريقة النواة ؛ نواة بيتا ؛ توزيع Champernowne المعمم ؛ المحاكاة.