République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université M. Mammeri de Tizi-Ouzou
Faculté des Sciences
Département Mathématiques

THÈSE

EN VUE DE L'OBTENTION DU DIPLÔME DE DOCTORAT LMD

Domaine : Mathématiques et Informatique **Filière :** Mathématiques

Spécialité : Recherche Opérationnelle et Optimisation

Présentée par

AFROUN Faïrouz

Thème

Estimation non paramétrique dans les processus Markoviens

Soutenue le 15 mars 2022 devant le jury composé de :

M. Aïdene Mohamed	**Professeur**	**Univ. de Tizi-Ouzou**	**Président**
M. Aïssani Djamil	**Professeur**	**Univ. de Béjaia**	**Rapporteur**
M. Hamadouche Djamel	**Professeur**	**Univ. de Tizi-Ouzou**	**Co-Rapporteur**
M. Adjabi Smaïl	**Professeur**	**Univ. de Béjaia**	**Examinateur**
M. Fellag Hocine	**Professeur**	**Univ. de Tizi-Ouzou**	**Examinateur**
M. Yousfate Abderrahmane	**Professeur**	**Univ. de Sidi Bel Abbes**	**Examinateur**

Année Universitaire : 2021/2022.

Estimation non paramétrique dans les processus Markoviens

Faïrouz AFROUN, Djamil AÏSSANI et Djamel HAMADOUCHE.

⁂ *Remerciements* ⁂

Je tiens à exprimer toute ma reconnaissance et mes vifs remerciements à mes encadreurs Mr Djamil AÏSSANI (Professeur à l'Université de Béjaïa) et Mr Djamel HAMADOUCHE (Professeur à l'Université de Tizi-Ouzou) pour la confiance qu'ils m'ont témoigné en acceptant de diriger ce travail, pour leurs encouragements et l'aide qu'ils m'ont apporté durant ces années de recherche. Qu'ils trouvent ici ma profonde gratitude.

Je remercie vivement Mr Mohamed AÏDENE (Professeur à l'Université de Tizi-Ouzou) pour m'avoir honoré en acceptant de présider le jury de cette thèse. Je remercie également : Mr Smaïl ADJABI (Professeur à l'Université de Béjaïa), Mr Hocine FELLAG (Professeur à l'Université de Tizi-Ouzou) et Mr Abderrahmane YOUSFATE (Professeur à l'Université de Sidi Bel Abbes) pour avoir accepté de juger ce travail.

Je tiens à adresser toute ma gratitude à l'ensemble des membres et personnel de l'unité de recherche LaMOS (Laboratoire de Modélisation et d'Optimisation des Systèmes) de l'université de Béjaïa et du Laboratoire des Mathématiques Pures et Appliquées (LMPA) de l'université de Tizi-Ouzou.

Je remercie vivement mon cher mari pour l'aide, le soutien, l'amour et les encouragements dont il a toujours fait preuve à mon égard.

Mes pensées vont enfin vers tous ceux qui ont contribué de près ou de loin à l'aboutissement de ce travail. Je tiens donc à adresser notamment un grand merci à mes parents, toute ma famille, tous mes professeurs, tous mes amis et collègues.

⁂ *Dédicaces* ⁂

A ma très chère et adorable fille Thiziri.
A mon très cher mari Mouloud.
A mes très chers parents.
Je dédie ce travail.

Faïrouz AFROUN

⁕ *Liste des travaux* ⁕

- **Articles de Revues :**
    1. F. Afroun, D. Aïssani, D. Hamadouche and M. Boualem. *Q-matrix method for the analysis and performance evaluation of unreliable $M/M/1/N$ queuing model*. Mathematical Methods in the Applied Sciences, Wiley edition, 2018 ; 41(18) : 9152-9163.
    2. F. Afroun, D. Aïssani and D. Hamadouche. *Nonparametric approximation of the characteristics of the $D/G/1$ queue with finite capacity*. International Journal of Computing Science and Mathematics, Inderscience Publishers Ltd edition, Ref : IJCSM 0061 Afroun et al. (3), 2020.
- **Articles de Proceedings :**
    1. F. Afroun, D. Aïssani et D. Hamadouche. *Estimation à noyau discret dans le modèle de stock de type $(R; s; S)$*. Dans les actes du Colloque sur l'Optimisation et les Systèmes d'Information, COSI'2019, 24 au 26 Juin 2019, UMMTO, Tizi Ouzou, Algérie. Pages : 71–82. ISBN : 978-9947-34-164-3.
    2. F. Afroun, D. Aïssani et D. Hamadouche, *Validation croisée non biaisée et estimation à noyau discret dans le modèle de stock de type $(R; s; S)$*. Dans les Actes du Colloque sur la Modélisation Stochastiques et Statistique, MSS'19, 24 au 26 Novembre 2019, USTHB, Alger, hal-02593238, 347–350, 2019.
- **Communications Internationales :**
    1. F. Afroun, D. Aïssani and D. Hamadouche, *Q-matrix method for the analysis and performance evaluation of unreliable $M/M/1/N$ queuing model*. The first international conference on the "Evolution of Contemporary Mathematics and their Impact in Sciences and Technology" (ECMI-SciTech 2017). Octobre 09-12, 2017. Frères Mentouri University-Constantine, Algeria (communication orale).
    2. F. Afroun, D. Aïssani and D. Hamadouche. *Estimation à noyau discret dans le modèle de stock de type $(R; s; S)$*. Colloque sur l'Optimisation et les Systèmes d'Information COSI'2019, 24 au 26 Juin 2019, UMMTO, Tizi Ouzou, Algérie (communication orale).
- **Communications Nationales :**
    1. F. Afroun, D. Aïssani et D. Hamadouche. *Estimation non paramétrique dans les processus Markoviens*. Doctoriales 2017, 01-06 Août 2017, Université de Bejaia, Algérie (communication orale).
    2. F. Afroun, D. Aïssani et D. Hamadouche. *Estimation non paramétrique dans les processus Markoviens*. Les premières doctoriales nationales de Mathématiques, 28-31 octobre 2017, Ecole Normale Supérieure Assia DJEBAR-Constantine, Algérie (communication orale + Poster).

3. F. Afroun, D. Aïssani and D. Hamadouche. *Analysis and performance evaluation of unreliable $M/M/1/N$ queueing model with various effects*. Les 4èmes Journées Scientifiques de LAROMAD "JSlaromad", 29 au 30 Novembre 2017, Université Mouloud Mammeri, Tizi-Ouzou, Algérie (communication orale).
4. F. Afroun, D. Aïssani and D. Hamadouche. *Discrete associated kernel estimation in $(R; s; S)$ inventory model*. Doctoriales de la Recherche opérationnelle, 12 et 13 Décembre 2018, université de Bejaia, Algérie (communication Poster).

- **Stage de perfectionnement :**
  - ✓ Stage de perfectionnement d'une durée d'un mois (Du 08/12/2019 Au 06/01/2020) à l'Institut gestion de la production, Département des systèmes d'information et des opérations, Université d'économie et de commerce, Vienne, Autriche. Sur la thématique "*Validation croisée non biaisée dans l'estimation à noyau de la matrice de transition d'une chaîne de Markov*".

Table des matières

Table des figures

Liste des tableaux

Introduction générale

Les processus et les modèles Markoviens constituent un sujet central dans la théorie de probabilités appliquées et de la statistique. En effet, ils peuvent donner une bonne description pour des phénomènes dont l'évolution dans le temps peut être décrite par des variables aléatoires faiblement dépendantes. De plus, plusieurs problèmes réels peuvent être modélisés par ce type de processus stochastiques à temps discret ou continu.

Théoriquement, l'évaluation des performances d'un système décrit par un processus stochastique, en général, et par un processus de Markov, en particulier, est basée sur les différents paramètres de départ qui le décrivent. Cependant en pratique, les valeurs des paramètres de départ d'un système ne sont connues que sous forme d'un échantillon d'observations (données). En ce sens, pour évaluer les performances du système considéré, l'utilisation des techniques d'estimation statistique (paramétriques et/ou non paramétriques), qui visent à fournir une approximation des valeurs des paramètres (inconnus) en exploitant les informations fournies par l'échantillon, est inévitable.

Les premiers travaux sur l'estimation statistique pour des modèles Markoviens à un espace d'état fini, particulièrement par la méthode du maximum de vraisemblance, ont été faits par Anderson et Goodman [8] en 1957 et Billingsley [11] en 1961. Ce dernier a étudié l'estimateur du maximum de vraisemblance et ses propriétés asymptotiques dans le cadre d'une observation de longue durée pour un modèle Markovien ergodique. Par contre, Anderson et Goodman dans [8] ont traité principalement l'estimation dans le cas d'une suite de copies indépendantes d'une chaîne Markovienne censurée à un instant fixe. Cependant, l'estimateur du maximum de vraisemblance pour des caractéristiques qui peuvent être exprimées en fonction des paramètres du modèle n'a été étudié que plus récemment. Nous citons dans ce contexte, le travail de Sadek et Limnios (2002) [83] dédié à l'estimation de la fiabilité, la disponibilité et le taux de défaillance dans des systèmes qui peuvent être modélisés par des chaînes de Markov finies et ergodiques. D'autres travaux dans le contexte d'estimation dans les chaînes de Markov sont focalisés sur d'autres méthodes d'estimation et le choix de l'estimateur, entre autres on cite : Lee et al. (1970) [59] ont étudié en détail les différentes méthodes d'estimation, outre celles basées sur les moindres carrés, ils ont envisagé la méthode du χ^2 minimum, celle du maximum de vraisemblance et celle basée sur la minimisation de la norme L_1 c'est-à-dire la minimisation de la somme des valeurs absolues des écarts. Par ailleurs, les auteurs ont envisagé aussi le problème sous l'angle bayésien. Bonnieux [13] s'est intéressé à l'estimation des caractéristiques qui découlent d'une chaîne de Markov finie, aux différentes représentations possibles pour ces caractéristiques et aux calculs de l'erreur qui leur est associée.

L'estimation paramétrique des caractéristiques pour les chaînes Markoviennes finies est importante pour les applications et elle n'a pas été beaucoup explorée. Ceci est dû au fait qu'il est difficile d'estimer avec précision les caractéristiques d'une chaîne Markovienne modélisant des phénomènes complexes. Afin de surmonter cette difficulté, on fait appel aux méthodes d'estimation non paramétriques. Dans la littérature, il existe plusieurs méthodes non paramétriques pour l'estimation, nous citons par exemple la méthode d'estimation par histogramme dont l'origine est attribuée à John Grant au $17^{\text{ème}}$ siècle, la méthode d'estimation par les séries orthogonales [82, 98], la méthode de lissage par les fonctions Splines [97]. La méthode qui a rencontré plus de succès est la méthode d'estimation par noyau qui a été proposée initialement par Rosenblatt (1956) [77] et Parzen (1962) [73] pour estimer des fonctions de densité f. Ce succès peut s'expliquer par au moins trois raisons : d'abord, l'expression théorique de l'estimateur est très simple, puisque il s'écrit comme la somme de n variables aléatoires indépendantes et identiquement distribuées, en utilisant la fonction noyau K et le paramètre de lissage h. Ensuite, il est convergeant en de nombreux sens. Enfin, l'estimateur à noyau est flexible, car il laisse à l'utilisateur une grande latitude dans le choix du noyau K et du paramètre de lissage h.

L'objectif visé par ce projet de thèse est, principalement, d'estimer les matrices de transition associées à des chaînes de Markov particulières à l'aide de la méthode du noyau. Le choix de l'estimation de cette composante, est motivé par le fait que la matrice des probabilités de transition associée à une chaîne de Markov a une grande importance dans l'analyse transitoire et stationnaire d'une chaîne de Markov et elle nous permet de déduire tout le reste de ses mesures de performance, (pour plus de détails sur la matrice des probabilités de transition et son utilité, le lecteur peut se référer par exemple à [69, 75]).

Historiquement, le premier travail consacré à l'estimation à noyau des probabilités de transition d'une chaîne de Markov est celui de Roussas [79] publié en 1969. Par la suite, plusieurs auteurs ont complété les résultats obtenus dans [79], on cite Masry et Gyor [64] et par Basu et Sahoo [10]. Par la suite, Laksaci et Yousfate [57] ont fournit une majoration de la vitesse de convergence (au sens de la norme L_p) de l'estimateur construit. Cependant, les différents résultats obtenus sont restreints au cadre théorique. L'utilisation de la méthode du noyau dans l'estimation d'une matrice de transition associée à une chaîne de Markov dans le cadre pratique, appliquée aux files d'attente, a été envisagée par Gontijo et al. [39], où les auteurs ont appliqué la méthode du noyau pour estimer les mesures de performance de la file d'attente $GI^{[X]}/M/s/C$. Récemment, Cherfaoui et al. [27] ont abordé le problème du choix du paramètre de lissage dans le cadre de l'estimation à noyau de la matrice de transition associée à une chaîne de Markov décrivant la file d'attente $GI/M/1/C$. Dans ce dernier travail, afin de prendre en compte l'interaction des différentes composantes du système (le processus inter-arrivées et le processus de service), les auteurs ont proposé des procédures de sélection du paramètre de lissage, basées sur la minimisation des normes matricielles. De plus, ils ont montré que l'estimateur du paramètre de lissage choisi en minimisant la norme matricielle $\|.\|_2$, donne de meilleurs résultats que les méthodes classiques.

Il est à noter que les travaux [27, 39] ont été réalisés via des estimateurs à noyaux asymétriques continus (pour estimer la distribution du temps des inter-arrivées qui est défini sur $\mathbb{R}^+$). Cependant en pratique, plusieurs situations sont modélisées par des chaînes de Markov gouvernées selon des distributions discrètes générales et inconnues. Dans ce cas, il est logique et naturel d'estimer ces distributions, générales et inconnues, à l'aide de noyaux discrets.

Pour estimer une fonction de densité discrète à l'aide de la méthode du noyau, l'estima-

teur à noyau de type Dirac, également appelé estimateur empirique, est souvent utilisé par les praticiens, en raison de sa simplicité et ses bonnes propriétés asymptotiques. Néanmoins, cet estimateur ne convient pas aux échantillons de petite ou moyenne taille. En plus de l'estimateur naïf, Aitchison et Aitken [5] ont proposé un autre noyau discret. Le problème de ce dernier noyau est qu'il n'a qu'une seule forme et ne convient qu'aux données catégorielles et aux distributions discrètes finies.

Récemment, d'autres noyaux discrets ont été proposés dans [50, 51] pour estimer les fonctions de densité à support discret. Les auteurs ont introduit la notion et la définition d'un noyau discret $K_{x,h}$, de la cible x et du paramètre de lissage h, conçu à partir d'une distribution de probabilité discrète. Cette dernière notion a permis un développement considérable de la méthode du noyau dans le cas discret, que ce soit dans le cadre d'estimation de la densité elle-même ou dans ces applications dans d'autres domaines. En effet, on peut citer, par exemple, les travaux [100, 109] qui sont consacrés au problème du choix du paramètre de lissage dans l'estimation à noyaux discrets d'une fonction de masse, et les travaux [32, 92, 109] où l'extension d'estimateur à noyau discret d'une fonction de masse au régresseur de Nadaraya-Watson, lorsque les variables explicatives sont discrètes, a été discutée.

L'objectif de notre travail est d'analyser numériquement l'impact de l'estimation des paramètres (estimation paramétrique) de départ décrivant une chaîne de Markov sur les propriétés statistiques des estimations des caractéristiques stationnaires de cette chaîne. D'autre part, vérifier la validité des conclusions atteintes dans [27] sur la chaîne de Markov en temps continu lorsque nous considérons une chaîne de Markov à temps discret. Pour cela, afin d'analyser l'impact des méthodes paramétriques et de mettre en évidence leurs limites dans l'estimation des caractéristiques d'une chaîne de Markov en fonction de la taille de l'échantillon et du paramètre de départ estimé ; nous avons mené des études de simulation sur la chaîne de Markov décrivant le modèle d'attente $M/M/1/N$ avec des vacances multiples, Bernoulli feedback, balking, reneging et rétention des clients impatients, ainsi que la possibilité d'une panne et de réparation du serveur [4]. Ensuite, nous avons abordé le problème du choix du paramètre de lissage via la minimisation des normes matricielles dans l'estimation à noyau discret des matrices de transition, P, correspondantes aux chaînes de Markov à temps discret décrivant les files d'attente $GI/D/1/C$ et $D/G/1/C$. Nous avons développé des formes explicites des expressions (résultantes des trois normes matricielles $\|.\|_1$, $\|.\|_2$ et $\|.\|_\infty$) à minimiser afin de sélectionner le paramètre de lissage lors de l'estimation des matrices P. Par ailleurs, pour analyser l'impact du choix du paramètre de lissage par la minimisation des normes matricielles sur la qualité des estimations des matrices de transitions, P, en fonction de la taille de l'échantillon et du noyau utilisé pour la construction de ces estimations, des applications numériques comparatives menées sur des échantillons simulés, ont été réalisées [3].

La présente thèse comprend une introduction générale, quatre chapitres, une conclusion générale et une liste bibliographique. Les deux premiers chapitres sont consacrés aux rappels de quelques aspects théoriques existants dans la littérature ayant un lien avec la thématique abordée tandis que les deux derniers chapitres sont réservés à la présentation de notre contribution dans le cadre de nos recherches [3, 4].

Le premier chapitre donne une review sur la notion de l'estimation à noyau d'une densité de probabilité avec ses différentes variantes, à savoir : estimateur à noyau classique, estimateur à noyau asymétrique et estimateur à noyau discret. De plus, pour chaque situation, les propriétés de l'estimateur conçu et le problème du choix du noyau et du paramètre de lissage ont été présentées.

Le deuxième chapitre est constitué de deux parties majeures. Nous avons introduit dans la première partie brièvement le principe de quelques méthodes paramétriques pour l'estimation dans les chaînes de Markov. Dans la deuxième partie, on s'est concentré sur la méthode non paramétrique du noyau adapté à l'estimation des probabilités de transition d'une chaîne de Markov, en particulier sur les résultats théoriques élaborés par Roussas [79, 80] et le problème du choix du paramètre de lissage par la minimisation des normes matricielles suggérée par Cherfaoui et al. [27].

Dans le troisième chapitre, nous avons considéré un système de file d'attente $M/M/1$ à capacité finie avec des vacances multiples, Bernoulli feedback, balking, reneging et rétention des clients impatients, ainsi que la possibilité d'une panne et de réparation du serveur. Dans un premier temps, à l'aide de la méthode Q-matrice, nous avons dégagé analytiquement les probabilités d'état de la chaîne de Markov bidimensionnelle décrivant le modèle à l'état stationnaire. De plus, à base d'une application numérique, nous avons illustré la sensitivité des mesures de performance de ce système à ses paramètres de départ. Dans un second temps, nous avons analysé par simulation, à l'aide de plusieurs outils statistiques, l'effet de l'estimation des paramètres de départ définissant ce même modèle sur les propriétés statistiques (variation, biais, MSE, distribution) de ses mesures de performance.

Dans le quatrième chapitre, nous avons proposé de considérer la file d'attente $D/G/1$ (respectivement, la file d'attente $GI/D/1$) à capacité finie que nous avons modélisé par une chaîne de Markov induite à temps discret. Par la suite, nous avons analysé le problème du choix du paramètre de lissage via la minimisation des normes matricielles, lorsque nous considérons l'estimateur à noyau discret des matrices de transition, P, correspondantes aux chaînes de Markov décrivant les deux files d'attente $GI/D/1/N$ et $D/G/1/N$. Pour ce faire, nous avons développé des formes explicites des expressions résultantes des trois normes matricielles $\|.\|_1$, $\|.\|_2$ et $\|.\|_\infty$, à minimiser afin de sélectionner le paramètre de lissage lors de l'estimation des matrices P. Par ailleurs, pour étayer et illustrer nos propositions, deux études de simulation comparatives sont présentées : La première concerne l'effet de la taille de l'échantillon tandis que la deuxième concerne l'effet de l'intensité du trafic.

On termine par une conclusion générale, relatant tous les résultats obtenus et applications entreprises dans cette thèse ainsi que les perspectives de recherche sur cette thématique.

Chapitre 1

Estimateur à noyau de la densité de probabilité

Introduction

En général, dans les statistiques, la densité f génère l'échantillon. Mais la question qui se pose, étant donnée un échantillon, est-ce-que nous pouvons approximativement recréer leur fonction de densité ?

Afin d'estimer la densité inconnue f, une première approche dite paramétrique consiste à supposer que f appartient à une famille de densités continues ou discrètes qui peuvent être décrites par un certain modèle $f(x, \theta)$ de paramètre $\theta \in \Theta \subseteq \mathbb{R}^d$, avec $d < \infty$ (le paramètre θ est de dimension fini). Le statisticien qui opte pour une telle approche possède une bonne connaissance a priori du phénomène aléatoire. Ces modèles paramétriques peuvent être modifiés, lorsque les données présentent des phénomènes spécifiques. Cependant, lorsqu'aucune information n'est disponible sur le phénomène étudié ou le paramètre θ est de dimension infini, l'application d'un modèle paramétrique n'est pas satisfaisant. Pour pallier les insuffisances et les défauts des familles paramétriques, une seconde approche dite non-paramétrique consiste à estimer, à partir des observations, une fonction inconnue sans spécifier préalablement la forme de cette fonction. Une petite recherche dans la littérature nous permet de rendre compte que de nombreuses approches non paramétriques ont été proposées pour l'estimation d'une densité de probabilité. Cependant, la technique qui a rencontré le plus de succès est bien la méthode du noyau vue la simplicité de sa forme, ses modes de convergence multiples et sa flexibilité lors du choix de ses paramètres.

Dans ce chapitre, après avoir décrit l'origine de l'estimateur à noyau introduit par Parzen [73] Rosenblatt [77] dédié spécialement pour des densités à support non borné ($x \in \mathbb{R}$), nous avons énoncé ses propriétés, les noyaux les plus usités et les techniques de sélection du paramètre de lissage dans ce cadre. Par la suite, notre attention sera portée sur les méthodes introduites pour remédier au problème d'effet du biais au bornes, engendré par les noyaux classiques lors de l'estimation de densités à support borné ou discret, en particulier sur les méthodes qui se basent sur les noyaux asymétriques continus et les noyaux discrets. En effet, nous allons présenter quelques noyaux asymétriques utilisés dans le cas de variables définies sur $[0; 1]$ et $\mathbb{R}_+$ et les noyaux discrets utilisés dans le cadre d'estimation de fonctions de masse (variables discrètes).

Enfin, la question du choix de noyau et du paramètre de lissage ainsi que les propriétés des estimateurs conçus dans ces deux dernières situations seront également présentées.

1.1 Estimateur à noyaux symétriques (classiques)

L'idée de construction de l'estimateur à noyau d'une densité consiste à évaluer la densité f au point x, en comptant le nombre d'observations tombées dans un certain voisinage de $x \in \mathbb{R}$. Plus précisément, le principe de sa construction se base sur le lien entre une densité de probabilité et la fonction de répartition qui lui est associée, tout en exploitant l'estimateur empirique de cette dernière fonction [90].

Définition 1.1. *Soit $X = (X_1, X_2, ..., X_n)$ une suite de variables aléatoires $i.i.d$, de densité de probabilité inconnue f à partir de l'échantillon X. On appelle estimateur à noyau de Parzen-Rosenblatt de $f : \mathbb{R} \to \mathbb{R}_+$ la variable aléatoire donnée par :*

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{1.1}$$

où $h = h(n)$ est appelé paramètre de lissage et la fonction K est appelée noyau. De plus, h et K vérifient respectivement les conditions (1.2) et (1.3).

$$h(n) \to 0 \; et \; nh \to \infty \; quand \; n \to \infty \tag{1.2}$$

$$K(u) = K(-u); \quad \int_{\mathbb{R}} K(u)du = 1; \quad \int_{\mathbb{R}} uK(u)du = 0; \quad \int_{\mathbb{R}} u^2 K(u)du = \sigma_K^2 < \infty. \tag{1.3}$$

1.1.1 Propriétés de l'estimateur à noyau classique

Juste après l'introduction de l'estimateur à noyau de la densité par Rosenblatt (1956) [77], en 1962 Parzen [73] a étudié les propriétés fondamentales de cet estimateur. Depuis, cet estimateur est devenu un objet classique étudié par les statisticiens. L'estimateur de la densité de probabilité par la méthode du noyau est le plus répandu aujourd'hui, car il répond au problème du choix des différents paramètres dans l'estimation à histogramme et possède de bonnes propriétés. Dans cette section, nous allons résumer quelques résultats obtenues sur les propriétés de l'estimateur en question.

Espérance, Biais et Variance de l'estimateur

Les expressions de l'espérance, biais et de la variance locales de l'estimateur à noyau, défini dans (1.1), sont données respectivement comme suit :

$$\begin{aligned}
\mathbb{E}\left(\hat{f}(x)\right) &= f(x) + \frac{h^2}{2} f''(x)\mu_2(K) + o(h^2), \\
Biais\left(\hat{f}(x)\right) &= \mathbb{E}\left(\hat{f}(x)\right) - f(x) = \frac{h^2}{2} f''(x)\mu_2(K) + o(h^2), \\
\mathbb{V}ar\left(\hat{f}(x)\right) &= \frac{f(x)}{nh}\int_{-\infty}^{\infty} K^2(y)dy - \frac{f'(x)}{n}\int_{-\infty}^{\infty} yK^2(y)dy - \frac{1}{n}\left(f(x) + biais\hat{f}(x)\right)^2,
\end{aligned}$$

où $\mu_2(K) = \int_{-\infty}^{\infty} y^2 K(y)dy$ (pour plus de détails voir Silverman [90]).

Comportement asymptotique de l'estimateur et ses propriétés

Parzen [73] a élaboré les conditions de plusieurs types de convergence de l'estimateur à noyau ainsi que la convergence de ses propriétés. Les principaux résultats obtenus par l'auteur, sont résumés dans le théorème ci-dessous :

Théorème 1.1. *(Parzen [73])*
Sous les conditions suivantes :

1. $\lim_{n\to+\infty} h(n) = 0$ *et* $\lim_{y\to+\infty} |yK(y)| = 0$,
2. $\sup_y |K(y)| < \infty$ *et* $\int_{-\infty}^{\infty} |K(y)|dy < \infty$,
3. $\int_{-\infty}^{\infty} K(y)dy = 1$.

on a :

- $\lim_{n\longrightarrow\infty} \mathbb{E}\left(\hat{f}(x)\right) = f(x)$ *et* $\lim_{n\longrightarrow\infty} nh\mathbb{V}ar\left(\hat{f}(x)\right) = f(x)\int_{-\infty}^{\infty} K^2(y)dy$.
  Si de plus, $\lim_{n\to\infty} nh(n) = \infty$, *alors*
- $\lim_{n\longrightarrow\infty} MSE(\hat{f}(x), f(x)) = \lim_{n\longrightarrow\infty} \mathbb{E}\left(\hat{f}(x) - f(x)\right)^2 = 0$,
  pour tout point x pour lequel la densité f est continue.
- $\lim_{n\longrightarrow\infty} MISE(\hat{f}, f) = \lim_{n\longrightarrow\infty} \int_{\mathbb{R}} MSE\left(f(x), \hat{f}(x)\right) dx = 0, \quad \forall f \in \mathbb{L}^p$.
- $\hat{f}(x) \xrightarrow{loi} \mathcal{N}\left(\mathbb{E}\left(\hat{f}(x)\right), \mathbb{V}ar\left(\hat{f}(x)\right)\right)$.
- $\forall \epsilon > 0, \ P\left(\sup_{x\in\mathbb{R}} |\hat{f}(x) - f(x)| < \epsilon\right) = 1$, *si la transformée de Fourier*
  $\tilde{K}(z) = \int_{-\infty}^{\infty} \exp(-izy)K(y)dy$ *est absolument intégrable.*

On note par $\mathbb{L}^p$: l'ensemble des fonctions f définies sur $\mathbb{R}$, telle que $\int |f(x)|^p dx < \infty$, $p > 0$.

Le Théorème suivant, élaboré par Nadaraya [68], concernant la convergence uniforme presque complète de l'estimateur à noyau classique.

Théorème 1.2. *(Nadaraya [68])*
Si K est un noyau positif à variation bornée et f est uniformément continue,

$$si \lim_{n\longrightarrow\infty} h(n) = 0 \ et \ \sum_{n=1}^{\infty} \exp(-\gamma n h(n)^2) < \infty, \ \forall \gamma > 0,$$

alors :

$$\sup_x |\hat{f}(x) - f(x)| \longrightarrow 0 \ avec \ une \ probabilité \ 1.$$

Silverman [91] a donné le même théorème sur la convergence presque complète en remplaçant la condition $\sum_{n=1}^{\infty} \exp(-\gamma n h^2) < \infty$, par les deux conditions suivantes :

$$\lim_{n\longrightarrow\infty} h(n) = 0 \ \text{et} \ \lim_{n\longrightarrow\infty} \frac{\log n}{nh(n)} = 0.$$

Théorème 1.3. *(Silverman [91])*
Si on a :

$$\lim_{n\longrightarrow\infty} h(n) = 0 \ \textit{et} \ \lim_{n\longrightarrow\infty} \frac{\log n}{nh(n)} = 0,$$

et K satisfait aux conditions suivantes :

- *K est uniformément continue et à variation bornée sur $\mathbb{R}$,*
- *f est uniformément continue,*
- *$\int_{-\infty}^{\infty} |K(y)|dy < \infty, \ \int_{-\infty}^{\infty} \sqrt{|y \log |y||} \ |dK(y)| < \infty$,*
- *$\int_{-\infty}^{\infty} K(y)dy = 1$,*

alors :

$$\lim_{n\longrightarrow\infty} \sup_x |\hat{f}(x) - f(x)| = 0 \ \textit{Presque Sûrement}.$$

Devroye [30] à montré que $\hat{f}$ est convergent également en norme L_1.

Théorème 1.4. *(Devroye [30])*
Si :

$$\lim_{n\longrightarrow\infty} h(n) = 0, \ \lim_{n\longrightarrow\infty} nh(n) = \infty,$$

alors

$$\forall (f \in \mathcal{F}), \lim_{n\longrightarrow\infty} \int |\hat{f}(x) - f(x)|dx = 0 \ \textit{Presque Complètement},$$

où $\mathcal{F}$: Ensemble des densités de probabilité.

Vitesse de convergence

Sous l'hypothèse que f est inconnue, Wahba [97] a montré qu'on ne peut améliorer indéfiniment la convergence de l'estimateur $\hat{f}$ vers f, même pour la fonction la plus régulière possible (indéfiniment dérivable et bornée), c'est-à-dire : $MSE(\hat{f}(x), f(x))$ ne peut tendre vers 0 que d'un ordre $\frac{c}{n}$, où c est une constante. Précisément :

$$\sup_{f\in W(r,m,M)} \left(\left|\hat{f}(x) - f(x)\right|^2 \right) \text{ ne peut tendre vers 0 que d'un ordre } \frac{1}{n^{G(m,r)}},$$

où $\frac{1}{n^{G(m,r)}} = \frac{2m - 2/r}{2m + 1 - 2/r}$ est une fonction croissante par rapport à m , r, et $\lim_{m\longrightarrow\infty} G(m, r) = 1$.

$f \in W(r, m, M)$ si :

a) les $(m-1)$ premières dérivées $f^{(i)}$ de f sont absolument continues ;
b) $\int \left|f^{(m)}(x)\right|^r dx < \infty$,
c) $\sup_x \left|f^{(m)}(x)\right| \leq M < \infty$,

1.1.2 Choix du noyau

En pratique, pour la mise en œuvre de l'estimateur défini dans (1.1), il faut au préalable fixer le noyau K et le paramètre de lissage h. Dans la littérature, il existe plusieurs fonctions qui jouent le rôle d'un noyau. La Table 1.1 résume les noyaux les plus usités dans la pratique, leurs formes sont illustrées dans la Figure 1.1.

Nom	Expression	Support
Noyau Uniforme (Rosenblatt)	$K(u) = \frac{1}{2}$	$\|u\| \leq 1$
Noyau Box (boite)	$K(u) = \frac{1}{2\sqrt{3}}$	$\|u\| \leq \sqrt{3}$
Noyau Triangulaire	$K(u) = (1 - \|u\|)$	$\|u\| \leq 1$
Noyau Cosine	$K(u) = \frac{\pi}{4}\cos(\frac{\pi u}{2})$	$\|u\| \leq 1$
Noyau Gaussien	$K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$	$u \in \mathbb{R}$
Noyau Biweight (Tukey)	$K(u) = \frac{15}{16}(1 - u^2)^2$	$\|u\| \leq 1$
Noyau Triweight	$K(u) = \frac{35}{32}(1 - u^2)^3$	$\|u\| \leq 1$
Noyau Epanechnikov	$K_E(u) = \frac{3}{4\sqrt{5}}\left(1 - \frac{u^2}{5}\right)$	$\|u\| \leq \sqrt{5}$

TABLE 1.1: Noyaux classiques usuels et leurs supports.

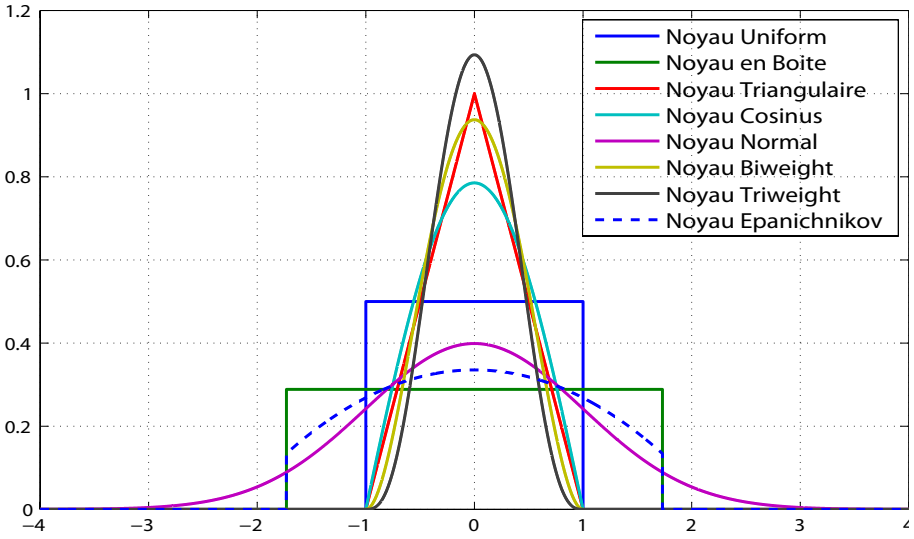

FIGURE 1.1 – La forme des noyaux usuels

Un noyau approprié aide à surmonter les problèmes des bosses (multi-modes) et de la discontinuité de la densité estimée. Par exemple, si K est une distribution gaussienne, alors la fonction de densité estimée $\hat{f}$ sera lisse et admet des dérivées de toutes ordres. Le choix du noyau K, en réalité, n'est pas un problème considérable. En effet, plusieurs étude ont montré que le noyau a peu d'influence sur la qualité de l'estimateur et ceci peut être expliquer clairement par le lemme de Bochner ci-après. De ce fait, les critères de sélection du noyau, dans ce cas, sont alors la simplicité et la rapidité des calculs.

Lemme 1.1. *(Bochner) :*

i) Soit K un noyau de Parzen-Rosenblatt et g une fonction de $\mathbb{L}^1$.
Alors, en tout point x, où g est continue, on a :

$$\lim_{h \longrightarrow 0} (K * g)(x) = g(x).$$

ii) Soit K un noyau quelconque et g une fonction de $\mathbb{L}^1$ uniformément continue, alors

$$\lim_{h \longrightarrow 0} \sup_x |(K * g)(x) - g(x)| = 0.$$

L'interprétation de ce lemme est que lorsque la fenêtre h est petite, la convolution d'une fonction g de $\mathbb{L}^1$ avec K perturbe peu cette fonction.

Remarque 1.1. *Il est à noter que les principaux résultats obtenus dans la littérature sur les convergences de l'estimateur à noyau classique sont basés sur le Lemme de Bochner ci-dessus.*

1.1.3 Choix du paramètre de lissage

D'après la formule (1.1), on constate que l'estimateur $\hat{f}(x)$ de $f(x)$ ne dépend pas seulement du noyau K, mais aussi du paramètre de lissage h. Dans la pratique, l'étape critique dans l'estimation d'une densité par la méthode du noyau est le choix du paramètre h, qui contrôle la qualité graphique de l'estimateur $\hat{f}$ [73, 90]. En effet, le choix de h est une étape importante lors de l'estimation par la méthode du noyau, dans le sens où une petite perturbation de h est suffisante pour que les caractéristiques de $\hat{f}$ changent complètement (performances numériques et/ou graphiques). Par ailleurs, si h est trop petit, le biais de l'estimateur devient petit devant sa variance et l'estimateur sera trop fluctuant. Ce qui engendre ce que nous appelons le phénomène de *sous-lissage*. Dans le cas contraire, lorsque h est trop grand, le biais prend l'ascendant sur la variance et l'estimateur varie peu, alors on obtient un phénomène de *sur-lissage*.

L'exemple présenté dans la Figure 1.2, réalisé dans le cadre d'estimation d'une densité d'une loi normale, centrée réduite, à partir d'un échantillon de taille $n = 200$, est une illustration de l'influence du choix du paramètre de lissage sur les caractéristiques graphiques de l'estimateur en question. Les graphes des trois estimateurs présentés, mis en évidence le phénomène de sur-lissage dans le cas $h = 0.8$ (trop grand), le phénomène de sous-lissage dans le cas $h = 0.1$ (trop petit) et l'estimation idéale dans le cas $h^* = 0.3334$ (h^* est l'optimal au sens du minimisation de ISE).

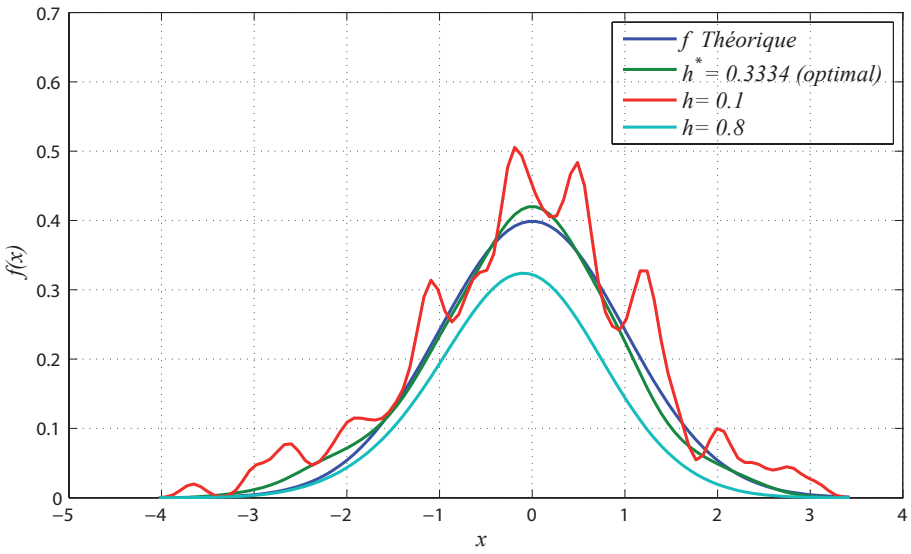

FIGURE 1.2 – Illustration graphique des phénomènes : sur-lissage $(h = 0.8)$, sous-lissage $(h = 0.1)$ et estimation idéal $(h^* = 0 : 3334)$

Il existe dans la littérature plusieurs techniques qui ont été proposées pour la sélection de ce paramètre que l'on peut regrouper en deux familles :

- Méthodes de plug-in (re-injection)
- Méthodes de Cross-Validation (Validation-croisée).

La multitude des méthodes de sélection du paramètre de lissage et leurs diversités du point de vue de leurs principes, est due au fait que ces méthodes restent incomplètes. Autrement dit, ces méthodes ont toujours des inconvénients [109], soit au sens de la qualité numériques de l'estimateur $\hat{f}$, par rapport à une norme d'erreur bien déterminée, soit au sens de la qualité graphique de l'estimateur (l'allure graphique de la courbe de $\hat{f}$ est sur-lissée ou sous-lissée).

Dans ce qui suit, nous allons présenter quelques méthodes de sélection de h les plus usuelles.

Choix optimal

La décision du choix du paramètre de lissage h suppose la spécification d'un critère d'erreur qui puisse être optimisé. De ce fait, il est claire que l'optimalité n'est pas un concept absolu, du fait que cette optimalité est étroitement liée au choix d'un critère, qui fait intervenir à la fois la densité inconnue f et l'estimateur $\hat{f}$ (c'est-à-dire le paramètre h et le noyau K).

Supposons que nous nous intéressons au choix du paramètre de lissage h qui minimise le $MISE$ (**Erreur Quadratique Intégrée Moyenne**), c'est-à-dire :

$$\begin{aligned} h_{opt} &= \arg\min_h MISE\left(f,\hat{f}\right) = \arg\min_h \int \mathbb{E}\left(\hat{f}(x)-f(x)\right)^2 dx \\ &= \arg\min_h \left[\int_R Biais\left(\hat{f}(x)\right)^2 dx + \int_R \mathbb{V}ar\left(\hat{f}(x)\right) dx\right]. \end{aligned}$$

Afin de déterminer le h optimal, au sens du $MISE$, nous allons exploiter le résultat suivant :

Théorème 1.5. *(Scott [86])*
Si f a une dérivée seconde absolument continue, $f^{(2)} \in \mathbb{L}^2$, le noyau $K \in \mathbb{L}^2$ est une densité de probabilité continue, symétrique et de variance $\sigma_K^2 < \infty$, alors, sous les conditions $h(n) \longrightarrow 0$ et $nh(n) \longrightarrow \infty$, on a le développement asymptotique :

$$MISE\left(f,\hat{f}\right) = \frac{h^4}{4}\sigma_K^4 \int \left(f''(x)\right)^2 dx \; + \; \frac{\int K^2(x)dx}{nh} \; + \; o\left(h^5+\frac{1}{n}\right), \tag{1.4}$$

où $\mathbb{L}^2$ est l'ensemble des fonctions f définies sur $\mathbb{R}$, tel que $\int |f(x)|^2 dx < \infty$.

A partir de l'expression (1.4) on définit la quantité suivante :

$$AMISE\left(f,\hat{f}\right) = \frac{h^4}{4}\sigma_K^4 \int \left(f''(x)\right)^2 dx \; + \; \frac{\int K^2(x)dx}{nh}, \tag{1.5}$$

tel que l'AMISE est **l'Erreur Quadratique Intégrée Moyenne Approchée**. On remarque que le premier terme du membre à droite du développement (1.5) est un terme de $Biais^2$, alors que le second est un terme de variance. De plus, on constate que, le terme du biais est une fonction croissante en h, alors que le terme de la variance est une fonction décroissante en h. C'est-à-dire, les deux termes varient dans le sens inverse par rapport à h. Une largeur de fenêtre h trop importante entraînera une augmentation du biais et une diminution de la variance (phénomène de sur-lissage), alors qu'une largeur de fenêtre trop petite provoquera une augmentation de la variance et une diminution du biais (phénomène de sous-lissage). De ce fait, le paramètre de lissage h^* optimal au sens du critère de l'$AMISE$, devra réaliser un compromis entre les valeurs de la variance et celle du biais (voir Figure 1.3).

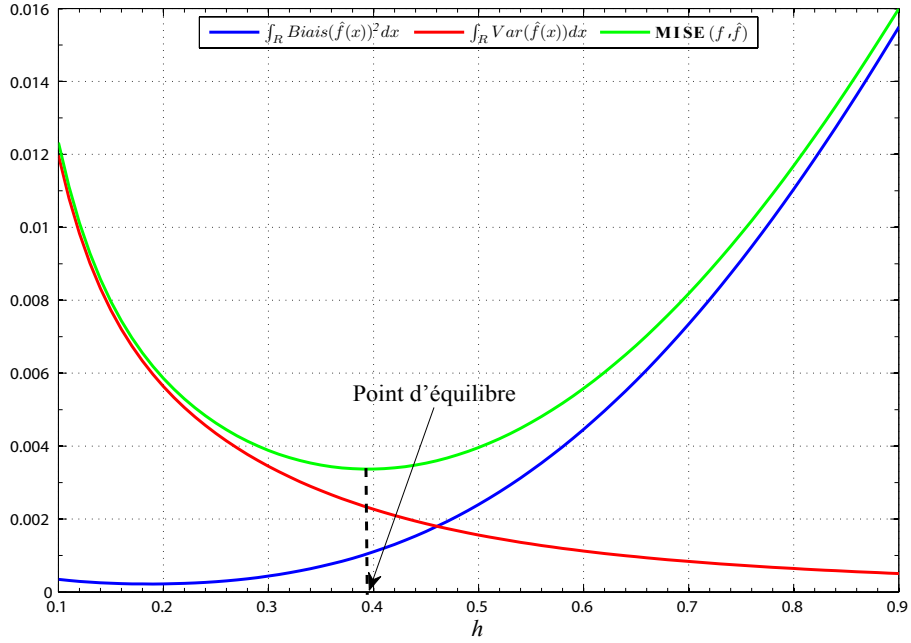

FIGURE 1.3 – Illustration du principe de l'équilibre biais-variance à l'aide d'une simulation.

Par ailleurs, pour obtenir le paramètre de lissage h^* qui minimise l'AMISE, il suffit de résoudre le système suivant :

$$\begin{cases} \frac{dAMISE}{dh} = 0, \\ \frac{d^2AMISE}{dh^2} > 0. \end{cases} \tag{1.6}$$

Ainsi, à partir de l'expression (1.5), on aura :

$$h^* = \left[\frac{R(K)}{\sigma_K^4 R(f'')}\right]^{1/5} n^{-1/5}, \tag{1.7}$$

avec $R(g) = \int (g(x))^2 dx$. Notons que h^* est une quantité déterministe qui dépend du nombre d'observations n. Ainsi, la valeur du $AMISE$ optimale ($AMISE^* = AMISE(h^*)$) est donnée par :

$$AMISE^* = \frac{5}{4}\left[\sigma_K^4 R^4(K) R(f'')\right]^{1/5} n^{-4/5}. \tag{1.8}$$

D'après l'expression (1.7), on constate qu'en plus de sa nature asymptotique, la largeur de la fenêtre optimale h^* dépend de la densité inconnue f au travers du paramètre $R(f'')$. Cette largeur de fenêtre "idéale" (relativement au critère d'erreur retenu) n'est donc pas directement calculable dans la pratique. Une façon classique de remédier à ce dernier problème consiste à remplacer la quantité $R(f'')$ inconnue par un estimateur approprié ou une quantité bien définie, d'où le principe des méthodes plug-in. Dans ce qui suit, nous allons présenter deux méthodes de plug-in, à savoir : la règle de référence de Silverman et la méthode de Sheather et Jones.

Méthode Rule Of Thumb (règle de référence)

L'estimateur "Rule Of Thumb" du paramètre de lissage, noté h_{rot} proposé par Silverman [90], suppose que nous utilisons le noyau gaussien pour estimer une densité f d'une distribution

normale centrée (la moyenne égale à 0) et de variance σ^2. De ce fait, la quantité $R(f'')$ est définie par :

$$R(f'') = \int (f''(x))^2 \, dx = \frac{3}{8}\sqrt{\pi}\sigma^{-5}. \tag{1.9}$$

En substituant $R(f'')$ et K par leurs expressions dans (1.7), on obtient :

$$h_{rot} = 1.06\sigma n^{-1/5}. \tag{1.10}$$

Il suffit donc d'estimer σ à partir de nos données et de le substituer dans la formule (1.7). D'après Silverman [90], la formule (1.10) donnera de bons résultats si la population est réellement normalement distribuée. Par contre, elle peut donner une distribution trop lissée si la population est multimodale. Dans ce cas, de meilleurs résultats peuvent être obtenus, si on utilise l'interquartile IQ à la place de l'acart-type σ. Dans le cas où X suit une loi normale centrée réduite, l'écart interquartile est $IQ = 1.394$ alors h_{rot} de l'équation (1.10) devient

$$h_{rot} = 1.06IQn^{-1/5}.$$

Cette dernière formule peut aussi donner une distribution trop lissée si la vraie densité est multimodale. Parfois cette dernière donne des résultats moins bons que si l'on avait utilisé l'écart type, d'où le meilleur des deux méthodes peut être obtenu en utilisant un estimateur adaptatif de l'étendue. C'est-à-dire, en utilisant A au lieu de σ dans la formule (1.10) où A est défini par $A = \min(\sigma, IQ)$, donc la formule pour h_{rot} devient :

$$h_{rot} = 1.06An^{-1/5}. \tag{1.11}$$

Méthode de Sheather et Jones

Sheather et Jones (1991) [89], recommandent l'utilisation de l'estimateur naturel $\hat{R}_a(f'')$, en faisant observer que le terme de biais $\frac{R(K'')}{na^5}$ est positif et peut donc servir à annuler le terme de biais (négatif) de l'erreur quadratique moyenne entre $\hat{R}_a(f'')$ et $R(f'')$. Afin de faire disparaître quelques effets indésirables du terme du biais, les deux auteurs sont contraints de mettre en place une procédure de type plug-in en trois étapes.
Pour cela, Sheather et Jones [89], proposent d'estimer $R(f'') = \int_{-\infty}^{+\infty}(f'')^2 dx$ par :

$$S(a) = \frac{1}{n(n-1)a^5}\sum_{i=1}^{n}\sum_{j=1}^{n} L^{(4)}\left(\frac{x_i - x_j}{a}\right), \tag{1.12}$$

et de le substituer dans la formule (1.7). Où $L^{(4)}$ désigne la dérivée quatrième du noyau suffisamment lisse L et a est un nouveau paramètre de lissage appelé paramètre de lissage pilote.

Validation croisée non biaisée (UCV)

Cette méthode a été proposée par Rudemo(1982)[81] et par Bowman(1984)[15]. Elle consiste à choisir le paramètre de lissage qui minimise un estimateur convenable de :

$$\begin{aligned} ISE &= \int_{\mathbb{R}} \left[\hat{f}(x) - f(x)\right]^2 dx, \\ &= \int_{\mathbb{R}} \hat{f}^2(x)dx - 2\int_{\mathbb{R}} \hat{f}(x)f(x)dx + \int_{\mathbb{R}} f^2(x)dx. \end{aligned} \tag{1.13}$$

Puisque $\int_{\mathbb{R}} f^2(x)dx$ ne dépend pas du paramètre de lissage h. On peut choisir alors le paramètre de lissage de façon à ce qu'il minimise un estimateur de :

$$\int_{\mathbb{R}} \hat{f}^2(x)dx - 2\int_{\mathbb{R}} \hat{f}(x)f(x)dx.$$

Maintenant, on veut trouver un estimateur de $\int_{\mathbb{R}} \hat{f}(x)f(x)dx$. Pour cela, remarquons que

$$\int_{\mathbb{R}} \hat{f}(x)f(x)dx = \mathbb{E}\left(\hat{f}(x)\right).$$

L'estimateur empirique de $\int_{\mathbb{R}} \hat{f}(x)f(x)dx$, par validation croisée est alors

$$\frac{1}{n}\sum_{i=1}^{n} \hat{f}_i(x_i),$$

où $\hat{f}_i$ est l'estimateur de la densité construit à partir de l'ensemble de points sauf au point x_i donné par :

$$\hat{f}_i(x_i) = \frac{1}{(n-1)h}\sum_{j=1,j\neq i}^{n} K\left(\frac{x_i - x_j}{h}\right). \tag{1.14}$$

Ainsi, le critère à minimiser est donné par :

$$UCV(h) = \int_{\mathbb{R}} \hat{f}^2(x)dx - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_i(x_i). \tag{1.15}$$

Nous notons par h_{ucv} l'estimateur de h qui minimise $UCV(h)$.

La popularité de cette méthode est due à la motivation intuitive et au fait que cet estimateur est asymptotiquement optimal sous de faibles conditions. L'optimalité asymptotique de la méthode validation croisée non biaisée a été obtenue par Stone [94]. Cependant, cette méthode présente deux problèmes majeurs. D'une part, son manque de robustesse par rapport aux changements de la taille de l'échantillon, c'est-à-dire, le résultat de simulation peut se révéler extrêmement variable d'un échantillon à un autre. D'autre part, la fonctionnelle à minimiser a souvent tendance à présenter plusieurs minimums locaux [45]. Pour d'autres études sur les limites de la méthode UCV le lecteur peut se référé à Hall [44], Burman [18] et Scott et Terrell [87].

Validation croisée biaisée (BCV)

Le critère de validation croisée biaisée, a été introduit, en 1987 par Scott et Terrell [87], pour remédier aux problèmes de la méthode "validation croisée non biaisée". Avant d'exposer cette technique, rappelons que l'Erreur Quadratique Intégrée Moyenne Approchée s'écrit sous la forme :

$$AMISE = \frac{h^4}{4}\sigma_K^4 R(f'') + \frac{R(K)}{nh}. \tag{1.16}$$

Le paramètre de lissage basé sur la méthode validation croisée biaisée est la valeur de h qui minimise un estimateur du $AMISE$. À partir de la formule (1.16), il est claire qu'afin d'estimer l'$AMISE$, il suffit d'estimer $R(f'')$. Un estimateur naturel de ce dernier terme est donné par $R(\hat{f}'')$. Finalement, Scott et Terrell [87] ont proposé la forme de l'estimateur, par validation croisée, du $AMISE$ à minimiser (critère de $BCV(h)$), donné par la proposition suivante.

Proposition 1.1. *(Scott et Trell [87])*
Soit $X_1, X_2, ..., X_n$ un n-échantillon $i.i.d$ issu d'une variable aléatoire X de fonction de densité f. Pour un noyau K, on obtient :

$$BCV(h) = \frac{R(K)}{nh} + h^4 \frac{\mu_2^2(K)}{4n^2} \sum_i \sum_{j,j\neq i} K_h^{(2)} * K_h^{(2)}(X_i - X_j). \tag{1.17}$$

Des résultats de simulations ont été obtenus pour la méthode de validation croisée biaisée dans le travail de Park et Marron [72]. Les auteurs ont constaté que la méthode validation croisée biaisée présente les mêmes inconvénients que la méthode validation croisée non biaisée.

Validation croisée par le maximum de vraisemblance (LCV)

Le paramètre de lissage basé sur la méthode de validation croisée par le maximum de vraisemblances est le paramètre qui maximise la fonction suivante :

$$LCV(h) = \left(\frac{1}{n} \sum_{i=1}^{n} \log \left(\hat{f}_i(x_i) \right) \right), \tag{1.18}$$

avec $\hat{f}_i(x_i)$ est donné par (1.14).

1.2 Problème d'effet du biais aux bornes

Soient $X_1, X_2, ..., X_n$ un n-échantillon issu d'une densité de probabilité inconnue f. L'estimateur classique de $f(x)$ obtenu par la méthode du noyau, cité dans la Section 1.1, s'écrit sous la forme :

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K \left(\frac{x - X_i}{h} \right), \tag{1.19}$$

où h représente le paramètre de lissage et K est un noyau qui vérifie les conditions (1.3).

L'estimateur à noyau continu (1.19) a été développé primairement pour les densités à supports continus et non-bornés. La fonction noyau K est classiquement symétrique (i.e. $K(-x) = K(x)$), elle est considérée comme moins importante que le paramètre de lissage h. Bien qu'un noyau symétrique soit approprié pour ajuster des densités à supports non-bornés, il ne l'est pas pour des densités à supports compacts ou bornés d'un côté et a fortiori à supports discrets. En effet, lorsque on veut estimer des densités à support borné au moins d'un seul coté, l'estimateur à noyau classique devient non consistant, à cause des effets du biais aux bornes. Ce problème est dû à l'utilisation d'un noyau symétrique qui assigne des poids positifs en dehors du support lorsque le lissage est pris en compte près de la borne.

Dans ce qui suit, nous allons présenter quelques techniques de correction d'effet du biais aux bornes.

1.3 Quelques techniques de correction d'effet du biais

1.3.1 Noyau Miroir (Schuster)

L'idée de cette méthode, développée par Deheuvels et Hominal (1979) [29] et ensuite par Schuster (1985) [85], est d'ajouter une "masse manquante" par réflexion de l'échantillon et qui concerne les données aux frontières. Elles se focalisent sur le cas où les variables sont positives $(x \in [0, \infty[)$. Formellement et sous sa forme plus simple, il consiste à remplacer $K\left(\frac{x-X_i}{h}\right)$ par $K\left(\frac{x-X_i}{h}\right) + K\left(\frac{x+X_i}{h}\right)$. L'estimateur de la densité devient alors :

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n}\left[K\left(\frac{x-X_i}{h}\right) + K\left(\frac{x+X_i}{h}\right)\right]. \tag{1.20}$$

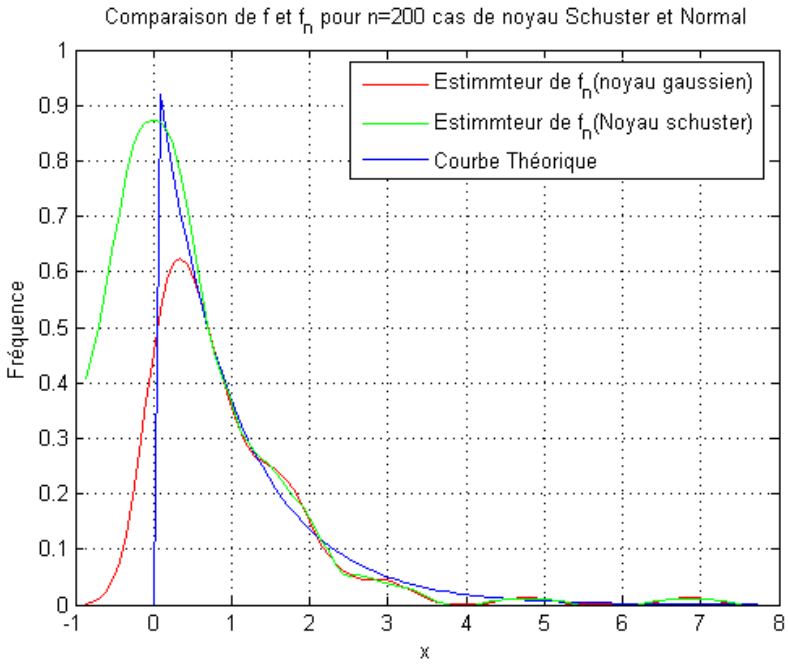

FIGURE 1.4 – Exemple illustratif du noyau Miroir (Schuster).

1.3.2 Noyaux tronqués et normalisés

Une autre technique de correction d'effet du biais aux bornes est proposée, en 1979, par Gasser et Müller [38]. Cette technique appelée méthode du *noyau tronqué et normalisé* (Cut-and-Normalized Kernel). Contrairement au noyau définis dans (1.1), la construction du noyau "Cut-and-Normalized" dépend du point x où la densité est estimée.

Étant donnés un noyau K et une densité cible f, définie sur un support positif $[0, \infty[$, le noyau tronqué et normalisé, noté (K_{cn}), est construit en premier lieu par la troncation du noyau K au point c et ensuite de le normalisé à l'unité de masse, c'est-à-dire $\int_{-\infty}^{c} K_{cn}(u) = 1$. Ici, nous avons deux exemples de noyaux tronqués et normalisés au point $x = ch,\ c \geq 0$:

1. Le noyau exponentielle tronqué et normalisé

$$K_{cn} = \frac{e^{-|t|}}{\int_{-\infty}^{c} e^{-|t|}dt}\mathbf{1}_{]-\infty,c]}, \tag{1.21}$$

2. Le noyau Epanechnikov tronqué et normalisé

$$K_{cn} = \frac{(1-t^2)}{\int_{-1}^{c}(1-t^2)dt}\mathbf{1}_{[-1,c]}, \tag{1.22}$$

où, $\mathbf{1}$ est la fonction indicatrice.
On peut montrer que l'estimateur à noyau tronqué et normalisé est équivalent à l'estimateur à noyau miroir de Schuster [105]. En remplaçant K dans la formule (1.1) par K_{cn}, et notons par f_{cn} l'estimateur résultant, on aura

$$Biais(f_{cn}(x)) = -hf'(x)\int_{-\infty}^{c} tK_{cn}dt + \frac{h^2}{2}f''(x)\int_{-\infty}^{c} t^2K_{cn}dt + o(h^2). \tag{1.23}$$

Dans le cadre de la comparaison de ce biais avec celui de l'estimateur classique pour des densité à support positif aux points $x \in [0, h[$ définie par :

$$Biais(f_{cn}(x)) = -f(x)\int_{c}^{1} K(t)dt - hf'(x)\int_{-1}^{c} tK(t)dt + \frac{h^2}{2}f''(x)\int_{-1}^{c} t^2K(t)dt + o(h^2), \tag{1.24}$$

Zhang et al. [106] ont montré que le noyau tronqué et normalisé élimine parfaitement la non-consistance du l'estimateur à noyau classique. Cependant, l'existence du terme $-hf'(x)\int_{-\infty}^{c} tK_{cn}dt$ dans la formule (1.23) indique que, pour $x = c/h\,;\, c \geq 0$, le biais de l'estimateur à noyau tronqué et normalisé ne converge vers zéro qu'avec un ordre $o(h)$ qu'est petit comparativement au taux de convergence de l'estimateur classique qu'est d'ordre $o(h^2)$, sauf dans le cas où $f'(x) = 0$.

1.3.3 Noyaux aux bornes

Une technique plus générale, que les deux précédentes, pour la correction du biais aux bornes est l'estimateur du noyau aux bornes. La méthode du noyau aux bornes est proposée initialement en 1979 par Gasser et Müller [38], suivie par Cheng et al. (1997) [26] et Zhang et Karunamuni (1998, 2000) [106, 107]. Notons que, à partir de la formule (1.24) le problème du biais au borne est causé par le fait que $\int_{c}^{1} K(u)du \neq 0$ et $\int_{-1}^{c} uK(u)du \neq 0$. L'idée de la présente technique est d'utiliser un noyau, noté K_c, qui élimine ce biais. Dans ce cas, il suffit de sélectionner K_c parmi les fonctions K, non nécessairement non négatives, satisfaisants les conditions suivantes :

$$\int_{c}^{1} K(u)du = 0, \quad \int_{-1}^{c} uK(u)du = 0 \quad \text{et} \quad \int_{-1}^{c} u^2K(u)du < \infty. \tag{1.25}$$

Ces dernières conditions s'écrivent souvent dans la littérature sous leurs formes équivalentes :

$$\int_{-1}^{c} K(u)du = 1, \quad \int_{-1}^{c} uK(u)du = 0 \quad \text{et} \quad \int_{-1}^{c} u^2K(u)du < \infty. \tag{1.26}$$

1.3.4 Estimateur de Jones-Foster

La méthode du noyau au borne est plus générale que la méthode de réflexion dans le sens où elle peut s'adapter à toutes les formes de densités. Cependant, les résultats fournis par cette technique peuvent être négatives. Plusieurs travaux sont dédiés pour corriger ce problème, parmi ces travaux on cite le travail de Jones et Foster [48]. En se basant sur la technique "Jackknife généralisée", Jones et Foster [48] ont proposé un estimateur, noté $\hat{f}_{JF}$, non négatif pour la correction du biais, dont la forme est donnée par :

$$\hat{f}_{JF}(x) = \hat{f}_{cn}(x)\ \exp\left(\frac{\hat{f}_b(x)}{\hat{f}_{cn}(x)} - 1\right),\ \ x \in [0, \infty[. \tag{1.27}$$

Une forme plus générale de cet estimateur a été proposée par les mêmes auteurs [48] :

$$\hat{f}_{JF}(x) = \hat{f}_{cn}(x)\ g\left(\frac{\hat{f}_b(x)}{\hat{f}_{cn}(x)} - 1\right),\ \ x \in [0, \infty[, \tag{1.28}$$

à condition que $\forall u$, pour un ϵ suffisamment petit et $G(u) \geq 0$ la fonction $g(\epsilon) \simeq 1 + \epsilon$.

Exemple 1.1. *Une famille d'exemples pour le choix de G est :* $G(u) = \left(1 + \frac{u}{k}\right)^k,\ \ \forall k.$

Quoique les méthodes précédentes diminuent le biais, aux bornes, mais elle reste peu efficace car le biais, qui est de l'ordre $o(h)$, reste considérable, si on le compare au biais de l'intérieur du support qui de l'ordre $o(h^2)$. Pour obtenir un biais aux bornes de même ordre que celui de l'intérieur, Devroye et Györfi (1985) [31] et Marron et Ruppert (1994) [63], ont proposé d'appliquer une transformation sur les données originales de telle façon que la dérivée d'ordre 1 de la densité des variables transformées soit égale à zéro. Ensuite d'utiliser la méthode de réflexion pour estimer la densité des données transformées.

L'objectif étant de trouver, cette fois, un biais du même ordre mais sans transformation des données. Plusieurs autres auteurs, ont proposé d'utiliser les noyaux adaptés dans la région des bornes et le noyau standard à l'intérieur du support (voir Müller (1991) [67] pour l'estimateur à noyau optimal aux bornes, Lejeune et Sarda (1992) [60] pour l'estimation linéaire local et Jones (1993) [47], pour l'estimation à noyau aux bornes).

L'inconvénient de ces estimateurs est qu'ils attribuent des poids négatifs aux valeurs du voisinage des bornes. Plusieurs solutions ont été proposées dans la littérature à cette difficulté, la solution la plus simple est de remplacer le noyau symétrique par un noyau asymétrique adaptés, qui n'assigne pas un poids en dehors du support de la densité que l'on veut estimer. Cette idée est due à l'origine aux travaux de Chen [23, 25]. Ainsi, la naissance de la notion des noyaux asymétriques où l'expression de l'estimateur à noyau d'une densité f, sera écrite alors sous la forme :

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{x,h}(X_i) \quad x \in \mathbb{R}, \tag{1.29}$$

avec h est le paramètre de lissage et $K_{x,h}$ sera dit alors "noyau asymetrique" de cible x et de fenêtre h.
Dans la section suivante des exemples des noyaux asymétriques continus seront présentés.

1.4 Estimation à noyaux asymétriques continus

1.4.1 Noyaux Beta

Le noyau beta a été proposé par Brown et Chen (1999) [16], et Chen (1999,2000) [23, 24] pour l'estimation non paramétrique de la courbe de régression et des densités unidimensionnelles défini sur un support compact.

L'idée dans [16, 23, 24] est d'utiliser le noyau beta pour estimer la densité de probabilité à support compact $[0,\ 1]$ et ainsi de régler le problème du biais aux bornes. L'estimateur de la densité sera alors de la forme :

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(X_i, \frac{x}{h}+1, \frac{(1-x)}{h}+1\right), \tag{1.30}$$

où $K(.,\alpha,\beta)$ représente la densité de la distribution $Beta$ de paramètres α et β, donnée par :

$$K(x,\alpha,\beta) = \frac{x^{\alpha}(1-x)^{\beta}}{\mathcal{B}(\alpha,\beta)}, \quad x \in [0,1],$$

avec

$$\mathcal{B}(\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)},$$

et

$$\Gamma(p) = \int_0^{\infty} x^{p-1}e^{-x}dx, \ \ p>0.$$

Le noyau beta a deux avantages, premièrement il peut parfaitement estimer les densités à support compact $[0,\ 1]$ 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 des bornes et fourni une réduction de la variance (voir figure (1.5)). Charpentier et al. [22] ont montré par simulation que l'estimateur à noyau beta est plus performant quand on le compare à d'autre estimateurs avec des noyaux standards.

1.4.2 Noyaux Gamma

Soit $X_1,\ X_2, ...,\ X_n$ un n-échantillon issu d'une densité de probabilité inconnue f qui est définie sur un support positif ([0, ∞[) et deux fois continûment derivable ($f \in \mathcal{C}^2([0,\infty[)$). Parmi les estimateurs proposés pour cette famille de distributions, on cite l'estimateur de Chen [25] qui suggère de remplacer l'estimateur classique par :

$$\hat{f}_G(x) = \frac{1}{n}\sum_{i=1}^{n} K_{(\rho_h(x),h)}(X_i), \tag{1.31}$$

où $h = h(n)$ représente le paramètre du lissage satisfaisant les conditions $h \to 0$ et $nh \to \infty$ lorsque $n \to \infty$ et $K_{(\rho_h(x),h)}$ est la fonction de densité de la distribution gamma, de paramètres $(\rho_h(x), h)$, donnée par la formule suivante :

$$K_{(\rho_h(x),h)}(t) = \frac{t^{\rho_h(x)-1}e^{-t/h}}{h^{\rho_h(x)}\Gamma(\rho_h(x))}, \tag{1.32}$$

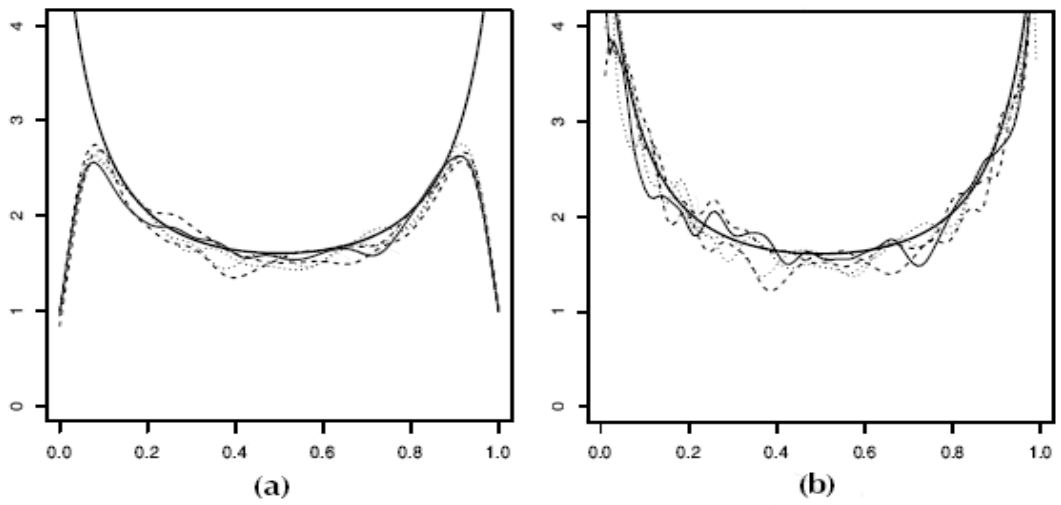

FIGURE 1.5 – Estimation d'une densité de probabilité à support compact $[0,1]$, pour $n = 10000$:(a) quand on utilise le noyau standard (gaussien), (b) quand on utilise le noyau beta.

avec

$$\Gamma(p) = \int_0^\infty x^{p-1}e^{-x}dx, \;\; p > 0.$$

La première version de l'estimateur à noyau gamma, notée $\hat{f}_{G_1}(x)$, est obtenue en remplaçant $\rho_h(x)$ par $x/h + 1$, dans la formule (1.31). Le biais et la variance asymptotiques corespondants aux estimateurs conçus à l'aide des noyaux gamma sont donnés respectivement par :

$$Biais(\hat{f}_{G_1}(x)) = h\left\{f'(x) + \frac{1}{2}xf''(x)\right\} + o(h), \tag{1.33}$$

$$Var\left(\hat{f}_{G_1}(x)\right) = n^{-1}\frac{h^{-1}\Gamma\left(2x/h+1\right)}{2^{2x/h+1}\Gamma^2(x/h+1)}f(x) + o(n^{-1}). \tag{1.34}$$

En raison de la contribution indésirable de f' dans le biais de l'estimateur $\hat{f}_{G_1}(x)$ (voir l'expression du biais donnée par la formule (1.33)) ; une autre version de $\hat{f}_{G_1}(x)$ appelée estimateur à noyau gamma modifié, notée $\hat{f}_{G_2}(x)$, est obtenue en remplaçant $\rho_h(x)$, dans la formule (1.31), par la nouvelle quantité suivante :

$$\begin{aligned}\rho_h(x) &= (x/h)\mathbf{1}_{\{x\geq 2h\}} + \left(\frac{1}{4}(x/h)^2 + 1\right)\mathbf{1}_{\{x\in[0,2h[\}} \\ &= \begin{cases} x/h, & \text{si } x \geq 2h, \\ \frac{1}{4}(x/h)^2 + 1, & \text{si } x \in [0,2h[. \end{cases}\end{aligned} \tag{1.35}$$

Le biais asymptotique de $\hat{f}_{G_2}(x)$ est exprimé par la formule suivante :

$$Biais\left(\hat{f}_{G_2}(x)\right) = \begin{cases} \frac{h}{2}xf''(x), & \text{si } x \geq 2h, \\ h\xi_h(x)f'(x), & \text{si } x \in [0,2h[, \end{cases} \tag{1.36}$$

où, $\xi_h(x) = (1-x)\left\{\rho_h(x) - x/h\right\} / \left\{1 + h\rho_h(x) - x\right\}$, tandis que sa variance est similaire à celle de $\hat{f}_{G_1}(x)$ [25].

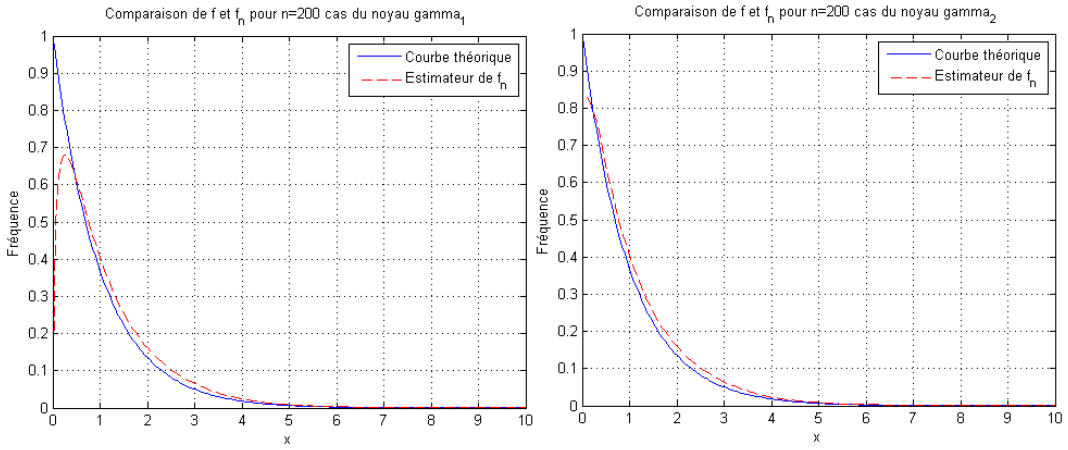

FIGURE 1.6 – Noyaux Gamma : (a) Noyau gamma et (b) Noyau gamma modifié.

L'avantage de ces deux noyaux est que la forme et la qualité du lissage des estimateurs qu'ils nous fournis changent selon la position où la densité est estimée, ce qui fait la différence avec les noyaux symétriques. Ils sont exempts du problème de biais aux bornes, non négatif et ils atteignent le taux de convergence optimal pour les variables $i.i.d$ au sens de MISE dans la classe des estimateurs à noyaux non négatifs. De plus, ils permettent une réduction de la variance lors du lissage en s'éloignant de la borne.

D'autres propriétés des estimateurs à noyaux gamma sont bien documentées dans la littérature. Bouezmarni et Scaillet [14] ont établi les conditions de convergence faible, de ces estimateurs, sur un support $[0, +\infty[$, lorsque f est continue sur ce support ainsi que la convergence faible au sens $MIAE$ (Mean Integer Absolute Error). Pour les densités non bornées à l'origine (au voisinage de zéro), les mêmes auteurs ont examiné les performances de cet estimateur par simulation et ont prouvé la convergence en probabilité. Fernandez et Monteira [36] ont établi le théorème central limite pour l'estimateur fonctionnel à noyaux gamma.

Le $MISE$, le h optimal (au sens du MISE) ainsi que le $MISE$ associe a ce dernier correspondants aux deux noyaux Gamma sont respectivement comme suit :

MISE :

$$\begin{aligned} MISE\left(\hat{f}_{G_1}\right) &= h^2 \int_0^\infty \{f'(x) + \frac{1}{2} x f''(x)\}^2 dx + \frac{n^{-1} h^{\frac{-1}{2}}}{2\sqrt{\pi}} \int_0^\infty x^{\frac{-1}{2}} f(x) dx + \circ(n^{-1} h^{\frac{-1}{2}} + h^2) \\ MISE\left(\hat{f}_{G_2}\right) &= \frac{1}{4} h^2 \int_0^\infty \{x f''(x)\}^2 dx + \frac{1}{2\sqrt{\pi}} n^{-1} h^{\frac{-1}{2}} \int_0^\infty x^{\frac{-1}{2}} f(x) dx + \circ(n^{-1} h^{\frac{-1}{2}} + h^2) \end{aligned}$$

Paramètre de lissage optimal :

$$\begin{aligned} h^*_{G_1} &= 4^{\frac{-2}{5}} \left[\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{\frac{-1}{2}} f(x) dx\right]^{\frac{2}{5}} \left[\int_0^\infty \{f'(x) + \frac{1}{2} x f''(x)\}^2 dx\right]^{\frac{-2}{5}} n^{\frac{-2}{5}}. \\ h^*_{G_2} &= \left[\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{\frac{-1}{2}} f(x) dx\right]^{\frac{2}{5}} \left[\int_0^\infty \{x f''(x)\}^2 dx\right]^{\frac{-2}{5}} n^{\frac{-2}{5}}. \end{aligned}$$

MISE Optimal :

$$MISE^*\left(f^*_{G_1}\right) = \frac{5}{4^{\frac{4}{5}}}\left[\frac{1}{2\sqrt{\pi}}\int_0^\infty x^{\frac{-1}{2}}f(x)dx\right]^{\frac{4}{5}} \quad \left[\int_0^\infty \{f'(x)+\frac{1}{2}xf''(x)\}^2dx\right]^{\frac{1}{5}} n^{\frac{-4}{5}}.$$

$$MISE^*\left(\hat{f}_{G_2}\right) = \frac{5}{4^{4/5}}\left[\frac{1}{2\sqrt{\pi}}\int_0^\infty x^{-1/2}f(x)dx\right]^{4/5}\left[\int_0^\infty \{xf''(x)\}^2\,dx\right]^{1/5} n^{-4/5}.$$

1.4.3 Noyaux inverse gaussien et réciproque inverse gaussien [84]

Soit $X_1, X_2, ..., X_n$ un échantillon aléatoire d'une distribution de densité de probabilité inconnue f, définie sur $[0,\infty[$. On suppose que $f \in C^2$ et $\int_0^\infty (x^3f''(x))^2dx < \infty$. Soit $K_{IG(m,\lambda)}$ la densité de probabilité Inverse Gaussien de paramètres m et λ, noté $IG(m,\lambda)$, de la variable aléatoire Y, définie par :

$$K_{IG(m,\lambda)}(y) = \frac{\sqrt{\lambda}}{\sqrt{2\pi y^3}}\exp\left(\frac{-\lambda}{2m}\left(\frac{y}{m}-2+\frac{m}{y}\right)\right), \quad y>0.$$

L'espérance et la variance de Y valent $E(Y)=m$ et $Var(Y)=\frac{m^3}{\lambda}$.
Posons $Z=\frac{1}{Y}$, alors Z suit une loi Réciproque Inverse Gaussien, notée $RIG(m,\lambda)$ de densité :

$$K_{RIG(m,\lambda)}(Z) = \frac{\sqrt{\lambda}}{\sqrt{2\pi Z}}\exp\left(\frac{-\lambda}{2m}\left(mZ-2+\frac{1}{mZ}\right)\right), \quad Z>0.$$

L'espérance et la variance de Z valent : $E(Z)=\frac{1}{m}+\frac{1}{\lambda}$ et $Var(Z)=\frac{1}{\lambda m}+\frac{2}{\lambda^2}$.
On considère les deux classes des noyaux suivantes :

$$K_{IG(x,\frac{1}{h})}(u) = \frac{1}{\sqrt{2\pi h u^3}}\exp\left(\frac{-1}{2hx}\left(\frac{u}{x}-2+\frac{x}{u}\right)\right), \tag{1.37}$$

et

$$K_{RIG\left(\frac{1}{x-h},\frac{1}{h}\right)}(u) = \frac{1}{\sqrt{2\pi h u}}\exp\left(-\frac{x-h}{2h}\left(\frac{u}{x-h}-2+\frac{x-h}{u}\right)\right), \tag{1.38}$$

où h doit satisfaire $h+\frac{1}{hn}\to 0$ $(h\to 0 \ \ et \ \ nh\to\infty)$ quand $n\to\infty$.
L'estimateur $\hat{f}_{IG}(x)$ est défini alors comme suit :

$$\hat{f}_{IG}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{IG(x,\frac{1}{h})}(X_i). \tag{1.39}$$

L'estimateur $\hat{f}_{RIG}(x)$ est de la forme suivante :

$$\hat{f}_{RIG}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{RIG\left(\frac{1}{x-h},\frac{1}{h}\right)}(X_i). \tag{1.40}$$

Il est facile de mettre en application ces deux estimateurs qui sont semblables à ceux des noyaux gamma. Il convient à remarquer que $K_{IG\left(x,\frac{1}{h}\right)}(u)$ tend vers 0, pour tout u lorsque x s'approche de la borne. Ceci induira la contrainte suivante $\hat{f}_{IG}(0)=0$, ce qui peut être indésirable dans certains cas.

Remarque 1.2. *Le noyau gamma modifié et RIG sont très semblables, excepté au point $x = 0$, tandis que la différence entre le noyau IG et le noyau gamma est plus remarquée (voir Scaillet [84]).*

Le biais, la variance, le paramètre de lissage optimale et le ISE optimale associes à ce dernier paramètre correspondant aux estimateurs conçus à l'aide des noyaux IG et RIG sont donnés, respectivement, comme suit :

Biais :

$$Biais\{\hat{f}_{IG}(x)\} = \frac{1}{2}x^3 f''(x)h + \circ(h). \tag{1.41}$$

$$Biais\{\hat{f}_{RIG}(x)\} = \frac{1}{2}x f''(x)h + \circ(h). \tag{1.42}$$

On constate que le biais est grand (respectivement petit) pour l'estimateur à noyau IG pour $x > 1$ (respectivement $x < 1$), par rapport au biais de l'estimateur RIG.

Variance : En prenant $x \in]0, \infty[$ et C strictement positif.

- Pour $\frac{x}{h} \to \infty$ (x à l'intérieur) les variances ont les formes suivantes :

$$Var[\hat{f}_{IG}(x)] = \frac{1}{2\sqrt{\pi}} n^{-1} h^{\frac{-1}{2}} x^{\frac{-3}{2}} f(x) + o(n^{-1}h^{\frac{-1}{2}}), \tag{1.43}$$

$$Var[\hat{f}_{RIG}(x)] = \frac{1}{2\sqrt{\pi}} n^{-1} h^{\frac{-1}{2}} x^{\frac{-1}{2}} f(x) + o(n^{-1}h^{\frac{-1}{2}}). \tag{1.44}$$

- Pour $\frac{x}{h} \to C$ (x au borne) les variances sont égales à :

$$Var[\hat{f}_{IG}(x)] = \frac{1}{2\sqrt{\pi}} n^{-1} h^{-2} C^{\frac{-3}{2}} f(x) + o(n^{-1}h^{-2}), \tag{1.45}$$

$$Var[\hat{f}_{RIG}(x)] = \frac{1}{2\sqrt{\pi}} n^{-1} h^{-1} \left(C^{\frac{-1}{2}} + \frac{7}{16} C^{\frac{-3}{2}} \right) f(x) + o(n^{-1}h^{-1}). \tag{1.46}$$

L'expression de la variance obtenue en utilisant le noyau RIG est égal à celle obtenue dans le cas des deux estimateurs à noyau gamma sous la condition $\frac{x}{h} \to \infty$. Au bornes de x, la variance de l'estimateur RIG a le même ordre que celle des estimateurs gamma.

Paramètres de lissage optimaux et leurs MISE associés :

$$h^*_{IG} = \frac{\left(\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{\frac{-3}{2}} f(x)dx \right)^{\frac{2}{5}}}{\left(\int_0^\infty \left(x^3 f''(x) \right)^2 dx \right)^{\frac{2}{5}}} n^{\frac{-2}{5}}. \tag{1.47}$$

$$h^*_{RIG} = \frac{\left(\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{\frac{-1}{2}} f(x)dx \right)^{\frac{2}{5}}}{\left(\int_0^\infty \left(x f''(x) \right)^2 dx \right)^{\frac{2}{5}}} n^{\frac{-2}{5}}. \tag{1.48}$$

$$MISE^*_{IG} = \frac{5}{4}\left(\frac{1}{2\sqrt{\pi}}\int_0^\infty x^{\frac{-3}{2}} f(x)dx\right)^{\frac{4}{5}}\left(\int_0^\infty (x^3 f''(x))^2 dx\right)^{\frac{1}{5}} n^{\frac{-4}{5}}. \tag{1.49}$$

$$MISE^*_{RIG} = \frac{5}{4}\left(\frac{1}{2\sqrt{\pi}}\int_0^\infty x^{\frac{-1}{2}} f(x)dx\right)^{\frac{4}{5}}\left(\int_0^\infty (xf''(x))^2 dx\right)^{\frac{1}{5}} n^{\frac{-4}{5}}. \tag{1.50}$$

1.4.4 Choix du paramètre de lissage

La méthode proposée ci-dessus qui consiste à choisir le paramètre de lissage h de sorte à minimiser le $MISE$, a un intérêt purement théorique. La difficulté majeure de cette méthode réside en fait dans les applications, car l'expression du h optimal dépend principalement de trois quantités inconnues $f,\ f',\ f''$. Ceci rend plus difficile le choix du paramètre de lissage. A cet effet, pour la sélection du paramètre de lissage, l'idée la plus naturelle est d'adopter les mêmes techniques exposées dans le cas des noyaux symétriques, à savoir : les méthodes plug-in et les méthode de validation croisée. Cependant, comme on la cité auparavant, l'$AMISE$ dépend principalement des quantités inconnues $f,\ f',\ f''$. Ceci rend la conception des méthodes plug-in pour le choix du paramètre de lissage plus difficile ; c'est l'une des raisons pour laquelle les méthodes les plus répandues dans le cas d'estimation à noyaux asymétriques est les méthodes de validation croisée.

Considérons un n-échantillon $X_1, X_2, ..., X_n$ $i.i.d.$ issue de la variable aléatoire X, et un noyau asymétrique $K_{x,h}$. La méthode la plus utilisée est celle qui minimise un estimateur convenable du $ISE\,(h)$. Ainsi, avec le même raisonnement et la même démarche que dans le cas de noyaux symétriques on peut montrer que le critère à minimiser dans le cas de noyaux asymétriques est bien que la fonctionnelle suivante :

$$UCV(h) = \int\left\{\frac{1}{n}\sum_{i=1}^{n} K_{x,h}\,(x_i)\right\}^2 dx - \frac{2}{n\,(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} K_{x_i,h}\,(x_j)\,. \tag{1.51}$$

On peut également sélectionner le paramètre de lissage en utilisant le maximum de vraisemblance par la validation croisée. Le h optimal, dans ce cas, est donné alors par :

$$h_{lcv} = \arg\max_{h>0}\left(\frac{1}{n}\sum_{i=1}^{n}\log\left(\hat{f}_i(x_i)\right)\right), \tag{1.52}$$

avec

$$\hat{f}_i(x_i) = \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} K_{x,h}(x_j). \tag{1.53}$$

1.5 Estimateur à noyaux discrets

Les définitions suivantes expliquent la notion du noyau discret et de l'estimateur à noyau discret pour la fonction de densité inconnue f définie sur le support discret $\aleph$.

Définition 1.2. *Soit $x \in \aleph$ et $h > 0$. On appelle **"noyau discret"** $K_{x,h}(.)$ toute fonction de masse de probabilité liée à une variable aléatoire discrète $\mathcal{K}_{x,h}$ de support $\aleph_x$, contenant au moins x et indépendant de h, vérifiant les quatres conditions suivantes :*

$$\bigcup_x \aleph_x \supseteq \aleph, \tag{1.54}$$

$$E(\mathcal{K}_{x,h}) \sim x \text{ lorsque } h \to 0, \tag{1.55}$$

$$Var(\mathcal{K}_{x,h}) < +\infty, \tag{1.56}$$

$$Var(\mathcal{K}_{x,h}) \to 0 \text{ lorsque } h \to 0, \tag{1.57}$$

Définition 1.3. *Soit $X_1,\ X_2, ...,\ X_n$ un n-échantillon $i.i.d.$ issu d'une variable aléatoire X de la fonction de masse de probabilité inconnue f sur $\aleph$. L'**estimateur à noyau discret** de f est défini par :*

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_{x,h}\left(X_i\right), \quad x \in \aleph \tag{1.58}$$

où $h > 0$ est le paramètre de lissage et $K_{x,h}$ est dit le noyau discret de cible x et de paramètre de lissage h définie sur le support $\aleph_{x,h} = \aleph_x$ (ne dépend pas de h).

1.5.1 Propriétés de l'estimateur à noyaux discrets

Dans cette section, nous allons introduire quelques propriétés de l'estimateur à noyau discret, qui ont été établis principalement par Senga Kiessé [49] et Kokonendji et Senga Kiessé [50], ainsi que les conditions qui assurent la convergence de cet estimateur en moyenne quadratique et en moyenne quadratique intégrée.

Proposition 1.2. *(Senga Kiessé[49]) Soit $X_1,\ X_2, ...,\ X_n$ un n-échantillon $i.i.d.$ issu d'une variable aléatoire X de la fonction de masse de probabilité inconnue f sur $\aleph$. Si $\hat{f}_h$ est l'estimateur à noyau discret de f, alors, pour $x \in \aleph$ et $h > 0$, on a :*

$$E\left(\hat{f}_h(x)\right) = E(\mathcal{K}_{x,h}),$$

où $\mathcal{K}_{x,h}$ est la variable aléatoire de loi $K_{x,h}$ sur $\aleph_x$. De plus, on a $\hat{f}_h(x) \in [0,1]$ pour $x \in \aleph$ et

$$\sum_x \hat{f}_h(x) = C,$$

où C est une constante strictement positive et finie.

Le résultat suivant garantit que l'estimateur à noyau discret est asymptotiquement sans biais en tout point x.

Proposition 1.3. *(Kokonendji et Senga Kiessé [50]) Soit $X_1,\ X_2, ...,\ X_n$ un n-échantillon $i.i.d.$ issu d'une variable aléatoire X de la fonction de masse de probabilité inconnue f sur $\aleph$. Si $\hat{f}_h$ est l'estimateur à noyau discret de f, alors, pour $x \in \aleph$ et $h > 0$, on a :*

$$E\left(\hat{f}_h(x)\right) = \sum_{y \in \aleph \cap \aleph_x} f(y)K_{x,h} \to f(x) \quad \textit{quand } h \to 0 \textit{ et } n \to +\infty.$$

•**L'erreur quadratique moyenne** (MSE) :

$$\begin{aligned} MSE(\hat{f}_h(x)) &= E\left[\hat{f}_h(x) - f(x)\right]^2 = Var(\hat{f}_h(x)) + Biais^2(\hat{f}_h(x)), \\ &= \frac{1}{n} f(x)\left[(\Pr(\mathcal{K}_{x,h} = x))^2 - f(x)\right] \\ &\quad + \left[f(E(\mathcal{K}_{x,h})) - f(x) + \frac{1}{2} Var(\mathcal{K}_{x,h}) f^{(2)}(x)\right]^2 + o\left(\frac{1}{nh} + h^2\right) \end{aligned} \tag{1.59}$$

avec $f^{(2)}(x)$ est la différence finie d'ordre 2 donnée par :

$$f^{(2)}(x) = \begin{cases} \{f(x+2) - 2f(x) + f(x-2)\}/4, & \text{si } x \in \mathbb{N}/\{0,1\}\,; \\ \{f(3) - 3f(1) + 2f(0)\}/4, & \text{si } x = 1\,; \\ \{f(2) - 2f(1) + f(0)\}/2, & \text{si } x = 0. \end{cases} \tag{1.60}$$

•**L'erreur quadratique moyenne intégrée** $(MISE)$ **:**

$$\begin{aligned} MISE(\hat{f}_h) &= \sum_{x\in\aleph} MSE(f(x), \hat{f}_h(x)) = \sum_{x\in\aleph} Var(\hat{f}_h(x)) + \sum_{x\in\aleph} Biais^2(\hat{f}_h(x)), \\ &= MISE(n, h, K, f). \end{aligned} \tag{1.61}$$

1.5.2 Noyaux discrets de premier ordre (standards)

Nous présentons dans cette section la première classe des noyaux discrets, dite, classe des **noyaux discrets standards** ou encore **noyaux discrets de premier ordre** proposés par Senga kiessé [49]. La particularité de ce type de noyaux est qu'ils ne vérifient pas la condition (1.57). Ici, nous présentons trois exemples de noyaux discrets standards et leurs caractristiques.

1. Le noyau Poissonnien

Pour un type de noyau Poissonnien $\mathcal{P}(\lambda)$, on considère le noyau discret $P_{x,h}$ de loi de Poisson, $\mathcal{P}(x+h)$, sur $\aleph_x = \mathbb{N}$ avec $x \in \mathbb{N}$ et $h > 0$, tel que :

$$P_{x,h}(y) = e^{-(x+h)} \frac{(x+h)^y}{y!}, \quad y \in \mathbb{N}$$

Notons que pour une cible $x \in \mathbb{N}$ et pour tout $h > 0$, le noyau $P_{x,h}$ est de support $\mathbb{N}$, dont la moyenne et la variance sont égaux et elles vaux $(x+h)$.
Soit $\hat{f}_h(x)$ l'estimateur d'une fonction de masse. Si l'estimateur $\hat{f}_h(x)$ est construit à l'aide d'un noyau **Poissonnien**, alors l'expression de cet estimateur, et celle de son biais ponctuelle et de sa variance ponctuelle sont données respectivement par :

$$\begin{aligned} \hat{f}_h(x) &= \frac{1}{n}\sum_{i=1}^{n}\left(e^{-(x+h)}\frac{(x+h)^{X_i}}{X_i!}\right), \\ Biais(\hat{f}_h(x)) &= f(x)\{P_{x,h}(x) - 1\} + \sum_{y\in\mathbb{N}/\{x\}} f(y) P_{x,h}(y), \\ Var(\hat{f}_h(x)) &= \frac{1}{n}\left(f(x)P_{x,h}^2(x) + \sum_{y\in\mathbb{N}/\{x\}} f(y)P_{x,h}^2(y) - \left\{f(x) + \sum_{y\in\mathbb{N}}[f(y) - f(x)]\, P_{x,h}(y)\right\}^2\right). \end{aligned}$$

2. Le noyau Binomial

Pour un type de noyau Binomial $\mathcal{B}(N,p)$, on considère le noyau discret $B_{x,h}$ de loi Binomiale, $\mathcal{B}(x+1,(x+h)/(x+1))$, définie par :

$$B_{x,h}(y)=\frac{(x+1)}{y!\,(x+1-y)!}\left(\frac{x+h}{x+1}\right)^{y}\left(\frac{1-h}{1+x}\right)^{x+1-y},\quad y\in\aleph_x\subseteq\mathbb{N},$$

tel que $\aleph_x=\{0,\ 1,...,\ x+1\}$, $x\in\mathbb{N}$, $h\in]0,1]$ et $\bigcup_x \aleph_x=\mathbb{N}$.
Ce noyau est à support $\aleph_x=\{0,1,....,x+1\}$, de moyenne $(x+h)$ et de variance $(x+h)(1-h)/(x+1)$.
Soit $\hat{f}_h(x)$ l'estimateur d'une fonction de masse. Si l'estimateur $\hat{f}_h(x)$ est construit à l'aide d'un noyau **Binomial**, alors l'expression de cet estimateur, et celle de son biais ponctuelle et de sa variance ponctuelle sont données respectivement par :

$$\begin{aligned}
\hat{f}_h(x) &= \frac{1}{n}\sum_{i=1}^{n}\frac{(x+1)}{X_i!\,(x+1-X_i)!}\left(\frac{x+h}{x+1}\right)^{X_i}\left(\frac{1-h}{1+x}\right)^{x+1-X_i},\\
Biais(\hat{f}_h(x)) &= f(x)\left\{B_{x,h}(x)-1\right\}+\sum_{y\in\aleph_x-\{x\}}f(y)B_{x,h}(y),\\
Var\left(\hat{f}_h(x)\right) &= \frac{1}{n}\left(f(x)B_{x,h}^2(x)+\sum_{y\in\aleph_x/\{x\}}f(y)B_{x,h}^2(y)-\left\{f(x)+\sum_{y\in\aleph_x}\left[f(y)-f(x)\right]B_{x,h}(y)\right\}^2\right).
\end{aligned}$$

3. Le noyau Binomial Négatif

Pour un type de noyau Binomial Négatif $\mathcal{BN}(\lambda,p)$, on considère le noyau discret $BN_{x,h}$ de loi Binomiale Négative, $\mathcal{BN}(x+1,(x+1)/(2x+1+h))$, sur $\aleph_x=\mathbb{N}$ avec $x\in\mathbb{N}$ et $h>0$, tel que :

$$BN_{x,h}(y)=\frac{(x+y)!}{y!x!}\left(\frac{x+h}{2x+1+h}\right)^{y}\left(\frac{x+1}{2x+1+h}\right)^{x+1},\quad y\in\mathbb{N}$$

Le noyau Binomial Négatif $BN_{x,h}$ est à support $\mathbb{N}$, de moyenne $(x+h)$ et de variance $(x+h)\left(1+(x+h)/(x+1)\right)$.
Soit $\hat{f}_h(x)$ l'estimateur d'une fonction de masse. Si l'estimateur $\hat{f}_h(x)$ est construit à l'aide d'un noyau **Binomial Négatif**, alors l'expression de cet estimateur, et celles de son biais ponctuelle et de sa variance ponctuelle sont données respectivement par :

$$\begin{aligned}
\hat{f}_h(x) &= \frac{1}{n}\sum_{i=1}^{n}\frac{(x+X_i)!}{X_i!x!}\left(\frac{x+h}{2x+1+h}\right)^{X_i}\left(\frac{x+1}{2x+1+h}\right)^{x+1},\\
Biais(\hat{f}_h(x)) &= f(x)\left\{BN_{x,h}(x)-1\right\}+\sum_{y\in\mathbb{N}/\{x\}}f(y)BN_{x,h}(y),\\
Var(\hat{f}_h(x)) &= \frac{1}{n}\left(f(x)BN_{x,h}^2(x)+\sum_{y\in\mathbb{N}/\{x\}}f(y)BN_{x,h}^2(y)-\left\{f(x)+\sum_{y\in\mathbb{N}}\left[f(y)-f(x)\right]BN_{x,h}(y)\right\}^2\right).
\end{aligned}$$

Remarque 1.3.

1. *Le fait qu'on dispose de l'expression du biais ponctuelle et celle de la variance ponctuelle des estimateurs conçus par les trois noyaux précédents, alors la forme de leurs MSE et $MISE$ peut être déduite facilement en utilisant respectivement* (1.59) *et* (1.61).
2. *Les estimateurs d'une densité discrète conçus à l'aide des trois noyaux précédents ne converge pas au sens du MSE et du $MISE$ sous la condition $h \to 0$ lorsque $n \to \infty$. En effet, pour un x fixé, la limite du biais de $\hat{f}_h(x)$ quand $h \to 0$ est donnée, respectivement pour le noyau Poissonnien, Binomial et Binomial Négatif, par :*

$$\lim_{h\to 0} Biais(\hat{f}_h(x)) = f(x)\left\{\frac{x^x e^{-x}}{x!} - 1\right\} + e^{-x}\sum_{y\in\mathbb{N}/\{x\}} f(y)\frac{x^y}{y!} \neq 0$$

$$\lim_{h\to 0} Biais(\hat{f}_h(x)) = f(x)\left\{\left(\frac{x}{x+1}\right)^x - 1\right\} + \frac{x!}{(x+1)^x}\sum_{y\in\mathbb{N}/\{x\}} f(y)\frac{x^y}{y!\,(x+1-y)} \neq 0$$

$$\lim_{h\to 0} Biais(\hat{f}_h(x)) = f(x)\left\{\frac{(2x)!x^x\,(x+1)^{x+1}}{(x!)^2\,(2x+1)^{2x+1}} - 1\right\} + \left(\frac{x+1}{2x+1}\right)^{x+1}\sum_{y\in\aleph_x/\{x\}} f(y)\frac{(x+y)!}{x!y!}\left(\frac{x}{2x+1}\right)^y \neq 0$$

Pour plus de détails sur ce constat, le lecteur peut se référer à [50].

1.5.3 Noyaux discrets de deuxième ordre

Nous présentons dans cette section la deuxième classe des noyaux discrets, dite, classe des noyaux discrets de **deuxième ordre**. La caractéristique fondamentale des noyaux de cette classe est qu'ils vérifient la totalité des conditions (1.54)–(1.57). Ci-dessous deux exemples de noyaux de deuxième ordre.

1. Le noyau Dirac

Soit le noyau Dirac noté $D_{x,0}$ lié à la variable aléatoire $\mathcal{D}_{x,h} = \mathcal{D}_{x,0}$ pour $x \in \mathbb{N}$ et $h = 0$ donné par :

$$D_{x,0}(y) = \mathbf{1}_{\{y=x\}} = \begin{cases} 1, & \text{si } y = x\,; \\ 0, & \text{sinon.} \end{cases},$$

où $\mathbf{1}$ est la fonction indicatrice, avec $\aleph_x = \{x\}$, $E(\mathcal{D}_{x,0}) = x$ et $Var(\mathcal{D}_{x,0}) = 0$.
Soit $\hat{f}_h(x)$ l'estimateur d'une fonction de masse. Si l'estimateur $\hat{f}_h(x)$ est construit à l'aide d'un noyau **Dirac**, alors l'expression de cet estimateur, et celle de son biais ponctuelle, de sa variance

ponctuelle et de son $MISE$ sont données respectivement par :

$$\begin{aligned}
\hat{f}_0(x) &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{X_i=x\}},\\
Biais(\hat{f}_0(x)) &= E(\hat{f}_0(x)) - f(x) = E(D_{x,0}) - f(x) = \sum_{y\in\aleph_x} f(y)\Pr(\mathcal{D}_{x,0}=y) - f(x) = 0,\\
Var(\hat{f}_0(x)) &= \frac{1}{n}f(x)(1-f(x)),\\
MISE(\hat{f}_0) &= \frac{1}{n}\sum_{x\in\mathbb{N}} f(x)(1-f(x)) = \frac{1}{n}\left(1-\sum_{x\in\mathbb{N}} f^2(x)\right).
\end{aligned}$$

D'après l'expression du $MISE$ de l'estimateur à noyau du Dirac, il est clair que l'estimateur en question converge en moyenne quadratique intégrée, c'est-à-dire $MISE(\hat{f}_0)\to 0$ quand $n\to\infty$ et cela le fait que $0\leq\sum\limits_{x\in\mathbb{N}} f^2(x)<1$ [50, 52].

2. Le noyau Triangulaire

Les noyaux discrets triangulaires ont été proposés par Kokonendji et al. [51] et Kokonendji and Zocchi [52]. Dans ce qui suit, avant de présenter l'estimateur à noyau Triangulaire, nous allons d'abord rappelé la définition d'une variable aléatoire triangulaire symétrique.

Définition 1.4. *Une variable aléatoire discrète $\mathcal{T}_{a,c}$ est dite triangulaire symétrique de centre $c\in\mathbb{N}$ et de bras $a\in\mathbb{N}$, si pour y dans son support $\aleph_c=\{c, c\pm1, ..., c\pm a\}$ sa probabilité individuelle s'écrit :*

$$\Pr(\mathcal{T}_{a,c}=y) = \frac{(a+1)-|y-c|}{(a+1)^2},$$

Définition 1.5. *Soit $h>0$ et $(a,c)\in\mathbb{N}^2$. Une variable aléatoire discrète $\mathcal{T}_{a,c,h}$ est dite triangulaire d'ordre h, de centre $c\in\mathbb{N}$ et de bras $a\in\mathbb{N}$, si pour y dans son support $\aleph_c=\{c, c\pm1, ..., c\pm a\}$ sa fonction de masse de probabilité s'écrit :*

$$\Pr(\mathcal{T}_{a,c,h}=y) = \frac{(a+1)^h-|y-c|^h}{P(a,h)},$$

où

$$P(a,h) = (2a+1)(a+1)^h - 2\sum_{k=0}^{a} k^h, \tag{1.62}$$

est la constante de normalisation.

Proposition 1.4. (*Kokonendji et al. [51]*) *Soit $\mathcal{T}_{a,c,h}$ la variable aléatoire triangulaire discrète d'ordre $h>0$, de centre $c\in\mathbb{N}$ et de bras $a\in\mathbb{N}$. Alors $\mathcal{T}_{a,c,h}$ est symétrique autour de sa moyenne $c=E\{\mathcal{T}_{a,c,h}\}$ et sa variance $Var\{\mathcal{T}_{a,c,h}\}=Var(a,h)$ ne depend pas de centre c avec :*

$$Var(\mathcal{T}_{a,c,h}) = \frac{1}{3P(a,h)}\left[a(a+1)^{h+1}(2a+1)\right] - 2\sum_{k=0}^{a} k^{h+2},$$

avec $P(a,h)$ est définie dans (1.62).

Définition 1.6. *Soit f une fonction de masse de probabilité sur $\aleph$. Soit $h > 0$ le paramètre de lissage et $a \in \mathbb{N}$ un entier fixé. Le noyau discret triangulaire $T_{a,x,h}$ associé à la variable aléatoire $\mathcal{T}_{a,x,h}$ d'ordre h, de centre x et de bras a définie sur $\aleph_x = \{x, x \pm 1, ...x \pm a\}$ est donné par :*

$$T_{a,x,h}(y) = \Pr(\mathcal{T}_{a,x,h} = y) = \frac{(a+1)^h - |y-x|^h}{P(a,h)}, \ \ \forall y \in \aleph_x,$$

avec $P(a,h)$ est définie dans (1.62).

Soit $\hat{f}_h(x)$ l'estimateur à noyau discret en utilisant le noyau triangulaire discret, alors l'erreur quadratique moyenne intégrée de cet estimateur est donnée comme suit :

$$MISE\left(\hat{f}_h\right) = \frac{1}{n}\sum_{x\in\aleph} f(x)\left[\left\{\frac{(a+1)^h}{P(a,h)}\right\}^2 - f(x)\right] + \frac{1}{4}\left\{Var(a,h)\right\}^2 \sum_{x\in\aleph}\left\{f^{(2)}(x)\right\}^2 + o\left(\frac{1}{n} + h^2\right).$$

avec $P(a,h)$ est définie dans (1.62).
De plus, $MISE\left(\hat{f}_h\right) \to 0$ lorsque $n \to \infty$ et $h > 0$, et cela le fait que $\lim_{h\to 0}(a+1)^h/P(a,h) = 1$, $0 \leq \sum_{x\in\aleph} f(x)\{1-f(x)\} < 1$, $\lim_{h\to 0} Var(a,h) = 0$ et $\sum_{x\in\aleph}\left\{f^{(2)}(x)\right\}^2$ est finie. Ceci se traduit par la convergence de l'estimateur à noyau triangulaire en moyenne quadratique intégrée.

1.5.4 Choix du paramètre de lissage

Le rôle du paramètre de lissage, $h > 0$, dans le cas discret, reste semblable au cas continu, où il permet de tenir compte des observations x_i qui sont proches de la cible $x \in \aleph$. Cependant, la dispersion locale en tout point d'estimation x se traduit par l'importance du noyau discret $K_{x,h}$ choisi. Ainsi, le choix d'un type de noyau discret s'oriente vers les distributions de $K_{x,h}$ qui soient moins dispersées autour de $x \in \aleph$ pour h fixe.
Dans cette section nous présentons quelques méthodes classiques pour le choix global du paramètre de lissage dans l'estimation des fonctions de densité discrètes.

1. **Minimisation du MISE**
   Soit $X = (X_1, \ldots, X_n)$ un n échantillons $i.i.d.$ de densité inconnue f, alors l'erreur quadratique intégrée (ISE) est donné par :

   $$ISE = \sum_{x\in\aleph}(\hat{f}(x) - f(x))^2 = ISE(n, X, h, K, f), \tag{1.63}$$

   ainsi (1.63) conduit à choisir une fenêtre adéquate :

   $$h^* = \arg\min_{h>0} ISE(n, X, h, K, f), \tag{1.64}$$

   pour laquelle la mesure est sur un seul échantillon et la fenêtre optimale h^*_{opt} peut être obtenue, dans le cas de plusieurs échantillons, à travers

   $$h^*_{opt} = \arg\min_{h>0} E(ISE(n, X, h, K, f)). \tag{1.65}$$

   Ces techniques ont été développées et détaillées par Kokonendji et Senga Kiessé en 2011 [50].

2. **Validation croisée**
   Nous exposons ici deux techniques qui se basent sur la méthode de validation croisée.

   (a) Validation croisée par les moindres carrées (UCV) :
   Le principe de la méthode dans le cas de variables discrètes est le même que dans le cas de variables continues, élaboré dans [81, 15]. La méthode consiste à estimer le ISE par la technique validation croisée et par la suite de sélectionner le paramètre de lissage qui minimise cet estimateur. La fenêtre optimale, dans ce cas, s'obtient par :

$$\hat{h}_{ucv} = \arg\min_{h>0} CV\,(h) = \arg\min_{h>0} \left[\sum_{x\in\mathbb{N}} \left\{ \hat{f}(x) \right\}^2 - \frac{2}{n} \sum_{i=1}^{n} \hat{f}_i(X_i) \right] \tag{1.66}$$

$$= \arg\min_{h>0} \left[\sum_{x\in\mathbb{N}} \left\{ \frac{1}{n} \sum_{i=1}^{n} K_{x,h}(X_i) \right\}^2 - \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i\neq j}}^{n} K_{X_i,h}(X_j) \right].$$

   (b) Validation croisée par le maximum de vraisemblance :
   Ce critère consiste à choisir h qui maximise la fonctionnelle

$$LCV(h) = \frac{1}{n} \sum_{i=1}^{n} \log\left(\hat{f}_i(X_i) \right),$$

   c'est-à-dire, on détermine une fenêtre optimale h_{LCV} par :

$$\hat{h}_{LCV} = \arg\max_{h>0} LCV(h). \tag{1.67}$$

3. **Excès des zéros**
   Cette technique repose sur une particularité des données de comptage qui n'est autre que l'excès des zéros dans l'échantillon, c'est-à-dire de choisir une fenêtre adaptés $h_0 = h_0(X, K)$ tel que h satisfait :

$$\sum_{i=1}^{n} \Pr(\mathcal{K}_{X_i,h_0} = 0) = n_0, \tag{1.68}$$

   où $n_0 = card\{X_i = 0\}$ désigne le nombre des zéros dans l'échantillon $X_1, ..., X_n$ à condition que $n_0 > 0$. Ci-dessous quelques exemples de h_0 selon le noyau :
   – Si le noyau utilisé est Poissonnien alors

$$h_0 = \log\left(\frac{1}{n_0} \sum_{i=1}^{n} e^{-X_i} \right).$$

   – Si le noyau utilisé est Binomial alors le h_0 est la solution de l'équation

$$n_0 = \sum_{i=1}^{n} \left(\frac{1-h}{X_i+1} \right)^{X_i+1}.$$

– Si le noyau utilisé est Binomial Négative alors le h_0 est la solution de l'équation

$$n_0 = \sum_{i=1}^{n} \left(\frac{X_i + 1}{2X_i + 1 + h} \right)^{X_i+1}.$$

Remarque 1.4. *Le paramètre de lissage sélectionné par la méthode "Excès des zéros" n'existe pas toujours. En effet, pour certains noyaux l'équation* (1.68) *n'admet pas de solution. A titre d'exemple, on peut cité le cas du noyau triangulaire (pour plus de détails voir Kokonendji et al. [51] et Senga Kiessé [49]).*

Conclusion

Dans ce chapitre, nous nous sommes intéressés à la méthode d'estimation à noyau de la densité de probabilité et au problème du choix de ses paramètres pour sa mise en œuvre, à savoir : le choix du noyau K et le choix du paramètre de lissage h.

Nous avons constaté que le choix du noyau K dans le cas de densités réelles à supports non bornés est très peu influent et les critères du choix sont alors la simplicité et la vitesse de calcul. Les noyaux employés ici sont symétriques (dit aussi classiques). Cependant, lorsqu'on veut estimer des densités à support borné au moins d'un côté, l'estimateur à noyau classique devient non consistant, à cause des effets du biais au borne. Ce problème est dû à l'utilisation d'un noyau symétrique qui assigne un poids en dehors du support lorsque le lissage est pris en compte près du bord. La solution la plus simple proposée dans la littérature est de remplacer le noyau symétrique par un noyau adapté au support de la fonction inconnue à estimer.

En revanche, le choix du paramètre de lissage est un facteur important et crucial, ceci est dû au fait que de petites perturbations de ce paramètre peuvent changer totalement la qualité des estimations. Deux catégories de méthodes classiques ont été proposées dans la littérature pour choisir ce paramètre. La première catégorie repose sur la minimisation de l'erreur quadratique moyenne intégrée ($MISE$). Cette classe de méthodes est intéressante en théorie, mais sa difficulté majeure réside dans la pratique, car le paramètre de lissage optimal dépend d'une ou de plusieurs quantités inconnues. La deuxième catégorie est de type validation croisée, elle est intéressante en pratique car elle se laisse guider seulement par les observations. Cependant, cette classe de méthodes présente deux problèmes majeurs. D'une part, elles manquent de robustesse par rapport aux changements de la taille de l'échantillon. D'autre part, la fonctionnelle à optimiser a souvent tendance à présenter plusieurs optimums locaux.

Chapitre 2

Estimation dans les processus stochastiques Markoviens

Introduction

Très souvent, lorsque nous étudions un phénomène qui dépend du hasard, il y a lieu de prendre en compte l'évolution de ce phénomène au cours du temps. De plus, chaque observation d'un phénomène réel est modélisé par une variable aléatoire réelle ; l'étude de phénomènes évoluant dans le temps va donc être modélisée par une famille de variables aléatoires, appelée processus stochastique.

Les premiers processus étudiés sont, bien-sûr, les suites de variables aléatoires indépendantes, ce qui a conduit à la loi des grands nombres et au théorème central limite. Le mathématicien russe Andreï Andreïevitch Markov poussé vers la théorie des probabilités par son maître, Pafnoutï Lvovitch Tchebychev, qui démontra, sous des conditions assez générales, le théorème central limite, chercha à généraliser ce théorème à des suites de variables aléatoires dépendantes. Il est amené ainsi à considérer des variables faiblement dépendantes, c'est-à-dire que l'évolution future ne dépend que de l'état présent, d'où le fondement de la théorie des processus auxquels fût donné son nom (Processus Markoviens). Depuis, le formalisme des chaînes de Markov est devenu le principal instrument d'analyse, que ce soit d'une manière directe ou d'une manière indirecte, des systèmes dynamiques utilisés aussi bien par l'industrie que par la recherche scientifique, des mathématiques aux sciences sociales,...

Dans ce chapitre, nous nous positionnons dans le cadre d'estimation dans les chaînes de Markov. En effet, premièrement nous introduisons un bref rappel sur la notion de chaînes de Markov au sens probabiliste, sa présentation, et quelques définitions liées à la théorie des chaînes de Markov. Par la suite, nous introduisons les méthodes paramétriques qui visent à estimer les probabilités de transition et les caractéristiques d'une chaîne de Markov. Enfin, nous allons exposer la méthode non paramétrique du noyau adapté à l'estimation des probabilités de transition d'une chaîne de Markov.

Nous nous limitons ici, uniquement à deux classes de processus Markoviens à espace d'état discret : la première est celle des processus à temps discret tandis que la deuxième est celle des

processus à temps continu. Les modèles de ces deux classes fourniront des outils simples de modélisation et d'analyse d'une classe particulière de systèmes à événements discrets.

2.1 Quelques notions sur les processus de Markov

Dans cette section, nous se limitons uniquement à la présentation des notions utiles pour la compréhension de la suite du présent Chapitre.

2.1.1 Définition d'un processus de Markov

On considère un processus stochastique $\{X_n\}_{n\in\mathbb{N}}$ à espace d'état discret, à temps discret et S l'espace d'état, où il peut être de dimension finie ou infinie ; Le plus souvent S sera un sous-ensemble des entiers non négatifs, c'est-à-dire $S = \{0, 1, 2, ..., N\}$.

Nous considérons pour cela une suite de variables aléatoires $\{X_n; n = 0, 1, 2, ...\}$ dans l'espace d'état S, et nous supposons que l'on connaît, pour chaque paire d'états (i, j) et pour chaque instant n la probabilité $p_{ij}(n)$ qui fait référence à la probabilité que le processus soit dans l'état j à l'instant $n + 1$ étant donné qu'il se trouve dans l'état i à l'instant n :

$$p_{ij}(n) = P\left[X_{n+1} = j/X_n = i\right].$$

C'est la probabilité conditionnelle pour que le processus fasse une transition de l'état i vers l'état j au cours de l'intervalle temporel $[n, n + 1]$; on l'appellera par la suite la ***probabilité de transition à une étape***.

Définition 2.1. *Le processus stochastique $\{X_n\}_{n\in\mathbb{N}}$ est une chaîne de Markov à temps discret ssi :*

$$p_{ij}(n) = P[X_{n+1} = j/X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0] = P[X_{n+1} = j/X_n = i], \qquad (2.1)$$

pour toute $n \geq 0$ et pour des états quelconque $(j, i, i_{n-1}, ..., i_0 \in S)$.

Cette définition peut être traduite comme suit : "La probabilité pour que la chaîne soit dans un certain état à la $(n + 1)^{ieme}$ étape du processus dépend uniquement de l'état du processus à l'étape précédente (la n^{ieme} étape)".

Une suite de variable aléatoires $\{X_n; n = 0, 1, 2, ...\}$ qui satisfait la condition (2.1) est appelée *chaîne de Markov à temps discret*. On dira aussi qu'un tel processus stochastique est sans mémoire.

$$p_{ij}(n) = P\left(X_{n+1} = j/X_n = i\right), n \in \mathbb{N}. \qquad (2.2)$$

On se bornera par la suite à étudier des chaînes de Markov homogènes dans le temps, c'est-à-dire pour lesquelles les probabilités de transition sont indépendantes du temps

$$p_{ij}(n) = p_{ij}, \ \ \forall\, n \in \mathbb{N}, \qquad (2.3)$$

tel que $\sum_{j\in S} p_{ij} = 1$ (les p_{ij} forme une distribution de probabilité sur i) et $p_{ii} \geq 0$ (il est possible de rester dans un certain état i entre deux étapes consécutives).

2.1.2 Matrice des probabilités de transition P

On peut présenter les probabilités conditionnelles p_{ij}, définie dans (2.3), sous forme matricielle ou sous forme de graphe. La *matrice des probabilités de transition* ou simplement *matrice de transition* $P = [p_{ij}]_{i,j\in S}$ possède clairement les propriétés suivantes :

- tous les termes sont positifs ou nuls $(\forall i, j \in S,\ p_{ij} \geq 0)$,
- la somme des termes de chaque ligne égale 1 $\left(\sum_{j\in S} p_{ij} = 1,\ \forall i \in S\right)$.

La matrice $P = [p_{ij}]_{i,j\in S}$ est une matrice carrée d'ordre fini ou infini, selon le cardinal de l'espace d'état S (fini ou infini) :

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1j} & \cdots \\ p_{21} & p_{22} & \cdots & p_{2j} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \\ p_{i1} & p_{i2} & \cdots & p_{ij} & \cdots \\ \vdots & \vdots & & \vdots & \ddots \end{pmatrix}.$$

Remarque 2.1. *Généralement lorsque la matrice P est d'ordre* infini *on l'appelle* **opérateur de transition** *plutôt que* **matrice de transition**.

2.1.3 Distribution des probabilités d'état

Nous introduisons les probabilités d'état :

$$\pi_k(n) = P(X_n = k), \quad \text{pour } n \in \mathbb{N}, \quad \text{et } k \in S. \tag{2.4}$$

La distribution de X_n peut alors être écrite sous forme de vecteur-ligne $\pi(n) = (\pi_0(n), \pi_1(n), \pi_2(n), ...)$ dont la somme des termes vaut 1. Pour calculer $\pi(n)$, il faut connaître soit la valeur prise par X_0, c'est-à-dire l'état initial du processus, soit sa distribution initiale $\pi(0)$. D'après le théorème des probabilités totales, on a alors :

$$\pi_k(n) = \sum_{i\in S} \pi_i(0) p_{ik}^n, \quad \forall n \in \mathbb{N}. \tag{2.5}$$

En notation matricielle, cette relation s'écrit

$$\pi(n) = \pi(0) P^n, \quad \forall n \in \mathbb{N}. \tag{2.6}$$

De façon tout à fait analogue(par récurrence), on obtient :

$$\pi(n+1) = \pi(n) P, \quad \forall n \in \mathbb{N}. \tag{2.7}$$

Les résultats établis ci-dessus affirment qu'une chaîne de Markov est complètement définie si l'on connaît sa matrice des probabilités de transition à une étape, ainsi que la distribution de X_0. C'est la simplicité de la structure mathématique de ce type de processus stochastiques qui explique son grand intérêt à la fois théorique et pratique.

2.2 Estimation paramétrique de la matrice de transition P

Les processus et les modèles Markoviens constituent un sujet central dans la théorie de probabilités appliquées et de la statistique. En effet, ils peuvent donner parfois une bonne description pour des phénomènes dont l'évolution dans le temps peut être décrite par des variables aléatoires faiblement dépendantes. De plus, plusieurs problèmes réels peuvent être modélisés par ce type de processus stochastiques à temps discret ou continu. Cependant, dans la pratique, la difficulté fondamentale pour utiliser une chaîne de Markov comme modèle concerne l'estimation de sa matrice de transition P. L'estimation des caractéristiques pour les chaînes Markoviennes finies est importante pour les applications. De plus, les différentes représentations pour les caractéristiques permettent d'obtenir des expressions facilement programmables et calculables. Les premiers travaux sur l'estimation statistique pour des modèles Markoviens à un espace d'état fini, particulièrement par la méthode du maximum de vraisemblance, ont été réalisés par Anderson et Goodman [8] en 1957 et Billingsley [11] en 1961. Ce dernier a étudié l'estimateur du maximum de vraisemblance et ses propriétés asymptotiques dans le cadre d'une observation de longue durée pour un modèle Markovien ergodique. Par contre, Anderson et Goodman dans [8] ont traité principalement l'estimation dans le cas d'une suite de copies indépendantes d'une chaîne Markovienne censurée à un instant fixe. Cependant, l'estimateur du maximum de vraisemblance pour des caractéristiques qui peuvent être exprimées en fonction des paramètres du modèle n'a été étudié que plus récemment. Nous citons dans ce contexte, le travail de Sadek et Limnios [83] (2002) dédié à l'estimation de la fiabilité, la disponibilité et le taux de défaillance dans des systèmes qui peuvent être modélisés par des chaînes de Markov finies et ergodique. D'autres travaux dans le contexte d'estimation dans les chaînes de Markov sont focalisés sur d'autres méthodes d'estimation et au choix de l'estimateur, entre autre on cite : Lee et al. (1970) [59] étudient en détail les différentes méthodes d'estimation, outre celles basées sur les moindres carrés, ils envisagent la méthode du χ^2 minimum, celle du maximum de vraisemblance et celle basée sur la minimisation de la norme L_1, c'est-à-dire la minimisation de la somme des valeurs absolues des écarts. Par ailleurs, les auteurs envisagent le problème sous l'angle bayésien. Bonnieux [13] s'est intéressé à l'estimation des caractéristiques qui découlent d'une chaîne de Markov finie, aux différentes représentations possible pour ces caractéristiques et aux calculs de l'erreur leurs associées.

2.2.1 Définition du modèle statistique

On considère une suite d'événements indicés par t $(t = 0, 1, ..., T)$ ayant chacun r issues possibles S_i $(i = 1, 2, ..., r)$. Ce processus peut être représenté par une variable aléatoire X_t $(t = 0, 1, ..., T)$ prenant ses valeurs dans $\{1, ..., r\}$ et telle que :

$$X_t = i, \tag{2.8}$$

si S_i est l'issue du $t^{\text{ième}}$ événement.

Les événements sont liés de telle sorte que la probabilité pour que X_t soit égale à i est entièrement définie par sa valeur à l'instant précédent, ainsi :

$$P(X_t/X_{t-1}, X_{t-2}, ..., X_0) = P(X_t/X_{t-1}). \tag{2.9}$$

D'où, d'après le principe de la fonction de vraisemblance, la probabilité d'avoir la suite $(X_0, X_1, ..., X_t)$ est donnée par :

$$P(X_0, X_1, ..., X_t) = P(X_0) \prod_{t=1}^{T} P(X_t/X_{t-1}) \tag{2.10}$$

Le processus est donc défini par la distribution initiale $P(X_0)$ et les probabilités conditionnelles $P(X_t/X_{t-1})$. Posons :

$$p_{ij} = p_{ij}(t) = P(X_t = j/X_{t-1} = i). \tag{2.11}$$

Les p_{ij} sont appelées probabilités de transition associées aux passages de S_i à S_j. Nous les supposons indépendantes de t, le processus est donc supposé homogène. On définit la matrice carrée d'ordre r,

$$P = [p_{ij}], \tag{2.12}$$

elle est appelée matrice de transition et :

$$\begin{cases} 0 \leq p_{ij} \leq 1, \\ \sum\limits_{j} p_{ij} = 1. \end{cases} \tag{2.13}$$

Le processus que nous venons de définir est alors une chaîne de Markov homogène à r états.

Connaissant l'état initial X_0 et la matrice de transition P, on peut déterminer la distribution de chaque variable aléatoire X_t, ou encore rechercher s'il existe une loi de X_t lorsque t tend vers l'infini. Dans ce contexte, on dit que la chaîne est ergodique s'il existe un vecteur :

$$\pi = (\pi_1, ..., \pi_r), \tag{2.14}$$

tel que

$$\lim_{t \to \infty} p_{ij}^{(t)} = \pi_j, \tag{2.15}$$

où $p_{ij}^{(t)}$ est l'élément d'ordre (i, j) de p^n, ayant les propriétés suivantes :

1. $0 \leq \pi_j \leq 1$,
2. $\sum\limits_{j} \pi_j = 1$,
3. $\pi = \pi P$.

2.2.2 Méthode du maximum de vraisemblance

Considérons n individus dont l'évolution est régie par une chaîne de Markov ergodique de matrice P inconnue. On désigne par $n_i(t)$ le nombre d'individus dans l'état i à l'instant t, et par $n_{ij}(t)$ le nombre d'individus qui passent de l'état i à l'état j dans la transition qui s'effectue de l'instant $t-1$ à l'instant t. L'état de l'échantillon de n individus peut être repérer à chaque instant par le vecteur d'état :

$$(n_1(t), n_2(t), ..., n_r(t)),$$

et les transitions par les matrices des comptes de transition d'élément général $[n_{ij}(t)]$. On appellera microdonnées les $n_{ij}(t)$, par opposition aux macrodonnées qui désigneront les $n_i(t)$. Si l'on connaît les microdonnées il est possible d'estimer P en utilisant la méthode du maximum

de vraisemblance envisagée en 1957 par Anderson et al. [8].
Posons :

$$n_{ij} = \sum_t n_{ij}(t), \tag{2.16}$$

les n_{ij} forment une statistique exhaustive minimale [13]. La fonction de vraisemblance associée à un échantillon de taille n observé pendant T périodes s'écrit :

$$\left[\prod_{t,i} \frac{n_i(t-1)!}{\prod_j n_{i,j}(t)!}\right] \times \left[\prod_{i,j} p_{i,j}^{n_{i,j}}\right]. \tag{2.17}$$

Pour dériver les estimateurs des p_{ij}, on suppose fixés $n_i(0)$ et on maximise le logarithme de la fonction de vraisemblance sous les contraintes :

$$\sum_j p_{ij} - 1 = 0, \quad \text{pour } i = 1, 2, ..., r. \tag{2.18}$$

Ainsi, on aura l'unique solution donnée par :

$$\widehat{p}_{ij} = \frac{n_{ij}}{\sum_j n_{ij}} \geq 0. \tag{2.19}$$

Ces estimateurs sont convergents, sans biais de plus ils suivent asymptotiquement la loi multi-normale (lorsque $n \to \infty$ ou lorsque $T \to \infty$). La généralisation de ce dernier résultat a été envisagée pour le cas de chaînes de Markov non homogène par Bonnieux [12].

Le plus souvent en pratique les microdonnées (c'est-à-dire les trajectoires individuelles) sont inconnues, l'information porte uniquement sur les macrodonnées (c'est-à-dire les vecteurs d'état). Dans telles situations, le problème de l'estimation de P est examiné à l'aide de la méthode des moindres carrés. Dans les deux sections suivantes nous allons présenter brièvement les méthodes des moindres carrés (ordinaire et généralisé) appliquées à l'estimation de la matrice P.

2.2.3 Méthode des moindres carrés

L'estimation habituelle des moindres carrés de la matrice de transition d'un processus de Markov fini est donnée pour le cas où, pour chaque instant, seules les proportions (fréquences) de l'échantillon dans chaque état sont connues. Pour cela, définissons les probabilités absolues :

$$q_j(t) = P(X_t = j), \tag{2.20}$$

on a :

$$q_j(t) = \sum_i q_i(t-1) p_{ij}. \tag{2.21}$$

Nous observons les vecteurs d'état d'éléments $n_j(t)$ et divisons les par l'effectif total n afin de définir le vecteur des fréquences observées d'éléments $y_j(t)$. On a alors la relation

$$y_j(t) = \sum_i y_i(t-1)p_{ij} + u_j(t), \tag{2.22}$$

où les y_i sont observés, p_{ij} désignent les paramètres à estimer et $u_j(t)$ une erreur aléatoire.
Le problème peut se présenter sous forme matricielle suivante :

$$y_j = Xp_j + u_j, \tag{2.23}$$

avec y_j et u_j sont des vecteurs colonnes d'ordre T, p_j est un vecteur colonne d'ordre r et X est une matrice d'ordre $T \times r$, tel que :

$$X = \begin{bmatrix} y_1(0) & y_2(0) & \cdots & y_r(0) \\ \vdots & \vdots & & \vdots \\ y_1(t-1) & y_2(t-1) & \cdots & y_r(t-1) \\ \vdots & \vdots & & \vdots \\ y_1(T-1) & y_2(T-1) & \cdots & y_r(T-1) \end{bmatrix},$$

En regroupant les r relations on obtient :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_r \end{bmatrix} = \begin{bmatrix} X & 0 & \cdots & 0 \\ 0 & X & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_r \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_r \end{bmatrix},$$

ou encore :

$$y = \tilde{X}p + u, \tag{2.24}$$

supposons que le rang de X égale à r et

$$\begin{aligned} E(u) &= 0, \ \text{(Un bruit blanc)} \\ E(uu') &= \Sigma, \end{aligned}$$

où Σ est une matrice carrée d'ordre Tr et singulière (Lee et al. [59], page 34). On constate que le modèle (2.24) n'est rien d'autre que le problème de **régression multilinéaire** dont la résolution peut se faire par la méthode des moindres carrés. En effet, Miller [65] a suggéré d'estimer les paramètres de ce modèle par la méthode des moindres carrés ordinaire dont son application nous fournira la solution suivante :

$$\hat{p} = \left(\tilde{X}' \, \tilde{X}\right)^{-1} \tilde{X}' \, y. \tag{2.25}$$

On peut vérifier, dans le travail de Lee et al. [58], que les estimateurs (2.25) satisfaites la relation :

$$\sum_j \hat{p}_{ij} = 1, \tag{2.26}$$

par contre la contrainte de la non-négativité des éléments n'est pas nécessairement satisfaite.

En 1959, Madansky [62] a montré que les u_j ne sont pas corrélés à X, il en résulte que $\hat{p}$ converge vers p. Cependant, les erreurs sont hétéroscédastiques donc l'estimateur n'est pas efficace. Pour surpasser ce problème, une solution avait été proposée en 1966 par Theil et Rey [95] et en 1970 par Lee et al. [58]. La solution consiste à minimiser la fonctionnelle :

$$uu' = (y - \tilde{X}p)'(y - \tilde{X}p) \tag{2.27}$$

sous les contraintes

$$p \geq 0, \tag{2.28}$$
$$Gp = \eta_r, \tag{2.29}$$

avec G est la matrice $r \times r^2$ définie par

$$G = (i_r, \ ..., \ i_r)\,, \tag{2.30}$$

où i_r est la matrice identité d'ordre r, et η_r le vecteur à r éléments tous égaux à un.
Il est à remarquer que la contrainte (2.28) assure la non-négativité des éléments de la matrice $\hat{p}$ tandis que la contrainte (2.29), elle assure que la matrice $\hat{p}$ est une matrice stochastique. Sous cette dernière forme, le problème se ramène à la résolution d'un programme quadratique.

2.2.4 Méthode des moindres carrés généralisés

Une autre technique proposée dans la littérature pour l'estimation d'une matrice de transition associée à une chaîne de Markov est bien la méthode des moindres carrés généralisés. Reprenant le modèle (2.24), donné par :

$$y = \tilde{X}p + u, \tag{2.31}$$

où

$$E(u) = 0, \tag{2.32}$$
$$E(uu') = \Sigma, \tag{2.33}$$

avec Σ est une matrice non-diagonale, singulière. Appliquons au modèle une matrice de pondération $H(rT \ \times \ rT)$, nous obtenons :

$$Hy = H\tilde{X}p + Hu, \tag{2.34}$$

d'où :

$$\hat{p} = (X'H'HX)^{-1}\,(X'H'Hy)\,. \tag{2.35}$$

Le problème posé, à ce niveau, est donc celui du choix de H. Comme Σ est singulière, alors il n'existe pas de matrice H qui transforme les erreurs hétéroscédasticité en erreurs homocédastiques. Cependant, on peut définir des matrices de pondération qui conduisent à des estimateurs plus efficaces que celui des moindres carrés ordinaire. On trouvera dans Theil et Rey [95], Madansky [62] et Lee et al. [58] différentes pondérations.

Nous allons illustrer que la singularité de Σ résulte d'une redondance parmi les paramètres. Après l'avoir éliminé on pourra expliquer la méthode des moindres carrés généralisés. Auparavant précisons les hypothèses et supposons que u est indépendant de $\tilde{X}$ ce qui entraîne que y et u ont la même matrice des variances-covariances. On peut par ailleurs préciser la distribution de y. Supposons que les fréquences observées y sont engendrées par une distribution multinomiale d'espérances $q_j(t)$, de variances $\frac{q_j(t)[1-q_j(t)]}{N(t)}$ et de covariances $\frac{-q_i(t)q_j(t)}{N(t)}$.

Soit Σ la matrice variances-covariances donnée par :

$$\Sigma \quad = \quad (\Sigma_{ij}) \;\; pour \;\; i,j = 1, \; ..., \; r, \tag{2.36}$$

tel que pour $i \neq j$, on a :

$$\Sigma_{ij} = \begin{bmatrix} \frac{-q_i(1)q_j(1)}{N(1)} & & & \\ & \frac{-q_i(2)q_j(2)}{N(2)} & & \\ & & \ddots & \\ & & & \frac{-q_i(T)q_j(T)}{N(T)} \end{bmatrix},$$

et pour $i = j$, on a :

$$\Sigma_{ii} = \begin{bmatrix} \frac{q_i(1)q_i(1)}{N(1)} & & & \\ & \frac{q_i(2)q_i(2)}{N(2)} & & \\ & & \ddots & \\ & & & \frac{q_i(T)q_i(T)}{N(T)} \end{bmatrix},$$

avec $N(t)$ désigne le nombre d'observations à la période t.
Cette structure de Variances-Covariances a les propriétés suivantes :

- chaque équation du modèle à des erreurs hétéroscédastiques,
- à une étape donnée les erreurs de deux équations sont liées,
- les erreurs ne sont pas autocorrelées.

La singularité de Σ résulte des relations :

$$\sum_{i=1}^{r} q_i(t) = 1, \quad \text{pour } t = 1, 2, ..., T. \tag{2.37}$$

Le modèle peut être réécrit sous sa forme détaillée suivante :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_r \end{bmatrix} = \begin{bmatrix} X & 0 & \cdots & 0 \\ 0 & X & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_r \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_r \end{bmatrix},$$

où on constate que les r équations sont linéairement dépendantes, nous écrirons donc le modèle sous la forme :

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{r-1} \end{bmatrix} = \begin{bmatrix} X & 0 & \cdots & 0 \\ 0 & X & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{r-1} \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{r-1} \end{bmatrix},$$

ce qui nous donne la forme condensée suivante :

$$y_* = X_* \, p_* + u_*, \tag{2.38}$$

avec

$$\begin{aligned} E(u_*) &= 0 \\ E(u_* u_*') &= \Sigma_* \end{aligned}$$

où Σ_* est une sous matrice régulière de Σ.
Nous sommes alors amener à résoudre le problème suivant :
Minimiser,

$$(y_* - X_* p_*)' \, \Sigma_*^{-1} (y_* - X_* p_*),$$

sous les contraintes :

$$\begin{cases} R p_* \leq \eta_r, \\ p_* \geq 0, \end{cases} \tag{2.39}$$

où R désigne une sous matrice de G définie dans (2.30). Pour le faire, on peut utiliser un algorithme de programmation quadratique. Notons qu'en l'absence des contraintes (2.39), nous aurions l'estimation suivante :

$$\hat{p}_* = (X_*' \ \Sigma_*^{-1} \ X_*)^{-1} X_*' \ \Sigma_*^{-1} \ y_* \tag{2.40}$$

$$V(\hat{p}_*) = (X_*' \ \Sigma_*^{-1} \ X_*)^{-1}, \tag{2.41}$$

d'où

$$\hat{p}_r = \eta_r - R\hat{p}_*. \tag{2.42}$$

Toutefois rien ne nous assure que cet estimateur satisfait aux contraintes. Pour estimer Σ on remplace les vrais fréquences $q_j(t)$ par les fréquences observées $y_j(t)$ ce qui conduit à une estimation convergente.

2.2.5 Autres méthodes d'estimation et choix d'un estimateur

En 1970 Lee et al. [59] étudient en détail les différentes méthodes d'estimation, outre celles basées sur les moindres carrés, ils envisagent la méthode du χ^2 minimum, celle du maximum de vraisemblance et celle basée sur la minimisation de la somme des valeurs absolues des écarts. Par ailleurs ils envisagent le problème sous l'angle bayesien. Les études de Monte-Carlo peuvent servir de base pour choisir une méthode d'estimation. Lorsque les microdonnées existent, la méthode du maximum de vraisemblance fournit les meilleurs résultats. Dans le cas général où seules les macrodonnées existent, les méthodes sous contraintes sont meilleures que les autres, on peut retenir celle des moindres carrés généralisés, toutefois lorsqu'on dispose d'informations a priori, il faut choisir l'approche bayésienne. Lee et al. [58] ont tenté de généraliser cette approche au cas d'une chaîne non homogène.

Le problème d'inférence statistique dans les processus de Markov a reçu une attention considérable, en particulier au cours des années 50 et 60. Une grande partie du travail consiste à reporter dans le cadre des chaînes de Markov les méthodes du maximum de vraisemblance et du χ^2 sur des processus à variables aléatoires indépendantes identiquement distribuées (voir, par exemple, [11] et d'autres références citées dans la Section 2.2). Cependant, quoi que l'estimation paramétrique des caractéristiques pour les chaînes Markoviennes finies est importante pour les applications, elle n'a pas été beaucoup explorée. Ceci est dû au fait qu'il est difficile d'estimer avec précision les caractéristiques d'une chaîne Markovienne modélisant des phénomènes complexes.

Pour surpasser le problème, d'autres approches alternative ont également été adoptées [78], dont certaines [20, 21] se réfèrent à des inférences statistiques dans des processus plus généraux. Mais, c'est l'apparition du travail de Rosenblatt [77], où il a considéré une nouvelle technique

d'estimation non paramétrique de la densité dans le cas de variables aléatoires indépendantes et identiquement distribution, qui permis un essor aux problèmes d'estimation dans plusieurs autres champs. En effet, plusieurs autres travaux [61, 101, 102] ont suivi, où en utilisant des méthodes similaires ou différentes ont obtenu des résultats supplémentaires et/ou complémentaires.

Historiquement, dans le cadre d'estimation non paramétrique dans les chaînes de Markov Roussas [79] fus le premier a considéré la méthode non paramétrique du noyau. En 1969 Roussas [79] a utilisées la méthode du noyau pour l'estimation la densité initiale et la densité jointe (bidimensionnelle) associées à une chaîne de Markov. Ce qui lui a permis de construire par la suite l'estimateur à noyau de la densité de transition (conditionnelle) d'une chaîne de Markov. Dans ce qui suit, nous présentons un résumé des principaux résultats obtenus par Roussas1969 dans [79, 80] sur l'estimation à noyau dans les chaînes de Markov.

2.3 Estimation à noyau des probabilités de transition des CM

Dans [79], le problème de l'estimation non paramétrique dans les processus de Markov a été considéré. Des estimations des densités initiale et jointe d'un processus de Markov, satisfaisant un certain nombre de propriétés optimales, ont été obtenues. Sous la même configuration non paramétrique exposé dans [79, 80] l'objectif de l'auteur a été principalement centré sur l'estimation à noyau d'autres caractéristiques d'un processus Markovien défini sur l'espace de probabilité (Ω, A, P) et prenant des valeurs dans $\mathbb{R}$, et ceci sous la condition de stationnarité de ce processus.

Soit $p(.)$, $q(.,.)$ et $t(.|x)$, $x \in \mathbb{R}$ représentent respectivement la densité initiale, la densité conjointe (bidimensionnelle) et la densité de transition.

2.3.1 Construction des estimateurs et leurs propriétés [79, 80]

Soit $X_j, j = 1, 2, ..., n+1$ les $n+1$ premières variables aléatoires issues du processus de Markov qui régis selon les densités inconnues $p(x)$ et $q(y)$ (avec : $x \in \mathbb{R}$ et $y \in \mathbb{R} \times \mathbb{R}$). Roussas [79] a proposé d'estimer $p(x)$ et $q(y)$, respectivement par :

$$p_n(x) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right), \tag{2.43}$$

$$q_n(y) = q_n(x, x') = \frac{1}{nh^2} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right) K\left(\frac{x' - X_{j+1}}{h}\right), \tag{2.44}$$

ou encore

$$p_n(x) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right), \tag{2.45}$$

$$q_n(y) = q_n(x, x') = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{x - X_j}{\sqrt{h}}\right) K\left(\frac{x' - X_{j+1}}{\sqrt{h}}\right), \tag{2.46}$$

où le noyau K et le paramètre de lissage h satisfaisaient aux conditions usuelles imposées sur le paramètre de lissage et le noyau (voir section1.1 du chapitre 1).

D'après les expressions des estimateurs de p et q, on constate que la construction de ces deux estimateurs n'a rien de nouveauté et d'exceptionnel. En effet, les estimateurs en question ne sont qu'un résultat immédiat du travail de Parzen [73] pour p_n, et le travail de Cacoullos [19] pour q_n. La contribution effectif de [79], réside dans les différents résultats théoriques dégagés par l'auteur sur les propriétés asymptotiques des deux estimateurs en question. Dans un premier lieu, l'auteur a démontré que les deux estimateurs p_n et q_n sont asymptotiquement sans biais (voir Théorème 2.1). Par la suite, la consistance simple et uniforme en moyenne quadratique ont été prouvé (voir Théorème 2.2). En fin, la convergence en loi des deux estimateurs a été mise en évidence (voir Théorèmes 2.3 et 2.4). Par ailleurs, l'auteur a démontrer, également, dans [80] que

$$\int |q_n(x,x') - q(x,x')|dx' \longrightarrow 0,$$

en probabilité pour tout $x \in \mathbb{R}$ lorsque n tend vers l'infinie (voir Lemme 2.1).

Théorème 2.1. *(Roussas [79]) Si les paramètres K et h satisfont aux conditions usuelles, sur le noyau et le paramètre de lissage, alors les deux variables $p_n(x)$ et $q_n(y)$ sont des estimateurs asymptotiquement sans biais de $p(x)$ et $q(y)$, respectivement, c'est-à-dire :*

$$E(p_n(x)) \to p(x), \qquad lorsque\ \ n \to \infty, \qquad pour\ \ x \in \mathbb{R},$$

et

$$E(q_n(y)) \to q(y), \qquad lorsque\ \ n \to \infty, \qquad pour\ \ y \in \mathbb{R} \times \mathbb{R}.$$

Théorème 2.2. *(Roussas [79]) Sous les conditions usuelles sur le noyau K et le paramètre de lissage h, les deux variables $p_n(x)$ et $q_n(y)$ sont consistantes en moyenne quadratique pour $x \in \mathbb{R}$ et $y \in \mathbb{R} \times \mathbb{R}$ et elles sont uniformément consistantes en moyenne quadratique sur un ensemble compact $x \in E_x \subset \mathbb{R}$ et $y \in E_y \subset \mathbb{R} \times \mathbb{R}$.*

Théorème 2.3. *(Roussas [79]) Si*

1. *Le noyau K et le paramètre de lissage h satisfont aux conditions usuelles ;*
2. *Les densités conjointes de X_1, X_i et X_1, X_i, X_j sont bornées par M $(< \infty)$ pour tout i et j, tel que*
$$1 < i \leq n, \quad 1 < i < j \leq n, \quad n = 2, 3, ...;$$
3. *Il existe trois entiers positifs α, β et μ, tel que*

$$\beta\mu\alpha^{-1} \to 0 \qquad \textit{et} \qquad \alpha h_n n^{-1} \to 0, \ \textit{lorsque}\ n \to \infty;$$

alors pour tout $x \in \mathbb{R}$, on a

$$(nh)^{\frac{1}{2}} \left[p_n(x) - E(p_n(x))\right] \xrightarrow{loi} \mathcal{N}(0, \sigma^2(x)), \ \textit{lorsque}\ n \to \infty;$$

avec $\sigma^2(x) = p(x) \int K^2(t)dt$.

Théorème 2.4. *(Roussas [79]) Si*

1. *Le noyau K et le paramètre de lissage h satisfont aux conditions usuelles ;*
2. *Les densités conjointes de Y_1, Y_i et Y_1, Y_i, Y_j sont bornées par M $(< \infty)$ pour tout i et j tel que*
$$1 < i \leq n, \quad 1 < i < j \leq n, \quad n = 2, 3, ...;$$

3. Il existe trois entiers positifs α, β et μ tel que

$$\beta\mu\alpha^{-1} \to 0 \qquad \textit{et} \qquad \alpha h_n n^{-1} \to 0, \ \textit{lorsque } n \to \infty;$$

alors pour tout $y \in \mathbb{R} \times \mathbb{R}$ tel que $q_n(y) > 0$, on a

$$(nh)^{\frac{1}{2}} \left[q_n(y) - E(q_n(y))\right] \xrightarrow{\ loi\ } \mathcal{N}(0, \sigma^2(y)), \ \textit{lorsque } n \to \infty;$$

avec $\sigma^2(y) = q(y) \int K^2(t)dt$.

Lemme 2.1. *(Roussas [80]) Si*

1. $K(x) \leq M, \ \ x \in \mathbb{R}$ et $|x|K(x) \to 0$ lorsque $|x| \to \infty$,

2. $nh \to \infty$ lorsque $n \to \infty$,

3. Les deux densités sont des fonctions continues et $p(x) > 0$, $x \in \mathbb{R}$,

alors, lorsque $n \to \infty$, on a :

$$E\left(\int |q_n(x, x') - q(x, x')|dx'\right) \longrightarrow 0, \ \ \forall x \in \mathbb{R}.$$

2.3.2 Construction d'estimateurs d'autres caractéristiques [79, 80]

L'estimation d'autres caractéristiques d'une chaîne de Markov, par la méthode du noyau, a été également considéré par Roussas dans [79, 80]. Dans [79], l'auteur a proposé un estimateur à noyau de la densité de transition $t(x'/x)$ tandis que dans [80], l'auteur s'est intéressé à l'estimation : des fonctions de répartition associées aux densités p et q, du moment d'ordre r de l'espérance conditionnelle de X_{n+1} sachant X_n ($n \geq 1$), ainsi que le quantile d'ordre p ($0 < p < 1$) de la distribution conjointe.

En se basant sur la définition d'une densité conditionnelle et les expressions des estimateurs p_n et q_n, Roussas à proposé dans [79] d'estimer $t(x'/x)$ ($x', \ x \in \mathbb{R}$) par :

$$t_n(x'|x) \ = \ q_n(x, x')/p_n(x). \tag{2.47}$$

Il est à souligner que $t_n(x'|x)$ est un estimateur convergent en probabilité lorsque n tend vers l'infinie et il converge uniformément sur un sous ensemble compact de $\mathbb{R}$ également lorsque n tend vers l'infinie (voir Corolaire 3.1., pages 77–78 de [79]). De plus, sous certaines conditions, la variable aléatoire $t_n(./x)$ converge vers une loi normale lorsque n tend vers l'infinie (voir Théorème 4.3. , pages 79–80 de [79]). Au moyen de $p_n(x)$ et $q_n(x, x')$, dans [80] l'auteur a définit les deux variables aléatoires suivantes :

$$F_n(x) \ = \ \int_{-\infty}^{x} p_n(z)dz, \tag{2.48}$$

$$G_n(z|x) \ = \ \int_{-\infty}^{x} t_n\left(dx'|x\right), \tag{2.49}$$

où $F_n(.)$ et $G_n(./x)$ sont l'estimateur à noyau, respectivement, de la fonction de répartition initiale F et de transition $G(.|x)$, $x \in \mathbb{R}$ du processus. Le résultat suivant porte sur la convergence des deux estimateurs en probabilité.

Théorème 2.5. *(Roussas[80]) Si le noyau K est continu et $|x|K(x) \to 0$ lorsque $|x| \to \infty$ alors, lorsque $n \to \infty$ on a :*

$$P\left(\sup_{x\in\mathbb{R}}\{|F_n(x) - F(x)| \to 0\}\right) = 1,$$

si de plus, on a :

1. $K(x) \leq M, \ \ x \in \mathbb{R}$,
2. $nh \to \infty$ *lorsque* $n \to \infty$,
3. *Les deux densités sont des fonctions continues et* $p(x) > 0$, $x \in \mathbb{R}$,

alors, lorsque $n \to \infty$ on a :

$$P\left(\sup_{z\in\mathbb{R}}\{|G_n(z/x) - G(z/x)| \to 0\}\right) = 1, \ \ \forall x \in \mathbb{R}.$$

Pour estimer le moment conditionnelle d'ordre k défini par

$$m(k,x) = E(X_{n+1}/X_n) = E(X_2/X_1) = \int t^k t(dt/x), \qquad \text{pour } k = 1, 2, ..., r.$$

Roussas [80] a proposé, sous l'hypothèse que les premiers moments abosolu d'ordre r ($r = 1, 2, ...$) de X_1 exist, ce qui suit :

$$m_n(k,x) \ = \ \frac{1}{(nh)p_n(x)}\sum_{j=1}^{n} X_{j+1}^k K\left(\frac{(x - X_j)}{h}\right), \qquad \text{pour } k = 1, 2, ..., r. \qquad (2.50)$$

De plus, il a montré que la variable aléatoire $m_n(k,x)$ converge vers $m(k,x)$ en probabilité, pour $k = 1, 2, ..., r$ et $x \in \mathbb{R}$ lorque $n \to \infty$ (voir Théorème 4.1, page 1392 de [80]).
Dans la section 5 du papier de [80], le problème de l'estimation de quantile d'ordre p de $G(.|x)$ a été examiné. Deux résultats importants ont été dérivés, à savoir :

- $\xi_n(p,x) \to \xi_(p,x)$ en probabilité, lorsque $n \to \infty$.
- $(\sqrt{nh})[\xi_n(p,x) - \xi_(p,x)] \to N(0, \tau(\xi,x))$ en loi lorsque $n \to \infty$,

où $\xi_n(p,x)$, qui représente l'estimateur du quantile d'ordre p, $\xi(p,x)$, est la plus petite racine de l'équation $G_n(Z|x) = p$ et la variance $\tau(\xi,x))$ est donnée dans [80] au niveau du Théorème 5.2, page 1399.

Les résultats obtenus dans [79, 80] ont été complétés par Masry et Györfi [64] et par Basu et Sahoo [10], parmi d'autres. Par la suite, Laksaci et Yousfate [57] ont étudié un estimateur à noyau de la densité de l'opérateur de transition, vu comme un endomorphisme de L^p, $p \in [1;\infty[$. Cet estimateur permet de construire un estimateur fonctionnel aussi bien pour l'opérateur de transition que pour son adjoint. Leur principal résultat fournit une majoration de la vitesse de convergence (au sens de la norme L^p) de l'estimateur construit. Cependant, les différents résultats obtenus jusqu'au là sont restreints dans un cadre théorique seulement. L'utilisation de la méthode du noyau dans les chaînes de Markov dans le cadre pratique, n'est appliquée que récemment. Entre autres, on peut citer : Bareche et Aïssani (2008) [9], qui ont appliqué la méthode du noyau pour mesurer les performances de la méthode de stabilité forte dans l'étude des systèmes d'attente classiques quand l'une des lois les régissant est générale et inconnue. Gontijo et al. (2011) [39], qui ont appliqué la méthode de noyau pour estimer les mesures de performance du système $GI^{[X]}/M/C/N$. Cherfaoui et al. (2015) [27] qui ont abordé le problème du choix du paramètre de lissage dans le cadre d'estimation à noyau gamma d'une matrice de transition d'une chaîne de Markov continue à un espace d'états fini,...

2.3.3 Estimation de P et correction d'effet du biais aux bornes [9, 39]

Dans [9], les auteurs ont évalué l'approximité des systèmes $GI/M/1$ et $M/M/1$, lorsque la densité des temps des inter-arrivées est estimée par la méthode du noyau. Dans [39] les auteurs ont appliqué la méthode de noyau sur le système $GI^X/M/C/N$ afin d'estimer ses mesures de performance. L'objectif essentiel de ces travaux est basé sur l'estimation de la matrice de transition, $\mathbb{P}$, associée au système considéré et cela en substituant la densité g, de la distribution générale supposée inconnue, dans les éléments de la matrice de transition par son estimateur à noyau, g_h. La démarche suivie dans ces deux travaux peut être résumée comme suit :

1. Considérer un n-échantillon $T_1, T_2, ..., T_n$ issu d'une variable aléatoire positive T ayant une densité de probabilité inconnue g.
2. Estimer le paramètre de lissage optimale, h^* par les méthodes classiques (voir chapitre 1).
3. Utiliser l'expression de l'estimateur à noyau pour quantifier g_{h^*}.
4. Substituer la densité inconnue g par son estimateur g_{h^*} dans les expressions des probabilités de transition, P_{ij}, afin de calculer l'estimateur de la matrice de transition , $\hat{\mathbb{P}}$, associe au système considéré.

Par ailleurs, il est à noter que dans [9, 39], les auteurs s'est intéressés à l'estimation de la distribution du temps des inter-arrivées qui est défini sur $\mathbb{R}^+$. Ce qui fait, la construction de l'estimateur de g à l'aide des noyaux classiques est à éviter (voir Chapitre 1). A cet effet, les auteurs ont fait recours aux méthodes de correction du biais aux bornes. En effet, dans [9] pour estimer la densité g les auteurs ont utilisé l'estimateur à noyau de Schuster, noyaux asymétriques et histogrammes lissés. Par contre, dans [39] les auteurs ont utilisé uniquement les noyaux gamma. Pour le choix du paramètre de lissage, dans les deux travaux en question, les auteurs ont utilisé les méthodes classiques (UCV, BCV, règle de référence,..).

2.3.4 Choix du paramètre de lissage dans l'estimation à noyau de P [27]

Soit n-échantillon $T_1, T_2, ..., T_n$ qui représente les durées des inter-arrivées, dans un système $GI/M/1/N$, ayant comme densité de probabilité inconnue g. L'estimation de la matrice de transition, $\mathbb{P}$, de la chaîne de Markov induite associée au système $GI/M/1/N$, donnée par :

$$P_{ij} = \begin{cases} \int_0^\infty \frac{e^{-\mu t}(\mu t)^{i-j+1}}{(i-j+1)!} g(t)dt, & \text{si } 1 \leq j \leq i+1 \leq N, \\ \int_0^\infty \frac{e^{-\mu t}(\mu t)^{N-j}}{(N-j)!} g(t)dt, & \text{si } 1 \leq j \leq N \text{ et } i = N, \\ 1 - \sum_{k=1}^{N} P_{ik}, & \text{si } j = 0, \\ 0, & \text{sinon,} \end{cases} \tag{2.51}$$

consiste à évaluer la densité inconnue g et de substituer son estimateur, noté $\hat{g}$, dans les P_{ij}. Supposons qu'on opte pour le critère $MISE$ et le premier noyau gamma proposé par Chen [25]. Alors, la formule explicite de $\hat{g}$ est donnée par :

$$\hat{g}(t) = \frac{1}{n}\sum_{i=1}^{n} K(t,h)(T_i) = \frac{1}{n}\sum_{i=1}^{n} \frac{T_i^{t/h} e^{-T_i/h}}{h^{(t/h)+1}\Gamma((t/h)+1)}, \tag{2.52}$$

où K est une densité de la loi Gamma de paramètres $(\frac{t}{h}+1, h)$ et le paramètre de lissage optimal h_1^* se calcule comme suit :

$$\begin{aligned} h_1^* &= \arg\min_h MISE(g,\hat{g}) = \arg\min_h \int_0^\infty \mathbb{E}\left(g(t)-\hat{g}(t)\right)^2 dt \\ &= \arg\min_h \left[h^2 \int_0^\infty \left\{ g'(t) + \frac{1}{2} t g''(t) \right\}^2 dt + \frac{1}{2\sqrt{\pi}} n^{-1} h^{-1/2} \int_0^\infty t^{-1/2} g(t) dt + o(n^{-1}h^{-1/2} + h^2) \right] \\ &= \left[\frac{1}{2\sqrt{\pi}} \int_0^\infty t^{-1/2} g(t) dt \right]^{2/5} \left[\int_0^\infty \left\{ g'(t) + \frac{1}{2} t g''(t) \right\}^2 dt \right]^{-2/5} 4^{-2/5} n^{-2/5}. \end{aligned} \tag{2.53}$$

Dans le but de prendre en considération les pondérations de l'estimateur $\hat{g}$ dans l'expression de $\hat{P}_{ij}$, Cherfaoui et al. [27] ont proposé l'utilisation des normes matricielles qui ont un impact sur la qualité de l'estimateur de $\mathbb{P}$, noté $\hat{\mathbb{P}}$. L'idée de l'utilisation des normes matricielles et quelles permettent d'inclure les pondérations, qui sont d'une loi de Poisson de paramètre $\mu * t$, de la quantité $g(t)$ dans l'expression de P_{ij} lors de l'estimation de $\mathbb{P}$. Les auteurs ont proposé de sélectionner le paramètre de lissage optimal selon l'une des trois expressions suivantes :

$$h_2^* = \arg\min_h \|\hat{\mathbb{P}} - \mathbb{P}\|_1 = \arg\min_h \left[\max_j \left(\sum_{i=0}^N \left| \hat{P}_{ij} - P_{ij} \right| \right) \right], \tag{2.54}$$

$$h_3^* = \arg\min_h \|\hat{\mathbb{P}} - \mathbb{P}\|_2 = \arg\min_h \left[\sum_{i=0}^N \sum_{j=0}^N \left(\hat{P}_{ij} - P_{ij} \right)^2 \right]^{1/2}, \tag{2.55}$$

$$h_4^* = \arg\min_h \|\hat{\mathbb{P}} - \mathbb{P}\|_\infty = \arg\min_h \left[\max_i \left(\sum_{j=0}^N \left| \hat{P}_{ij} - P_{ij} \right| \right) \right], \tag{2.56}$$

où $\hat{P}_{ij}$ est l'estimateur de P_{ij} lorsque on remplace $g(t)$ par son estimateur $\hat{g}(t)$ donné par (2.52).

Énonçons à présent quelques résultats théoriques concernant les propriétés de l'estimateur $\hat{P}_{ij}$.

Proposition 2.1. *(Cherfaoui et al. [27]) Soient $g \in \mathcal{C}^2([0,\infty[)$ une densité de probabilité de la distribution générale des durées des inter-arrivées du système $GI/M/1/N$ et $\hat{g}$ son estimateur à noyau Gamma. Sous la condition $\lim_{n\to+\infty} h = 0$, le biais asymptotique de l'estimateur $\hat{P}_{ij}$ s'écrit comme suit :*

$$Biais(\hat{P}_{ij}) = \begin{cases} h \int_0^\infty \frac{(\mu t)^{i-j+1}}{(i-j+1)!} e^{-\mu t} \left\{ g'(t) + \frac{1}{2} t g''(t) \right\} dt + o(h), & \text{si } 1 \le j \le i+1 \le N, \\ h \int_0^\infty \frac{(\mu t)^{N-j}}{(N-j)!} e^{-\mu t} \left\{ g'(t) + \frac{1}{2} t g''(t) \right\} dt + o(h), & \text{si } 1 \le j \le N \text{ et } i = N, \\ -\sum_{k=1}^N Biais(\hat{P}_{ik}), & \text{si } j = 0, \\ 0, & \text{sinon.} \end{cases} \tag{2.57}$$

Théorème 2.6. *(Cherfaoui et al. [27]) Soient $g \in \mathcal{C}^2([0,\infty[)$ une densité de probabilité de la distribution générale des durées des inter-arrivées du système $GI/M/1/N$ et $\hat{g}$ son estimateur à noyau Gamma.*

$$Si\left(\lim_{n\to+\infty} h = 0 \ \textit{et} \ \lim_{n\to+\infty} nh^2 = +\infty\right), alors\left(|\hat{P}_{ij} - P_{ij}| \xrightarrow{P} 0 \ \textit{lorsque} \ h \to 0\right). \quad (2.58)$$

Il a été démontré par simulation, également, dans [27] qu'il est préférable dans l'estimation d'une matrice de transition, d'utiliser les normes matricielles pour le choix du paramètre de lissage plutôt que les autres méthodes classiques de sélection et ceci dans le sens de la vitesse de convergence de l'erreur quadratique et de l'erreur d'approximation des caractéristiques stationnaires du système considéré. Dans [28], les auteurs ont montré que l'application des méthodes classiques d'une manière usuelles peuvent nous fournir des estimateurs non consistants voir même erronés.

Conclusion

Dans le présent chapitre, nous nous sommes intéressés à l'estimation des probabilités de transition d'une chaîne de Markov, en particulier par la méthode non paramétrique du noyau.

Il est facile de remarquer que les méthodes paramétriques exposées considèrent uniquement le cas de CMTD (Chaînes de Markov à Temps Discret). Mais, il est à noter que ces méthodes sont applicables également au cas de Chaînes de Markov à Temps Continue (CMTC), et ceci est dû au fait que dans la pratique, lors de l'échantillonnage, en particulier dans le cas de macrodonnées, la discrétisation du temps est inévitable. L'analyse intuitive de ces méthodes nous laisse prédire que leur mise en œuvre nécessite des échantillons de tailles considérablement grandes.

Les différents travaux (références) cités et exposés dans ce chapitre, nous permet de conclure que les études effectuées dans le cadre d'estimation à noyau dans les chaînes de Markov n'ont été considérées que dans le cas de CMTC. C'est la raison pour laquelle, nous allons considérer dans la suite de ce document le cas de CMTD (voir Chapitre 4).

Chapitre 3

Performances du modèle d'attente $M/M/1/N$ non fiable à plusieurs variantes

Introduction

L'origine des études sur les phénomènes d'attente remonte aux années 1909–1920 avec les travaux d'A.K. Erlang [35] concernant le réseau téléphonique de Copenhague. La théorie mathématique s'est ensuite développée notamment grâce aux contributions de Palm, Kolmogorov, Khintchine, Pollaczek,... et actuellement s'est étendue à de nombreux champs d'application comme la gestion de stocks, les télécommunications en général, la fiabilité de systèmes complexes,... Et ceci est dû à la qualité des résultats fournis par cette théorie et au fait que les problèmes liés à l'attente dans un centre de service sont omniprésents dans nos jours dont les exemples ne manquent pas :

- attente à un guichet (caisse dans un supermarché, administration),
- trafic urbain ou aérien,
- réseaux téléphoniques,
- circulation de pièces dans un atelier,
- programmes dans un système informatique,
- ...

Dans ce chapitre, nous considérons un modèle d'attente qui peut être très approprié pour la modélisation de plusieurs situations réelles de divers domaines, tels que les systèmes de communication et de télécommunications, les systèmes de fabrication, l'informatique, etc. Le modèle en question est le modèle $M/M/1$ à capacité finie, avec vacances multiples du serveur, Bernoulli feedback du client servis, balking, reneging et rétention des clients impatients, ainsi que la possibilité de panne et réparation du serveur. Notre contribution réside dans l'analyse probabiliste (analyse exacte) et statistique (estimation paramétrique) de ce modèle.

Dans la première partie du présent chapitre, nous nous sommes intéressés à l'analyse exacte du modèle d'attente en question. En effet, à l'aide de la méthode $Q-$matrice (matrice génératrice infinitésimale) nous avons dégagé la forme exacte de la distribution des probabilités d'état stationnaire de la chaîne de Markov bidimensionnelle, décrivant le modèle d'attente considéré. Ces dernières probabilités nous ont permis, par la suite, de décrire plusieurs mesures de performance du système à l'état stationnaire. De plus, l'analyse du comportement de ces mesures de performance en fonction des paramètres fiabilistes (taux des pannes, taux des réparations et taux des

vacances, vu en tant que maintenances préventives), a été mise en évidence numériquement.

Dans la deuxième partie, sous l'hypothèse que les paramètres de départ du système ne sont disponibles que sous forme d'un ensemble d'observations et à l'aide de plusieurs outils statistiques (histogramme, box plot, test de conformité,...), nous avons analysé l'impact de l'estimation des paramètres départ, définissant le modèle d'attente ci-dessus, sur les propriétés statistiques (biais, variance, risque quadratique,...) des estimateurs de ses mesures de performance stationnaires.

3.1 État de l'art

Dans la théorie des files d'attente, une file d'attente classique peut être décrite comme un système dans lequel les clients arrivent selon un processus d'arrivées, pour être servis par une installation de service selon un processus de service. Cependant, en pratique différents comportements du (des) serveur(s) et des clients peuvent être identifiés. Un client peut quitter définitivement le système sans être servi pour diverses raisons. Dans un scénario de balking, les clients refusent d'entrer dans la file d'attente étant donné qu'elle a atteint une certaine longueur. Un autre cas est le reneging du client impatient. Autre que le balking du client, un client reneging se joint à la file d'attente en attente de service. Si le temps d'attente perçu dépasse les attentes du client, alors ce client quitte le système. Un autre comportement du client que l'on peut distinguer en pratique est celui décrit par la notion de "feedback", introduite en général pour exploiter les situations d'attente où tous les clients demandent le service principal et quelques-uns ont besoin de demander un autre service supplémentaire. Concernant le(s) serveur(s), autre que la période d'activité du serveur, le serveur peut être indisponible pendant une période aléatoire résultant de nombreux facteurs. Dans certains cas, l'indisponibilité peut être le résultat d'une panne du serveur, ce qui signifie que le système doit être réparé et remis en service. Cela peut également être une action délibère d'utiliser le temps d'inactivité du serveur à différentes fins et dans ce cas, le serveur est dit être en période de vacances.

De nombreuses études sur des systèmes régissent selon l'une ou plusieurs des variantes citées ci-haut, sont parues dans la littérature, on peut citer entre autres : Haight (1957) [42] a d'abord considéré une file d'attente $M/M/1$ avec balking, par la suite, il a considéré la file d'attente $M/M/1$ avec de clients impatients (reneging) [43]. La combinaison des deux variantes balking et reneging dans une file d'attente $M/M/1/N$ a été étudiés par Ancker et Gafarian (1963) [7]. Abou-EI-Ata et al. (1992) [33] ont considéré le système d'attente à plusieurs serveurs $M/M/c/N$ avec balking et reneging. Wang et Chang (2002) [99] ont étendu ce dernier travail au cas du système $M/M/c/N$ avec balking, reneging et possibilité de pannes de serveurs. Zhang et al. (2005) [108] ont présenté une analyse complète pour un système de file d'attente $M/M/1/N$ avec balking, reneging et vacances de serveur. El-Paoumy et Nabwey (2011) [34] ont obtenu la solution analytique de la file d'attente $M/M/2/N$ avec reneging, deux serveurs hétérogènes et une fonction de balking générale. Kumar et al. [54] ont étudié analytiquement une file d'attente $M/M/1/N$ avec feedback et rétention des clients impatients. Kumar (2013) [53] a présenté une analyse économique du modèle d'attente $M/M/c/N$ avec balking, reneging et rétention de clients impatients. Kumar et Sharma (2014) [55] ont étudié le modèle de file d'attente Markovienne à plusieurs serveurs, à capacité finie, Bernoulli feedback, balking et reneging et rétention de clients impatients. Kumar et al. (2014) [56] ont présenté une analyse économique d'une file d'attente $M/M/1/N$ avec feedback et rétention des clients impatients où ils ont obtenu la capa-

cité et le taux du service optimal, au sens économique, du système.

Misra et Goswami (2015) [66] ont analysé un trafic de classe II à économie d'énergie dans IEEE 802.16E avec plusieurs états en veille et de balking. Panda et Goswami (2016) [71] ont analysé les stratégies de balking d'équilibre pour une file d'attente $GI/M/1$ avec des vacances Bernoulli (vacances actives) et une interruption de vacances, dans le cas où un client ne peut qu'observer l'état du serveur (files d'attente observables) et lorsqu'il n'y a pas d'informations disponibles pour un client avant de prendre la décision de rejoindre le système ou non (files d'attente totalement inobservables).

Cependant, à notre connaissance, il n'y a pas de travaux où les diverses variantes (balking, reneging et rétention de clients impatients, vacances du serveur, possibilité de pannes et réparation de serveur) ont été considérés ensemble. Ceci nous a motivé à considérer l'étude du système d'attente qui regroupe tous les précédentes variantes simultanément. C'est-à-dire que nous avons envisagé un système de file d'attente $M/M/1/N$ Markovien avec des vacances multiples, Bernoulli feedback, balking, reneging et rétention des clients impatients, ainsi que la possibilité de la panne et de la réparation du serveur pour lequel nous avons dégagé les probabilités stationnaires et quelques de ses mesures de performance.

Il est à noter que ces variantes peuvent se présenter sous différentes formes en pratique, et ce en fonction du domaine de la problématique. Ci-dessous deux exemples qui mettent en évidence certaines formes des concepts introduits en haut.

Exemple 3.1. *Une entreprise de production :*

Notion	*Sens pratique*
Client	*Un ordre de fabrication (une commande)*
Serveur et service	*Une machine et le produit*
Balking	*Après l'analyse de la situation, au sens du nombre d'ordres de fabrication en attente, le client peut juger qu'il n'aura pas son produit à l'heure espérée, il sera donc découragé et décide de ne pas faire une demande de production.*
Vacances Multiples	*Effectuer une maintenance préventive sur la machine, lorsqu'il n'y a pas d'ordres de fabrication en attente.*
Pannes et réparations	*Pannes et réparations (maintenances correctives) de la machine.*
Reneging et rétention	*Après un certain délai d'attente, le client peut procéder à l'annulation de son ordre de fabrication ou il maintient sa commande après une hésitation.*
Bernoulli feedback	*Le produit nécessite un service supplémentaire (le produit n'est pas finalisé, le produit a des défauts et nécessite des ajustements,...) ou il y a un nouvel ordre de fabrication qui est passé par le même client juste au moment de la finalisation de sa commande précédente.*

Exemple 3.2. *Centre d'appels :*
Un autre exemple très fréquent et bien détaillé dans la littérature est celui du centre d'appels. En effet, il existe une littérature abondante, consacrée à ce sujet, qui a mis en évidence le sens et l'importance pratique des différents effets précédents dans ce domaine, nous citons l'exemple de [104, 93].

3.2 Description du modèle

Considérons un système de file d'attente $M/M/1/N$ avec vacances multiples, Bernoulli feedback, balking, reneging et rétention des clients impatients, et la possibilité de la panne et de la réparation du serveur. Supposons que le système fonctionne sous les hypothèses suivantes :

- Les arrivées se produisent suivant un flux de Poisson avec un taux d'arrivée moyen λ. Ainsi les temps inter-arrivées sont indépendants et identiquement distribués selon une loi exponentielle de paramètre λ.
- À l'arrivée, le client décide soit de rejoindre la file d'attente avec une probabilité δ_n, qui dépend du nombre de clients dans le système, soit de s'abstenir avec une probabilité $\delta'_n = 1 - \delta_n$ (Balking). Notez que si le système est vide, alors le client arrivant rejoint le système avec une probabilité 1, c'est-à-dire $\delta_0 = 1$ et si le système est plein alors il rejoint le système avec une probabilité 0, c'est-à-dire $\delta_N = 0$. Alternativement, le découragement des clients peut être modélisé par une fonction monotone décroissante δ_n.
- Le client qui rejoint la file d'attente attendra un certain temps (temps de patience) pour que son service commence. S'il n'a pas commencé à ce moment-là, alors avec une probabilité α, il quitte le système définitivement sans être servi et avec une probabilité complémentaire $\alpha' = 1 - \alpha$, il reste dans la file et le processus d'impatience sera déclenché à nouveau. Les temps d'impatience suivent une distribution exponentielle de paramètre ξ.
- Les clients sont servis selon le principe du premier arrivé, premier servi $(FCFS)$. Une fois que le service d'un client commence, le service se poursuit toujours jusqu'à son achèvement. Les temps de service suivent une distribution exponentielle de paramètre μ.
- Chaque fois que le système est vide, le serveur prend des vacances pendant une période de temps aléatoire T. Si le serveur revient d'une vacance et ne trouve aucun client en attente, il commencera immédiatement une autre vacance. On suppose que T suit une distribution exponentielle de paramètre θ.
- Après avoir obtenu un service, avec une probabilité $\beta' = 1 - \beta$, le client peut rejoindre le système en tant que client Bernoulli feedback pour recevoir un autre service supplémentaire. Sinon, il quitte définitivement le système, avec une probabilité β, (où $\beta' + \beta = 1$). On ne fait pas la distinction entre une arrivée régulière et une arrivée du feedback.
- Le système (le serveur) est également sujet à des pannes actives et conservatrices. Ces pannes se produisent selon un processus de Poisson de taux η. Les temps de réparations du serveur en panne sont distribués de façon exponentielle de paramètre γ.

Les temps d'inter-arrivées, les temps de service, les temps de vacances, les temps d'impatience, les temps des inter-arrivées des pannes, les temps de réparation sont tous supposés indépendants les uns des autres. Une partie du système peut être illustrée par le schéma présenté dans la Figure 3.1.

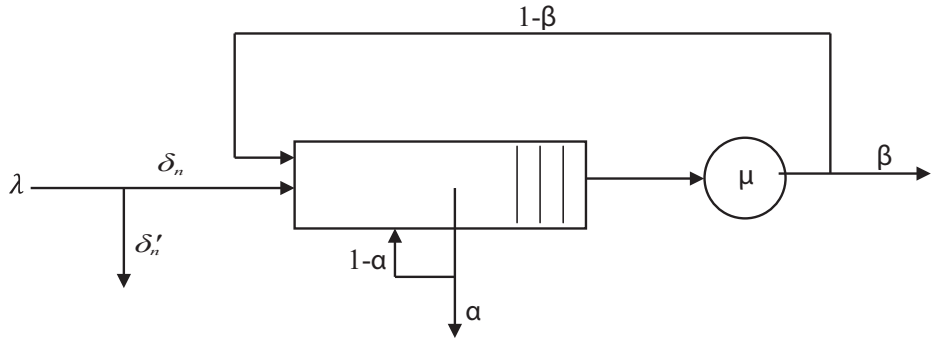

FIGURE 3.1 – File d'attente $M/M/1/N$ non fiable avec vacances multiples, Bernoulli feedbacks, balking, reneging et rétention de clients impatients.

3.3 Probabilités d'états en régime stationnaire

Dans cette section, nous allons introduire les probabilités d'état du système en régime stationnaire et ceci en utilisant la $Q-$matrice (matrice génératrice infinitésimale). Afin de répondre à notre objectif, nous allons considérer les notations suivantes qui seront utilisées tout au long du présent chapitre.

- $\xi_n = n\alpha\xi$, $\lambda_n = \lambda\delta_n$ et $\mu_n = \mu\beta + n\xi\alpha$,
- $X(t)$: le nombre de clients dans le système à l'instant t,
- $S(t)$: désigne l'état du système à l'instant t et défini par

$$S(t) = \begin{cases} 0, & \text{lorsque le serveur est en période de vacances ;} \\ 1, & \text{lorsque le serveur est en période d'activité ;} \\ 2, & \text{lorsque le serveur est en période de panne (période de réparation).} \end{cases}$$

- $P_{n,s}(t)$ désigne la probabilité d'avoir n clients dans le système à l'instant t et le serveur à l'état s :

$$P_{n,s}(t) = \mathbb{P}\left(X(t) = n,\ S(t) = s\right),\ n = 0,\ 1, ...,\ N,\ et\ s = 0,\ 1,\ 2.$$

Ainsi au régime stationnaire, $\lim\limits_{t\to\infty} P_{n,s}(t)$ seront comme suit :

- $P_{n,0}$; $0 \leq n \leq N$, est la probabilité qu'il y a n clients dans le système et le serveur est en période de vacances,
- $P_{n,1}$; $1 \leq n \leq N$, est la probabilité qu'il y a n clients dans le système et le serveur est en période d'activité,
- $P_{n,2}$; $1 \leq n \leq N$, est la probabilité qu'il y a n clients dans le système et le serveur est en période de panne ou de réparation.

En appliquant la théorie des processus de Markov, nous pouvons obtenir le diagramme de probabilités de transition, correspondant à notre système présenté dans la Figure 3.2.

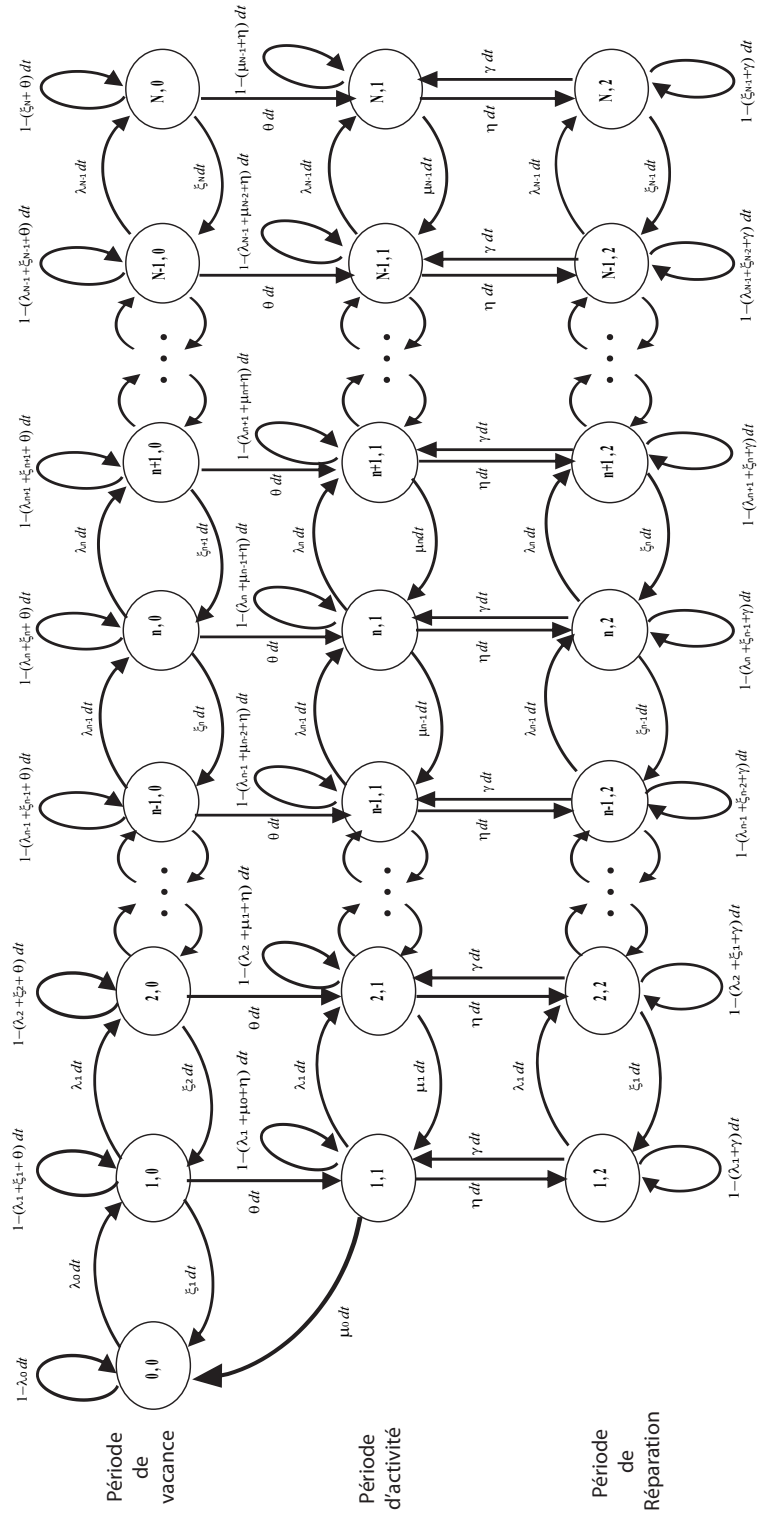

FIGURE 3.2 – Diagramme des probabilités de transition.

A base du graphe de transition présenté dans la Figure 3.2, il est facile de déduire les équations d'équilibre qui gouvernent le processus $\{(X(t), S(t)), t \geq 0\}$ dont leurs expressions en régime stationnaire sont données par :

$$\begin{aligned}
\lambda_0 P_{0,0} &= \xi_1 P_{1,0} + \mu_0 P_{1,1}, && (3.1)\\
(\lambda_n + \xi_n + \theta) P_{n,0} &= \lambda_{n-1} P_{n-1,0} + \xi_{n+1} P_{n+1,0}, \ \ 1 \leq n \leq N-1, && (3.2)\\
(\xi_N + \theta) P_{N,0} &= \lambda_{N-1} P_{N-1,0}, && (3.3)
\end{aligned}$$

$$\begin{aligned}
(\lambda_1 + \mu_0 + \eta) P_{1,1} &= \theta P_{1,0} + \mu_1 P_{2,1} + \gamma P_{1,2}, && (3.4)\\
(\lambda_n + \mu_{n-1} + \eta) P_{n,1} &= \lambda_{n-1} P_{n-1,1} + \theta P_{n,0} + \mu_n P_{n+1,1} + \gamma P_{n,2}, \ \ 2 \leq n \leq N-1, && (3.5)\\
(\mu_{N-1} + \eta) P_{N,1} &= \lambda_{N-1} P_{N-1,1} + \theta P_{N,0} + \gamma P_{N,2}, && (3.6)
\end{aligned}$$

$$\begin{aligned}
(\lambda_1 + \gamma) P_{1,2} &= \eta P_{1,1} + \xi_1 P_{2,2}, && (3.7)\\
(\lambda_n + \xi_{n-1} + \gamma) P_{n,2} &= \lambda_{n-1} P_{n-1,2} + \eta P_{n,1} + \xi_n P_{n+1,2}, \ \ 2 \leq n \leq N-1, && (3.8)\\
(\xi_{N-1} + \gamma) P_{N,2} &= \lambda_{N-1} P_{N-1,2} + \eta P_{N,1}. && (3.9)
\end{aligned}$$

Maintenant, notre objectif principal est d'obtenir les expressions analytiques des probabilités stationnaires $P_{n,s}$. D'après les équations (3.1)–(3.9), la matrice génératrice infinitésimale Q est une matrice définie en bloc et qui est donnée par :

$$Q = \begin{pmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{pmatrix},$$

où

$$A_1 = \begin{pmatrix}
-\lambda_0 & \lambda_0 & & & & & \\
\xi_1 & L_1 & \lambda_1 & & & & \\
& \xi_2 & L_2 & \lambda_2 & & & \\
& & & \ddots & \ddots & \ddots & \\
& & & & & & \\
& & & & \xi_{N-1} & L_{N-1} & \lambda_{N-1} \\
& & & & 0 & \xi_N & L_N
\end{pmatrix}, \qquad A_2 = \begin{pmatrix}
0 & & & & \\
\theta & & & & \\
& \theta & & & \\
& & \ddots & & \\
& & & & \theta
\end{pmatrix},$$

$B_1 = Diag\,(\mu_0, 0, \ldots, 0)$, $B_3 = Diag\,(\eta, \eta, \ldots, \eta)$, $C_2 = Diag\,(\gamma, \gamma, \ldots, \gamma)$, A_3 et C_1 sont des matrices dont tous les éléments sont des zéro,

$$B_2 = \begin{pmatrix} D_1 & \lambda_1 & & & & & \\ \mu_1 & D_2 & \lambda_2 & & & & \\ & \mu_2 & D_3 & \lambda_3 & & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & \mu_{N-2} & D_{N-1} & \lambda_{N-1} \\ & & & & 0 & \mu_{N-1} & D_N \end{pmatrix},$$

et

$$C_3 = \begin{pmatrix} K_1 & \lambda_1 & & & & & \\ \xi_1 & K_2 & \lambda_2 & & & & \\ & \xi_2 & K_3 & \lambda_3 & & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & \xi_{N-2} & K_{N-1} & \lambda_{N-1} \\ & & & & 0 & \xi_{N-1} & K_N \end{pmatrix}.$$

Avec $L_i = -(\lambda_i+\xi_i+\theta)$, $D_i = -(\lambda_i+\mu_{i-1}+\eta)$ et $K_i = -(\lambda_i+\xi_{i-1}+\gamma)$, pour $i = 1, 2, \cdots, N$.

Il est à noter que, A_1 est une matrice carrée d'ordre $N+1$, A_2 et A_3 sont des matrices d'ordre $(N+1) \times N$, B_1 et C_1 sont des matrices d'ordre $N \times (N+1)$ et B_2, B_3, C_2 et C_3 sont des matrices carrées d'ordre $N \times N$.

Soit $P = (P_0, P_1, P_2)$ le vecteur des probabilités correspondantes à l'état stationnaire du système associé à la matrice Q, où $P_0 = (P_{0,0}, P_{1,0}, \cdots, P_{N,0})$, $P_1 = (P_{1,1}, P_{2,1}, \cdots, P_{N,1})$ et $P_2 = (P_{1,2}, P_{2,2}, \ldots, P_{N,2})$. Le vecteur des probabilités P doit satisfaire les conditions suivantes :

$$\begin{cases} PQ &= 0, \\ Pe &= 1, \end{cases} \tag{3.10}$$

où $e = (e_0, e_1, e_2)$ est un vecteur colonne d'ordre $3N+1$ avec :

- e_0 est un vecteur colonne d'ordre $N+1$ dont tous les éléments sont égaux à 1,
- e_1 et e_2 sont des vecteurs colonne d'ordre N dont tous les éléments sont égaux à 1.

A partir du système d'équations (3.10), après quelques substitutions, nous obtenons le système suivant :

$$\begin{cases} P_0A_1 + P_1B_1 + P_2C_1 &= 0, \\ P_0A_2 + P_1B_2 + P_2C_2 &= 0, \\ P_0A_3 + P_1B_3 + P_2C_3 &= 0, \\ P_0e_0 + P_1e_1 + P_2e_2 &= 1. \end{cases} \tag{3.11}$$

Le fait que A_3 et C_1 sont des matrices dont tous les éléments sont nuls alors le système (3.11) peut être réécrit comme suit :

$$\begin{aligned} P_0A_1 + P_1B_1 &= 0, && (3.12)\\ P_0A_2 + P_1B_2 + P_2C_2 &= 0, && (3.13)\\ P_1B_3 + P_2C_3 &= 0, && (3.14)\\ P_0e_0 + P_1e_1 + P_2e_2 &= 1. && (3.15) \end{aligned}$$

Rappelons que la technique Q-matrice, nous permet d'obtenir la forme de toutes les probabilités d'état stationnaire d'une chaîne de Markov en fonction de la probabilité d'un de ses états particuliers, qui est généralement la probabilité que le système soit vide ou la probabilité que le système soit plein. D'après le système d'équations (3.12)–(3.15), on constate qu'il est préférable d'envisager l'écriture des probabilités stationnaires du système en fonction de la probabilité $P_{1,1}$ plutôt que en fonction d'autres probabilités d'états particuliers du système. En effet, afin de réduire la complexité des calculs (complexité algorithmique) il est claire qu'il vaut mieux d'écrire les probabilités d'état du système en fonction $P_{1,1}$ plutôt qu'avec les probabilités $P_{0,0}$, $P_{0,s}$ $(s = \overline{1,2})$ ou $P_{N,s}$ $(s = \overline{1,3})$.

La résolution du système d'équations (3.12)–(3.15), nous fournis les probabilités d'état du système en régime stationnaire données par le théorème ci-dessous.

Théorème 3.1. *(Afroun et al. [4]) Les probabilities stationnaire du système sont données comme suit :*

$$\begin{aligned} P_{n,0} &= (-\mu_0 a_n)\, P_{1,1}, \ \ pour\ \ 0 \le n \le N, && (3.16)\\ P_{n,1} &= \left(\mu_0\theta \sum_{i=1}^{N} a_{i+1}\tilde{b}_{in}\right) P_{1,1}, \ \ pour\ \ 1 \le n \le N, && (3.17)\\ P_{n,2} &= \left(-\mu_0\theta\eta \sum_{j=1}^{N}\sum_{i=1}^{N} a_{i+1}\tilde{b}_{ij}\tilde{c}_{jn}\right) P_{1,1}, \ \ pour\ \ 1 \le n \le N, && (3.18) \end{aligned}$$

et

$$P_{1,1} = (\Psi_0 + \Psi_1 + \Psi_2)^{-1},$$

où

$$\begin{aligned} \Psi_0 &= -\mu_0 \sum_{n=0}^{N} a_n,\\ \Psi_1 &= \mu_0\theta \sum_{n=1}^{N}\sum_{i=1}^{N} a_{i+1}\tilde{b}_{in},\\ \Psi_2 &= -\mu_0\theta\eta \sum_{n=0}^{N}\sum_{j=1}^{N}\sum_{i=1}^{N} a_{i+1}\tilde{b}_{ij}\tilde{c}_{jn}, \end{aligned}$$

$\tilde{c}_{ij}$ sont les éléments de la matrice $\tilde{C} = C_3^{-1}$, et $\tilde{b}_{ij}$ sont les éléments de la matrice $\tilde{B} = \left(B_2 - \eta\gamma C_3^{-1}\right)^{-1}$.

A_1^{-1} et C_3^{-1} les matrices inverse de A_1 et C_3, respectivement. Soit aussi $a = (a_0, a_1, \cdots, a_N)$ le premier vecteur ligne de la matrice A_1^{-1}.

Démonstration. B_1, A_2, B_3 et C_2 peuvent être réécrites sous forme de matrices en bloc comme suit :

$B_1 = \begin{pmatrix} \mu_0 & O_1 \\ O_2 & O_3 \end{pmatrix}$, $A_2 = \begin{pmatrix} O_1 \\ \theta I_N \end{pmatrix}$, $B_3 = \begin{pmatrix} \eta I_N \end{pmatrix}$, $C_2 = \begin{pmatrix} \gamma I_N \end{pmatrix}$, où O_1 est une matrice d'ordre $1 \times (N+1)$, O_2 est une matrice d'ordre $(N-1) \times 1$, O_3 est une matrice d'ordre $(N-1) \times N$, et I_N est une matrice d'identité d'ordre N. Après quelques manipulations algébriques, on obtient

$$P_0 = -P_1 B_1 A_1^{-1} = -P_1 \begin{pmatrix} \mu_0 a \\ O_4 \end{pmatrix} = -P_{1,1}\mu_0 a, \tag{3.19}$$

$$P_2 = -P_1 B_3 C_3^{-1} = -P_1 \eta C_3^{-1}, \tag{3.20}$$

où O_4 est la matrice d'ordre $(N-1) \times (N+1)$ dont toutes ses élément sont des zéros. Soit $\tilde{a} = (a_1, a_2, \cdots, a_n)$ un vecteur ligne d'ordre N, c'est-à-dire $a = (a_0, \tilde{a})$. En substituant les équations (3.19) et (3.20) dans (3.13), on obtient

$$-P_{1,1}\mu_0\theta\tilde{a} + P_1\left(B_2 - \eta\gamma C_3^{-1}\right) = 0, \tag{3.21}$$

et ceci le fait que

$$P_1 B_1 A_1^{-1} A_2 = P_1 \begin{pmatrix} \mu_0 a_{11}^{-1} & \mu_0\tilde{a} \\ O_2 & O_5 \end{pmatrix} \begin{pmatrix} O_1 \\ \theta I_N \end{pmatrix} = P_{1,1}\mu_0\theta\tilde{a},$$

et

$$P_1 B_3 C_3^{-1} C_2 = \gamma\eta P_1 C_3^{-1},$$

où O_5 est la matrice d'ordre $(N-1) \times (N)$. Comme les matrices B_2 et C_3^{-1} sont des matrices carrées d'ordre N, alors la matrice $\tilde{B} = \left(B_2 - \eta\gamma C_3^{-1}\right)^{-1}$ existe et elle est une matrice carrée d'ordre N. En utilisant l'équation (3.21) on obtient

$$P_1 = \left(\mu_0\theta\tilde{a}\tilde{B}\right) P_{1,1}, \tag{3.22}$$

cela nous permet de déduire que :

$$P_2 = \left(-\mu_0\theta\tilde{a}\tilde{B}B_3 C_3^{-1}\right) P_{1,1}, \tag{3.23}$$

et

$$P_{n,0} = (-\mu_0 a_n)\, P_{1,1}, \tag{3.24}$$

$$P_{n,1} = \left(\mu_0\theta \sum_{i=1}^{N} a_{i+1}\tilde{b}_{in}\right) P_{1,1}, \tag{3.25}$$

$$P_{n,2} = \left(-\mu_0\theta\eta \sum_{j=1}^{N}\sum_{i=1}^{N} a_{i+1}\tilde{b}_{ij}\tilde{c}_{jn}\right) P_{1,1}. \tag{3.26}$$

Afin d'obtenir l'expression de $P_{1,1}$, nous substituons les équations (3.24)–(3.26) dans la condition de normalisation suivante

$$\sum_{n=0}^{N} P_{n,0} + \sum_{n=1}^{N} P_{n,1} + \sum_{n=1}^{N} P_{n,2} \;=\; 1,$$

où nous obtenons,

$$\begin{aligned} 1 \;=\; & \left[-\mu_0 \sum_{n=0}^{N} a_n\right] P_{1,1} \\ & + \left[\mu_0\theta \left(\sum_{n=1}^{N}\sum_{i=1}^{N} a_{i+1}\tilde{b}_{in}\right)\right] P_{1,1} \\ & + \left[-\mu_0\theta\eta \left(\sum_{n=1}^{N}\sum_{j=1}^{N}\sum_{i=1}^{N} a_{i+1}\tilde{b}_{ij}\tilde{c}_{jn}\right)\right] P_{1,1}. \end{aligned}$$

Ainsi, la formule de la probabilité $P_{1,1}$ est donnée par :

$$P_{1,1} \;=\; \left(\Psi_0 + \Psi_1 + \Psi_2\right)^{-1},$$

où

$$\begin{aligned} \Psi_0 &= -\mu_0 \sum_{n=0}^{N} a_n, \\ \Psi_1 &= \mu_0\theta \sum_{n=1}^{N}\sum_{i=1}^{N} a_{i+1}\tilde{b}_{in}, \\ \Psi_2 &= -\mu_0\theta\eta \sum_{n=0}^{N}\sum_{j=1}^{N}\sum_{i=1}^{N} a_{i+1}\tilde{b}_{ij}\tilde{c}_{jn}, \end{aligned}$$

Fin de la démonstration. □

3.4 Mesures de performance du système

Dans cette section, nous nous présentons quelques expressions des mesures de performance, plus usuelles, du système décrit dans la section précédente :

- P_B : La probabilité que le système soit en période d'activité.
- P_V : La probabilité que le système soit en période de vacances.
- P_R : La probabilité que le système soit en période de réparation.
  Qui sont obtenus comme suit :

$$P_V = \sum_{n=0}^{N} P_{n,0} = \Psi_0 P_{1,1}, \tag{3.27}$$

$$P_B = \sum_{n=1}^{N} P_{n,1} = \Psi_1 P_{1,1}, \tag{3.28}$$

$$P_R = \sum_{n=1}^{N} P_{n,2} = \Psi_2 P_{1,1}, \tag{3.29}$$

où $\Psi_i,\ i = \overline{1,3}$ sont définis dans le Théorème 3.1.

- Une autre performance intéressante que l'on peut déduire en utilisant les probabilités ci-dessus est la fiabilité du système. En effet, la fiabilité du système est donnée par :

$$R = 1 - P_R = 1 - (P_V + P_B). \tag{3.30}$$

- Le nombre moyen de clients dans le système (L_s) et le nombre moyen de clients dans la file d'attente (L_q) sont donnés, respectivement, par :

$$L_s = \sum_{n=1}^{N} n\left(P_{n,0} + P_{n,1} + P_{n,2}\right), \tag{3.31}$$

$$L_q = \sum_{n=1}^{N} nP_{n,0} + \sum_{n=2}^{N}(n-1)\left(P_{n,1} + P_{n,2}\right). \tag{3.32}$$

- Le taux moyen de joindre le système λ_e, appelé aussi taux d'arrivée effectif, est donné par :

$$\begin{aligned}\lambda_e &= \sum_{n=0}^{N} \lambda_n P_{n,0} + \sum_{n=1}^{N} \lambda_n P_{n,1} + \sum_{n=1}^{N} \lambda_n P_{n,2}, \\ &= \lambda P_{0,0} + \sum_{n=1}^{N} \lambda_n \left(P_{n,0} + P_{n,1} + P_{n,2}\right).\end{aligned} \tag{3.33}$$

- Le taux moyen du balking B_r est :

$$B_r = \lambda - \lambda_e. \tag{3.34}$$

- La charge du système est définie par :

$$\rho = \frac{\lambda_e}{\beta\mu R + L_q \alpha\xi}. \tag{3.35}$$

- En utilisant les formules de Little, on peut obtenir d'autres mesures de performance comme le temps moyen de séjour dans le système W_s et le temps moyen d'attente dans la file W_q qui sont donnés par :

$$W_s = L_s/\lambda_e, \tag{3.36}$$

$$W_q = L_q/\lambda_e. \tag{3.37}$$

- Le taux moyen de Reneging des clients $(R_{reneging})$ et le taux moyen de Rétention des clients $(R_{retention})$ sont donnés respectivement par :

$$\begin{aligned} R_{reneging} &= \sum_{n=1}^{N} n\xi\alpha P_{n,0} + \sum_{n=1}^{N}(n-1)\xi\alpha P_{n,1} + \sum_{n=1}^{N}(n-1)\xi\alpha P_{n,2}, \\ &= \xi\alpha\left(\sum_{n=1}^{N} nP_{n,0} + \sum_{n=1}^{N}(n-1)P_{n,1} + \sum_{n=1}^{N}(n-1)P_{n,2}\right), \\ &= \xi\alpha L_q, \end{aligned} \tag{3.38}$$

et

$$\begin{aligned} R_{retention} &= \sum_{n=1}^{N} n\xi\alpha' P_{n,0} + \sum_{n=1}^{N}(n-1)\xi\alpha' P_{n,1} + \sum_{n=1}^{N}(n-1)\xi\alpha' P_{n,2}, \\ &= \xi\alpha' L_q. \end{aligned} \tag{3.39}$$

3.5 Illustrations numériques

L'objectif principal de cette section est de vérifier numériquement l'efficacité des résultats analytiques, obtenus dans les sections précédentes, et d'analyser l'impact des paramètres de fiabilité sur les mesures de performance de notre système. Pour ce faire, nous avons conçu un programme sous l'environnement Matlab qui nous visualise deux types de résultats : des résultats numériques et des résultats graphiques.

En pratique, le décideur peut parfois intervenir pour modifier un ou plusieurs paramètres de contrôle du système, tels que le taux de service, la capacité du système, le nombre de serveurs, ... Afin de prendre en compte de telles situations dans notre application numérique, nous proposons les deux scénarios suivants :

Premier scénario : Le contrôle du taux de service μ.

Deuxième scénario : Le contrôle de la capacité du système N.

De plus, nous avons supposé que :

1. Les périodes de vacances sont des maintenances préventives.
2. Les périodes de réparations sont des maintenances correctives
3. Le fait que le comportement de certaines caractéristiques puisse être déduit du comportement des autres, nous nous sommes limités à l'analyse des principales caractéristiques seulement.

3.5.1 Premiers scénario : variation du taux de service

Afin d'étudier l'effet de la variation de μ, sur les caractéristiques de notre système, nous avons fixé : $\lambda = 3$, $\alpha = 0.8$, $\xi = 1$, $\beta = 0.6$, $N = 5$ et une probabilité de balking linéaire $\delta' = \frac{n}{N}$, et nous avons considéré les cas suivants :

Cas 1. $\mu = [2\ ;\ 3\ ;\ 4\ ;\ 5],\ \eta = 0.01 : 0.01 : 5,\ \gamma = 0.5$, et $\theta = 2$.

Cas 2. $\mu = [2\ ;\ 3\ ;\ 4\ ;\ 5],\ \eta = 0.1,\ \gamma = 0.01 : 0.01 : 5$, et $\theta = 2$.

Cas 3. $\mu = [2\ ;\ 3\ ;\ 4\ ;\ 5],\ \eta = 0.1,\ \gamma = 0.5$, et $\theta = 0.01 : 0.01 : 5$.

Les Figures 3.3-3.7, sont un échantillon des résultats graphiques, obtenus par l'exécution de notre programme, pour les paramètres précédents correspondants au premier scénario. Les résultats illustrent le lien étroit entre les caractéristiques du système et les paramètres de départ qui le régissent où on constate que :

La fiabilité du système R :

- La fiabilité du système est inversement proportionnelle au temps moyen de service($1/\mu$). En effet, on peut voir sur la Figure 3.3, que la diminution du temps de service engendre une augmentation de la fiabilité du système. Ceci peut s'expliquer par le fait que si le temps de service est plus petit alors le nombre de clients servis augmente, ce qui signifie que le système a tendance à être vide (passage à la période de vacances) et comme on a considéré que les pannes sont actives alors le système ne peut pas être en panne à cette période de vacances, d'où l'augmentation de R.
- À mesure que le taux de défaillances η augmente, le système a tendance à être toujours en période de panne, ce qui entraîne une diminution de la fiabilité R du système (voir Figure 3.3.(a)).
- L'augmentation du taux de réparation γ signifie que le système, lorsqu'il est en panne sera réparé rapidement, par conséquent il revient rapidement à l'état actif, ce qui entraîne une réduction de la durée d'indisponibilité du système, d'où l'augmentation de la fiabilité du système (voir Figure 3.3.(b)).
- Pour de petites valeurs du taux de vacances θ, la fiabilité du système est élevée, cela est dû au temps considérable que le système prend en état de vacances (vacances de longues durées) donc il n'y a pas de pannes dans cette période à cause du type des pannes considérées. En revanche, si le taux θ augmente, le système a une probabilité considérable de passer de l'état de vacances à l'état actif s'il y a une arrivée, et d'être sujet à des pannes actives, d'où la réduction de sa fiabilité (voir Figure 3.3.(c)).

Le nombre moyen de clients dans le système L_s :

- Le nombre moyen de clients dans le système, L_s, est proportionnel au temps moyen de service $1/\mu$ (voir la Figure 3.4). Ce comportement peut être interprété par : la faible durée du service permet de réduire le nombre de clients en attente, car le serveur les serves rapidement et ils quittent le système, ce qui engendre la réduction de L_s.
- Le nombre L_s est proportionnel au taux de pannes η, ceci est le fait que, lorsque le taux η augmente, le système a tendance à être en panne (ce qui a été expliqué précédemment dans le cas de la Figure 3.3.(a)). Par conséquent, le nombre de clients s'accumule dans le système dans l'espoir d'être servi, d'où l'augmentation de L_s.
- Le nombre L_s est inversement proportionnel au taux de réparation γ (voir Figure 3.4.(b)). Le fait que le système retourne rapidement de l'état de réparation à l'état de service, le nombre de clients servis augmente cela conduit à la diminution de L_s.

- Le nombre L_s est considérable pour les petites valeurs du taux, θ, (voir Figure 3.4.(c)). Ceci peut s'expliquer par le fait que lorsque le système prend de vacances de longues durées, le taux de départ des clients diminue car il n'y a pas de service. Cependant, lorsque le taux θ est assez grand, c'est-à-dire que les vacances sont de petites durées, le taux de départ augmente (il y a un service) ce qui entraîne la diminution de L_s.

Le nombre moyen de clients dans la file d'attente L_q :

- Pour de petite valeur de θ ($\theta \in [0; 0.5]$ dans notre cas), signifie que le système a tendance à prendre de longues périodes de vacances avant de revenir à l'état actif. Pour ce fait, si le temps de service moyen $(1/\mu)$ est petit alors les clients en attente seront servis rapidement ce qui conduit le système à se vider. Par conséquent, prendre des vacances avec un temps considérable, d'où l'augmentation à nouveau du nombre de clients en attente (voir Figure 3.16.(c)).
- En revanche, si la durée moyenne de service est importante, alors la durée moyenne de la période d'activité sera longue, ce qui allonge la date de début des vacances. C'est la raison pour laquelle l'augmentation du nombre moyen de clients en attente n'est pas considérable comme dans le cas de petites valeurs des durées de service. En résumé, L_q est inversement proportionnel au temps de service moyen $(1/\mu)$ seulement si θ est grand ($\theta > 0.5$ dans notre cas).
- Le comportement du nombre moyen de clients en attente L_q en fonction des paramètres de fiabilité du système ($\eta,\ \gamma$ et θ) reste le même que celui de L_s.

Il est à souligner que le comportement du taux $R_{reneging}$ et $R_{retention}$ par rapport au temps de service moyen $(1/\mu)$, le taux de panne η, le taux de réparation γ et le taux de vacances θ, est similaire au comportement du nombre moyen de clients dans le système, L_q qu'on vient de discuter et ceci peut être justifier par le lien linéaire des deux taux moyens $R_{reneging}$ et $R_{retention}$ avec L_q exposé dans les formules (3.38) et (3.39).

Le taux moyen de rejoindre le système λ_e et le taux moyen de balking B_r :

Pour la probabilité de Balking retenu dans notre application numérique, l'expression de λ_e et B_r peuvent être simplifier comme suit :

$$\lambda_e = \lambda - \frac{\lambda}{N} L_s, \tag{3.40}$$

$$B_r = \frac{\lambda}{N} L_s. \tag{3.41}$$

Ainsi, d'après ces deux dernières expressions, on peut conclure que :

- le comportement B_r en fonction des paramètres fiabiliste du système, que nous faisons variés, doit être exactement le même que le comportement du nombre moyen de client dans le système L_s en fonction de ces paramètres.
- le comportement λ_e en fonction des paramètres fiabiliste du système, que nous faisons variés, doit être le contraire du comportement du nombre moyen de client dans le système L_s en fonction de ces paramètres.

Il est à noter que nos résultats numérique et graphique, obtenus dans l'application numérique, coïncident parfaitement avec ces deux précédentes remarques (voir Figures 3.4 et 3.6).

Les probabilités de l'état du système P_B, P_V et $R = 1 - P_R$

- Les probabilités P_B, P_V, R sont inversement proportionnelles au taux de pannes η. Ceci peut être s'expliqué par le fait que l'augmentation de η entraîne des pannes répétitives du système. Par conséquent, le système sera en permanence en état de réparation plutôt que en vacance ou en état d'activité (voir Figure 3.7.(a)).
- Les probabilités P_B, P_V et R sont proportionnelles au taux de réparation γ, car lors d'une éventuelle panne du système, ce dernier sera remis rapidement en état de marche vu que le taux de réparation est considérable (voir Figure 3.7.(b)).
- Les probabilités P_v et R sont inversement proportionnelles au taux de vacances θ. Ceci, peut être expliqué par le fait que l'augmentation du taux de vacances (c'est-à-dire les durées des vacances sont petites), garanti que le système sois en permanence en état de service ce qui justifier l'augmentation de P_B. Par conséquence, le système est plus susceptible à tomber en panne (pannes actives) et d'être en état de réparation d'où la diminution de R et l'augmentation de P_R. Donc il est toute évident que P_v sera décroissante en fonction de θ, car $P_v = 1 - (P_B + P_R)$ (voir Figure 3.7.(c)).

3.5.2 Deuxième scénario : variation de la capacité du système

Afin d'étudier l'effet de la variation de la capacité du système, N, sur ses caractéristiques, nous avons fixé ce qui suit : $\lambda = 3$, $\alpha = 0.8$, $\xi = 1$, $\beta = 0.6$, $\mu = 4$ et une probabilité de balking linéaire $\delta' = \frac{n}{N}$. De plus, nous avons considéré les trois cas suivants :

Case 1. $N = [\,4\,;\,6\,;\,8\,;\,10]$, $\eta = 0.01 : 0.01 : 5$, $\gamma = 0.5$, et $\theta = 2$.

Case 2. $N = [\,4\,;\,6\,;\,8\,;\,10]$, $\eta = 0.1$, $\gamma = 0.01 : 0.01 : 5$, et $\theta = 2$.

Case 3. $N = [\,4\,;\,6\,;\,8\,;\,10]$, $\eta = 0.1$, $\gamma = 0.5$, et $\theta = 0.01 : 0.01 : 5$.

Les Figures 3.8-3.11 représentent un échantillon des résultats graphiques, obtenus par l'exécution de notre programme, pour les paramètres correspondants au deuxième scénario. Les résultats obtenus misent en évidence l'effet de la capacité du système sur les mesures de performance de ce dernier. De plus, ces résultats nous laissent conclure que l'effet de la variation de la capacité du système sur ces mesures est similaire aux cas de variation de taux moyen de service μ (voir Section 3.5.1).

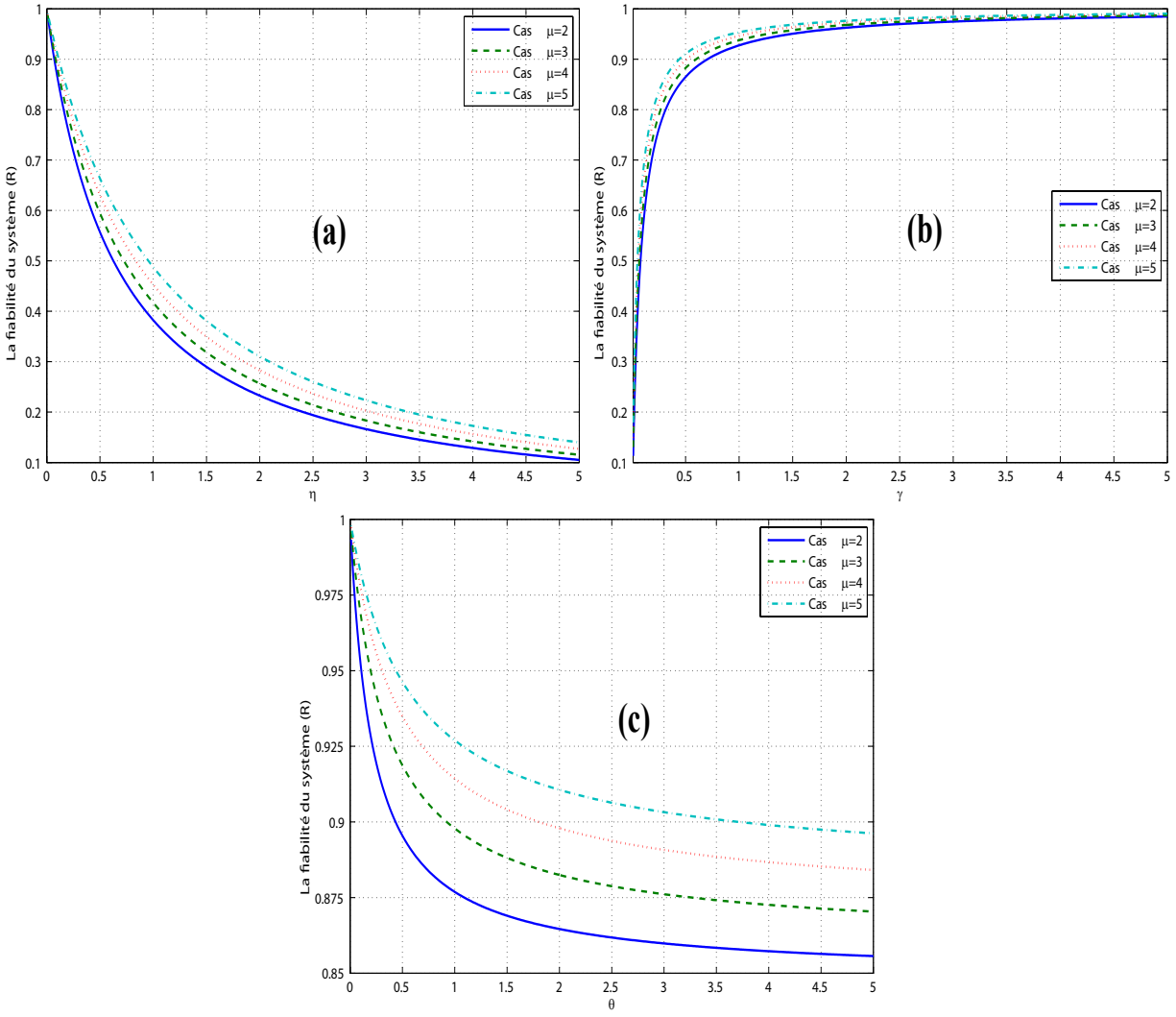

FIGURE 3.3 – Variation de R par rapport aux paramètres fiabilistes $\eta,\ \gamma$ et θ.

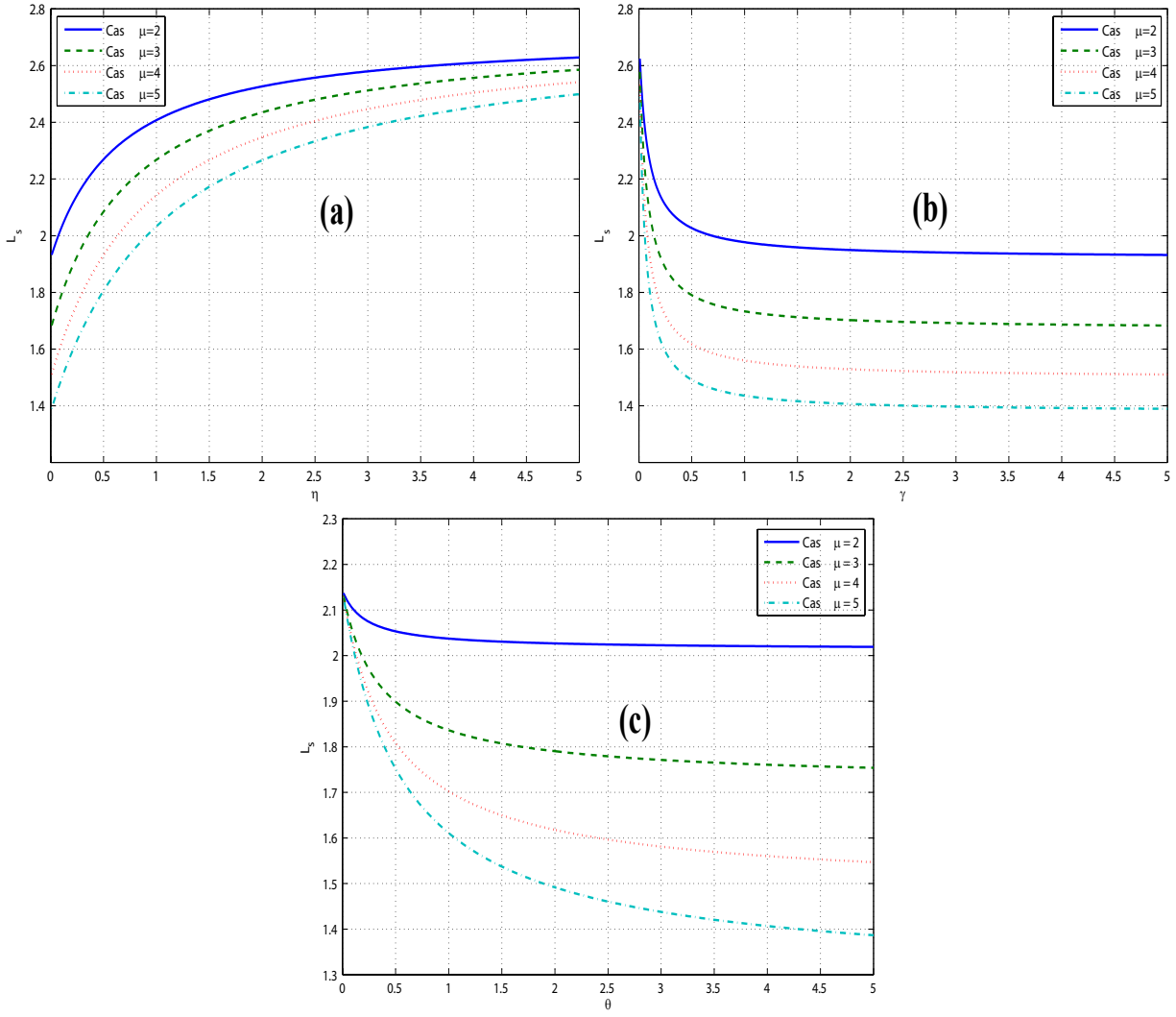

FIGURE 3.4 – Variation de L_s par rapport aux paramètres fiabilistes $\eta,\ \gamma$ et θ.

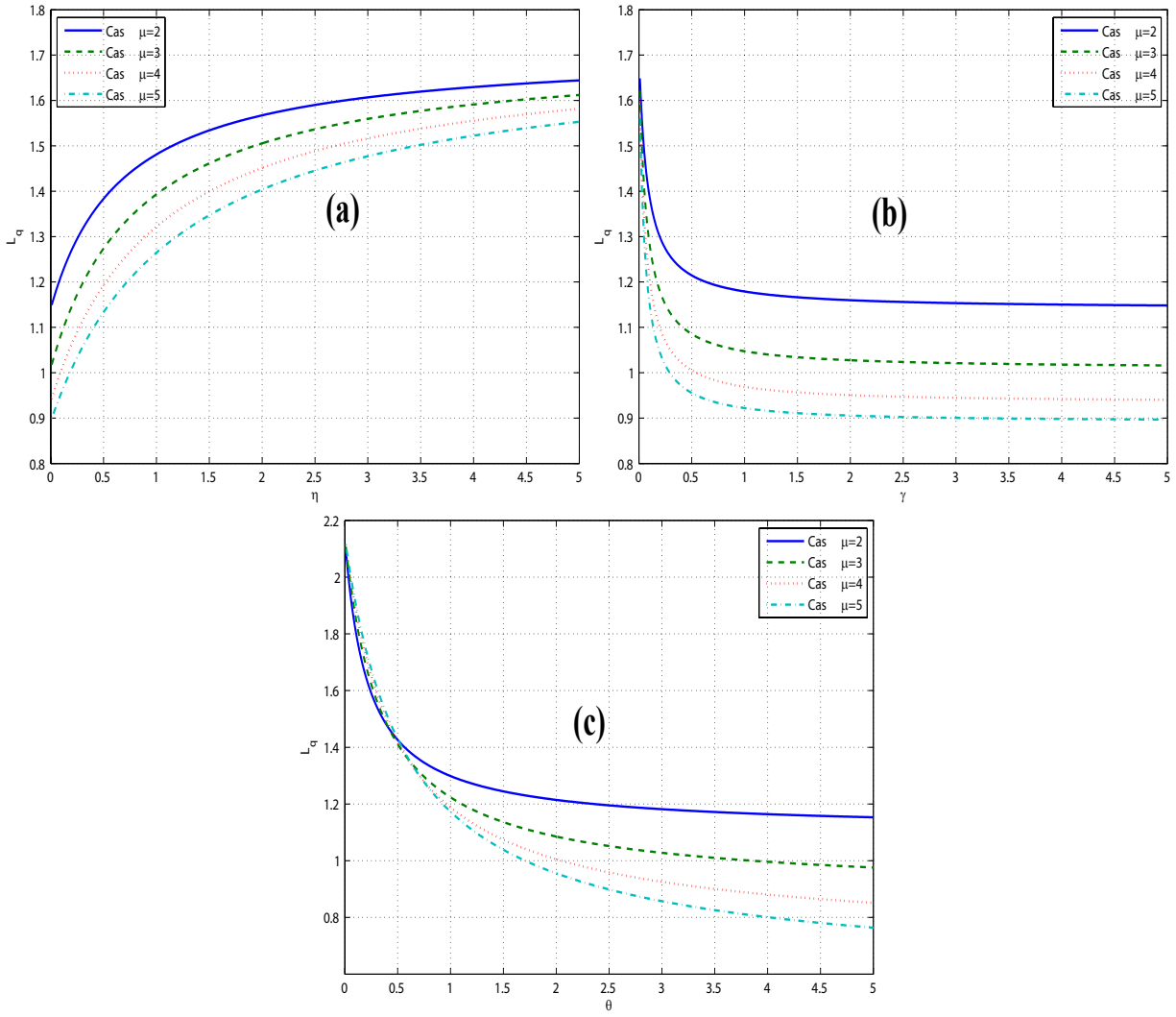

FIGURE 3.5 – Variation de L_q par rapport aux paramètres fiabilistes $\eta,\ \gamma$ et θ.

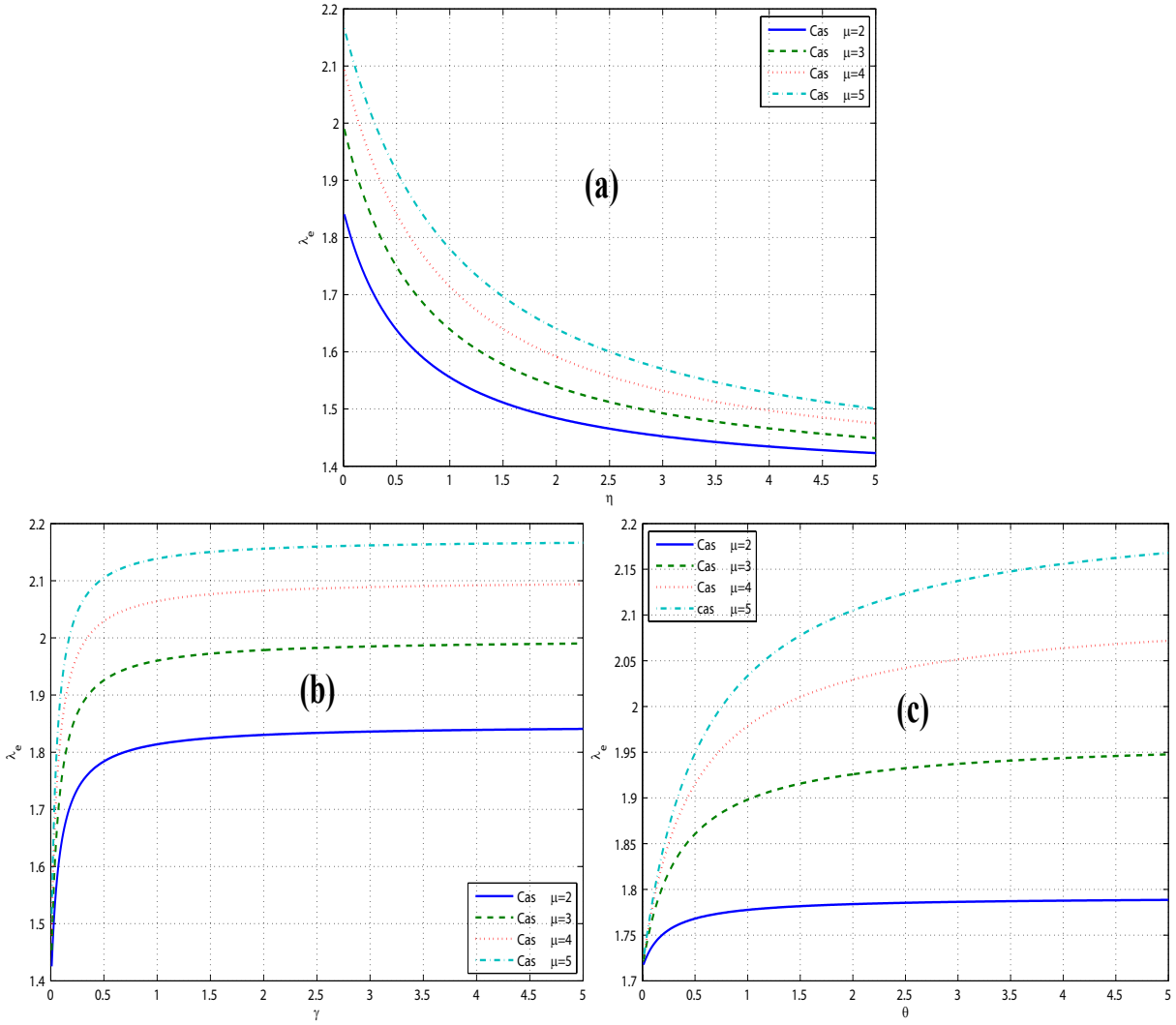

FIGURE 3.6 – Variation de λ_e par rapport aux paramètres fiabilistes $\eta,\ \gamma$ et θ.

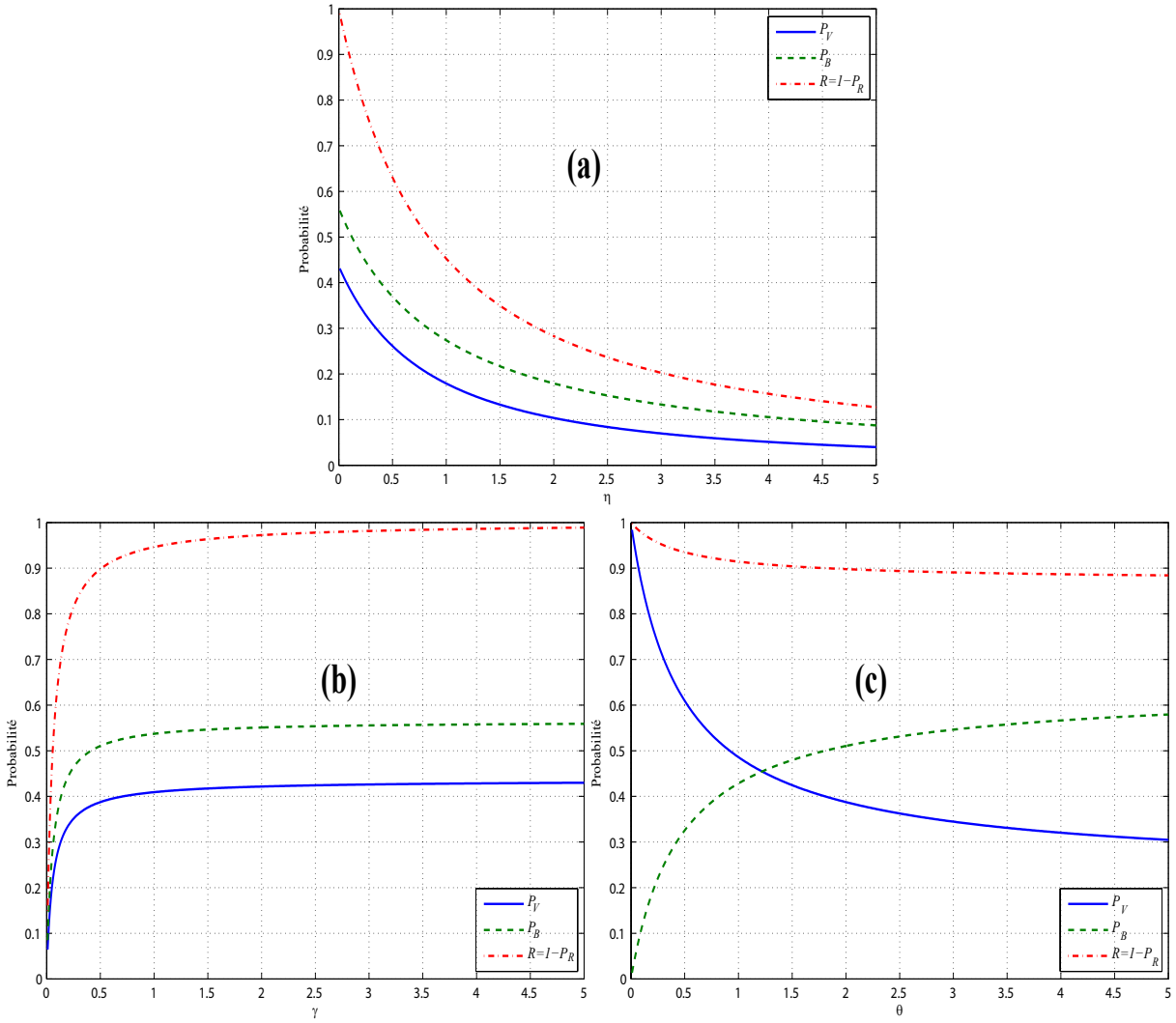

FIGURE 3.7 – Variation de P_V, P_B et R par rapport aux paramètres fiabilistes η, γ et θ.

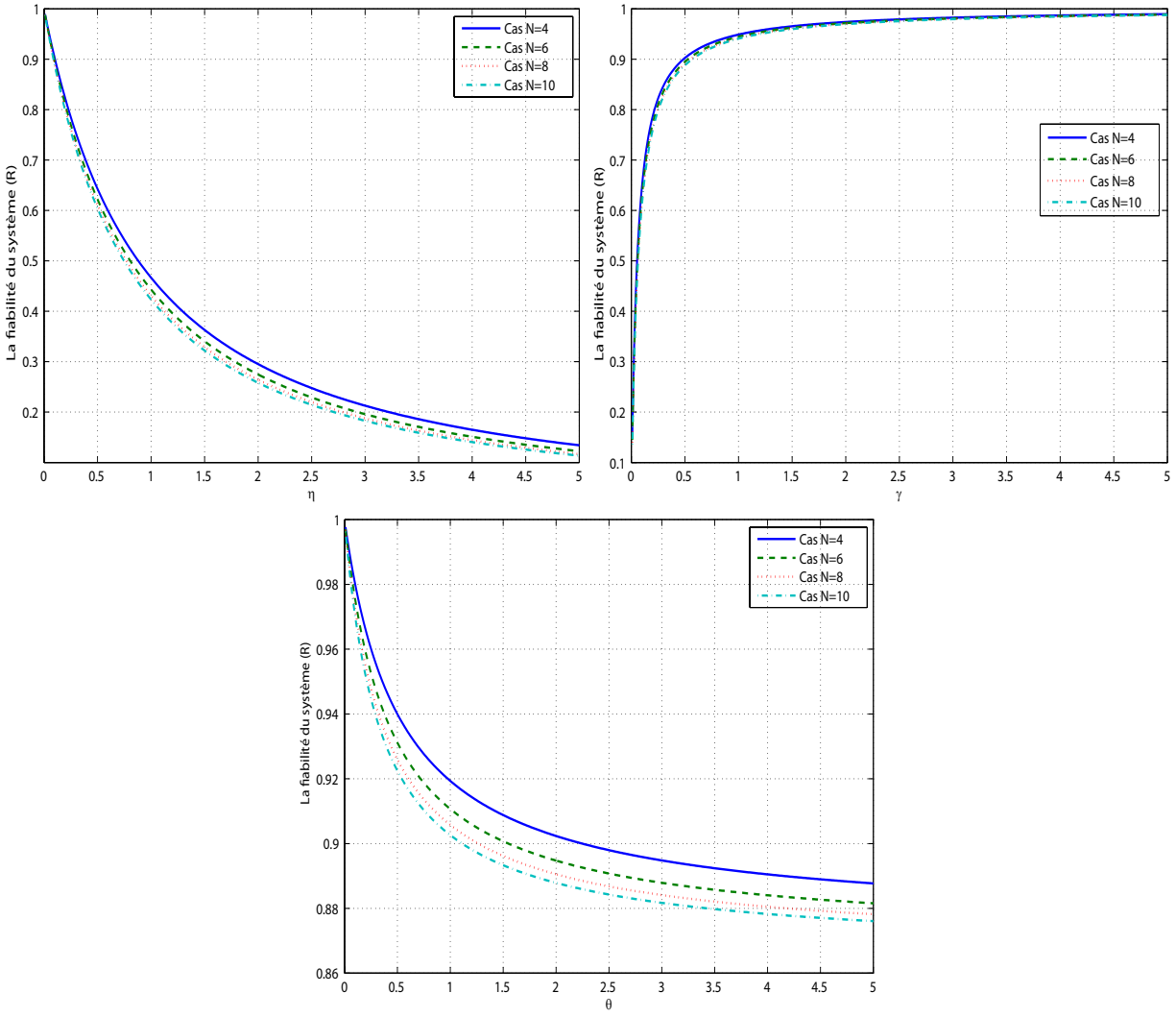

FIGURE 3.8 – Variation de R par rapport aux paramètres fiabilistes $\eta,\ \gamma$ et θ.

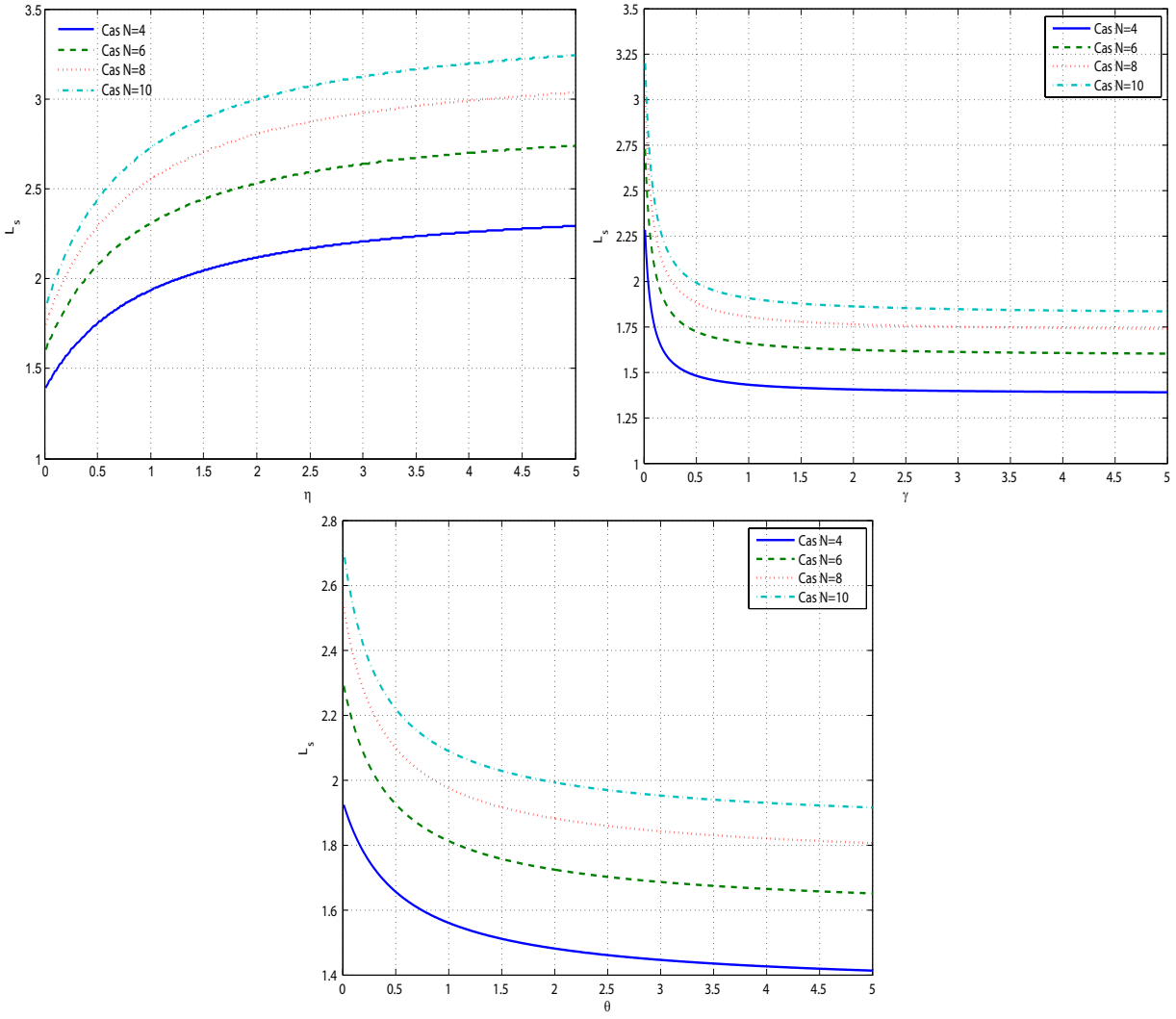

FIGURE 3.9 – Variation de L_s par rapport aux paramètres fiabilistes $\eta,\ \gamma$ et θ.

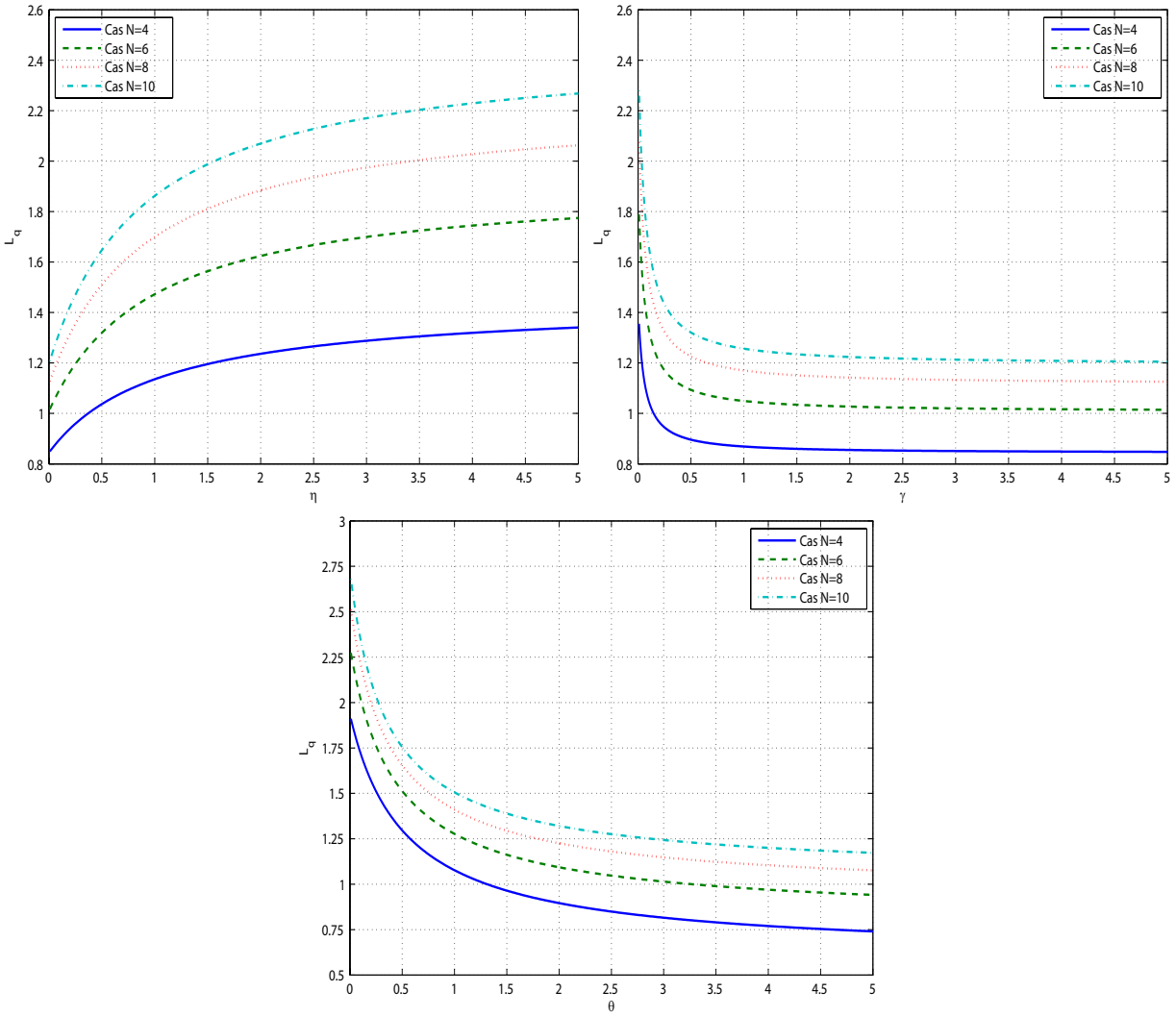

FIGURE 3.10 – Variation de L_q par rapport aux paramètres fiabilistes $\eta,\ \gamma$ et θ.

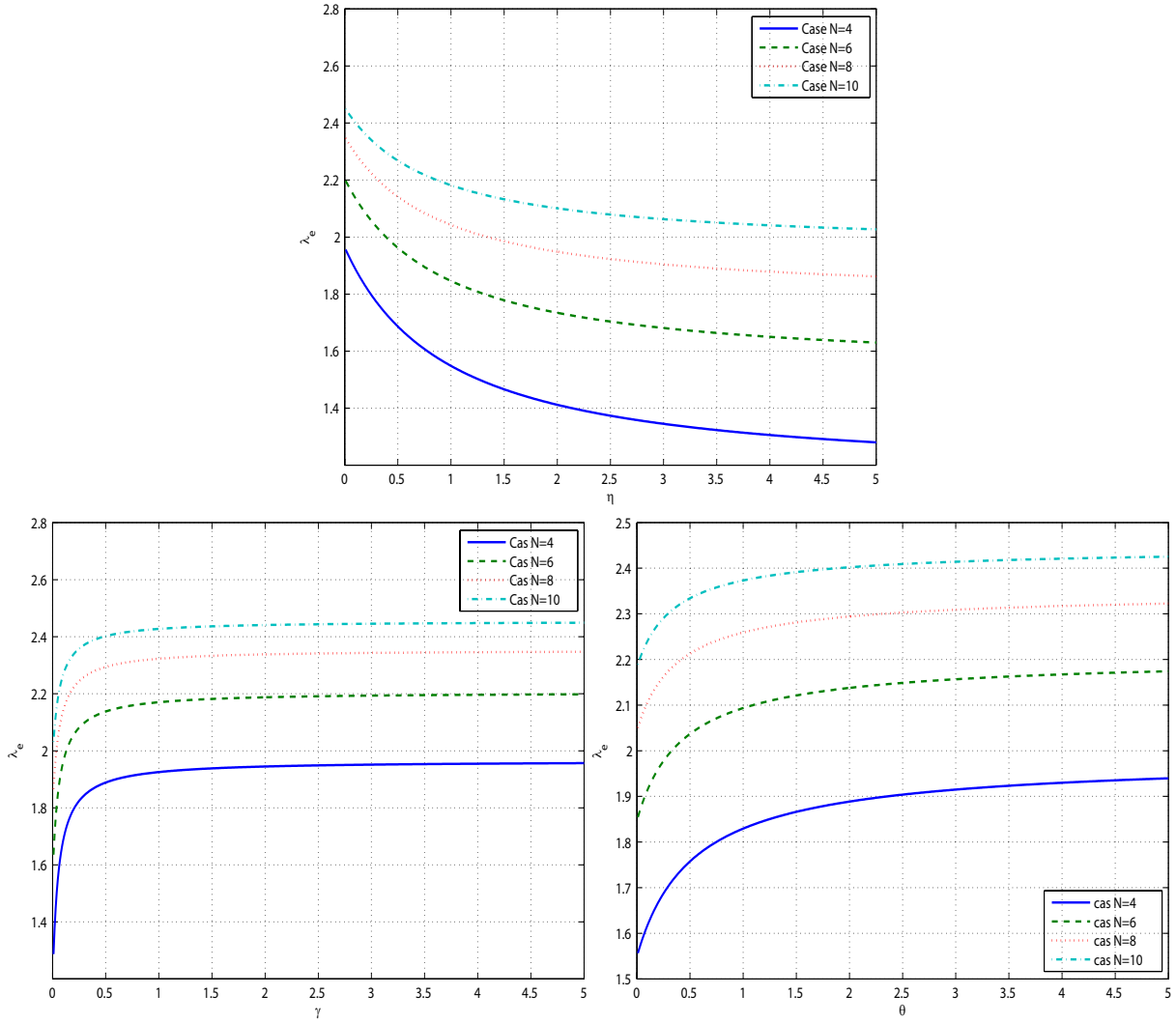

FIGURE 3.11 – Variation de λ_e par rapport aux paramètres fiabilistes $\eta,\ \gamma$ et θ.

3.6 Propriétés statistiques des performances de la file d'attente $M/M/1/N$

Dans cette section, nous considérons le même système d'attente décrit précédemment dans la Section 3.2, et nous supposons que les paramètres de départ (taux d'arrivée λ, taux de service μ, taux d'impatience ξ, taux de panne γ, taux de vacances θ, taux de réparation η) du système ne sont connus qu'à partir d'un échantillon d'observations. Dans une telle situation, il est clair que la substitution de ces paramètres par leurs estimateurs dans les expressions des mesures de performance du système ne nous fournira que des estimations pour ces mesures de performance (principe des méthode plug-in).

Notre idée est d'analyser, par simulation, l'impact de l'estimation des paramètres de départ décrivant le modèle en question sur les propriétés statistiques (distribution, biais, variance et erreur quadratique) des estimateurs de ses mesures de performance.

Certes, on peut considérer qu'un ensemble de paramètres de départ (> 1), définissant le modèle considéré, sont simultanément inconnus (se présente sous la forme d'un ensemble d'échantillons d'observation). Cependant, lors de l'analyse et de la discussion des résultats qu'on obtiendra dans cette étude, si l'estimation de ces paramètres a un impact considérable sur les propriétés statistiques des estimateurs de mesures de performance, alors on ne peut prédire avec précision quel paramètre de départ est le plus influent. C'est pourquoi dans l'étude numérique nous avons fixé tous les paramètres sauf un, qui est sous forme d'un échantillon. Pour atteindre notre objectif, nous avons eu recours à plusieurs techniques statistiques, à savoir : l'estimation paramétrique, les tests de conformité et la présentation graphique en Box plot et en histogramme.

Pour répondre à notre objectif nous avons implémenté un programme sous Matlab dont la principale tâche est d'exécuter les étapes suivantes.

Étape 0. Fixer le paramètre de départ à simuler.

Étape 1. Générer mc échantillons de taille n d'une loi exponentielle du paramètre de départ supposé inconnu.

Étape 2. Estimer le paramètre inconnu pour chacun des mc échantillons.

Étape 3. Calculer les caractéristiques stationnaires du système à l'aide de l'estimateur obtenu à l'étape 2.

Étape 4. Présentez les propriétés statistiques (numériques et/ou graphiques) des caractéristiques obtenues en 3.

Pour l'application numérique, nous fixons les différents paramètres associés au système d'attente considéré comme suit : $\lambda = 6$, $\mu = 8$, $\alpha = 0.8$, $\xi = 4$, $\beta = 0.6$, $N = 5$, $\eta = 1,\ \gamma = 2,$ et $\theta = 5$ et une probabilité de balking linéaire $\delta' = \frac{n}{N}$.

3.6.1 Test de conformité d'une moyenne

L'objectif de ce passage est de vérifier si la valeur moyenne d'une caractéristique du système quantifiée à base de l'estimateur d'un paramètre de départ, coïncide (au sens statistique) avec la valeur recherchée. Cette vérification se fait à l'aide du test de conformité d'une moyenne.

En effet, le test de conformité, au sens générale, a pour but de vérifier si un échantillon peut être considéré comme un extrait d'une population donnée ou représentatif de cette population, par rapport à un paramètre ou une caractéristique bien déterminé comme la moyenne, la variance,...

Ainsi, pour confirmer un résultat théorique, il suffit de simuler mc échantillons de taille n et de comparer les résultats obtenus avec les résultats théoriques correspondants.

Soit m_0 la valeur théorique d'une caractéristique de notre système (R, L_q,...) et m la moyenne de cette même caractéristique fournie par notre simulateur dont la quantification de la ième observation se fait à base de l'estimateur, du paramètre de départ supposé inconnu, évalué à travers le ième échantillon. Le problème revient donc à tester l'hypothèse suivante :

$$\text{"}H_0 : m = m_0\text{" contre "}H_1 : m \neq m_0\text{"}.$$

A la fin de la simulation, on obtient un estimateur de m d'une moyenne empirique $\hat{m}$ et de variance empirique S^2. La confirmation des résultats se fait par le test bilatéral sur la statistique

$$T_{mc} = \frac{\hat{m} - m_0}{S}\sqrt{mc - 1}.$$

Ceci implique que pour prendre une décision sur le rejet ou le non rejet de l'hypothèse H_0 la connaissance de la loi de cette statistique est nécessaire. Malheureusement, dans notre cas nous ne disposons d'aucune information sur la distribution de T_{mc} et sa démonstration est très difficile, voire impossible vue la complexité des expressions des caractéristiques (voir la section 3.4). Afin de surpasser ce problème, la solution est de générer un nombre mc suffisamment grand d'échantillons et de considérer que T_{mc} suit une loi normale centrée et réduite qu'on justifiera par le théorème centrale limite.

Ainsi, la région critique du test, pour un seuil de risque α, est donnée par :

$$|T_{mc}| > \Phi^{-1}(1 - \alpha/2),$$

avec $\Phi^{-1}(1 - \alpha/2)$ est le quantile d'ordre $1 - \alpha/2$ d'une loi normale centrée et réduite.

Des exemples de résultats obtenus sur 100 $(mc = 100)$ échantillons de taille $n = 200$ sont résumés dans la Table 3.1 où IC représente l'intervalle de confiance de la moyenne de la caractéristique considérée à un seuil $1 - \alpha = 95\%$, t_{mc} est la réalisation de la statistique T_{mc} et p est la probabilité de signification du test donnée par $p = P(T_{mc} > t_{mc})$ avec T_{mc} est une variable aléatoire qui suit une loi normale centrée et réduite.

		λ	μ	ξ	η	γ	θ
$P_{1.1}$	IC	[0.183 ; 0.185]	[0.183 ; 0.185]	[0.184 ; 0.186]	[0.183 ; 0.185]	[0.184 ; 0.185]	[0.184 ; 0.1860]
	t_{mc}	-1.39	-1.388	0.370	-1.251	-0.905	0.644
	p	0.171	0.171	0.713	0.217	0.370	0.522
P_V	IC	[0.381 ; 0.394]	[0.387 ; 0.399]	[0.387 ; 0.393]	[0.386 ; 0.390]	[0.387 ; 0.390]	[0.386 ; 0.3910]
	t_{mc}	-0.566	1.127	0.682	-1.251	-0.905	-0.584
	p	0.574	0.265	0.498	0.217	0.370	0.562
P_B	IC	[0.404 ; 0.412]	[0.401 ; 0.409]	[0.405 ; 0.409]	[0.404 ; 0.408]	[0.405 ; 0.408]	[0.406 ; 0.4094]
	t_{mc}	0.566	-1.127	-0.682	-1.257	-0.920	0.584
	p	0.574	0.265	0.498	0.215	0.362	0.562
L_q	IC	[0.744 ; 0.772]	[0.748 ; 0.755]	[0.743 ; 0.759]	[0.752 ; 0.755]	[0.752 ; 0.755]	[0.750 ; 0.7546]
	t_{mc}	0.737	-0.944	-0.534	1.259	0.925	-0.584
	p	0.465	0.350	0.596	0.214	0.359	0.562
W_q	IC	[0.172 ; 0.173]	[0.171 ; 0.173]	[0.170 ; 0.174]	[0.172 ; 0.173]	[0.172 ; 0.173]	[0.172 ; 0.1729]
	t_{mc}	0.631	-0.927	-0.469	1.277	0.941	-0.584
	p	0.531	0.358	0.642	0.208	0.351	0.562
R_{ret}	IC	[0.595 ; 0.617]	[0.599 ; 0.604]	[0.598 ; 0.610]	[0.602 ; 0.604]	[0.602 ; 0.604]	[0.600 ; 0.6037]
	t_{mc}	0.737	-0.944	0.619	1.259	0.925	-0.584
	p	0.465	0.350	0.539	0.214	0.359	0.562

TABLE 3.1: Résultats du Test de conformité de moyenne.

D'après les résultats obtenus par le test de conformité d'une moyenne rangés dans la Table 3.1, on remarque que si on fixe le seuil de risque $\alpha = 5\%$, alors dans la totalité des cas considérés on ne rejette pas l'hypothèse H_0 et ceci le fait que $t_{mc} < \Phi^{-1}(1-\alpha/2) = \Phi^{-1}(0.975) = 1.96$ $(\alpha < p)$.

Ce constat signifie que, dans toutes les situations considérées, la moyenne de la caractéristique recherché est égale significativement à la valeur exacte (théorique) de cette caractéristique et ceci quel que soit le paramètre de départ estimé. Ce constat signifie également que les estimateurs des caractéristiques du système sont, au moins, asymptotiquement sans biais.

L'autre propriété au quelle on s'intéresse lors de l'analyse de la qualité d'un estimateur est bien sa variation (variance), ceci fait l'objet de la Section suivante.

3.6.2 Variation des estimateurs des caractéristiques (box plot)

Afin d'analyser la variation des estimations des caractéristiques stationnaires du système d'attente en question nous allons faire recours à la technique Box plot. Nous avons opté pour cette technique car elle nous permet la représentation de la position de la variable en faisant apparaître sur un graphique la médiane, les quartiles et les valeurs adjacentes. Elle donne une bonne impression de la forme de la variable entre les valeurs adjacentes. On peut ainsi juger la symétrie en étudiant la position de la médiane dans le corps de la boîte de la position des valeurs adjacentes. L'étendue de la boîte, la longueur des pattes, et la mise en évidence de valeurs aberrantes apportent des renseignements sur la dispersion. De plus, lors de la comparaison de plusieurs variables, elle nous permet de voir rapidement comment évoluent la position, la dispersion et la forme de la variable quantitative selon les modalités de la variable qualitative, par l'étude des positions relatives des boîtes, de leur taille et des observations aberrantes.

Un échantillon des résultats graphique fournis par la technique Box plot, sur 100 ($mc = 100$) échantillons de taille $n = 200$, pour les valeurs des paramètres fixées précédemment est présenté dans les Figures 3.12–3.16.

Les diagrammes donnés dans les Figures 3.12–3.16, nous donnent les tendances globales de la variation des estimateurs de quelques caractéristiques stationnaires du système et elles apportent des réponses graphiques sur la façon dont l'estimation des paramètres du départ influent sur les caractéristiques du système considéré où on constate que :

• Quel que soit le paramètre de départ estimé, les distributions des estimateurs des caractéristiques, du système d'attente retenu dans l'étude, sont pratiquement symétriques.

• Pour une même caractéristique, la valeur médiane de ses estimations est pratiquement la même quel que soit le paramètre de départ estimé.

• La distribution de l'estimateur $\hat{R}$ est beaucoup plus dispersée et étendue dans le cas d'estimation du taux de pannes η et le taux de réparation γ. Les meilleures estimations de la fiabilité R du système, au sens les moins dispersées, sont obtenues dans le cas d'utilisation des estimations du taux des vacances θ et du taux d'impatience ξ où la variation de $\hat{R}$ est très faible. Tandis qu'une situation intermédiaire est obtenue dans le cas où nous avons utilisé les estimations du

taux d'arrivées λ et du taux de service μ. Les représentations graphiques montrent également que l'estimateur de la fiabilité du système R peut prendre des valeurs aberrantes, du côté inférieur, dans le cas de l'estimation du taux d'arrivées λ et du taux de pannes η (voir Figure 3.12 à gauche).

• L'impact de l'estimation de chacun des paramètres de départ est pratiquement le même sur la distribution de l'estimateur $\hat{P}_{1,1}$. En effet, on constate que les présentations graphiques des box plot sont similaires dans les six situations considérées. Les représentations graphiques montrent également que l'estimateur de la probabilité $P_{1,1}$ peut prendre des valeurs aberrantes, du côté inférieur, dans le cas de l'estimation du taux d'arrivées λ et le taux de pannes η (voir Figure 3.12 à droite).

• La distribution de l'estimateur $\hat{L}_s$ du nombre moyen de clients dans le système est beaucoup plus étendue et dispersée dans le cas d'utilisation de l'estimateur du paramètre λ pour la quantification de cette caractéristique, suivi par le cas de l'estimation des paramètres μ et ξ. Les estimateurs $\hat{L}_s$ les plus performants, au sens de la variation, sont obtenus dans le cas d'estimation du paramètre θ où la distribution de l'estimateur est aplatie (très faible dispersion). Des situations médianes sont obtenues lorsque nous utilisons les estimateurs du taux de pannes et du taux de réparation pour estimer L_s. La représentation graphique montre également que l'estimateur du nombre moyen de clients dans le système L_s peut prendre des valeurs aberrantes en excès dans le cas de l'estimation du taux d'arrivée λ et le taux η (voir Figures 3.13 à gauche). Ces discussions restent valables sur l'effet de l'estimation des paramètres de départ sur la variation (distribution) de l'estimateur du taux de Balking B_r, mais ce dernier est légèrement moins variable que dans le cas d'estimation de L_s (voir Figure 3.14 à gauche).

• La distribution de l'estimateur du nombre moyen de clients en attente L_q est beaucoup plus étendue et dispersée dans le cas de l'estimation du paramètre λ suivi par le cas de l'estimation de ξ. Les meilleurs résultats ont obtenu dans le cas d'estimation des paramètres fiabiliste γ et η où la variation de l'estimateur $\hat{L}_q$ est presque aplatie (très faible dispersion). Tandis que des situations intermédiaires sont obtenues dans le cas d'utilisation des estimateurs de μ et θ. La représentation graphique montre également que l'estimateur du L_q peut prendre des valeurs aberrantes en excès dans le cas de l'estimation du taux d'arrivée λ (voir Figures 3.14 à droite). Il est à noter que, des résultats similaire, que ces derniers, sont constatés aussi dans le cas d'estimation du taux moyen $R_{renenging}$ et $R_{retention}$ (voir Figures 3.13 à droite).

• La distribution de l'estimateur W_s du temps moyen de séjour d'un client dans le système est beaucoup plus étendue et dispersée dans le cas d'estimation du taux d'impatience ξ suivi par le cas d'estimation du paramètre μ. Les meilleurs résultats, au critère retenu, sont obtenus dans le cas de substitution du paramètre λ ou du paramètre θ par leurs estimateur dans l'expression de W_s, où la distribution de l'estimateur $\hat{W}_s$, dans ces deux situations, est presque aplaties (très faible dispersion). Des estimateurs de qualité intermédiaire sont obtenus dans le cas d'estimation du paramètre γ ou du paramètre η. Mais dans le cas de l'estimation de ce dernier paramètre, il est à souligner que $\hat{W}_s$ peut prendre des valeurs aberrantes en excès (voir Figure 3.15 à gauche).

• La distribution de l'estimateur du temps moyen d'attente d'un client dans la file, W_q, est beaucoup plus étendue et dispersée dans le cas d'estimation du paramètre ξ suivi par le cas d'estimation de μ. L'estimation du reste des paramètres (λ, γ, η et θ) nous fournis des estimateurs de W_q ayant pratiquement la même qualité au sens de la dispersion. Cependant, dans le cas d'estimation du paramètre λ ou du paramètre η, il s'avère que $\hat{W}_q$ peut prendre des valeurs aberrantes en excès (voir Figure 3.15 à droite).

• La distribution de l'estimateur de la probabilité P_B est beaucoup plus étendu et dispersé dans le cas d'estimation des paramètres μ et λ. Le reste des situations ($\theta,\ \eta,\ \gamma$ et ξ) nous fournis des estimateur de P_B ayant pratiquement les mêmes variation. La représentation graphique montre également que l'estimateur de la probabilité P_B peut prendre des valeurs aberrantes en excès dans le cas de l'estimation du taux d'arrivées λ et peut prendre des valeurs aberrantes en valeur inférieur lorsque le taux d'arrivée des pannes η est substitué par son estimateur (voir Figure 3.16 à gauche).

• La distribution de l'estimateur de la probabilité P_v est beaucoup plus étendue et dispersée dans le cas d'estimation du paramètre μ suivi par le cas de l'estimation du paramètre λ. Les estimateurs les moins dispersés sont obtenues dans le cas d'estimation des paramètres fiabiliste du système, c'est-à-dire lorsque le paramètre γ ou le paramètre η est remplacé par son estimateur dans l'expression de la probabilité P_v. Tandis que des situations intermédiaires sont obtenus dans le cas d'utilisation des estimateurs de ξ et θ. La représentation graphique montre également que l'estimateur de la probabilité P_v peut prendre des valeurs aberrantes, du côté inférieur, dans le cas de l'estimation du taux d'arrivées λ et du taux d'arrivées des pannes η (voir figure 3.16 à droite).

En guise de conclusion, les résultats obtenus indiquent d'une part que d'une manière générale c'est l'estimation du taux d'arrivée λ et le taux d'impatience ξ qui nous fournis des estimations moins performantes, au sens de la variation, des caractéristiques stationnaires du système d'attente considéré dans l'étude. En effet, l'estimation de ces deux paramètres nous fournis des estimations ayant une grande variation voir mêmes des estimations aberrantes. D'autre part, quel que soit le paramètre de départ estimé, les distributions des estimateurs des caractéristiques du système d'attente retenu dans l'étude sont relativement symétriques. De plus, la valeur médiane des estimations d'une même caractéristique est pratiquement la même.

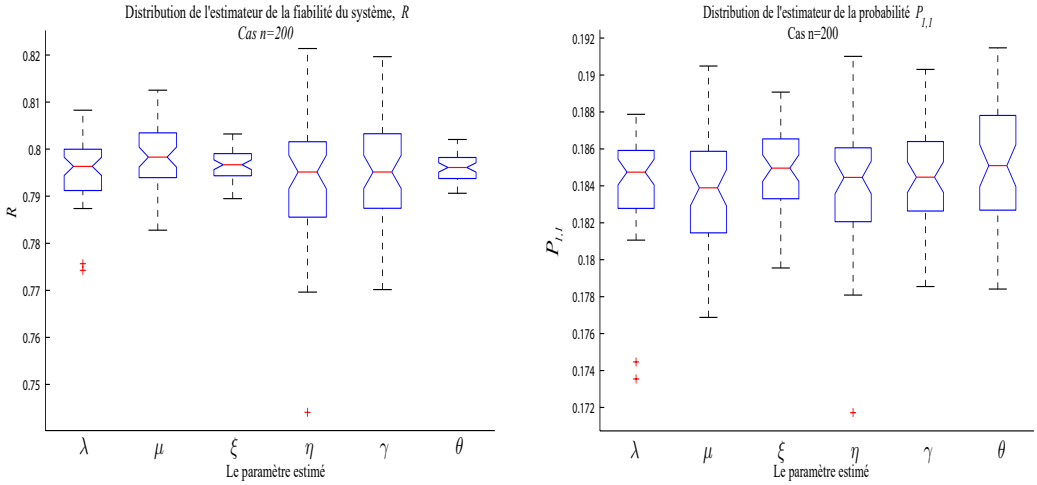

FIGURE 3.12 – Variation (Box plot) des estimateurs de la fiabilité du système R (à gauche) et de la probabilité $P_{1,1}$ (à droite), en fonction du paramètre de départ estimé.

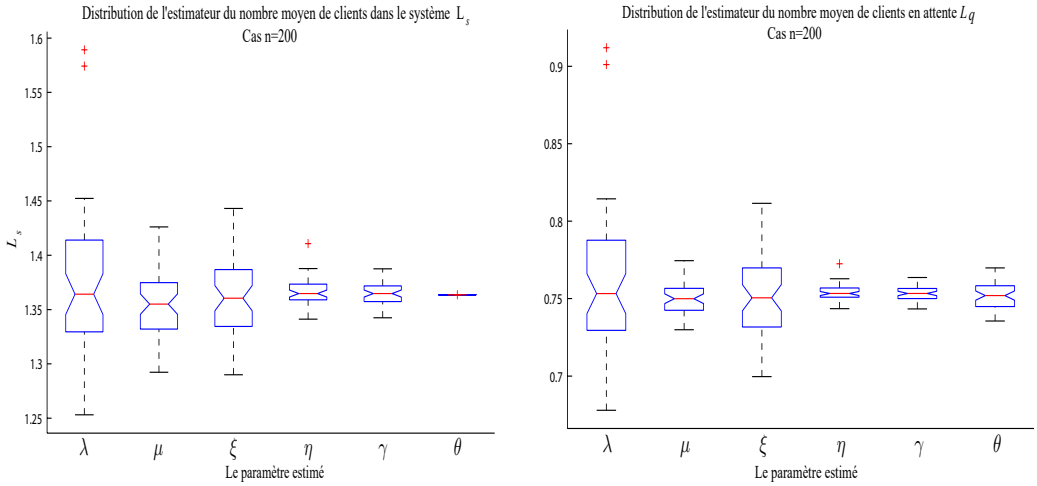

FIGURE 3.13 – Variation (Box plot) des estimateurs du nombre moyen de clients dans le système L_s (à gauche) et du moyen moyen de clients en attente L_q (à droite), en fonction du paramètre de départ estimé.

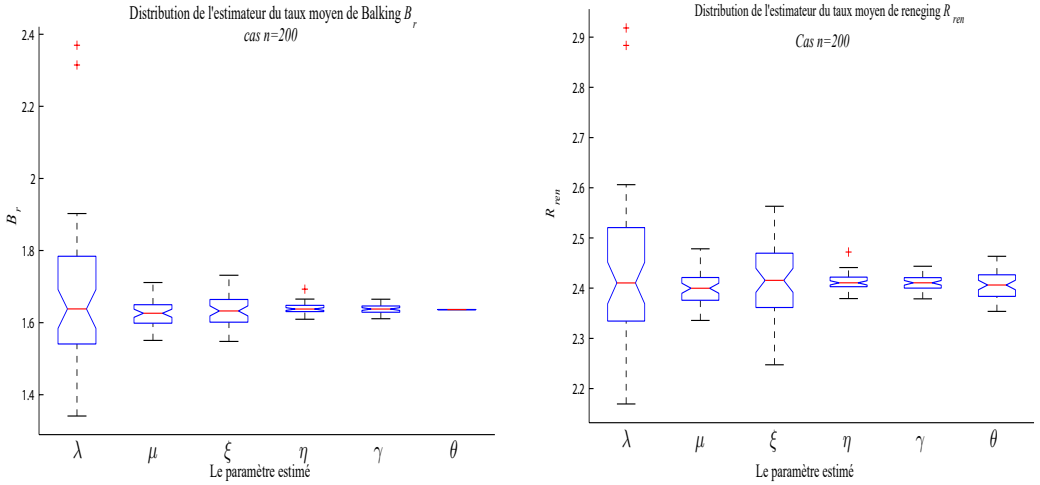

FIGURE 3.14 – Variation (Box plot) des estimateurs du taux moyen de balking B_r (à gauche) et le taux moyen de renenging $R_{reneging}$ (à droite), en fonction du paramètre de départ estimé.

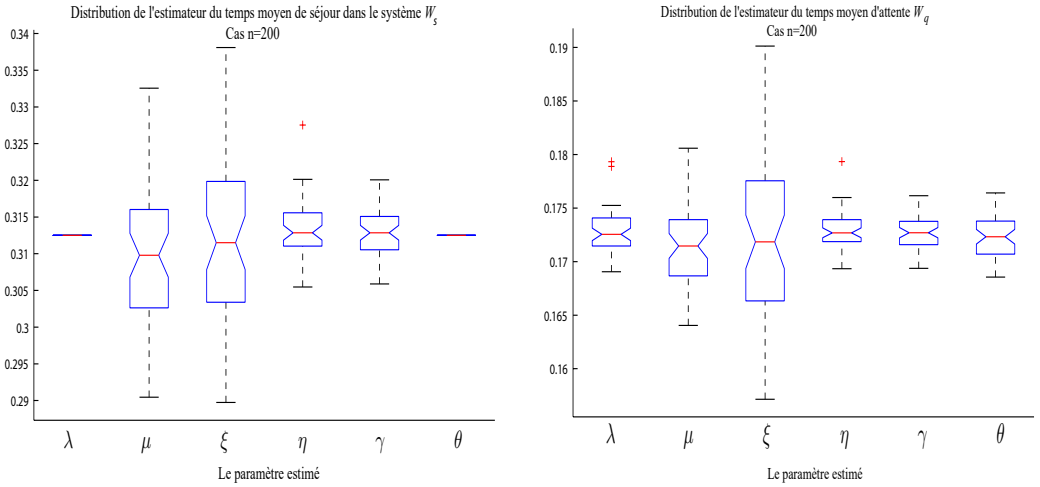

FIGURE 3.15 – Variation (Box plot) des estimateurs du temps moyen de séjour dans le système W_s (à gauche) et du temps moyen d'attente W_q (à droite) en fonction du paramètre de départ estimé.

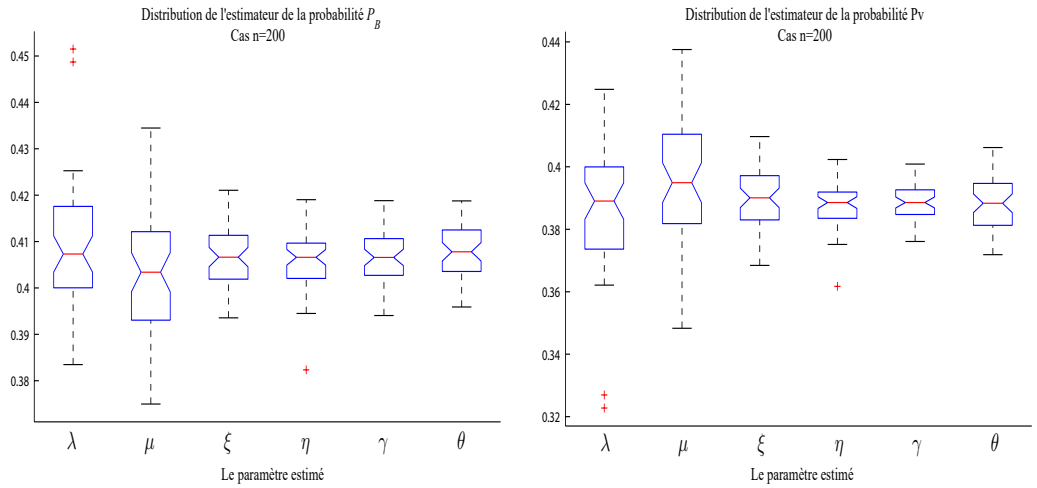

FIGURE 3.16 – Variation (Box plot) des estimateurs de la probabilité P_B (à gauche) et de la probabilité P_V (à droite), en fonction du paramètre de départ estimé.

3.6.3 Comportement asymptotique des estimateurs des caractéristiques

L'objectif de la présente partie est d'analyser le comportement du biais, de la variance et du MSE (l'Erreur Quadratique Moyenne) des estimations des mesures de performance du système en fonction du paramètre de départ estimé et de la taille de l'échantillon et afin de conclure sur leurs convergences. Pour cela, nous reprenons la même démarche que dans les deux études précédentes mais cette fois-ci nous faisons varier la taille de l'échantillon. Les résultats obtenus sont présentés dans les Figures 3.17–3.20.

Les résultats obtenus dans cette application montrent que :

– Tous les estimateurs des caractéristiques stationnaire du système d'attente considéré convergent en biais, en variance et en MSE quel que soit le paramètre de départ estimé. En effet, les graphiques montrent clairement que :

$$\lim_{n\to\infty} Biais^2(\hat{m}) = \lim_{n\to\infty} Var(\hat{m}) = \lim_{n\to\infty} MSE(\hat{m}) = 0.$$

– Pour toutes les tailles d'échantillon considérées, le comportement de la variance des estimateurs des caractéristiques du système en fonction du paramètre de départ estimé est identique aux résultats fournis par la technique box plot (voir Figures 3.12–3.16 et leurs discussions dans la section 3.6.2).

– Même pour des échantillons de petites taille ; le comportement, voir mêmes les valeurs, des variances et des MSE associées à l'estimateur d'une même caractéristique sont pratiquement identiques ($Var \simeq MSE$) et cela quel que soit le paramètre de départ estimé. En effet, dans toutes les situations considérées le carré du biais ($biais^2$) des caractéristiques est négligeable par rapport à leurs variances. Ceci signifié que les estimateurs des caractéristiques, qu'on quantifies via les estimateurs des paramètres de départ, sont des estimateurs sans biais.

– L'impact de l'estimation de chacun des paramètres de départ sur le biais des caractéristiques du système est le même que sur leurs variances. C'est-à-dire, le biais se comporte d'une manière similaire que la variance en fonction du paramètre de départ estimé.

– En générale, l'estimation des paramètres de départ qui a un impacts considérable (la plus influente) sur les biais, la variance et le MSE des estimateurs des caractéristique du système est bien que celle du paramètre λ et du paramètre ξ, ce qui coïncide avec les résultats obtenus précédemment.

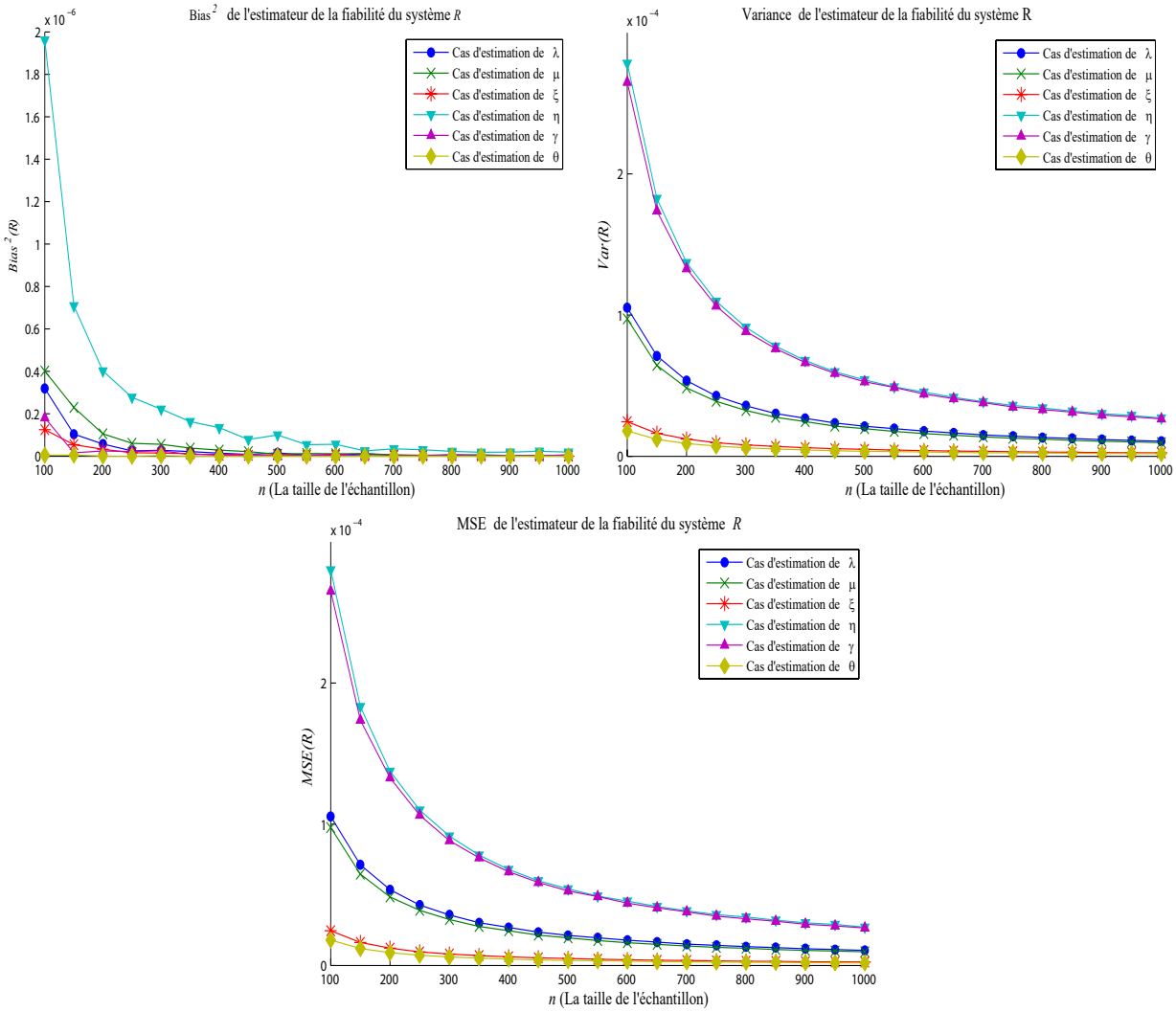

FIGURE 3.17 – Variation du biais^2, de la variance et du MSE de l'estimateur de R, en fonction de la taille de l'échantillon n.

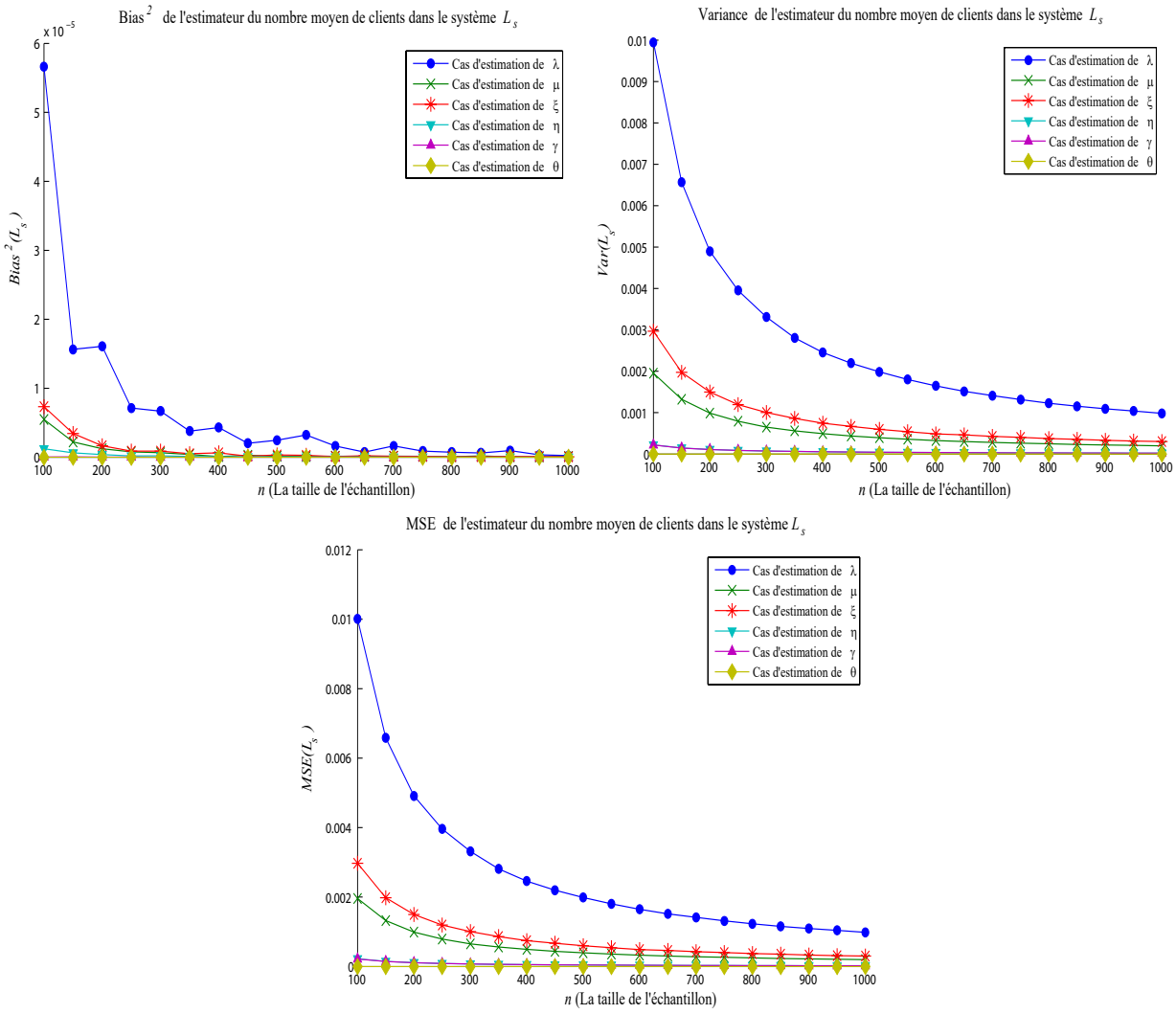

FIGURE 3.18 – Variation du biais^2, de la variance et du MSE de l'estimateur de L_s, en fonction de la taille de l'échantillon n.

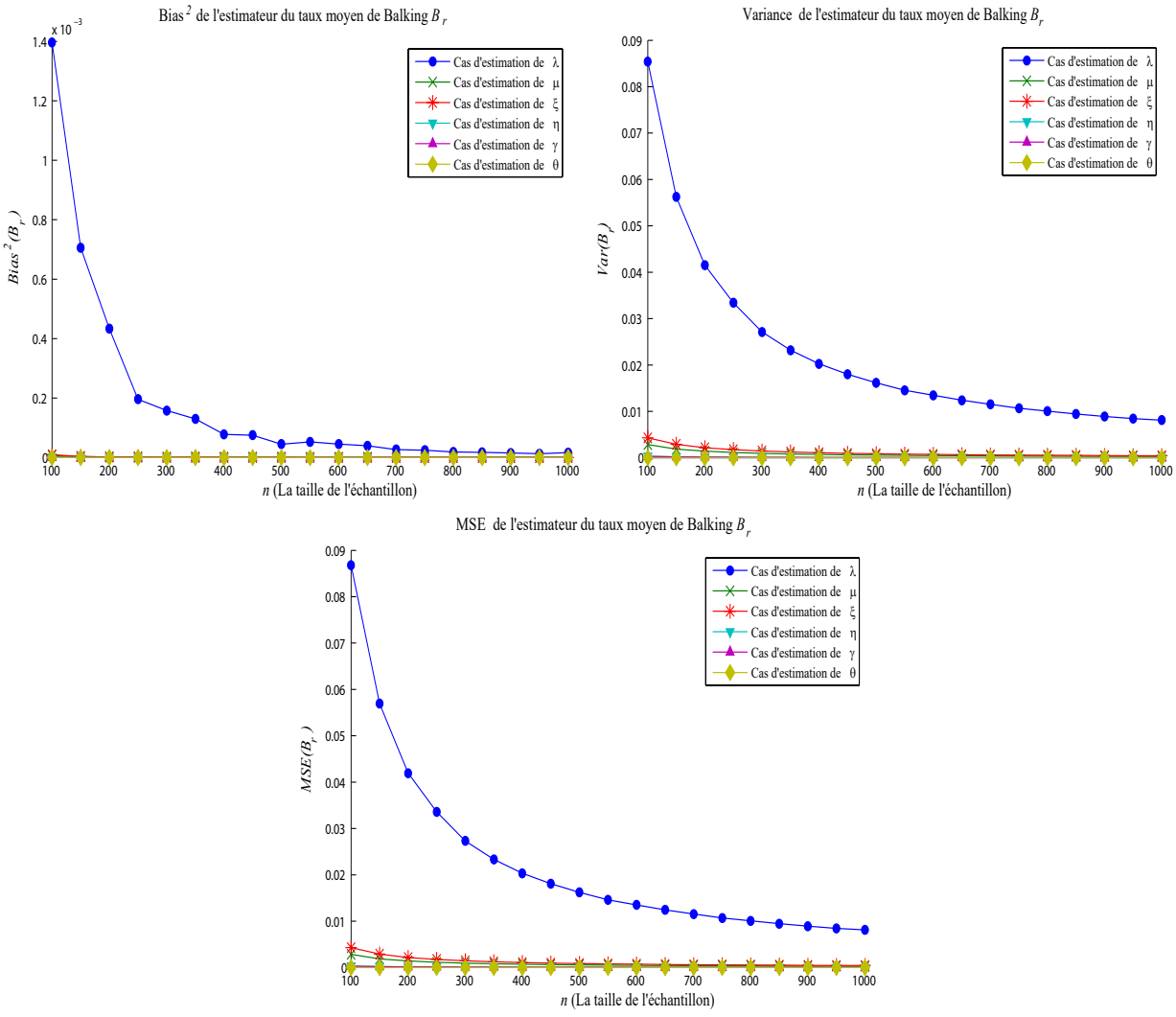

FIGURE 3.19 – Variation du biais^2, de la variance et du MSE de l'estimateur de B_r, en fonction de la taille de l'échantillon n

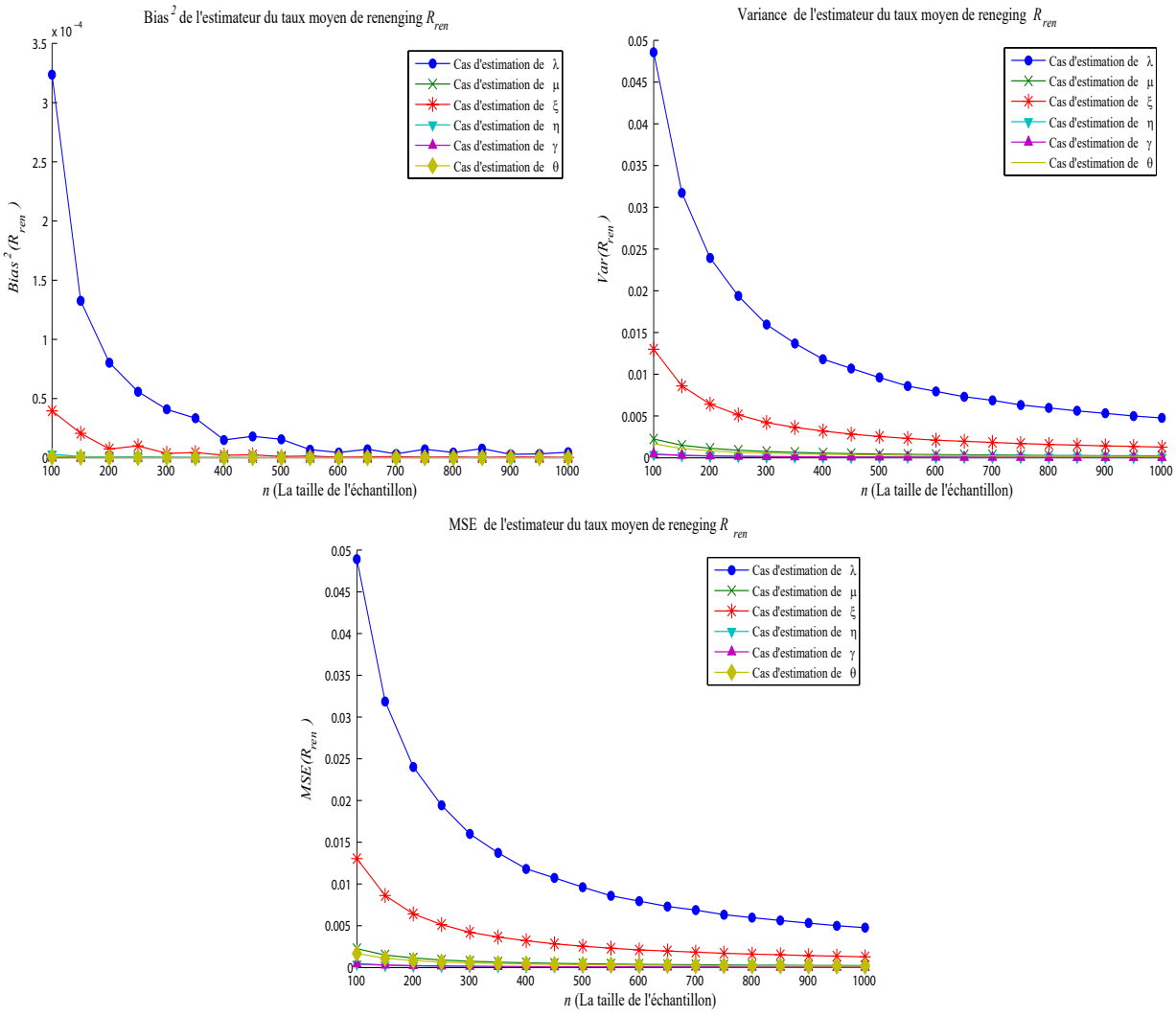

FIGURE 3.20 – Variation du biais^2, de la variance et de l'MSE de l'estimateur de $R_{reneging}$, en fonction de la taille de l'échantillon n.

3.6.4 Convergence en loi des estimateurs des caractéristiques

Dans ce passage, nous allons montrer, par simulation, que les estimateurs des caractéristiques du système convergent en loi vers une loi normale et ceci quel que soit le paramètre de départ estimé. Certes on peut justifier la convergence en question en faisant recours au théorème central limite. Mais dans notre cas, nous allons voir que les résultats de simulation obtenus nous permettent de prédire la normalité asymptotique des estimations des caractéristiques du système, sans faire recours à la théorie des grands nombres.

Généralement lors de la démonstration qu'un estimateur converge en loi, ce n'est que pour faciliter la tâche, qu'on fait recours au théorème centrale limite. Dans notre cas, indépendamment de ce théorème on peut prédire, pour n suffisamment grand, que l'estimateur $\hat{m}$ de la caractéristique m du système suit une loi normale de moyenne m (valeur exacte de la caractéristique) et de variance $var(\hat{m})$. En effet, les différents résultats obtenus précédemment et ceux des Figures 3.21–3.22 et de la table 3.2, indiquent que :

1. la distribution des estimateurs des caractéristiques est symétrique (voir section 3.6.2 et Figures 3.21–3.22),
2. la répartition des quantiles de la distribution des estimateurs des caractéristiques ressemble à la répartition de ceux d'une loi normale (voir Figures 3.12–3.16),
3. la distribution des estimateurs des caractéristiques est unimodale (voir Figures 3.21–3.22),
4. l'allure des polygones des histogrammes présentés dans les Figures 3.21–3.22 ont la forme d'une cloche,
5. la moyenne, la médiane et le mode des estimateurs d'une même caractéristique sont pratiquement identiques : $mode(\hat{m}) = mediane(\hat{m}) = moyenne(\hat{m})$ (voir Table 3.2).

Ces cinq derniers constats ne sont que les caractéristiques d'une loi normale. De plus, d'après le test de conformité réaliser dans la section 3.6.1 on a $E(\hat{m}) = m$, par conséquent

$$\hat{m} \rightsquigarrow \mathcal{N}(m, var(\hat{m})).$$

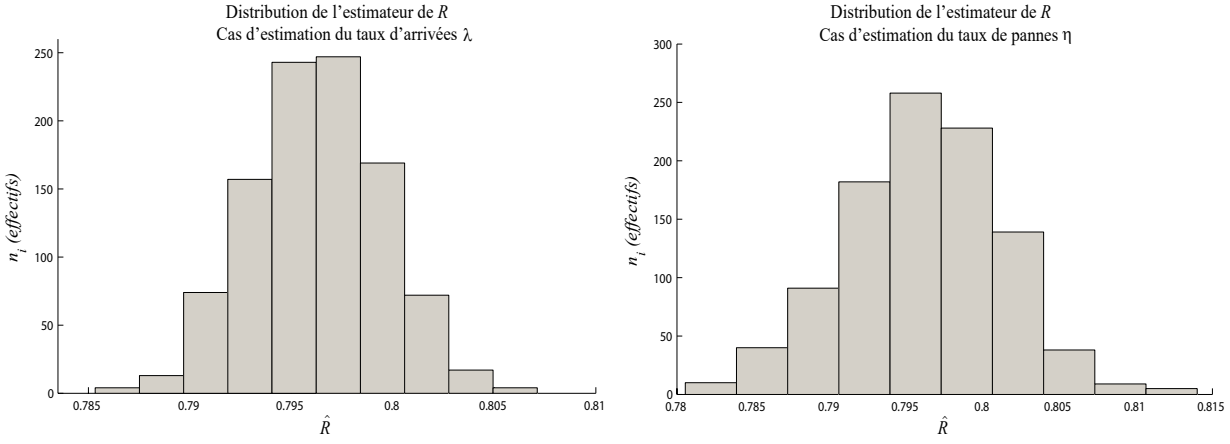

FIGURE 3.21 – Distribution de l'estimateur de la fiabilité du système, en fonction du paramètre de départ estimé.

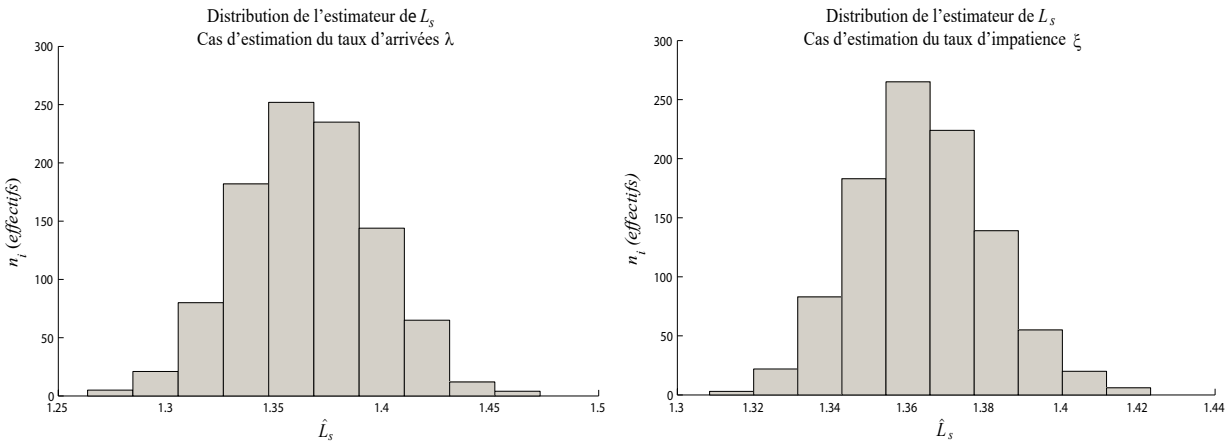

FIGURE 3.22 – Distribution de l'estimateur du nombre moyen de clients dans le système, en fonction du paramètre de départ estimé.

		$P_{1,1}$	P_V	P_B	R	L_q	λ_{eff}	B_r	W_s	R_{ren}
μ	Médiane	0.1848	0.3889	0.4074	0.7963	0.7530	4.3631	1.6369	0.3127	2.4096
	Mode	0.1801	0.3583	0.3874	0.7861	0.7384	4.3072	1.5833	0.2987	2.3628
	Moyenne	0.1848	0.3890	0.4074	0.7963	0.7530	4.3631	1.6369	0.3127	2.4097
η	Médiane	0.1847	0.3891	0.4071	0.7961	0.7529	4.3634	1.6366	0.3126	2.4094
	Mode	0.1805	0.3802	0.3991	0.7792	0.7477	4.3451	1.6214	0.3086	2.3925
	Moyenne	0.1847	0.3891	0.4071	0.7961	0.7529	4.3633	1.6367	0.3126	2.4094
θ	Médiane	0.1848	0.3891	0.4073	0.7964	0.7527	4.3636	1.6364	0.3125	2.4087
	Mode	0.1801	0.3766	0.3990	0.7922	0.7403	4.3636	1.6364	0.3125	2.3689
	Moyenne	0.1847	0.3893	0.4072	0.7964	0.7529	4.3636	1.6364	0.3125	2.4093

TABLE 3.2: Médiane, mode et moyenne des estimateurs de quelques caractéristiques du système, en fonction du paramètre estimé.

Conclusion

Dans ce chapitre, nous avons présenté une analyse stationnaire détaillée (analyse analytique et numérique) d'un système d'attente Markovien à capacité finie avec plusieurs variantes. En effet, nous avons envisagé un système de file d'attente avec vacances multiples, Bernoulli feedback, balking, reneging et rétention des clients impatients et un serveur sujet à des pannes avec réparations. Premièrement, en utilisant la méthode Q-matrice (matrice génératrice infinitésimale), nous avons obtenu les probabilités d'état du système en régime stationnaire. Par la suite, nous avons donné quelques mesures de performance du système. Enfin, pour mettre en évidence le lien entre les performances du système et ses paramètres de fiabilité, nous avons présenté quelques exemples numériques.

Dans l'analyse probabiliste (analytique) d'un modèle de file d'attente, c'est-à-dire lors de l'évaluation analytique de ses mesures de performance, on suppose souvent que les paramètres de départ sont fixes et exacts. Cependant, dans le cadre pratique, les valeurs de ces paramètres ne

sont connues que sous forme d'échantillons de données. A cet effet et afin d'évaluer les performances du système considéré, l'utilisation de techniques d'estimation statistiques (paramétriques et/ou non paramétriques), qui visent à fournir une approximation des valeurs des paramètres (inconnus) en exploitant les informations fournies par l'échantillon, est inévitable.

L'étude de simulation qu'on a réalisé dans ce sens, sur le modèle d'attente considéré, indique que l'intensité de l'impact de l'estimation des paramètres de départ sur les estimations des mesures de performance dépend étroitement du paramètre de départ estimé et de la taille de l'échantillon. En effet, les résultats montrent que :

- Toutes les estimations des mesures de performance convergent en MSE (par conséquence, convergent en biais et en variance) et sont asymptotiquement gaussiennes, convergence en loi, lorsque la taille de l'échantillon est assez grande.
- Pour une même taille d'échantillon, l'estimation des paramètres de départ peuvent être regroupés dans trois ensembles :
  1. Ceux qui nous fournissent des estimations des mesures de performance, très variables d'un échantillon à un autre (mauvaise qualité) voir même des estimations aberrantes dans certaines situations et afin de surpasser ce problème, des échantillons de grandes tailles sont nécessaire.
  2. Ceux qui nous fournissent des estimations des mesures de performance, moyennement variables d'un échantillon à un autre (qualité moyenne) et afin de surpasser ce problème des échantillons de grandes tailles sont nécessaire également.
  3. Ceux qui nous fournissent des estimations des mesures de performance, faiblement variables d'un échantillon à un autre (bonne qualité), même pour des échantillons de petite taille.

En conclusion, pour l'utilisation des méthodes paramétriques dans l'estimation dans des chaînes de Markov, nous sommes contraints de disposer d'échantillons de grande taille afin d'obtenir des estimations d'une qualité raisonnable, ce qui coïncide avec la remarque faite dans le chapitre précèdent.

Dans le chapitre suivant, nous allons analyser la qualité des estimations des mesures de performance d'une chaîne de Markov discrète à temps discret obtenues à l'aide de la méthode non paramétrique du noyau.

Chapitre 4

Estimation à noyau des caractéristiques des systèmes d'attente $D/G/1/N$ et $GI/D/1/N$

Introduction et position du problème

Dans la pratique, la notion **déterministe** dans les modèles de files d'attente n'est pas toujours utilisée en son sens absolu, mais elle peut également désigner les processus où leurs fluctuations aléatoires (la variance) sont faibles, c'est-à-dire se référer aux processus presque déterministes. C'est pourquoi, par exemple, dans l'évaluation des performances des systèmes de télécommunications modernes, la notion déterministe est souvent utilisée. En effet, pour exprimer les faibles fluctuations des temps de service ou/et des temps inter-arrivées, les temps de service ou/et les temps inter-arrivées déterministes, sont fréquemment appliqués pour la modélisation des systèmes dans ce domaine (voir [6, 41, 70, 76, 96]).

Dans ce chapitre, nous avons proposé de considérer la file d'attente $D/G/1$ (respectivement, la file d'attente $GI/D/1$) à capacité finie que nous allons modéliser par une chaîne de Markov induite à temps discret et ceci sous l'hypothèse que la distribution du nombre de clients servis entre deux arrivées consécutives (respectivement, le nombre d'arrivées pendant le temps de service) est une fonction de masse inconnue. De plus, notre objectif est d'estimer les matrices de transition, associées aux chaînes de Markov décrivant les deux files d'attentes en questions, à l'aide de la méthode du noyau. Le choix de l'estimation de la matrice de transition, est motivé par le fait que cette composante a une grande importance dans l'analyse transitoire et stationnaire d'un modèle d'attente et nous permet de déduire tout le reste de ses mesures de performance (pour plus de détails sur la matrice des probabilités de transition et son utilité, le lecteur peut se référer par exemple à [40, 69, 75]).

Rappelons que les travaux existant dans la littérature, dans le cadre d'estimation à noyau dans les modèles d'attente, ont été réalisés via des estimateurs à noyaux asymétriques continus (pour estimer la distribution des temps qui est définie sur $\mathbb{R}^+$). Cependant, en pratique, plusieurs situations sont modélisées par des chaînes de Markov gouvernées selon des distributions discrètes générales et inconnues. Dans ce cas, il est logique et naturel d'estimer ces distributions générales inconnues à l'aide de noyaux discrets.

L'objectif du travail présenté dans ce chapitre est de vérifier la validité des résultats obtenus

dans [27] sur la chaîne de Markov discrète à temps continu lorsque nous considérons une chaîne de Markov discrète à temps discret. Plus précisément, nous analysons le problème du choix du paramètre de lissage via la minimisation des normes matricielles lorsque nous considérons l'estimateur à noyaux discrets des matrices de transition, P, correspondantes aux Chaînes de Markov induites décrivant les deux files d'attente $GI/D/1/N$ et $D/G/1/N$ et qui sont discrètes et à temps discret. Pour ce faire, nous avons développé des formes explicites des expressions, résultantes des trois normes matricielles $\|.\|_1$, $\|.\|_2$ et $\|.\|_\infty$, à minimiser afin de sélectionner le paramètre de lissage lors de l'estimation des matrices P. Par ailleurs, pour étayer et illustrer nos propositions, deux applications numériques comparatives, basées sur des échantillons simulés, sont présentées.

4.1 Description des modèles

4.1.1 La file d'attente $D/G/1/N$

Considérons la file d'attente $D/G/1/N$ où N $(N > 1)$ est la capacité du système y compris celui qui est en service. On suppose que la distribution du temps de service $S(t)$ est une distribution générale de paramètre μ et la distribution du temps inter-arrivées $A(t)$ est une fonction de distribution déterministe de moyenne λ^{-1}, (c'est-à-dire $A(t) = \mathbf{1}_{\{t=\lambda^{-1}\}}$, avec $\mathbf{1}_{\{.\}}$ est la fonction indicatrice). Nous supposons également que le processus des durées des inter-arrivées et le processus de service sont indépendants.

Considérons le processus $\{L(t),\ t \geq 0\}$, où $L(t)$ représente le nombre de clients dans le système à la date t. Évidemment, $\{L(t),\ t \geq 0\}$ n'est pas un processus de Markov. Pour cela, nous considérons la chaîne de Markov induite définie aux instants d'arrivée de nouveaux clients dans le système. Soit $t_n = n\lambda^{-1}$ une séquence qui représente les instants d'arrivée des clients, c'est-à-dire l'instant de l'arrivée du n^{ieme} client. Et soit $N_n = N(t_n)$ le nombre de clients dans le système à la date t_n, alors l'état du système, N_{n+1}, à l'arrivée du $(n+1)^{ieme}$ client est donné par :

$$N_{n+1} = min\{N_n + 1, N\} - X_n, \tag{4.1}$$

où $X_n \equiv X_n(\lambda^{-1})$ représente le nombre de clients servis (départ) entre deux arrivées consécutives (voir la Figure 4.1).

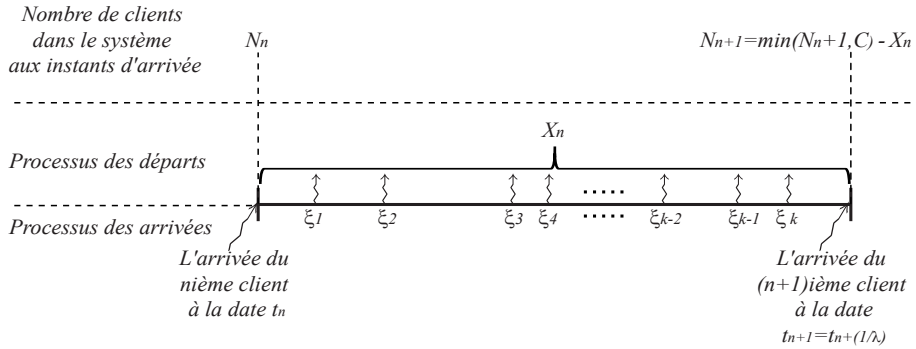

FIGURE 4.1 – Exemple illustratif du processus $\{N_n,\ n \in \mathbb{N}\}$ décrivant la file d'attente $D/G/1/C$.

Notons que la variable aléatoire N_{n+1} ne dépend que de N_n et X_n. De plus, les variables aléatoires $\{X_n,\ n \in \mathbb{N}\}$, sont des variables indépendantes et identiquement distribuées $(i.i.d.)$ d'une loi commune :

$$a_x = f(x) = Pr(X_n = x),\ \ x = 0, 1, 2, 3, ... \tag{4.2}$$

de moyenne $\rho^{-1} = \mu\lambda^{-1}$ (l'inverse de l'intensité du trafic) et sont indépendantes de n et de l'état du système avant t_n. Ainsi, il est clair que le processus $\{N_n,\ n \in \mathbb{N}\}$ est une chaîne de Markov discrète homogène à temps discret, avec un espace d'états $E = \{0, 1, ..., N\}$.
Supposant que le nombre de clients dans le système à la date t_n et à la date t_{n+1} est respectivement $N_n = i$ et $N_{n+1} = j$. Alors, les probabilités de transition $P_{ij} = Pr(N_{n+1} = j \mid N_n = i)$ de cette chaîne sont définies par :

Cas : $i + 1 < j \leq N$ **et** $0 \leq i \leq N - 1$

$$\Pr(N_{n+1} = j \mid N_n = i) = 0. \tag{4.3}$$

Cas : $1 \leq j \leq i + 1$ **et** $0 \leq i \leq N - 1$

$$\Pr(N_{n+1} = j \mid N_n = i) = \Pr(X_n = i + 1 - j) = a_{i-j+1}, \tag{4.4}$$

Cas : $j = 0$ **et** $0 \leq i \leq N - 1$

$$\begin{aligned} \Pr(N_{n+1} = 0 \mid N_n = i) &= \Pr(N_n - X_n + 1 = 0 \mid N_n = i) \\ &= \Pr(X_n \geq i + 1) = \sum_{k \geq i+1} a_k, \end{aligned} \tag{4.5}$$

Cas : $1 \leq j \leq N$ **et** $i = N$

$$\begin{aligned} \Pr(N_{n+1} = j \mid N_n = N) &= \Pr(N_n - X_n = 0 \mid N_n = N) \\ &= \Pr(X_n = N - j) = a_{N-j}, \end{aligned} \tag{4.6}$$

Cas : $j = 0$ **et** $i = N$

$$\Pr(N_{n+1} = 0 \mid N_n = N) = \Pr(X_n \geq N) = \sum_{k \geq N} a_k. \tag{4.7}$$

avec a_k sont définis dans (4.2).
Finalement, les probabilités de transition $P_{ij} = Pr(N_{n+1} = j \mid N_n = i)$ peuvent être résumées comme suit :

$$P_{ij} = \begin{cases} a_{i-j+1}, & \text{si } 1 \leq j \leq i+1 \text{ et } 0 \leq i \leq N-1\,; \\ a_{N-j}, & \text{si } 1 \leq j \leq N \text{ et } i = N\,; \\ \sum\limits_{k \geq i+1} a_k, & \text{si } j = 0 \text{ et } 0 \leq i \leq N-1\,; \\ \sum\limits_{k \geq N} a_k, & \text{si } j = 0 \text{ et } i = N\,; \\ 0, & \text{sinon.} \end{cases} \tag{4.8}$$

Avec a_k sont définis dans (4.2). De plus, on sait que la matrice stochastique P est ergodique. Par conséquent, la distribution stationnaire $\pi = (\pi_0,\ \pi_1, ..., \pi_N)$ existe et elle correspond à l'unique solution du système suivant :

$$\begin{cases} \pi P &= \pi, \\ \sum\limits_{i=0}^{N} \pi_i &= 1. \end{cases} \tag{4.9}$$

Alors, le nombre moyen de clients dans le système est calculé comme suit :

$$L_s = \sum_{i=0}^{N} i\pi_i. \tag{4.10}$$

4.1.2 La file d'attente $GI/D/1/N$

Considérons la file $GI/D/1/N$ où N $(N > 1)$ est la capacité du système y compris celui qui est en service. Nous supposons que la distribution du temps des inter-arrivées $A(t)$ est une fonction de distribution générale de taux $\lambda > 0$ et la distribution du temps de service $S(t)$ est une fonction de distribution déterministe de moyenne μ^{-1}. Nous supposons également que le processus des arrivées et le processus de service sont indépendants.

Considérons le processus $\{L(t),\ t \geq 0\}$, où $L(t)$ représente le nombre de clients dans le système à la date t. De toute évidence, $\{L(t),\ t \geq 0\}$ n'est pas un processus de Markov. C'est pourquoi nous considérons la chaîne de Markov induite définie aux instants de fin de service. Soit t_n une séquence représentant les instants de départs des clients, c'est-à-dire les instants de fin de service du n^{ieme} client. Soit aussi $N_n = N(t_n)$ le nombre de clients dans le système à l'instant de départ du n^{ieme} client, $t_n = n\mu^{-1}$, alors l'état du système, N_{n+1}, à la fin du $(n+1)^{ieme}$ service est donné par :

$$N_{n+1} = min\left\{N_n + X_n, N\right\} - 1, \tag{4.11}$$

où $X_n \equiv X_n(\mu^{-1})$ représente le nombre de clients arrivant durant le n^{ieme} service (voir la Figure 4.2)

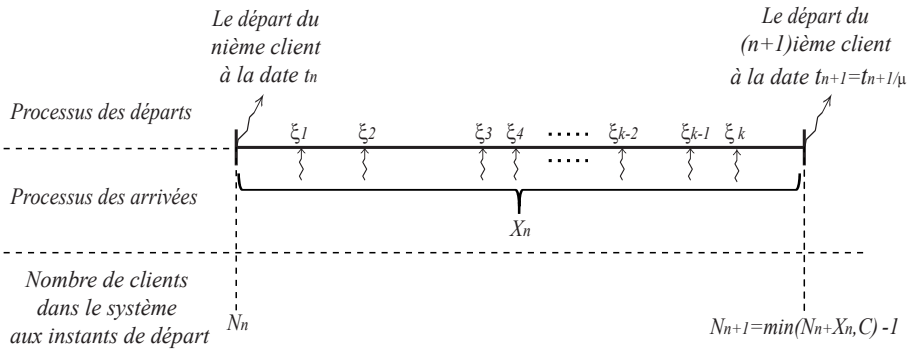

FIGURE 4.2 – Exemple illustratif du processus $\{N_n,\ n \in \mathbb{N}\}$ décrivant la file d'attente $GI/D/1/N$.

Notons que la variable aléatoire N_{n+1} ne dépend que de N_n et X_n. De plus, les variables aléatoires $\{X_n,\ n \in \mathbb{N}\}$, sont des variables $i.i.d.$ d'une loi commune :

$$a_x = P(X_n = x) = f(x),\ \ x = 0, 1, 2, 3, ... \tag{4.12}$$

de moyenne $\rho = \lambda\mu^{-1}$ (l'intensité du trafic), sont indépendants de n et de l'état du système avant t_n. Ainsi, le processus $\{N_n,\ n \in \mathbb{N}\}$ est une chaîne de Markov discrète homogène à temps discret, avec un espace d'états $E = \{0, 1, ..., N\}$.
Supposant que le nombre de clients dans le système à la date t_n et à la date t_{n+1} sont respectivement $N_n = i$ et $N_{n+1} = j$. Alors, avec le même raisonnement que dans le cas du premier modèle (voir section 4.1.1), nous pouvons montrer que les probabilités de transition $P_{ij} = Pr(N_{n+1} = j \mid N_n = i)$ de la chaîne de Markov induite, N, sont comme suit :

$$P_{ij} = \begin{cases} a_{j-i+1}, & \text{si } i-1 \leq j \leq N-1 \text{ et } 1 \leq i \leq N\,; \\ a_{j-i}, & \text{si } 0 \leq j \leq N-1 \text{ et } i = 0\,; \\ \sum\limits_{k \geq N-i+1} a_k, & \text{si } j = N \text{ et } 1 \leq i \leq N\,; \\ \sum\limits_{k \geq N} a_k, & \text{si } j = N \text{ et } i = 0\,; \\ 0, & \text{sinon,} \end{cases} \tag{4.13}$$

avec a_k sont donnés dans (4.12). De plus, le fait que la matrice stochastique P est ergodique alors la distribution stationnaire $\pi = (\pi_0,\ \pi_1, ..., \pi_N)$ existe et elle correspond à l'unique solution du système suivant :

$$\begin{cases} \pi P &= \pi, \\ \sum\limits_{i=0}^{N} \pi_i &= 1. \end{cases} \tag{4.14}$$

Alors, le nombre moyen de clients dans le système est calculé comme suit :

$$L_s = \sum_{i=0}^{N} i\pi_i. \tag{4.15}$$

Remarque 4.1.

1. *Dans le modèle $GI/D/1/N$, si nous supposons que $A(t)$ est une loi exponentielle (le modèle $M/D/1/N$), alors f est une distribution de Poisson de paramètre ρ (voir [17, 37]).*
2. *Dans le modèle $D/G/1/N$, si nous supposons que $S(t)$ est une loi exponentielle (le modèle $D/M/1/N$) alors f est une distribution de Poisson de paramètre ρ^{-1}.*

4.2 Deux exemples de systèmes réels modélisés à l'aide de modèles d'attente déterministes

Dans cette section, nous présentons deux exemples de systèmes réels qui peuvent être modélisés par les files d'attente précédentes, à savoir : un système d'ingénierie (mécanique) à redondance passive et un commutateur de données. Mais, il est à noter que l'objectif ici est de concrétiser les différents concepts liés aux deux systèmes en question et les deux modèles d'attente $GI/D/1/N$ et $D/G/1/N$ et non pas leurs analyse. C'est pourquoi les deux systèmes sont considérés sous leurs formes basiques et simplifiées.

4.2.1 Système d'ingénierie à redondance passive et le modèle $GI/D/1/N$

Il est bien connu que la fiabilité des systèmes d'ingénierie peut être améliorée en augmentant le niveau de redondance utilisé. L'intégration de la redondance dans les systèmes est particulièrement efficace lorsque les pannes aléatoires prédominent ou lorsque la fiabilité est cruciale pour la sécurité des personnes et de l'environnement, comme l'aéronautique [74, 103] ou le nucléaire [46].

Considérons un système à redondance passive composé, initialement, d'un composant actif et de $N-1$ composants inactifs (en veille) qui deviennent actifs en cas de panne du composant opérationnel. De plus, ce système fonctionne selon les hypothèses suivantes :

- Tous les composants sont identiques, au moins au sens de leurs caractéristiques fiabilistes.
- En cas de panne d'un composant actif, un réparateur procédera immédiatement à son remplacement par un composant équivalent en redondance passive.
- Le système est considéré en panne à la date t si et seulement si les N composants sont tous en panne à la date t.
- Le réparateur ne peut remplacer qu'un seul composant défectueux à la fois.
- Les composants défaillants sont remplacés à l'ordre chronologique de leurs défaillances.
- La durée de vie d'un composant opérationnel est une variable aléatoire X dont la distribution est une loi arbitraire (générale) $A(t)$ de moyenne $1/\lambda$ et les composants qui sont en veille sont supposés ne pas tomber en panne (pannes actives).
- Les temps de remplacement R des composants défectueux sont pratiquement égaux ($Var(R) \approx 0$), de temps moyen $1/\mu$ ($E(R) = 1/\mu$).
- Le système est équipé d'un détecteur de commutation fiable (sans défaillance) qui assure une commutation instantanée du composant défaillant au composant de veille (commutation parfaite).

La Figure 4.3 illustre la structure et le mécanisme simplifiés du système en question.

Sous les hypothèses ci-dessus et qu'il y a un nombre suffisant de composants de remplacement dans le stock, alors ce système peut être fidèlement modélisé par la file d'attente $GI/D/1/N$. La dualité entre les paramètres de départ du système et ceux du modèle d'attente $GI/D/1/N$ est résumée dans la Table 4.1.

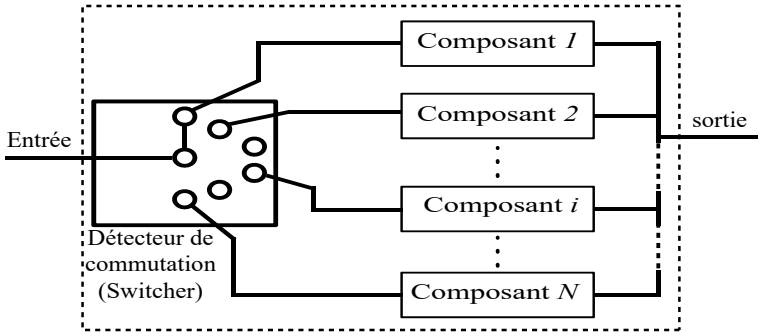

FIGURE 4.3 – Schéma illustratif d'une configuration d'un système à redondance passive avec un commutateur.

Paramètres d'un système à redondance passive	Paramètres du modèle de la file d'attente $GI/D/1/N$
• Un composant défaillant.	• Un client.
• Le réparateur (un seul réparateur).	• Le serveur (un seul serveur).
• Remplacement d'un composant.	• Le service.
• N : Le nombre total de composants dans le système.	• N : La capacité du système.
• λ : Le taux de panne ($1/\lambda$: durée de vie moyenne d'un composant).	• λ : Le taux d'arrivée ($1/\lambda$: Le temps moyen entre deux arrivées consécutives).
• $A(t)$: La distribution de la durée de vie d'un composant.	• $A(t)$: La distribution de temps des inter-arrivées.
• Temps de remplacement.	• Temps de service.
• Le temps de remplacement est une constante ($1/\mu$).	• La distribution du temps de service est une distribution déterministe $\left(S(t) = \mathbf{1}_{\{t=\mu^{-1}\}}\right)$.
• Les composants défaillants sont remplacés dans l'ordre chronologique de leurs défaillances.	• La discipline de la file d'attente est FIFO.

TABLE 4.1: Équivalence entre les paramètres de départ d'un système à redondance passive et ceux de la file d'attente $GI/D/1/N$.

A ce niveau, à l'aide de la théorie des files d'attente, plusieurs mesures de performance de la file d'attente $GI/D/1/N$ peuvent être quantifiées, par conséquent celles du système à redondance passive. La signification et l'équivalence des principales mesures de performance, en régime stationnaire, de la file d'attente $GI/D/1/N$ avec celles du système à redondance passive sont exposées dans la Table 4.2.

Notation	Théorie des files d'attente	Système à redondance passive
π_n	• La probabilité qu'il y ait n clients dans le système ($1 - \pi_n$: La probabilité qu'il y ait $C - n$ espace libre dans le système).	• La probabilité qu'il y ait n composants défectueux dans le système (La probabilité qu'il y ait $C - n$ composants en état de fonctionnement dans le système).
$1 - \pi_N$	• La probabilité que le système ne soit pas plein (π_N : La probabilité de perte)	• La fiabilité du système (π_N : La probabilité que le système soit en panne).
L_s	• Le nombre moyen de clients dans le système.	• Le nombre moyen de composants défaillant.
L_q	• Le nombre moyen de clients en attente.	• A la fin du remplacement, d'un composant défaillant, il reste en moyenne L_q composants qui doivent être remplacés à leur tour.
W_q	• Le temps moyen de séjour dans le système.	• Le temps moyen nécessaire pour remplacer un composant défaillant, à partir de la date de sa défaillance.
W_s	• Le temps moyen d'attente dans la file d'attente.	• Le temps moyen nécessaire pour procéder au remplacement d'un composant défaillant, à partir de la date de sa défaillance.

TABLE 4.2: Équivalence entre les mesures de performance de la file d'attente $GI/D/1/N$ et celles du système à redondance passive.

Le présent exemple peut facilement expliquer et justifier la modélisation du système $GI/D/1/N$ par la chaîne de Markov définie par l'expression (4.11). En effet, supposons que le réparateur désir estimer la fiabilité du système dont il s'occupe et cela à la fin de chaque remplacement d'un composant défaillant, sachant qu'il dispose du nombre de composants défaillants en attente qu'il doit remplacer à leurs tours. Cette dernière information correspond à la variable N_{n+1} $(n \geq 0)$ définie dans (4.11). Cependant, les observations $N_1,\ N_2, ...,\ N_n$, à l'état brut, ne lui permettent pas d'estimer la fiabilité du système en question. En réalité, ce qui lui permettra d'estimer la fiabilité du système, c'est la distribution de l'échantillon $N_{i+1} - N_i,\ (i = 1, ..., n)$ qui correspond au nombre de composants défaillants, lors du remplacement. Ce dernier échantillon correspond exactement aux variables X_i $(i = 1, ..., n)$ introduites dans l'expression (4.11).

Certes, le réparateur peut procéder autrement pour estimer la fiabilité du système et ce en échantillonnant sur les durées de vie des composants actifs afin d'estimer la fiabilité d'un composant, ce qui lui permettra de déduire celle du système. Cependant, si la distribution de la durée de vie du composant $A(t)$ n'est pas exponentielle, et qu'il retient le système d'attente $GI/D/1/N$ comme modèle mathématique décrivant le système, dans ce cas, contrairement à la première approche, même si la distribution de la durée de vie du composant $A(t)$ est connue avec exactitude (théoriquement), il n'aura qu'une valeur approximative de la quantité recherchée. En effet, dans la littérature, l'analyse du système $GI/D/1/N$ est généralement basée sur le modèle $GI/G/1/C$, hors les caractéristiques de ce dernier n'existent que sous leurs formes approchées, obtenues via le modèle $GI/M/1/C$ ou le modèle $M/G/1/C$. C'est ce qui nous a motivé à proposer les chaînes de Markov induites définies en (4.1) et (4.11) pour la modélisation des systèmes $D/G/1/N$ et $GI/D/1/N$, respectivement, qui nous permettent d'obtenir des résultats exacts.

4.2.2 Système de commutation de données et le modèle d'attente $D/G/1/N$

De nombreux systèmes de commutation de données ont des processus avec des flux d'arrivées de tâches régulièrement espacées (par exemple, des octets individuels) où le service de la file d'attente de tâches résultante par le processus peut consister à diriger les octets vers la ligne sortante appropriée. De plus, pour de nombreux systèmes physiques, l'horloge du processus du système quantifie intrinsèquement la synchronisation de tous les événements, rendant le système naturellement discret.

Supposons qu'un commutateur de données, d'une capacité mémoire N, reçoive des paquets à un débit (déterministe régulier) de taux λ par unité de temps et supposons que le temps de traitement d'un paquet (c'est-à-dire le temps de service) dépend des informations d'en-tête du paquet et de durée aléatoire discrète obéissant à une loi discrète arbitraire $S(k),\ k \geq 1$ de moyenne $1/\mu$. Cet exemple décrit implicitement un modèle $D/G/1/N$.

Pour plus de détails sur le présent exemple et sa modélisation par un système de files d'attente en présence d'autres variantes (capacité mémoire infinie ; le processus est contraint de gérer d'autres tâches, ...), le lecteur peut se référer à [88].

4.3 Estimation à noyau de la matrice de transition P

Rappelons que l'évaluation des performances des systèmes de files d'attente est basée sur ses différents paramètres de départ. Cependant, en pratique et en règle générale, les valeurs des paramètres de départ ne sont connues que sous forme d'un échantillon de données. En ce sens, supposons que dans le cadre des deux systèmes d'attente précédents nous sommes dans une situation où :

- Dans le cas de la file d'attente $D/G/1/N$: la distribution du nombre de départs durant la période de deux arrivées consécutives est générale et inconnue. C'est-à-dire que nous ne disposons que d'une information partielle sur la distribution du nombre de clients servis entre deux arrivées consécutives, qui est donnée sous forme d'un échantillon d'observations $x_1,\ x_2, ...,\ x_n$.
- Dans le cas de la file d'attente $GI/D/1/N$: la distribution du nombre d'arrivées pendant un temps de service est générale et inconnue. C'est-à-dire que nous n'avons qu'une information partielle sur la distribution du nombre d'arrivées pendant un temps de service, qui est donnée sous forme d'un échantillon d'observations $x_1,\ x_2, ...,\ x_n$.

De plus, notre intérêt est l'estimation des matrices de transition P associées aux chaînes de Markov induites N, définies dans (4.1) et (4.11), en utilisant la méthode du noyau définie par :

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{x,h}(X_i), \quad x \in \mathbb{N}; \tag{4.16}$$

où h est le paramètre de lissage et $K_{x,h}$ est le noyau discret de cible x et de paramètre de lissage h sur le support $\aleph_{x,h} = \aleph_x$ (indépendant de h).

La mise en œuvre de cette technique, comme nous l'avons mentionné dans le Chapitre 1, nécessite de fixer le noyau K et le paramètre de lissage h. Pour le choix du noyau K, le problème est a priori facile. D'une part, la sélection de K peut facilement s'adapter au support de la densité à estimer, ce qui suggère l'utilisation d'un noyau discret dans notre cas. D'autre part, la littérature existante sur l'estimation par noyau discret indique que le choix d'un noyau discret ne dépend que de la taille de l'échantillon n. Pour des échantillons de grande taille, le noyau Dirac est suffisant pour obtenir de bonnes estimations. Par contre dans le cas, d'échantillon de petites ou de moyennes tailles, le noyau Dirac ne convient pas. Il est intéressant, dans ce cas, de sélectionner K parmi les noyaux suivants : noyau de Poisson, noyau binomial, noyau binomial négatif et noyau triangulaire, qui sont les plus utilisés dans le cadre de l'estimation d'une densité discrète (voir Chapitre 1). Pour le choix du paramètre de lissage h, nous proposons d'utiliser deux approches, à savoir : les techniques classiques et les normes matricielles.

4.3.1 Choix du paramètre de lissage via les techniques classiques

Soit $X_1, \ldots,\ X_n$ un n-échantillon $i.i.d.$ à partir de la distribution générale inconnue f.
Dans ce cas, l'idée est d'estimer les éléments $f(x) = Pr(X = x)$ à partir de l'échantillon sans tenir compte de leurs répétitions dans la matrice P, c'est-à-dire qu'il suffit de quantifier $\hat{f}(x)$ par

$$\hat{f}(x) = \hat{Pr}(X = x) = \frac{1}{n}\sum_{i=1}^{n} K_{x,h^*}(X_i), \quad x \in \mathbb{N},$$

où h^* est le paramètre de lissage optimal sélectionné par l'une des procédures classiques présentées dans la Section 1.5.4 du Chapitre 1 et de remplacer par la suite les éléments $f(x)$ par leurs estimations $\hat{f}(x)$ dans la matrice P pour obtenir son estimateur $\hat{P}$.

4.3.2 Choix du paramètre de lissage via les normes matricielles

Dans cette approche, l'idée est de prendre en compte le nombre de répétitions des éléments $\hat{f}(x)$ dans les matrices $\hat{P}$. Notons que l'estimateur $\hat{P}$ est donné par :

$$\hat{P} = \begin{pmatrix} \hat{a}_0 & \hat{a}_1 & & \cdots & & \hat{a}_{N-2} & \hat{a}_{N-1} & \hat{b}_N \\ \hat{a}_0 & \hat{a}_1 & & \cdots & & \hat{a}_{N-2} & \hat{a}_{N-1} & \hat{b}_N \\ 0 & \hat{a}_0 & & \cdots & & \hat{a}_{N-3} & \hat{a}_{N-2} & \hat{b}_{N-1} \\ \vdots & \ddots & \ddots & & & & & \vdots \\ 0 & \cdots & 0 & \hat{a}_0 & \cdots & \hat{a}_{i-1} & \hat{a}_i & \hat{b}_{i+1} \\ \vdots & & & \ddots & \ddots & & & \vdots \\ 0 & & \cdots & & 0 & \hat{a}_0 & \hat{a}_1 & \hat{b}_2 \\ 0 & & \cdots & & 0 & 0 & \hat{a}_0 & \hat{b}_1 \end{pmatrix}, \tag{4.17}$$

pour le modèle $GI/D/1/N$, et

$$\hat{P} = \begin{pmatrix} \hat{b}_1 & \hat{a}_0 & 0 & 0 & & \cdots & & & 0 \\ \hat{b}_2 & \hat{a}_1 & \hat{a}_0 & 0 & & \cdots & & & 0 \\ \vdots & \vdots & & & \ddots & & & & \vdots \\ \hat{b}_{i+1} & \hat{a}_i & \hat{a}_{i-1} & \cdots & & \hat{a}_0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & & & \ddots & & \vdots \\ \hat{b}_{N-1} & \hat{a}_{N-2} & \hat{a}_{N-3} & & \cdots & & & \hat{a}_0 & 0 \\ \hat{b}_N & \hat{a}_{N-1} & \hat{a}_{N-2} & & \cdots & & & \hat{a}_1 & \hat{a}_0 \\ \hat{b}_N & \hat{a}_{N-1} & \hat{a}_{N-2} & & \cdots & & & \hat{a}_1 & \hat{a}_0 \end{pmatrix}, \tag{4.18}$$

pour le modèle $D/G/1/N$, où $\hat{a}_x = \hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{x,h}(X_i)$, $\hat{b}_y = \sum_{y \geq x} \hat{a}_x$ et $x \in \mathbb{N}$.

Afin de prendre en compte les répétitions des éléments $\hat{f}(x)$ dans $\hat{P}$, nous proposons d'utiliser les normes matricielles qui ont un impact sur la qualité de l'estimateur $\hat{P}$. En effet, l'utilisation des normes matricielles permet d'inclure les répétitions des éléments $\hat{f}(x)$ dans la matrice $\hat{P}$. Dans ce cas, le paramètre de lissage optimal peut être calculé, par exemple, avec l'une des trois expressions suivantes :

$$\begin{cases} h_1^* &= \arg\min_h \left\| \hat{P} - P \right\|_1 = \arg\min_h \left[\max_{0 \leq i \leq N} \left(\sum_{j=0}^{N} \left| \hat{P}_{ij} - P_{ij} \right| \right) \right], \\ h_2^* &= \arg\min_h \left(\left\| \hat{P} - P \right\|_2 \right) = \arg\min_h \left(\left\| \hat{P} - P \right\|_2 \right)^2 = \arg\min_h \left[\sum_{i=0}^{N} \sum_{j=0}^{N} \left(\hat{P}_{ij} - P_{ij} \right)^2 \right], \\ h_3^* &= \arg\min_h \left\| \hat{P} - P \right\|_\infty = \arg\min_h \left[\max_{0 \leq j \leq N} \left(\sum_{i=0}^{N} \left| \hat{P}_{ij} - P_{ij} \right| \right) \right], \end{cases}$$

Ainsi, si l'on considère l'estimation de la matrice de transition associée au modèle $GI/D/1/N$, puis en remplaçant les éléments de P et $\hat{P}$ par leurs expressions, les trois formules ci-dessus peuvent être réécrites comme suit :

$$
\begin{aligned}
h_1^* &= \arg\min_h \left[\max_{1\leq i\leq N} \left(\sum_{j=i-1}^{N} \left| \hat{P}_{ij} - P_{ij} \right| \right) \right], \\
&= \arg\min_h \left[\max_{1\leq i\leq N} \left(\sum_{x=0}^{N-i} |\hat{a}(x) - a(x)| + \left| \hat{b}_{N-i+1} - b_{N-i+1} \right| \right) \right], \\
&= \arg\min_h \left[\max_{1\leq i\leq N} \left(\sum_{x=0}^{N-i} \left| \hat{f}(x) - f(x) \right| + \left| \sum_{x>N-i} \left(\hat{f}(x) - f(x) \right) \right| \right) \right]. \qquad (4.19)
\end{aligned}
$$

$$
\begin{aligned}
h_2^* &= \arg\min_h \left[\sum_{x=0}^{N-1} (N-x+1) \left(\hat{a}_x - a_x \right)^2 + \sum_{i=1}^{N} \left(\hat{b}_i - b_i \right)^2 + (\hat{b}_N - b_N)^2 \right], \\
&= \arg\min_h \left[\sum_{x=0}^{N-1} (N-x+1) \left(\hat{f}(x) - f(x) \right)^2 + \sum_{i=1}^{N} \left(\sum_{x\geq i} \left(\hat{f}(x) - f(x) \right) \right)^2 \right. \\
&\qquad \left. + \left(\sum_{x\geq N} \left(\hat{f}(x) - f(x) \right) \right)^2 \right]. \qquad (4.20)
\end{aligned}
$$

$$
\begin{aligned}
h_3^* &= \arg\min_h \left[\max_{0\leq j\leq N} \left(\begin{array}{l} \left\{ \sum_{x=0}^{j} |\hat{a}_x - a_x| + |\hat{a}_j - a_j| \right\} \mathbf{1}_{\{j\neq N\}} \\ + \left\{ \sum_{i=1}^{N} \left| \hat{b}_i - b_i \right| + \left| \hat{b}_N - b_N \right| \right\} \mathbf{1}_{\{j=N\}} \end{array} \right) \right] \\
&= \arg\min_h \left[\max_{0\leq j\leq N} \left(\begin{array}{l} \left\{ \sum_{x=0}^{j} \left| \hat{f}(x) - f(x) \right| + \left| \hat{f}(j) - f(j) \right| \right\} \mathbf{1}_{\{j\neq N\}} \\ + \left\{ \sum_{i=1}^{N} \left| \sum_{x\geq i} \left(\hat{f}(x) - f(x) \right) \right| + \left| \sum_{x\geq N} \left(\hat{f}(x) - f(x) \right) \right| \right\} \mathbf{1}_{\{j=C\}} \end{array} \right) \right] \qquad (4.21)
\end{aligned}
$$

avec $\mathbf{1}_{\{.\}}$ est la fonction indicatrice.

Notons que, si l'on considère l'estimation de la matrice de transition associée au modèle $D/G/1/N$, excepté la fonction indicatrice $\mathbf{1}_{\{j=N\}}$ (respectivement $\mathbf{1}_{\{j\neq N\}}$) de l'expression (4.21) qui sera cette fois $\mathbf{1}_{\{j=0\}}$ (respectivement $\mathbf{1}_{\{j\neq 0\}}$), le reste des expressions (4.19)–(4.21) restent inchangées.

4.4 Application numérique

Le but de cette section est d'analyser numériquement l'impact du paramètre de lissage, choisi par la minimisation des normes matricielles, sur les performances de l'estimateur à noyau des matrices de transition associées aux chaînes de Markov induites décrivant les modèles d'attentes $GI/D/1/N$ et $D/G/1/N$. Pour cela, nous avons conçu un simulateur, sous l'environnement Matlab, dont les principales étapes sont les suivantes :

Étape 0. Fixer le modèle.

Étape 1. Définir les paramètres de départ du modèle : $\mu,\ \lambda,\ N,\ f$.

Étape 2. Générer m échantillons de taille n de distribution cible f.

Étape 3. Estimer h_{opt} en utilisant les expressions (1.65), (4.19)$-$(4.21) pour chaque échantillon.

Étape 4. Calculer $\hat{P}$, $\hat{\pi}_N$ et $\hat{L}_s$ pour chaque h_{opt} obtenu à l'étape 3.

Étape 5. Calculer h^*, $\overline{\pi}_N$ et $\overline{L}_s$ la moyenne des m estimateurs h_{opt}, $\hat{\pi}_N$ et $\hat{L}_s$ et leurs variances.

Les grandeurs $\hat{\pi}_N$ et $\hat{L}_s$ sont respectivement les estimateurs de la probabilité de perte (refus) et de la longueur de la file d'attente, qui sont calculées à partir les formules (4.9) et (4.10) dans le cas du modèle $D/G/1/N$, et à partir des formules (4.14) et (4.15) dans le cas du modèle $GI/D/1/N$ et ceci lorsque on remplace P par son estimateur $\hat{P}$.

Pour l'application numérique, nous avons considéré deux situations, à savoir : la variation de la taille de l'échantillon et la variation de l'intensité du trafic.

4.4.1 Effet de la taille de l'échantillon sur les performances de $\hat{P}$

Dans cette application, nous concentrons principalement sur l'effet du choix du paramètre de lissage, sur la qualité des estimateurs des caractéristiques stationnaires de nos modèles, en fonction de la taille de l'échantillon. Pour cela, nous fixons :

- la capacité du système $N = 10$.
- le nombre de simulations $m = 1000$.
- les tailles d'échantillon $n \in \{50,\ 100;\ 250;\ 500;\ 1000\}$.
- le taux de service $\mu = 1$, l'intensité du trafic (le taux d'arrivée) $\rho = \frac{\lambda}{\mu} = \lambda = 0.75$ et la distribution $f \in \{$Poisson(θ), Géométrique $(\frac{1}{1+\theta})$, Binomiale $(N, \theta/N)\}$ avec $\theta = 1/\rho$ dans le cas du modèle $D/G/1/N$ et $\theta = \rho$ dans le cas du modèle $GI/D/1/N$.
- le noyau $K \in \{$Poisson ; Binomial négatif ; Triangulaire$\}$ noté respectivement K_{P_o}, K_{NB} and K_T.

Les formes de ces derniers noyaux sont présentées dans la Section 1.5.4 du Chapitre 1. Pour le noyau triangulaire K_T, dans les calculs numérique, nous avons fixé le bras $a = 2$.

Il est à noter que le comportement des résultats de simulation obtenus dans le cas des deux modèles est similaire, c'est la raison pour laquelle, dans la suite de la présente Section et de la Section 4.4.2, nous nous concentrerons uniquement sur les résultats obtenus dans le cadre du modèle $D/G/1/N$.

Les caractéristiques exactes des deux modèles considérés dans cette application sont données dans la Table 4.3 et un échantillon de résultats de simulation obtenus pour les différents paramètres précédents est rangé dans les Tables 4.4–4.9 où les meilleurs résultats associés à un noyau bien défini sont présentés en gras tandis que les meilleurs résultats globaux (indépendamment du noyau utilisé) sont présentés en gras souligné.

Distribution	Poisson		Géométrique		Binomiale	
Modèle	$D/G/1/N$	$GI/D/1/N$	$D/G/1/N$	$GI/D/1/N$	$D/G/1/N$	$GI/D/1/N$
L_s	1.1813	1.8393	2.5149	2.5149	1.0140	1.7417
π_N ($\times 10^{-4}$)	6.38	25.00	146.99	146.99	3.05	16.62

TABLE 4.3: Valeurs exactes de L_s et π_N selon la distribution f.

D'après les résultats obtenus, nous remarquons que :

- Excepté le cas d'utilisation du noyau K_T où la propriété de convergence du paramètre de lissage optimal h^* en fonction de la taille de l'échantillon est satisfaite ($\lim_{n\to\infty} h(n) = 0$) lorsque on utilise les noyaux K_{P_0} ou K_{NB} la propriété en question n'est pas satisfaite et cela quel que soit la procédure de sélection utilisée et la distribution de f (voir Tables 4.4–4.6).
- Dans le cas des distributions Géométrique et Poisson, le comportement de la variance du paramètre de lissage optimal, $Var(h^*)$, n'est pas régulier en fonction de la taille de l'échantillon lorsque nous utilisons le noyau K_{P_0} pour construire l'estimateur de P. Tandis que, dans le cas de l'utilisation des noyaux K_{NB} et K_T, la variation de $Var(h^*)$ est sensible à la taille de l'échantillon (converge vers zéro en fonction de n), et cela indépendamment de la distribution f et de la procédure de sélection utilisée. En outre, la plus petite variance du paramètre de lissage optimal est observée lors de l'utilisation du K_T (voir Tables 4.4–4.6).
- Dans le cas où f est une distribution de Poisson, le paramètre de lissage optimal qui nous fournit un estimateur du nombre moyen de clients dans le système et de la probabilité de perte, avec une MSE plus petite, est celui qui a été sélectionné en minimisant la norme matricielle $\|.\|_2$, et ceci lorsque nous utilisons le noyau K_T. De plus, il est préférable d'éviter l'utilisation des noyaux K_{P_0} et K_{NB} (voir Tables 4.7 et 4.9). Cette constatation reste valable pour le cas f est une distribution binomiale.
- Dans le cas où f est une distribution géométrique, pour obtenir un estimateur plus performant du nombre moyen de clients dans le système, au sens du MSE, il est préférable de construire l'estimateur, $\hat{P}$, à l'aide du noyau K_T et de sélectionner le paramètre de lissage en minimisant la norme matricielle $\|.\|_\infty$ pour $(n \geq 250)$ et en utilisant le noyau K_{P_0} et choisir le paramètre de lissage par la minimisation de la norme matricielle $\|.\|_\infty$ pour $n < 250$ (voir Table 4.8). Cela reste valable pour l'estimateur de la probabilité de perte.
- Si on retiens la variance comme critère de comparaison des estimateurs, les pires estimation sont obtenus lorsque nous les construisons à l'aide du noyau K_T, tandis que les meilleurs estimateurs sont obtenus lorsque en utilisant du noyau K_{NB}, et ce qui est tout à fait le contraire au sens du MSE. Ce résultat peut s'expliquer par le biais significatif des estimateurs conçus à l'aide du K_{NB}, qui affecte considérablement le MSE ($MSE = biais^2 + Var$) au point que ces estimateurs deviennent les moins performants au sens de ce dernier critère, critère, par rapport aux restes des estimateurs.

En réalité, quelques remarques faites auparavant sur le biais et la variance des estimateurs de $\overline{L}_s$ et $\overline{\pi}_N$ sont prévues dès les premiers résultats qu'on a obtenus (voir Tables 4.4 –4.6). En effet, d'une part, il est bien connu dans la théorie de l'estimation d'une densité par la méthode du noyau, que le biais de l'estimateur est proportionnel au paramètre de lissage tandis que la variance de l'estimateur est inversement proportionnelle à la valeur du paramètre de lissage. D'autre part, d'après les résultats rangés dans les Tables 4.4–4.6, on remarque que les valeurs des paramètres

de lissage optimaux obtenus dans le cas des noyaux K_{P_0} et K_{NB} sont considérablement plus élevés, comparés à ceux obtenus dans le cadre de noyau K_T. Par conséquence, à ce niveau déjà, nous pouvons prévoir que les estimations que nous obtiendrons en utilisant les noyaux K_{NB} et K_{P_0} auront un biais (respectivement une variance) plus élevé (respectivement plus faible) que ceux que nous obtiendrons en utilisant le noyau K_T.

Enfin, nos résultats nous permettent de conclure que les paramètres de lissage choisis par la minimisation des normes matricielles nous fournissent, en général, des estimateurs plus performants et que le choix du noyau est d'une grande importance. En effet, nos résultats montrent que pour estimer, par la méthode du noyau, la matrice de transition associée à l'un des modèles de files d'attente considérés, il vaut mieux d'utiliser le noyau K_T et d'éviter le noyau K_{NB}.

Remarque 4.2. *L'interprétation de nos résultats au sens de l'exemple du système d'ingénierie à redondance passive, présenté dans la Section 4.2.1 du présent chapitre, est la suivante : Afin que le réparateur ait de bonnes estimations des caractéristiques du système, nos résultats lui suggèrent d'utiliser le paramètre de lissage minimisant la norme $\|.\|_2$ ou $\|.\|_\infty$ et le noyau triangulaire comme paramètres de l'estimateur à noyau défini dans l'expression (4.17).*

		ISE		$\|.\|_1$		$\|.\|_2$		$\|.\|_\infty$	
K	n	h^*	$Var(h^*)$	h^*	$Var(h^*)$	h^*	$Var(h^*)$	h^*	$Var(h^*)$
	50	0.5992	0.6493	0.7780	2.1306	0.7399	0.6528	2.2548	1.7485
	100	0.4508	0.1021	0.5681	2.3116	0.6046	0.1291	2.2733	1.3769
K_{P_o}	250	0.4341	0.0354	0.4227	1.9258	0.6072	0.0410	2.2856	1.6051
	500	0.5903	2.2176	0.5744	4.6277	0.7616	2.1495	2.3137	2.1100
	1000	0.4401	0.0116	0.2745	0.0082	0.6135	0.0133	2.1753	0.1825
	50	3.0893	1.0508	3.6595	1.3829	3.9222	1.0231	9.1221	1.2335
	100	3.1094	0.5194	3.7284	0.8306	3.9414	0.5117	9.1531	0.6418
K_{NB}	250	3.1120	0.2111	3.7070	0.3359	3.9330	0.2126	9.1272	0.2833
	500	3.1448	0.1022	3.7446	0.1822	3.9680	0.1020	9.1766	0.1216
	1000	3.1695	0.0430	3.7651	0.0747	3.9908	0.0429	9.1879	0.0553
	50	0.2404	0.0401	0.2256	0.0506	0.2162	0.0383	0.1933	0.0313
	100	0.1180	0.0103	0.1278	0.0126	0.1047	0.0093	0.1010	0.0079
K_T	250	0.0563	0.0031	0.0628	0.0038	0.0505	0.0028	0.0548	0.0037
	500	0.0396	0.0017	0.0480	0.0027	0.0362	0.0016	0.0377	0.0018
	1000	0.0259	0.0008	0.0278	0.0008	0.0239	0.0008	0.0240	0.0007

TABLE 4.4: Estimation des paramètres de lissage h^* : cas de la distribution de Poisson.

		ISE		$\|.\|_1$		$\|.\|_2$		$\|.\|_\infty$	
K	n	h^*	$Var(h^*)$	h^*	$Var(h^*)$	h^*	$Var(h^*)$	h^*	$Var(h^*)$
	50	12.8042	36.9091	13.0647	37.0867	12.9009	37.2185	12.1216	37.8620
	100	13.5294	29.4363	13.7931	29.2290	13.6178	29.2755	13.4730	30.5834
K_{P_o}	250	16.5537	40.1850	16.9919	38.4396	16.6223	39.2460	17.0359	35.5228
	500	15.6298	51.5259	15.5096	54.7955	15.3469	54.7303	16.6891	48.8362
	1000	16.1402	58.9441	16.4646	59.9760	16.3606	59.0005	16.7988	58.3202
	50	8.8605	2.2823	8.8284	2.1928	9.2378	2.6419	11.5852	4.8646
	100	8.6226	1.6838	8.6071	1.5924	9.0096	1.8044	11.4782	2.6132
K_{NB}	250	8.9698	0.5720	8.9480	0.5446	9.3882	0.6465	11.9838	1.1491
	500	8.7934	0.2640	8.7738	0.2533	9.1942	0.3019	11.7197	0.5581
	1000	8.7784	0.1477	8.7608	0.1419	9.1745	0.1693	11.6839	0.3141
	50	0.0772	0.0059	0.0817	0.0060	0.0667	0.0053	0.0643	0.0079
	100	0.0414	0.0027	0.0410	0.0033	0.0374	0.0026	0.0356	0.0035
K_T	250	0.0301	0.0012	0.0293	0.0014	0.0286	0.0011	0.0263	0.0011
	500	0.0207	0.0005	0.0179	0.0005	0.0198	0.0005	0.0193	0.0006
	1000	0.0114	0.0002	0.0104	0.0002	0.0111	0.0002	0.0110	0.0002

TABLE 4.5: Estimation des paramètres de lissage h^* : cas de la distribution Géométrique.

		ISE		$\|.\|_1$		$\|.\|_2$		$\|.\|_\infty$	
K	n	h^*	$Var(h^*)$	h^*	$Var(h^*)$	h^*	$Var(h^*)$	h^*	$Var(h^*)$
	50	0.3231	0.0951	0.2946	0.1637	0.4345	0.1240	1.4451	0.3863
	100	0.3450	0.0482	0.2454	0.0477	0.4731	0.0570	1.5245	0.1484
K_{P_o}	250	0.3013	0.0206	0.3046	1.7627	0.4382	0.0232	1.4793	0.0610
	500	0.3030	0.0112	0.1597	0.0080	0.4429	0.0126	1.4841	0.0323
	1000	0.3070	0.0058	0.1636	0.0045	0.4485	0.0064	1.4924	0.0106
	50	2.3894	1.4084	3.4413	1.7009	3.1967	1.1835	7.8601	1.0809
	100	2.5008	0.4686	3.4811	0.6580	3.2666	0.3926	7.9421	0.4766
K_{NB}	250	2.5872	0.1908	3.5462	0.2842	3.3351	0.1830	8.0214	0.1812
	500	2.5595	0.0733	3.5231	0.1047	3.3025	0.0728	7.9830	0.0835
	1000	2.5739	0.0424	3.5363	0.0645	3.3185	0.0421	8.0075	0.0458
	50	0.1707	0.0176	0.1548	0.0157	0.1549	0.0157	0.1445	0.0136
	100	0.1224	0.0147	0.1211	0.0195	0.1133	0.0136	0.1093	0.0123
K_T	250	0.0503	0.0024	0.0547	0.0026	0.0470	0.0024	0.0483	0.0024
	500	0.0247	0.0011	0.0313	0.0015	0.0233	0.0011	0.0209	0.0009
	1000	0.0196	0.0006	0.0216	0.0007	0.0186	0.0005	0.0173	0.0005

TABLE 4.6: Estimation des paramètres de lissage h^* : cas de la distribution Binomiale.

		ISE			$\|.\|_1$			$\|.\|_2$			$\|.\|_\infty$		
K	n	$\overline{L}_s$	Var	MSE	$\overline{L}_s$	Var	MSE	$\overline{L}_s$	Var	MSE	$\overline{L}_s$	Var	MSE
K_{P_o}	50	0.8354	0.0525	0.1721	0.8675	0.1785	0.2770	0.9112	0.0462	**0.1192**	1.6946	0.0712	0.3347
	100	0.7887	0.0051	0.1592	0.7871	0.1258	0.2812	0.8698	0.0035	**0.1006**	1.7185	0.0540	0.3425
	250	0.7731	0.0015	0.1681	0.7247	0.0748	0.2833	0.8641	0.0013	**0.1019**	1.7303	0.0610	0.3624
	500	0.7934	0.0533	0.2038	0.7417	0.1747	0.3680	0.8845	0.0490	**0.1371**	1.7351	0.0511	0.3577
	1000	0.7675	0.0003	0.1716	0.6823	0.0001	0.2491	0.8598	0.0003	**0.1036**	1.7276	0.0272	0.3256
K_{NB}	50	0.7961	0.0008	0.1492	0.8761	0.0039	0.0970	0.9122	0.0002	**0.0726**	2.5075	0.0139	1.7727
	100	0.7966	0.0002	0.1482	0.8828	0.0024	0.0915	0.9127	0.0001	**0.0723**	2.5172	0.0073	1.7917
	250	0.7951	0.0001	0.1493	0.8780	0.0009	0.0929	0.9115	0.0000	**0.0729**	2.5189	0.0030	1.7920
	500	0.7944	0.0000	0.1498	0.8779	0.0006	0.0927	0.9114	0.0000	**0.0729**	2.5238	0.0014	1.8036
	1000	0.7942	0.0000	0.1499	0.8770	0.0002	0.0928	0.9115	0.0000	**0.0728**	2.5219	0.0006	1.7978
K_T	50	0.8658	0.0668	0.1664	0.8606	0.0563	0.1591	0.8850	0.0651	**0.1529**	0.8995	0.0788	0.1582
	100	0.9109	0.0429	0.1160	0.9007	0.0481	0.1268	0.9276	0.0424	**0.1067**	0.9258	0.0452	0.1104
	250	1.0174	0.0320	0.0589	1.0127	0.0422	0.0707	1.0268	0.0295	**0.0534**	1.0189	0.0315	0.0579
	500	1.0634	0.0222	0.0361	1.0471	0.0262	0.0443	1.0704	0.0201	**0.0324**	1.0713	0.0285	0.0406
	1000	1.1163	0.0092	0.0135	1.1122	0.0125	0.0173	1.1219	0.0083	**0.0118**	1.1229	0.0110	0.0144

TABLE 4.7: Estimation du nombre moyen de clients dans le système : cas de la distribution de Poisson.

		ISE			$\|.\|_1$			$\|.\|_2$			$\|.\|_\infty$		
K	n	$\overline{L}_s$	Var	MSE	$\overline{L}_s$	Var	MSE	$\overline{L}_s$	Var	MSE	$\overline{L}_s$	Var	MSE
K_{P_o}	50	2.6871	0.0571	0.0867	2.6998	0.0445	0.0787	2.6452	0.0400	0.0570	2.3841	0.0242	**0.0413**
	100	2.5978	0.1570	0.1638	2.6105	0.1489	0.1581	2.5614	0.1391	0.1413	2.3643	0.0664	**0.0891**
	250	2.4318	0.2526	0.2596	2.4627	0.2280	0.2307	2.4167	0.2239	0.2335	2.3020	0.1233	**0.1686**
	500	2.2907	0.3730	0.4233	2.2990	0.3732	0.4198	2.2697	0.3483	0.4084	2.2340	0.1870	**0.2659**
	1000	2.1723	0.4257	0.5431	2.1935	0.4168	0.5201	2.1702	0.3955	0.5144	2.1149	0.2720	**0.4320**
K_{NB}	50	1.9509	0.0182	0.3364	1.9391	0.0165	0.3481	2.0677	0.0103	**0.2103**	2.9815	0.0116	0.2293
	100	1.9208	0.0082	0.3612	1.9147	0.0068	0.3671	2.0419	0.0050	**0.2288**	3.0021	0.0044	0.2417
	250	1.8963	0.0046	0.3873	1.8893	0.0040	0.3955	2.0248	0.0026	**0.2428**	3.0182	0.0027	0.2560
	500	1.9042	0.0021	0.3751	1.8980	0.0018	0.3825	2.0301	0.0012	**0.2363**	3.0115	0.0011	0.2477
	1000	1.9061	0.0011	0.3718	1.9006	0.0010	0.3784	2.0311	0.0007	**0.2347**	3.0077	0.0006	0.2434
K_T	50	1.9555	0.3649	0.6779	1.9622	0.4699	0.7754	2.0080	0.3466	0.6036	2.0349	0.3006	**0.5310**
	100	2.1619	0.2636	0.3882	2.1787	0.2793	0.3924	2.1887	0.2466	0.3530	2.2043	0.2060	**0.3024**
	250	2.3324	0.1326	0.1659	2.3431	0.1396	0.1691	2.3437	0.1243	0.1537	2.3601	0.1043	**0.1282**
	500	2.3696	0.0612	0.0824	2.3967	0.0651	0.0790	2.3771	0.0574	0.0764	2.3810	0.0488	**0.0667**
	1000	2.4297	0.0416	0.0489	2.4395	0.0406	0.0463	2.4324	0.0382	0.0450	2.4309	0.0289	**0.0359**

TABLE 4.8: Estimation du nombre moyen de clients dans le système : cas de la distribution Géométrique.

		ISE		$\|.\|_1$		$\|.\|_2$		$\|.\|_\infty$	
K	n	$\overline{\pi}_N$	MSE	$\overline{\pi}_N$	MSE	$\overline{\pi}_N$	MSE	$\overline{\pi}_N$	MSE
K_{P_o}	50	3.9593	0.0414	9.5954	0.1750	4.8105	**0.0403**	41.7484	0.2411
	100	1.4857	0.0025	6.6148	0.1435	2.3455	**0.0017**	42.9779	0.1992
	250	1.2581	0.0026	4.0257	0.1052	2.2109	**0.0018**	44.3390	0.2397
	500	3.7503	0.0649	8.7334	0.3592	4.6890	**0.0637**	44.2053	0.2138
	1000	1.1858	0.0027	0.6199	0.0033	2.1327	**0.0018**	42.2591	0.1474
K_{NB}	50	1.5845	0.0023	2.6623	0.0015	3.1301	**0.0011**	153.3177	2.2216
	100	1.5796	0.0023	2.7321	0.0014	3.1424	**0.0011**	154.8834	2.2385
	250	1.5600	0.0023	2.6142	0.0014	3.1172	**0.0011**	154.6869	2.2123
	500	1.5535	0.0023	2.6066	0.0014	3.1178	**0.0011**	155.5688	2.2316
	1000	1.5511	0.0023	2.5836	0.0014	3.1184	**0.0011**	155.0893	2.2140
K_T	50	2.8811	0.0022	2.6271	0.0022	3.0572	**0.0020**	3.5825	0.0030
	100	2.9368	0.0019	2.9682	0.0024	3.1272	**0.0017**	3.2056	0.0022
	250	4.1530	0.0013	4.3551	0.0020	4.2452	**0.0012**	4.1667	0.0015
	500	4.7239	0.0010	4.5481	0.0013	4.7968	**0.0009**	5.0260	0.0016
	1000	5.3219	0.0005	5.3239	0.0007	5.4061	**0.0004**	5.4968	0.0007

TABLE 4.9: Estimation de la probabilité de perte : cas de la distribution de Poisson ($\times 10^{-4}$).

4.4.2 Effet de l'intensité du trafic sur les performances de l'estimateur $\hat{P}$

Le but de cette deuxième application est de vérifier la validité des résultats obtenus dans la première application lorsque l'intensité du trafic varie. Autrement dit, notre objectif est d'analyser numériquement l'effet de l'intensité du trafic, ρ, sur la qualité des estimations h^*, $\hat{\pi}_N$ et $\hat{L}_s$. Pour les calculs numériques, nous fixons les différents paramètres comme suit :

- la capacité du système $N = 10$,
- le taux de service $\mu = 1$, le taux d'arrivée $\lambda \in [0.1\ ;\ 1.5]$ d'où l'intensité du trafic $\rho = \lambda$,
- la distribution $f \in \{\text{Poisson}(\theta)$, Géométrique $(\frac{1}{1+\theta})$, Binomiale $(N, \theta/N)\}$ avec $\theta = 1/\rho$ dans le cas du modèle $D/G/1/N$ et $\theta = \rho$ dans le cas du modèle $GI/D/1/N$,
- le noyau $K \in \{K_{P_o},\ K_{NB},\ K_T\}$.

Un échantillon de résultats obtenus sur 1000 échantillons de taille $n = 200$, lorsque on considère le modèle $D/G/1/C$, sont présentés dans les Figures 4.4–4.6.

D'après les résultats obtenus, nous constatons que :

- Pour de faibles intensités de trafic ($\rho < 0.6$), dans toutes les situations considérées dans cette seconde application, on obtient pratiquement les mêmes caractéristiques (biais, variance et MSE) des estimateurs $\hat{L}_s$ et $\hat{\pi}_N$ avec une légère préférence pour la technique ISE et la norme $\|.\|_1$, pour la sélection du paramètre de lissage optimal.
- Contrairement au cas du noyau K_T, où le paramètre de lissage optimal h^* est inversement proportionnel à l'intensité du trafic. Lors de l'utilisation du noyau K_{P_0} ou K_{NB}, le paramètre de lissage optimal h^* est proportionnel à l'intensité du trafic et cela pour toutes les situations analysées (voir Figure 4.4). On remarque également que le comportement de la variance $Var(h^*)$ en fonction de l'intensité du trafic ρ n'est pas régulière. En effet, dans toutes les situations considérées, la variance $Var(h^*)$ est une fonction croissante au départ, pour qu'elle devienne par la suite, à partir d'une certaine valeur de ρ, une fonction décroissante.
- Dans le cas où f est une distribution de Poisson ou Binomiale, les paramètres de lissage optimaux qui nous fournissent des estimateurs du nombre moyen de clients dans le système et de la probabilité de perte ayant de petites MSE selon le noyau utilisé sont : $(K_{P_0},\ h_3^*)$, $(K_{NB},\ h_3^*)$ et (K_T, h_2^*), mais ça reste que, globalement, pour une intensité de trafic moyenne ou élevée ($\rho \geq 0.6$), il est préférable d'utiliser le noyau triangulaire et de sélectionner le paramètre de lissage en minimisant la norme matricielle $\|.\|_2$ et d'éviter l'utilisation du noyau K_{NB} (voir Figures 4.5 et 4.6).
- Dans le cas f est une distribution géométrique, pour une forte intensité du trafic, les paramètres de lissage optimaux qui nous fournissent, des estimateurs du nombre moyen de clients dans le système et de la probabilité de perte, avec une petite MSE selon le noyau utilisé sont : $(K_{P_0},\ h_3^*)$, $(K_{NB},\ h_3^*)$, $(K_T,\ h_3^*)$ et $(K_T,\ h_2^*)$ le fait les paramètres de lissage sélectionnés par ces deux dernières expressions nous fournissent pratiquement les mêmes résultats lorsque on utilise le noyau K_T. Globalement, dans ce cas, il est préférable d'utiliser le noyau triangulaire et de sélectionner le paramètre de lissage par la formule (4.20) ou (4.21) (voir Figures 4.5 et 4.6).

Bien que le noyau binomial négatif nous fournis, dans certaines situations, de cette seconde application, de bons résultats, il est essentiel de souligner que la qualité de ses performances dépend strictement de la taille de l'échantillon. En effet, pour d'autres tailles d'échantillon, le noyau K_{BN} peut nous fournir les pires estimations au sens du critère MSE (voir la première application). Ce qui fait, nous devons être très prudents lors de l'utilisation de ce noyau.

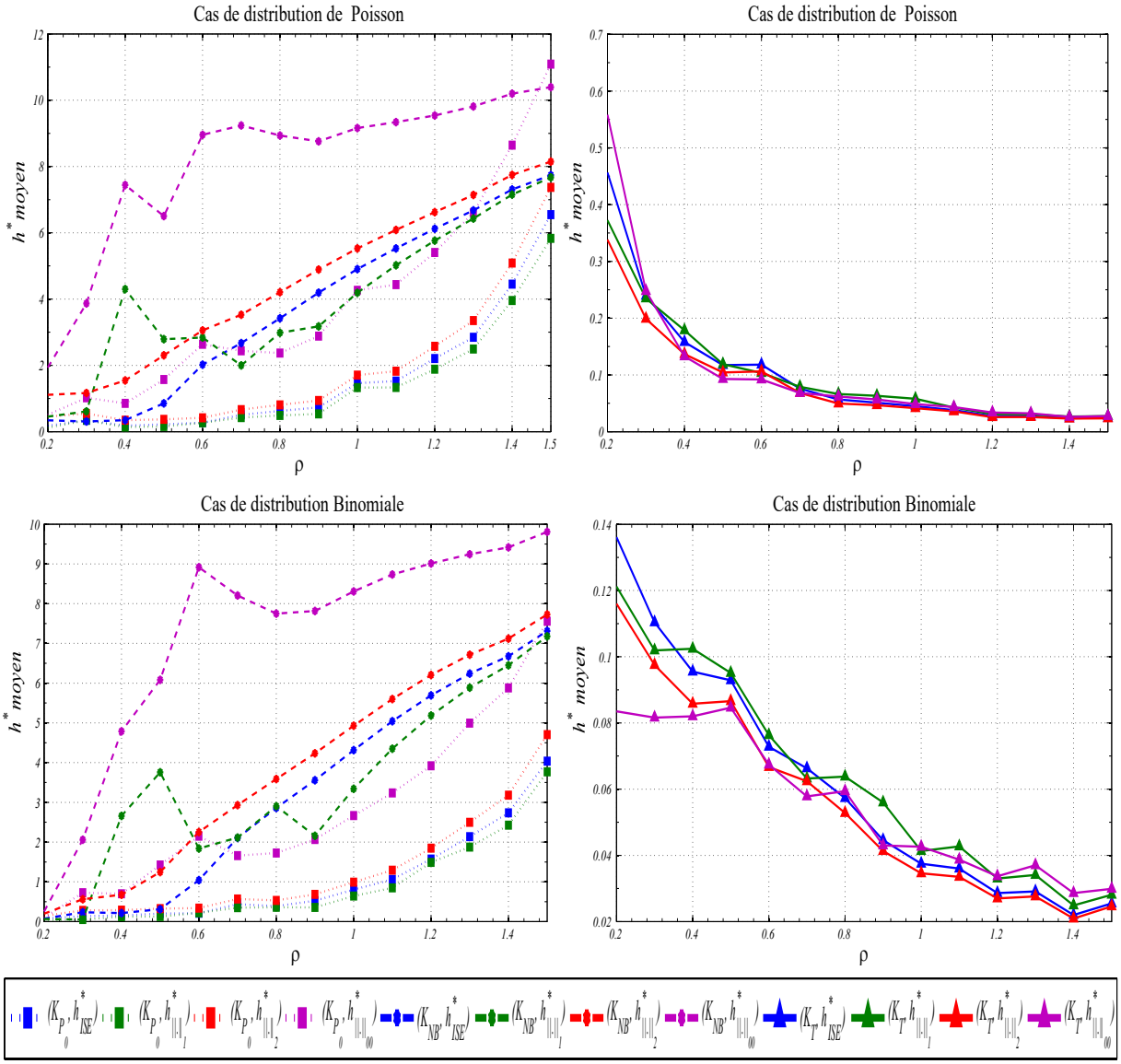

FIGURE 4.4 – Variation du h^* moyen en fonction de l'intensité du trafic ρ.

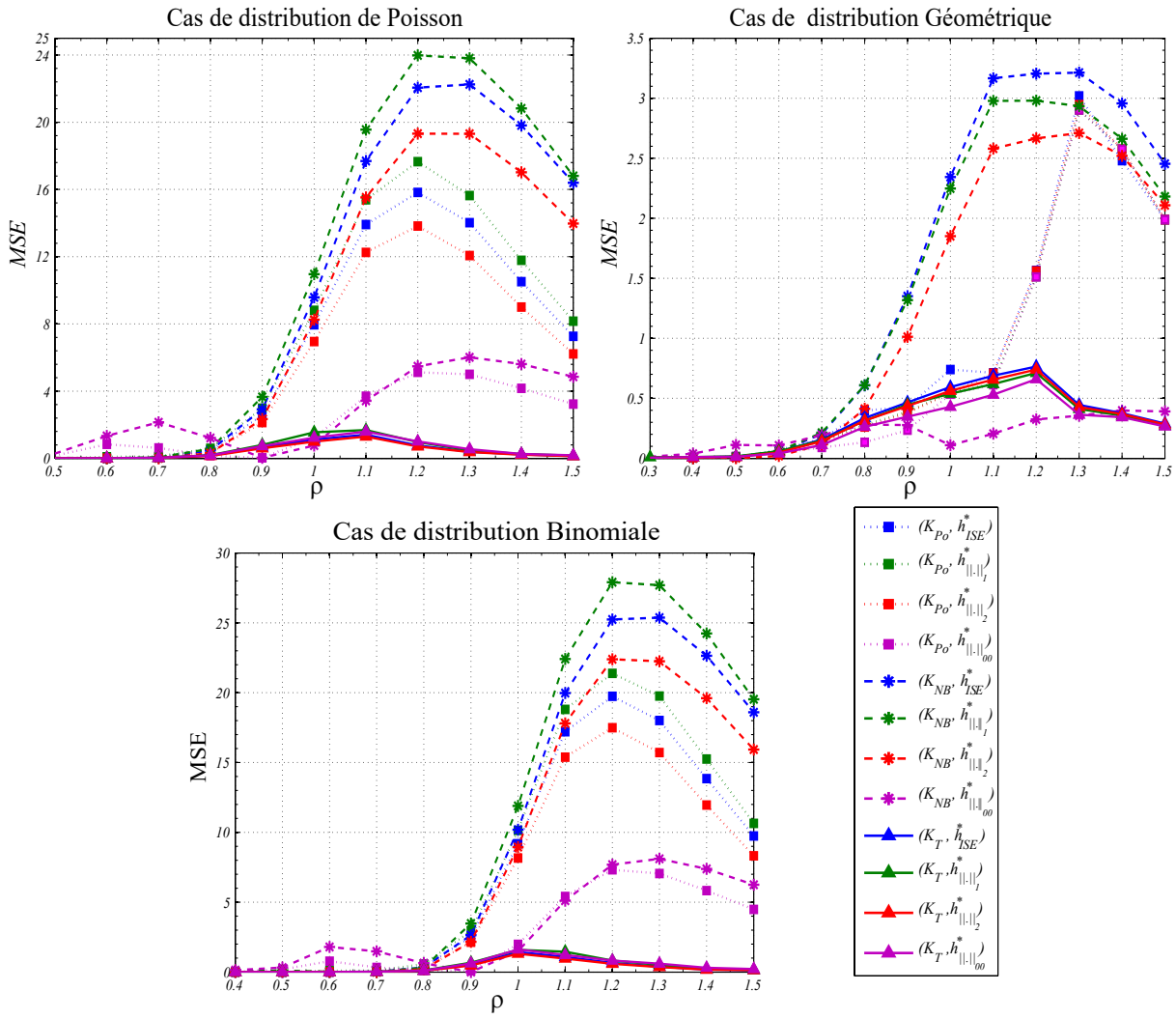

FIGURE 4.5 – Variation du MSE de l'estimateur de L_s en fonction de l'intensité du trafic ρ.

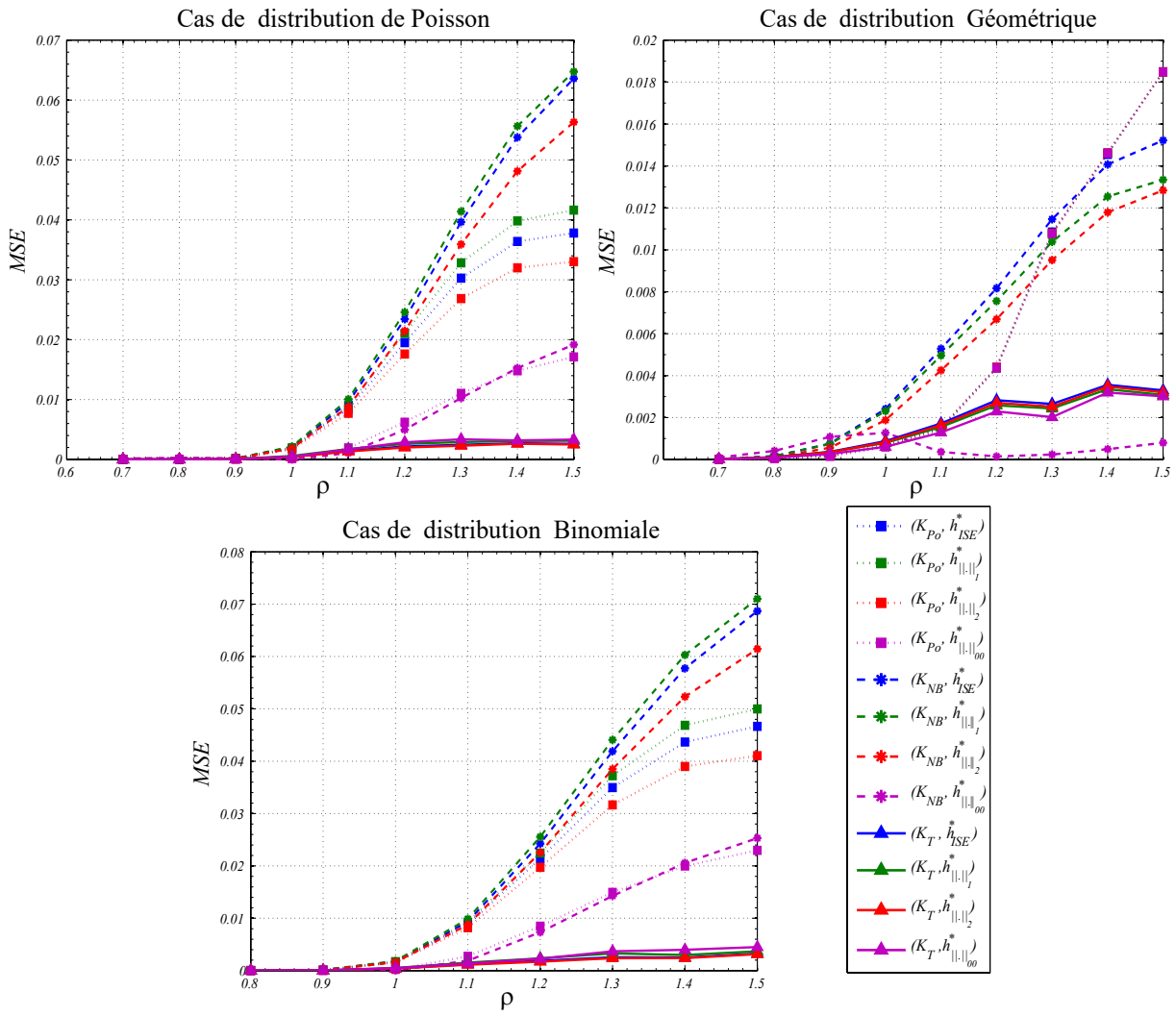

FIGURE 4.6 – Variation du MSE de l'estimateur de la probabilité de perte π_N en fonction de l'intensité du trafic ρ.

4.5 Remarques et Conclusions

Dans ce chapitre, nous avons considéré le choix du paramètre de lissage par des procédures basées sur la minimisation des normes matricielles dans l'estimation à noyaux discrets des matrices de transition associées aux chaînes de Markov, discrètes à temps discret, décrivant les files d'attente $GI/D/1/N$ et $D/G/1/N$. L'étude de simulation réalisée montre l'intérêt de la sélection du paramètre de lissage par la minimisation des normes matricielles, en particulier l'utilisation de la norme matricielle quadratique lorsqu'elle est combinée avec le noyau triangulaire. En revanche, elle nous suggère d'éviter l'utilisation du noyau Binomial Négatif dans le contexte abordé. Ces résultats ont été confirmé dans le cas de la chaîne de Markov décrivant le modèle de stock de type (R, s, S) [1, 2].

Afin de justifier et d'expliquer le comportement paradoxal de nos résultats (en particulier le comportement du paramètre de lissage optimal) en fonction de la taille de l'échantillon n dans certaines situations, rappelons d'abord que le biais global (bien connu, dans la littérature , sous la terminologie *le biais intégré*) et la variance globale (bien connue aussi sous la terminologie *la variance intégrée*) de l'estimateur d'une fonction de probabilité de masse f, notés respectivement $IBias$ et $IVar$, sont définis comme suit :

$$\begin{aligned} IBias(\hat{f}) &= h\, B\left(h, K, f, f''\right) \\ &= h\left\{\sum_{x\in\mathbb{N}}\left[f\left\{\mathbb{E}\left(\mathcal{K}_x, h(n)\right)\right\} - f(x) + \frac{V(K,x,h)}{2h} f''(x)\right]\right\}; \end{aligned} \tag{4.22}$$

$$IVar(\hat{f}) = \frac{1}{n}\sum_{x\in\mathbb{N}} f(x) Pr(\mathcal{K}_{x,h(n)} = x); \tag{4.23}$$

où $\mathcal{K}_{x,h}$ est la variable aléatoire de la loi $K_{x,h}$ défini sur $\aleph_{x,h}$ et $V(K,x,h)$ est donnée par :

$$V(K,x,h) = \begin{cases} x+h, & \text{si } K = K_{P_0}\,; \\ \frac{x-(x-1)h}{x+1}, & \text{si } K = K_{NB}\,; \\ \frac{(2x+1)x+(3x+1)h}{x+1}, & \text{si } K = K_T. \end{cases} \tag{4.24}$$

A partir de l'expression (4.23), on constate que lorsque la taille de l'échantillon tend vers l'infini, la variance de f tend vers zéro quel que soit le noyau et le paramètre de lissage utilisés utilisé pour la construction de $\hat{f}$, et ceci le fait que $\sum_{x\in\mathbb{N}} f(x) Pr(\mathcal{K}_{x,h(n)} = x)$ est une quantité finie (voir [51, 50]). Par conséquent, le comportement du paramètre de lissage optimal en fonction de la taille de l'échantillon n ne peut être justifié que par le comportement du biais global.

D'après l'expression (4.22), il est clair que le biais tend vers zéro si seulement $h \equiv h(n)$ tend vers zéro lorsque n tend vers l'infini et $B(h, K, f, f'')$ est une quantité finie. Cependant, l'expression (4.24), indique que, dans le cas des noyaux K_{P_0} et K_{NB}, la quantité $B(h, K, f, f'')$ est inversement proportionnelle à la valeur de h et ceci le fait que $V(x,h)/2h$ est décroissante en fonction de h (il est facile de vérifier que $((V(x,h)/2h)' < 0)$. Cela signifie que $B(h(n), K, f, f'')$ tend vers zéro lorsque n tend vers l'infini uniquement si le paramètre de lissage h est suffisamment grand, autrement dit, h doit tendre vers l'infini dans le cas des noyaux binomial négatif et de Poisson. A ce stade, nous concluons que le paramètre de lissage h doit être choisi de manière

à avoir un compromis entre h et $B(h, K, f, f'')$ et qui n'est pas forcément nul à la limite. Cette conclusion reflète parfaitement le comportement des résultats classés dans les Tables 4.4–4.6.

Concernant le noyau triangulaire, il faut noter que l'estimateur conçu via ce noyau est exempt du problème de convergence de son biais local et intégré. En effet, lorsque on considère le noyau triangulaire, l'expression (4.22) peut être réécrite comme suit :

$$IBias\left(\hat{f}\right) = \sum_{x\in\mathbb{N}} \left[f(x) \left\{ \frac{(a+1)^h}{P(a,h)} - 1 \right\} + \sum_{y\in\aleph_{x,h}\setminus\{x\}} f(y) Pr\left(\mathcal{K}_{T_{(a,h,x)}} = y\right) \right], \tag{4.25}$$

tendent vers zéro quand h tend vers zéro (n tend vers l'infini) et ceci le fait que $P(a,h) = (2a+1)(a+1)^h - 2\sum_{j=0}^{a} j^h$ tendent vers 1 et la probabilité $Pr\left(\mathcal{K}_{T_{(a,h,x)}} = y\right)$ tend vers zéro lorsque h tend vers zéro (pour plus de détails, voir [51]). Cette dernière conclusion coïncide également avec nos résultats.

L'impact négatif du biais de l'estimateur $\hat{f}$ sur la performance de l'estimateur à noyau d'une matrice de transition d'une chaîne de Markov a déjà été rapporté, par Cherfaoui et al. [28], dans le contexte de données continues ($x \in \mathbb{R}_+$). Par conséquent, à la lumière de nos résultats et de ceux de Cherfaoui et al. [28], il est naturel d'envisager l'introduction des techniques de réduction de biais pour une éventuelle amélioration des performances de l'estimateur à noyau (continu et discret) de la matrice de transition.

Conclusion générale

Dans l'analyse probabiliste d'un modèle Markovien décrivant un certain système, et l'évaluation analytique des mesures de performance de ce système, on suppose souvent que les paramètres de départ sont fixes et exacts. Cependant, en pratique, les valeurs de ces paramètres ne sont connues que sous la forme d'un échantillon de données. En ce sens, pour évaluer les performances du système considéré, l'utilisation de techniques statistiques d'estimation (paramétriques et/ou non paramétriques), qui visent à fournir une approximation des valeurs des paramètres (inconnus) en exploitant les informations fournies par l'échantillon, est inévitable.

Cette thèse porte sur l'application de la méthode d'estimation non paramétrique du noyau dans la théorie des chaînes de Markov. Plus précisément, on s'est intéressé à l'influence du couple (paramètre de lissage, noyau) sur les caractéristiques d'une chaîne de Markov, lorsque l'une des lois la régissant est inconnue et doit être remplacée par son estimateur à noyau. Afin de répondre à notre objectif, plusieurs notions et études ont été introduites dans le présent document.

Dans un premier lieu, pour une bonne prise en main des différentes notions liées à la problématique abordée, une synthèse sur la théorie de l'estimation à noyaux et sur les méthodes d'estimation paramétrique et non paramétrique dans les chaînes de Markov a été présentée.

Par la suite, afin de mettre en évidence les limites des méthodes paramétriques pour l'estimation dans les chaînes de Markov, nous avons analysé par simulation l'impact de l'estimation (paramétrique) des paramètres de départ d'une chaîne de Markov bidimensionnelle, décrivant le système d'attente $M/M/1$ à capacité finie (avec vacances multiples, Bernoulli feedback, balking, reneging et rétention des clients impatients et possibilité d'une panne et de réparation du serveur) sur les propriétés statistiques (biais, variance, risque quadratique,...) des estimateurs de ses mesures de performance stationnaires. Les résultats de simulation obtenus montrent d'une part, que la qualité des estimations des mesures de performance d'un système d'attente dépend du paramètre de départ estimé et d'autre part, afin d'obtenir des estimations des mesures de performance d'une qualité raisonnable, on doit disposer d'échantillons de grande taille, ce qui n'est pas évident dans la pratique.

Enfin, nous nous sommes intéressés au choix du paramètre de lissage dans l'estimation à noyau des matrices de transition associées aux chaînes de Markov décrivant les deux systèmes d'attente déterministes $D/G/1/N$ et $GI/D/1/N$. Afin de prendre en considération la forme des deux matrices, nous avons proposé d'utiliser les normes matricielles pour la sélection du paramètre de lissage. Pour cela, nous avons réalisé une étude comparative par simulation sur la qualité des estimateurs conçus, lorsque le paramètre de lissage est choisi par les méthodes

classiques et par d'autres méthodes qui se basent sur la minimisation des normes matricielles $\|.\|_1$, $\|.\|_2$ et $\|.\|_\infty$. Les résultats obtenus dans cette étude montrent l'intérêt de la sélection du paramètre de lissage par les méthodes basées sur la minimisation des normes matricielles ainsi que le choix minutieux du noyau pour la construction de l'estimateur des matrices de transition associées aux chaînes de Markov décrètes à temps discret. En effet, l'étude de simulation réalisée, sur des échantillons de différentes tailles et les différents noyaux discrets usuels, indique d'une part, qu'il est plus judicieux de sélectionner le paramètre de lissage par la minimisation des normes matricielles, en particulier l'utilisation de la norme matricielle quadratique lorsqu'elle est combinée avec le noyau triangulaire. D'autre part, elle nous suggère d'éviter l'utilisation du noyau Binomial Négatif dans le contexte abordé.

Parmi nos perspectives de recherche, nous citons :

- L'introduction des techniques de réduction du biais afin de réduire l'impact négatif du biais de l'estimateur $\hat{f}$ sur la performance de l'estimateur à noyau d'une matrice de transition d'une chaîne de Markov.
- La considération des processus Markoviens gouvernés par des distributions plus complexes (multimodales, à queue lourde, mélange de lois,...) pour généraliser et adapter la méthode d'estimation à noyaux dans la théorie des chaînes de Markov.
- Concrétiser l'idée des procédures de sélection du paramètre de lissage par la minimisation des normes matricielles en décortiquant leurs différentes composantes afin de dégager leurs propriétés théoriques (condition de convergence, vitesse de convergence,...).
- Élaborer des procédures explicites pour la sélection du paramètre de lissage par la minimisation des normes matricielles (Plug-in, validation croisée), de telle sorte qu'elles soient exploitables en pratique [2].

Bibliographie

[1] F. Afroun, D. Aïssani, and D. Hamadouche. Estimation à noyau discret dans le modèle de stock de type $(R; s; S)$. In *Proceedings : Colloque sur l'Optimisation et les Systèmes d'Information COSI'2019, ISBN : 978-9947-34-164-3*, pages 71–82, 2019.

[2] F. Afroun, D. Aïssani, and D. Hamadouche. Validation croisée non biaisée et estimation à noyau discret dans le modèle de stock de type $(R; s; S)$. In *Proceedings : K. Boukhetala et J.F. Dupuy, Modélisation Stochastique et Statistique, MSS'19, hal-02593238*, pages 347–350, 2019.

[3] F. Afroun, D. Aïssani, and D. Hamadouche. Nonparametric approximation of the characteristics of the $D/G/1$ queue with finite capacity. *Int. J. of Computing Science and Mathematics, Inderscience ed.*, pages 1–10, 2020. Ref : IJCSM 0061 Afroun et al. (3), 2020.

[4] F. Afroun, D. Aïssani, D. Hamadouche, and M. Boualem. Q–matrix method for the analysis and performance evaluation of unreliable $M/M/1/N$ queueing model. *Mathematical methods in the applied sciences, John Wiley ed.*, 41(18) : 9152–9163, 2018.

[5] J. Aitchison and C.G.G. Aitken. Multivariate binary discrimination by the kernel method. *Biometrika*, 63(3) : 413–420, 1976.

[6] A.S. Alfa. *Queueing theory for telecommunications : Discrete time modelling of a single node system*. Springer, New York, 2010.

[7] C.J.Jr Ancker and A.V. Gafarian. Some queuing problems with balking and reneging-I. *Operations Research*, 11(1) : 88–100, 1963.

[8] T.W. Anderson and L.A. Goodman. Statistical inference about Markov chains. *Institute of Mathematique Statistics*, 28(1) : 89–110, 1957.

[9] A. Bareche and D. Aïssani. Kernel density in the study of the strong stability of the $M/M/1$ queueing system. *Operations Research Letters*, 36(5) : 535–538, 2008.

[10] A.K. Basu and D.K. Sahoo. On berry–esséen theorem for nonparametric density estimation in Markov sequences. *Bull. Inform. Cybernet*, 30(1) : 25–39, 1998.

[11] P. Billingsley. *Statistical inference for Markov processes*. University of Chicago Press, 1961.

[12] F. Bonnieux. Modèle de conduite d'un élevage de veaux de boucherie. *Revue de statistique appliquée*, 20(3) : 71–84, 1972.

[13] F. Bonnieux. Estimation des probabilités de transition des chaînes de Markov. Technical Report 1–15, hal-01600049, Institut National de la Recherche Agronomique, 1974.

[14] 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, 2003.

[15] A.W. Bowman. An alternative method of cross-validation for the smoothing density estimates. *Biometrika*, 71(2) : 353–360, 1984.

[16] B.M. Brown and S. Chen. Beta-bernstein smoothing for regression curves with compact supports. *Scandinavian Journal of Statistics*, 21(1) : 47–59, 1999.

[17] O. Brun and J.M. Garcia. Analytical solution of finite capacity $M/D/1$ queues. *Journal of applied probability*, 39(4) : 853–864, 2000.

[18] P. Burman. A date dependent approach to density estimation. *Zeitschrift Für Wahrscheinlichkeitstheorie and Verwandte Gebiete*, 69(4) : 609–628, 1985.

[19] T. Cacoullos. Estimation of a multivariate density. *Ann. Inst. Statist. Math.*, 18(1) : 179–189, 1966.

[20] L. Le Cam. Locally asymptotically normal families of distributions. *University of California Publications in Statistics*, 3 : 37–98, 1960.

[21] L. Le Cam and G. Lo Yang. *Asymptotics in Statistics*, chapter Locally asymptotically normal families, pages 52–98. Springer Series in Statistics. Springer, New York, N, 1990.

[22] A. Charpentier, J.D. Fermanian, and O. Scaillet. *Copulas : from theory to application in finance*, volume ISBN : 190433945X, chapter The Estimation of Copulas : Theory and Practice, pages 35–64. London : Risk Books, 2007.

[23] S. Chen. Beta kernel estimators for density functions. *Computational Statistics & Data Analysis*, 31(2) : 131–145, 1999.

[24] S. Chen. Beta kernel for regression curve. *Statistica Sinica*, 10 : 73–92, 2000.

[25] S. Chen. Probability density function estimation using gamma kernels. *Annals of the Institute of Statistical Mathematics*, 52(3) : 471–480, 2000.

[26] M.Y. Cheng, J. Fan, and J.S. Marron. On automatic boundary corrections. *The Annals of Statistics*, 25(4) : 1691–1708, 1997.

[27] M. Cherfaoui, M. Boualem, D. Aïssani, and S. Adjabi. Choix du paramètre de lissage dans l'estimation à noyau d'une matrice de transition d'un processus semi-Markovien. *C. R. Acad. Sci. Paris, Ser. I.*, 353(3) : 273–277, 2015.

[28] M. Cherfaoui, M. Boualem, D. Aïssani, and S. Adjabi. Quelques propriétés des estimateurs à noyaux gamma pour des échantillons de petites tailles. *Afrika Statistika*, 10(1) : 763–776, 2015.

[29] P. Deheuvels and P. Hominal. Estimation non paramétrique de la densité compte tenu d'informations sur le support. *Revue de Statistique Appliquée*, 27 : 47–68, 1979.

[30] L. Devroye. The equivalence of weak, strong and complete convergence in $\mathbb{L}_1$ for kernel density estimates. *The Annals of Statistics*, 11(3) : 896–904, 1983.

[31] L. Devroye and L. Györfi. *Nonparametric Density Estimation : The* $\mathbb{L}_1$. View, New York ; John Wiley, 1985.

[32] L. Djerroud, T.S. Kiessé, and S. Adjabi. Semiparametric multiple kernel estimators and model diagnostics for count regression functions. *Communications in Statistics - Theory and Methods*, 49(9) : 2131–2157, 2020.

[33] M.O. Abou El-Ata and A.M.A. Hariri. The $M/M/c/N$ queue with balking and reneging. *Computers and Operations Research*, 19(13) : 713–716, 1992.

[34] M. S. El-Paoumy and H. A. Nabwey. The poissonian queue with balking function, reneging and two heterogeneous servers. *International Journal of Basic and Applied Sciences*, 11(6) : 149–152, 2011.

[35] A.K. Erlang. Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. *Elektroteknikeren*, 13 :5–13, 1917.

[36] M. Fernandez and P. Monteiro. Central limit theorem for asymmetric kernel functionals. *Annals of the Institute of Statistical Mathematics*, 57(3) : 425–442, 2005.

[37] J.M. Garcia, O. Brun, and D. Gauchard. Transient analytical solution of $M/D/1/N$ queues. *Journal of applied probability*, 39(4) : 853–864, 2002.

[38] T. Gasser and H.G. Müller. Kernel estimation of regression functions. *In : Gasser, T., Rosenblatt, M. (Eds.), Smoothing Techniques for Curve Estimation. In : Lecture Notes in Mathematics, Springer-Verlag, Heidelberg*, 757 : 23–68, 1979.

[39] G.M. Gontijo, G.S. Atuncar, F.R.B. Cruz, and L. Kerbache. Performance evaluation and dimensioning of $GI^{[X]}/M/c/N$ systems through kernel estimation. *Mathematical Problems in Engineering*, 2011 : 20 pages, 2011.

[40] C. Graham. *Markov Chains : Analytic and Monte Carlo Computations*. Wiley Series in Probability and Statistics. Wiley, 2014.

[41] A. Gravey, J.R. Louvion, and P. Boyer. On the $Geo/D/1$ and $Geo/D/1/N$ queues. *Performance Evaluation*, 11(2) : 117–125, 1990.

[42] F. A. Haight. Queueing with balking. *Biometrika*, 44(3-4) : 360–369, 1957.

[43] F. A. Haight. Queueing with reneging. *Metrika*, 2(1) : 186–197, 1959.

[44] P. Hall. Cross-validation in density estimation. *Biometrika*, 69(2) : 383–390, 1982.

[45] P. Hall and J.S. Marron. Local minima in cross-validation function. *Journal of the royal statistical society*, 90(1) : 149–173, 1991.

[46] M. Harunuzzaman and T. Aldemir. Optimization of standby safety system maintenance schedules in nuclear power plants. *Nuclear Technology*, 113(3) : 354–367, 1996.

[47] M.C. Jones. Simple boundary correction for kernel density estimation. *Statistical Computing*, 3(3) : 135–146, 1993.

[48] M.C. Jones and P. Foster. A simple nonnegative boundary correction method for kernel density estimation. *Statistica Sinica*, 6(4) : 1005–1013, 1996.

[49] T.S. Kiessé. Approche non-paramétrique par noyaux associés discrets des données de dénombrement. Thèse doctorat, Université de Pau et des Pays de l'Adour, 2008.

[50] C.C. Kokonendji and T.S. Kiessé. Discrete associated kernels method and extensions. *Statistical Methodology*, 8(6) : 497–516, 2011.

[51] C.C. Kokonendji, T.S. Kiessé, and S.S. Zocchi. Discrete triangular distributions and nonparametric estimation for probability mass function. *Journal of Nonparametric Statistics*, 19(6-8) : 241–254, 2007.

[52] C.C. Kokonendji and S.S. Zocchi. Extensions of discrete triangular distributions and boundary bias in kernel estimation for discrete functions. *Econometric Theory*, 80 : 1655–1662, 2010.

[53] R. Kumar. Economic analysis of an $M/M/c/N$ queuing model with balking, reneging and retention of reneged customers. *OPSEARCH*, 50(3) : 383–403, 2013.

[54] R. Kumar and S.K. Sharma. A markovian feedback queue with retention of reneged customers. *Advanced Modeling and Optimization*, 14(3) : 668–681, 2012.

[55] R. Kumar and S.K. Sharma. A multi-server markovian feedback queue with balking reneging and retention of reneged customers. *AMO-Advanced Modeling and Optimization*, (16)2 : 395–406, 2014.

[56] R. Kumar and S.K. Sharma. Optimization of an $M/M/1/N$ feedback queue with retention of reneged customers. *Operations research and decisions*, 24(3) : 45–58, 2014.

[57] A. Laksaci and A. Yousfate. Estimation fonctionnelle de la densité de l'opérateur de transition d'un processus de Markov à temps discret. *C. R. Acad. Sci. Paris*, 334(Ser. I) : 1035–1038, 2002.

[58] T.C. Lee and G.G. Judge. Estimation of transition probabilities in a non stationary finite Markov chains. *Metroeconomica*, 24(2) : 180–201, 1972.

[59] T.C. Lee, G.G. Judge, and A. Zellner. *Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data*. North-Holland Publ. Co, Amsterdam, 1970.

[60] M. Lejeune and P. Sarda. Smooth estimators of distribution and density functions. *Computational Statistics and Data Analysis*, 14 : 457–471, 1992.

[61] M. Loève. *Probability Theory, 3rd ed.* Van Nostrand, N. J., 1963.

[62] A. Madansky. Least squares estimation in finite Markov processes. *Psychometrika*, 24 : 137–144, 1959.

[63] J. S. Marron and D. Ruppert. Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society*, 56(Series B) : 653–671, 1994.

[64] E. Masry and L. Györfi. Strong consistency and rates for recursive probability density estimators of stationary processes. *J. Multivariate Anal.*, 22 : 79–93, 1987.

[65] G.A. Miller. Finite Markov processes in psychology. *Psychometrika*, 17 : 149–167, 1952.

[66] C. Misra and V. Goswami. Analysis of power saving class II traffic in IEEE 802.16E with multiple sleep state and balking. *Foundations of computing and decision sciences*, 40(1) : 53–66, 2015.

[67] H.G. Müller. Smooth optimum kernel estimators near endpoints. *Biometrika*, 78 : 521–530, 1991.

[68] E. Nadaraya. On nonparametric estimation density function and regression. *Theory Probab. P.P.L*, 10 : 186–190, 1965.

[69] J.R. Norris. *Markov chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1997.

[70] T.J. Ott. Simple inequalities for the $D/G/1$ queue. *Operations Research*, 35(4) : 589–597, 1987.

[71] G. Panda and V. Goswami. Equilibrium balking strategies in renewal input queue with bernoulli-schedule controlled vacation and vacation interruption. *Journal of industrial and management optimization*, 12(3) : 851–878, 2016.

[72] B.U. Park and S.J. Marron. Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, 85 : 66–72, 1990.

[73] E. Parzen. On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33 : 1065–1076, 1962.

[74] F.S. Pranoto and A. Wirawan. Preliminary design of redundancy management for LSA -02 automatic flight control system. *Journal of Physics : Conference Series*, 1130(1) : 012025, 2018.

[75] N. Privault. *Understanding Markov Chains : Examples and Applications*. Springer Undergraduate Mathematics Series. Springer Singapore, 2018.

[76] J. Roberts, U. Mocci, and J. Virtamo. Broadband network teletraffic. Final report of action COST 242, Springer, Berling, 1996.

[77] M. Rosenblatt. Remarks in some nonparametric estimates of a density function. *Ann. Math. Statist.*, 27 : 832–837, 1956.

[78] G.G. Roussas. Asymptotic inference in Markov processes. *Ann. Math. Statist.*, 36(3) : 978–992, 1965.

[79] G.G. Roussas. Nonparametric estimation in Markov processes. *Annals of the Institute of Statistical Mathematics*, 21(1) : 73–87, 1969.

[80] G.G. Roussas. Nonparametric estimation of the transition distribution function of a Markov processes. *Ann. Math. Statist.*, 40(4) : 1386–1400, 1969.

[81] M. Rudemo. Empirical choice of histogram and kernel density estimators. *Scandinavian Journal of Statistics*, 9 : 65–78, 1982.

[82] N. Saadi and S. Adjabi. On the estimation of the probability density by trigonometric series. *Communications in Statistics : Theory and Methods*, 38(19) : 3583–3595, 2009.

[83] A. Sadek and N. Limnios. Asymptotic properties for maximum likelihood estimators for reliability and failure rates of Markov chains. *Communication in Statistics-Theory and Methods*, 31(10) : 1837–1861, 2002.

[84] O. Scaillet. Density estimation using inverse and reciprocal inverse gaussian kernels. *Journal of Nonparametric Statistics*, 1–2(16) : 217–226, 2004.

[85] E. Schuster. Incorporating support constraints into nonparametric estimators of densities. *Communications in Statistics- Theory and Methods*, 14 : 1123–1136, 1985.

[86] D.W. Scott. Averaged shift histograms : effective nonparametric density estimators in several dimensions. *The annals of statistics*, 13 : 1024–1040, 1985.

[87] D.W. Scott and G.R. Terrell. Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, 82 : 1131–1146, 1987.

[88] L.D. Servi. $D/G/1$ queues with vacations. *Operations Research*, 34(4) : 619–629, 1986.

[89] S.J. Sheather and M.C. Jones. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(3) : 683–690, 1991.

[90] B. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London, 1986.

[91] B.W. Silverman. Weak and strong uniform consistency of the kernel estimate of density function and its derivates. *Ann. Statist.*, 6 : 177–184, 1978.

[92] S.M. Somé and C.C. Kokonendji. Effects of associated kernels in nonparametric multiple regressions. *Journal of Statistical Theory and Practice*, 10(2) : 456–471, 2016.

[93] R. Stolletz. *Performance Analysis and Optimization of Inbound Call Centers*. Berlin, Heidelberg : Springer Berlin Heidelberg, 2003.

[94] C. Stone. An asymptotically optimal window selection rule for kernel density estimates. *The Annals of Statistics*, 12 : 1285–1297, 1984.

[95] H. Theil and G. Rey. A quadratic programming approach to the estimation of transition probabilities. *Management Science*, 12 : 714–721, 1966.

[96] N. Vicari and P. Tran-Gia. A numerical analysis of the $Geo/D/N$ queueing system. Technical report 151, Institute of Computer Science, University of Würzburg, 1996.

[97] G. Wahba. Optimal convergence properties of variable knot, kernel, and orthogonal series methods for density estimation. *Ann. Stat.*, 3 : 15–29, 1975.

[98] G. Wahba. Data-based optimal smoothing of orthogonal series density estimates. *The Annals of Statistics*, 9(1) : 146–156, 1981.

[99] K.H. Wang and Y.C. Chang. Cost analysis of a finite $M/M/R$ queueing system with balking, reneging and server breakdowns. *Mathematical Methods of Operations Research*, 56 : 169–180, 2002.

[100] W.E. Wansouwé, S.M. Somé, and C.C. Kokonendji. An R package for discrete and continuous associated kernel estimations. *The R Journal*, 8(2) : 258–276, 2016.

[101] G. S. Watson and M. R. Leadbetter. On the estimation of the probability density,i. *Ann. Math. Statist.*, 34 : 480–491, 1963.

[102] P. Whittle. On the smoothing of probability density functions. *Journal of the Royal Statistical Society : Series B (Methodological)*, 20(2) : 334–343, 1958.

[103] S.K. Yang. *Handbook of Materials Failure Analysis with Case Studies from the Aerospace and Automotive Industries*, chapter 12- A failure-processing scheme based on Kalman prediction and the reliability analysis for 25 kVA generators used on IDF, pages 237–260. Elsevier Ltd., Butterworth-Heinemann, 2016.

[104] W. Yue, Y. Takahashi, and H. Takagi. Advances in queueing theory and network applications. *New York, N.Y*, 2009.

[105] S. Zhang. A note on the performance of the gamma kernel estimators at the boundary. *Statistics and Probability Letters*, 80 : 548–557, 2010.

[106] S. Zhang and R. J. Karunamuni. On kernel density estimation near endpoints. *Journal of Statistical Planning and Inference*, 70 : 301–316, 1998.

[107] S. Zhang and R.J. Karunamuni. On nonparametric density estimation at the boundary. *Nonparametric Statistics*, 12 : 197–221, 2000.

[108] Y. Zhang, D. Yue, and W.Yue. Analysis of an $M/M/1/N$ queue with balking, reneging and server vacations. *International Symposium on OR and Its Applications*, pages 37–47, 2005.

[109] N. Zougab. Aproche bayésienne dans l'estimation non paramétrique de la densité de probabilité et la courbe de régression de la moyenne. Thèse de doctorat, Université Abderrahmane Mira de Bejaïa, 2013.

RÉSUMÉ

L'objet de cette thèse est d'analyser l'influence du couple (paramètre de lissage, noyau) sur les estimations des caractéristiques d'une chaîne de Markov (modélisant un système de file d'attente) lorsque l'une des lois générales qui la régit est entièrement inconnue et remplacée par son estimateur à noyau. En effet, les études de simulation menées dans ce cadre ont mis en évidence les limites des méthodes d'estimation paramétriques dans les chaînes de Markov. Ainsi, les résultats obtenus dans l'analyse des propriétés statistiques (biais, variance, MSE, distribution) des mesures de performance de la chaîne de Markov bidimensionnelle, décrivant le système d'attente $M/M/1/N$ (avec vacances multiples, Bernoulli feedback, balking, reneging et rétention des clients impatients et possibilité de panne et réparation du serveur) montrent que la qualité des estimations de ces mesures dépend du paramètre de départ estimé et que des échantillons de grande taille sont nécessaires pour obtenir des estimations d'une qualité raisonnable. D'autre part, ces études de simulation ont également mis en évidence l'intérêt de la sélection du paramètre de lissage par les méthodes basées sur la minimisation des normes matricielles ainsi que le choix minutieux du noyau dans la construction de l'estimateur à noyau des matrices de transition associées aux chaînes de Markov discrètes (décrivant les deux files d'attente $D/G/1/N$ et $GI/D/1/N$). En effet, nos résultats indiquent qu'il est plus judicieux de sélectionner le paramètre de lissage en minimisant les normes matricielles, en particulier la norme $\|.\|_2$ combinée avec le noyau triangulaire, et éviter l'utilisation du noyau binomial négatif.

Mots clés : Estimation ; Chaînes de Markov ; Files d'attente ; Mesures de performance ; Paramètre de lissage ; Noyaux discrets ; Simulation.

ABSTRACT

The object of this thesis was to analyse the influence of the couple (smoothing parameter, kernel) on the estimates of a Markov chain characteristics when one of the general laws governing it is entirely unknown and replaced by its kernel estimator. The simulation studies carried out in this framework, on the one hand, have highlighted the limits of parametric estimation methods in Markov chains where the results obtained in the analysis of statistical properties (bias, variance, MSE, distribution) of the performance measures of the two-dimensional Markov chain, describing the $M/M/1/C$ queue with multiple vacations, Bernoulli feedback, balking, reneging and retention of impatient customers and the possibility of server failure and repair, show that the quality of the estimates of these measures depends on the estimated starting parameter and that large-sizes sample are necessary to obtain estimates of reasonable quality. On the other hand, the interest of the selection of the smoothing parameter by the methods based on the minimization of the matrix norms as well as the judicious choice of the kernel in the construction of the kernel estimator of the transition matrices associated with the discrete-times Markov chains describing the two $D/G/1/C$ and $GI/D/1/C$ queues. Indeed, our results indicate that it is more preferable to select the smoothing parameter by minimizing the matrix norms, in particular, the norm $\|.\|_2$ combined with the triangular kernel, and avoids the use of the negative binomial kernel.

Key words : Estimation ; Markov chains ; Queues ; Performance measures ; Smoothing parameter ; Discrete kernels ; Simulation.