

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA
RECHERCHE SCIENTIFIQUE
UNIVERSITÉ Abderrahmane Mira - Béjaia
FACULTÉ DES SCIENCES EXACTES
Département de Mathématiques

En vue de l'obtention du diplôme de Master en Mathématiques
Option : Statistique et Analyse Décisionnelle

THÈME

Estimation non paramétrique du taux de défaillance
par
la méthode des fonctions orthogonales

Présenté par:

- ACHERCHOUR Atika
- MAZOUZ Samia

Soutenu publiquement, le devant le jury composé de :

Mme K.LAGHA M.C.B Université A-Mira de Béjaia. Promotrice.
Université A-Mira de Béjaia.

Promotion: 2016/2017

Remerciements

Au terme de notre modeste travail, nous tenons à remercier tout d'abord Dieu, le tous puissant de nous avoir accordé la volonté, force, santé, détermination et courage afin d'accomplir notre mémoire.

Nous tenons à exprimer nos profondes gratitudees à notre promotrice Mme LAGHA.K pour la manière dont elle nous a encadré et pour nous avoir proposé ce sujet. Nous sommes également reconnaissantes pour la confiance qu'elle nous a accordé, ses remarques, sa compréhension, ses conseils et sa grande disponibilité qui nous a donné plus de volonté pour réaliser ce mémoire.

Nous adressons aussi nos remerciements à tous les membres du jury, qui nous font l'honneur de juger notre travail.

Tout particulièrement nous souhaitons remercier vivement Mr BOUMZAID Yacine et M^{elle} ZIANE Yasmina de l'unité de recherche LaMOS qui nous ont beaucoup aidé à réaliser ce travail.

Qu'il nous soit permis enfin, de remercier tous ceux et celles qui ont contribué de manière directe ou indirecte pour la réalisation de ce travail.

Dédicaces

C'est avec profonde gratitude et s'insère mots, que je dédie ce modeste travail de fin d'étude :

À mes chers parents qui ont sacrifié leurs vies pour notre réussite et nous éclairer le chemin par leur puissant amour et leurs conseils judicieux, j'espère qu'un jour je pourrais leur rendre un peu de ce qu'ils ont fait pour moi, que Dieu leur prête bonheur et long vie.

À mes chères sœurs Randja, Saida, Soraya, Hassina, Nassima et leur époux sans oublié ma petite sœur Amina.

À mes frères Mussa, Mahdi, Nassim et leur femme, Djamel et Boukhalfa.

À toutes mes nièces et neveux.

À tous mes oncles, tantes, cousins et cousines.

À mon âme Juba et toute sa famille.

À mon binôme Samia et toute sa famille.

Atika ACHERCHOUR.

Dédicaces

C'est avec profonde gratitude et s'insère mots, que je dédie ce modeste travail de fin d'étude :

À mes chers parents qui ont sacrifié leurs vies pour notre réussite et nous éclairer le chemin par leur puissant amour et leurs conseils judicieux, j'espère qu'un jour je pourrais leur rendre un peu de ce qu'ils ont fait pour moi, que Dieu leur prête bonheur et long vie.

À ma sœur Ouahiba et son mari Bouelem et toute sa famille.

À mon grand frère Nabil et sa femme Syhem et toute sa famille et a mon petit frère Ghiles

À mon petit poussin Aksel .

À tous mes oncles, tantes, cousins et cousines.

À mon chère Mouhoub ainsi qu'à tous mes amies sans exeption.

À mon binôme Atika et toute sa famille.

Samia MAZOUZ.

Table des matières

Introduction	1
1 Principales caractéristiques probabilistes de la fiabilité	4
1.1 Introduction	4
1.2 Objectifs et intérêts de la fiabilité	5
1.3 Fonction de défaillance	5
1.4 Fonction de fiabilité ou fonction de survie	6
1.5 Taux de défaillance (taux de hasard)	6
1.6 Calcul du taux défaillance	7
1.7 Principales lois de fiabilité	9
1.8 Fiabilité des systèmes	12
1.9 Conclusion	14
2 Estimation non paramétrique	15
2.1 Introduction	15
2.2 Critères d'erreurs et définitions	15
2.3 Estimation par Histogramme	17
2.4 Estimation par la méthode du noyau	18
2.5 Estimation par des fonctions orthogonales	23
2.5.1 Principe de la méthode	23
2.5.2 Estimation de la densité	24
2.5.3 Estimation de la fonction de répartition	28
2.5.4 Choix pratique de la base	32

2.5.5	Exemples d'estimateurs	33
2.6	Conclusion	37
3	Estimation du taux de défaillance: application à base de Dirichlet	38
3.1	Introduction	38
3.2	Estimation de la densité	39
3.2.1	Propriétés des $(\hat{a}_k)_{k \in \mathbb{N}}$	40
3.2.2	Propriétés de \hat{f}_{d_n}	44
3.3	La fonction de répartition	46
3.3.1	Propriétés de \hat{F}_{d_n}	47
3.4	Estimation de la fiabilité et du taux défaillance	49
3.5	Simulation	49
3.5.1	Position du problème	49
3.5.2	Algorithme de simulation	50
3.5.3	Résultats de simulation	51
3.5.4	Interprétation des résultats	62
3.6	Conclusion	63
	Conclusion générale	64
	Bibliographie	70

Introduction générale

La théorie de l'estimation est une des préoccupations majeures des statisticiens. Ainsi l'estimation non paramétrique réelle a reçu un intérêt croissant tant sur le plan théorique que pratique. Cette branche de la statistique ne se résume pas à l'estimation d'un nombre fini de paramètres réels associés à la loi de l'échantillon (comme cela est le cas pour la théorie de l'estimation paramétrique), elle consiste généralement à estimer à partir des observations une fonction inconnue, élément d'une certaine classe fonctionnelle (espaces fonctionnels, de dimension infinie), telle que la fonction de densité, la fonction de répartition, la fonction de régression, ...etc.

Parmi les méthodes les plus basiques de la statistique non paramétriques dédiée à l'estimateur de la fonction de densité de probabilité, on cite la méthode d'histogramme dont l'origine est attribuée à John Grant XVII ème siècle. Les propriétés des estimateurs par histogramme ont été initialement étudiées par Geffroy (1974) [22] et Abou-Jaoude (1976) [2]. Cette méthode d'estimation est naturelle très rependue car elle est facilement implémentable. Cependant, l'estimateur de densité obtenu par cette méthode ne peut pas être adapté à la situation assez courante où nous disposons d'une information à priori sur la régularité de la densité à estimer. Plus précisément, si l'on sait par avance que la densité de l'échantillon observé est, par exemple, deux fois continûment différentiable, on aurait naturellement envie d'estimer cette densité par une fonction qui, elle aussi, est deux fois continûment différentiable. Or, les histogrammes sont des fonctions qui ne sont même pas continus. L'idée de lisser les histogrammes a donné naissance à la classe d'estimateurs dite la classes des estimateurs à noyaux, introduite par Rosenblatt (1956) [37] et Parzen (1962) [36] pour une densité univariée. L'estimateur obtenu par cette méthode dépend

d'une fonction noyau et d'un paramètre de lissage. Une méthode alternative est donnée par la méthode des fonctions orthogonales appelée aussi méthode des séries orthogonale ou encore méthode des polynômes orthogonaux.

Dans ce mémoire, nous nous intéressons à la méthode des fonctions orthogonales pour l'estimation du taux de hasard.

Cette méthode a été développée à partir des travaux de Cencov (1962) [11] pour l'estimation de la densité. L'intérêt de cette méthode réside dans le caractère global de l'estimateur. Pour peu que le développement d'une fonction par rapport à une base orthogonale converge rapidement, en ramenant l'estimation de cette dernière à celle des premiers coefficients de ce développement, on obtiendra, une approximation globale de cette fonction qui reste satisfaisante même dans les zones où sont apparues peu d'observations, puisqu'on estime les coefficients du développement par l'ensemble de l'échantillon (pour plus de détails voir ([39], [50], [26], [27], [56], [28])). Il est intéressant de noter que les développements de la densité en série relativement à une base orthogonale sont beaucoup plus anciens que les travaux de Cencov, notamment pour les développements d'Edgeworth [18] et Gram-Charlier [9]. La différence essentielle entre la théorie moderne et la théorie anciennes réside dans le fait que ces dernières utilisaient la méthode des moments pour les estimations des coefficients. Concernant l'estimation de la fonction de répartition par la méthode des fonctions orthogonales plusieurs études ont été effectuées à savoir les études de Bosq et Lecoutre [4], les estimateurs par ondelettes, les méthodes combinatoires, développées par Devroye et Lugosi [24], apportent un regard nouveau sur le choix efficace d'estimateurs non paramétriques.

Le taux de défaillance est l'une des mesures largement étudiés dans plusieurs domaines à savoir : Statistique, mécanique, électronique...etc. L'estimation du taux de défaillance, de part les variétés de ses possibilités d'application, de nombreuses études ont été effectuées sur l'utilisation du taux de défaillance, on peut citer les travaux de : Ferraty, Rabhi, Vieu et Al (1981) [19], Los, Mack et Wang (1989) [15], Yazourh [57] et Nassiri [35]. Dans les travaux concernant le taux de défaillance quelques études ont néanmoins déjà utilisé la méthode alternative des fonctions orthogonales [57].

L'objectif de ce travail est d'appliquer la méthode des fonctions orthogonales pour estimer le taux de défaillance. Pour cela, nous nous utilisons deux estimateurs différents à savoir l'estimateur de la densité et celui de la fonction de répartition relativement à base de Dirichlet. Ces deux derniers seront combinés afin d'estimer la fonction du taux de hasard (taux de défaillance).

Ce mémoire est composé d'une introduction générale, de trois chapitres et d'une conclusion générale.

Le premier chapitre sera consacré aux principales caractéristiques probabilistes de la fiabilité et du taux de défaillance.

Le second chapitre aborde les différentes méthodes d'estimation non paramétrique de la densité f et de la fonction de répartition F , et met l'accent sur la méthode des séries orthogonales.

Dans le dernier chapitre, nous proposons d'appliquer la base de Dirichlet. Nous présentons enfin les résultats de calcul par simulation de la fiabilité et du taux de défaillance.

Ce document se termine sur une conclusion qui mentionne les objectifs atteints et ouvre des perspectives à notre travail.

Principales caractéristiques probabilistes de la fiabilité

1.1 Introduction

La fiabilité est l'une des composantes essentielles de la qualité d'un produit et elle est retenue en tant que critère fondamental pour leur élaboration. Elle est prise en considération dès le stade de la conception.

La performance d'une entreprise est dépendante de la fiabilité de son système, ce dernier subit une défaillance quand il ne peut plus délivrer le service attendu, alors la panne est l'état du système résultant d'une défaillance, qui se produit de façon aléatoire. Il est logique alors, de faire appel au calcul de probabilités pour étudier les problèmes de fiabilité. Ainsi, nous définissons la fiabilité d'un dispositif comme étant sa probabilité de fonctionner correctement pendant une durée donnée. Autrement dit la probabilité qu'aucune défaillance ne se produise pendant cette durée. Dans ce qui suit nous donnons un petit aperçu sur les principaux éléments de la théorie de fiabilité ainsi que sur leur propriétés, particulièrement "le taux de défaillance".

1.2 Objectifs et intérêts de la fiabilité

L'analyse de la fiabilité constitue une phase indispensable dans toute étude de sûreté de fonctionnement. A l'origine, la fiabilité concernait les systèmes à haute technologie (centrales nucléaires, aérospatial). Aujourd'hui, la fiabilité est devenue un paramètre clé de la qualité et d'aide à la décision, dans l'étude de la plupart des composants, produits et processus "grand public": Transport, énergie, bâtiments, composants électroniques, composants mécaniques. . . .

De nombreux industriels travaillent dans l'évaluation et l'amélioration de la fiabilité de leurs produits au cours de leur cycle de développement, de la conception à la mise en service (conception, fabrication et exploitation) afin de développer leurs connaissances sur le rapport Coût/Fiabilité et maîtriser les sources de défaillance [53].

1.3 Fonction de défaillance

Soit T la *v.a* représentant la durée de vie d'un dispositif (i.e, le temps qui sépare la mise en service et la défaillance). On note par $f(t)$ la densité de probabilité de la *v.a* T et $F(t)$ sa fonction de répartition. La densité de probabilité de la *v.a* T est appelée densité de défaillance ou probabilité de panne de l'élément à l'instant t [29]. Cette dernière est donnée par l'équation suivantes :

$$f(t) = \lim_{dt \rightarrow 0} \left[\frac{F(t + dt) - F(t)}{dt} \right]. \quad (1.1)$$

La fonction de défaillance notée $F(t)$ représente la fonction de répartition de la *v.a* T . C'est la probabilité qu'un dispositif prélevé au hasard dans la population considérée ait une défaillance avant l'instant t . Elle est définie par l'équation suivante :

$$F(t) = P(T \leq t) = \int_0^t f(t)dt \text{ pour } t \geq 0. \quad (1.2)$$

Tels que : $0 \leq F(t) \leq 1$ et $t \in [0, +\infty[$.

1.4 Fonction de fiabilité ou fonction de survie

La fiabilité d'un système non réparable est la fonction du temps R , représentant la probabilité qu'un dispositif prélevé au hasard dans la population considérée n'ait pas de défaillance avant l'instant t . Elle est définie par :

$$R(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t). \quad (1.3)$$

Soit R une fonction de fiabilité, alors nous avons les propriétés suivantes :

- 1) $R(0) = 1$.
- 2) $\lim_{t \rightarrow \infty} R(t) = 0$.
- 3) $P(t_1 \leq T \leq t_2) = F(t_2) - F(t_1) = R(t_1) - R(t_2)$.

De plus, la présentation graphique de cette fonction est donnée dans la figure (1.1)

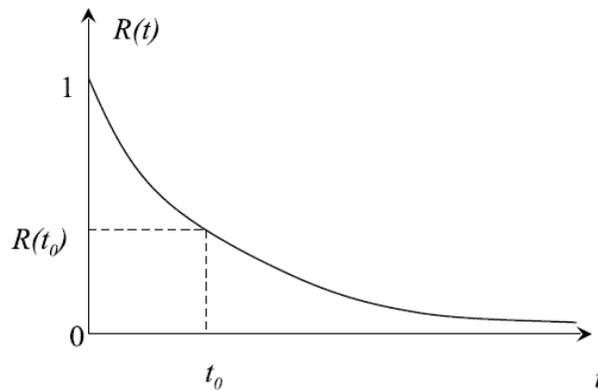


FIG.1.1 Courbe de survie ou de fiabilité.

1.5 Taux de défaillance (taux de hasard)

Le taux de défaillance appelé aussi taux de hasard ou taux de panne est la fréquence avec laquelle un système ou un composant défaillant est exprimé par exemple "en pannes par heure". Il est souvent indiqué par la lettre grecque λ (lambda) et il est d'une importance cruciale dans la théorie de la fiabilité.

Le taux de défaillance est une fonction de temps qui dépend d'un corollaire intuitif. Autrement dit, le taux de défaillance change au fil du temps par rapport au cycle de vie attendu d'un système.

Le taux de défaillance $\lambda(t)$ est la probabilité que l'équipement fonctionne encore après t unité de temps et tombe en panne durant les dt unités supplémentaires. C'est donc la probabilité conditionnelle que l'équipement tombe en panne entre t et $t + dt$, sachant qu'il a fonctionné sans défaillance entre 0 et t .

1.6 Calcul du taux défaillance

Le taux de défaillance d'un système non réparable est la fonction du temps λ définie par :

$$\lambda(t) = \lim_{dt \rightarrow 0} \frac{1}{dt} P(t < T \leq t + dt | T > t), \quad \forall t \geq 0.$$

L'interprétation du taux de défaillance est liée à celle de la densité de la façon suivante [19] :

$$\begin{aligned} f(t) = F'(t) &= \lim_{dt \rightarrow 0} \frac{F(t + dt) - F(t)}{dt} \\ &= \lim_{dt \rightarrow 0} \frac{1}{dt} [P(T \leq t + dt) - P(T \leq t)] \\ &= \lim_{dt \rightarrow 0} \frac{1}{dt} P(t \leq T \leq t + dt), \quad t > 0. \end{aligned}$$

On a donc, pour un dt assez petit :

$$f(t)dt \approx P(t \leq T \leq t + dt).$$

D'où,

$$\lambda(t)dt = P(t < T \leq t + dt | T > t).$$

Il est facile d'établir les liens entre le taux de défaillance et la fiabilité :

$$\begin{aligned}\lambda(t) &= \lim_{dt \rightarrow 0} \frac{1}{dt} P(t < T \leq t + dt | T > t) = \lim_{dt \rightarrow 0} \frac{1}{dt} \frac{P(t < T \leq t + dt)}{P(T > t)} \\ &= \frac{1}{R(t)} \lim_{dt \rightarrow 0} \frac{1}{dt} [F(t + dt) - F(t)] = \frac{f(t)}{R(t)}.\end{aligned}$$

D'où

$$\lambda(t) = \frac{-R'(t)}{R(t)} = -\frac{\partial}{\partial t} \ln R(t), \quad t > 0. \quad (1.4)$$

En intégrant et en prenant comme condition initiale $R(0) = 1$ (on suppose que le système fonctionne à l'instant initial), on obtient la formule d'exponentiation :

$$R(x) = e^{(-\int_0^x \lambda(t) dt)}, \quad x > 0. \quad (1.5)$$

Le taux de défaillance cumulé

Le taux de défaillance cumulé ou taux de hasard cumulé d'un système non réparable est la fonction du temps H , définie par :

$$\forall x \geq 0, \quad H(x) = \int_0^x \lambda(t) dt = -\ln(R(x)). \quad (1.6)$$

À partir des relations précédentes, on peut facilement montrer que :

- la fiabilité $R(x) = e^{(-H(x))}$, $t > 0$,
- la fonction densité $f(x) = \lambda(x)e^{(-\int \lambda(x) dx)}$, $t > 0$,
- la fonction de répartition $F(x) = 1 - e^{(-H(x))}$, $t > 0$.

La courbe en baignoire

La courbe en baignoire (voir la figure (1.2) ci dessous), donne l'évolution du taux de défaillance $\lambda(t)$ en fonction de l'âge du matériel. Elle comprend trois phases, chacune avec un sens de variation différent [20].

- La période de jeunesse : elle est caractérisée par des pannes précoces (défauts de conception et surproduction), dans laquelle $\lambda(t)$ est décroissant .
- La vie utile : pendant cette période, les défaillances sont purement accidentelles, et $\lambda(t)$ est pratiquement constant .

- Le vieillissement : $\lambda(t)$ se remet à croître car le risque de défaillance va finir par augmenter à cause de l'usure du système.

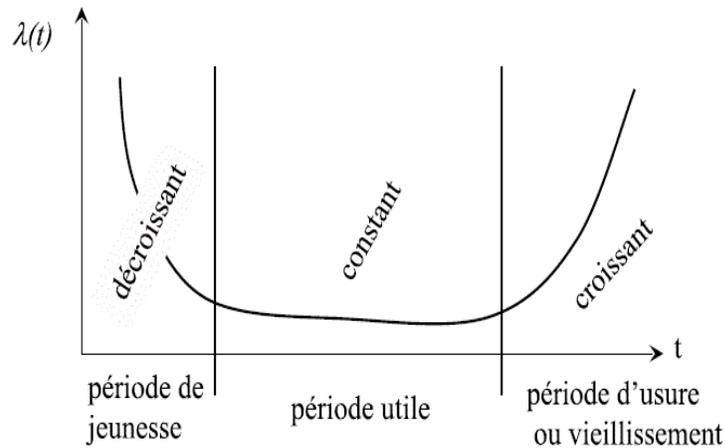


FIG.1.2 Courbe du taux de défaillance en baignoire.

Remarque 1.6.1 Lorsque la période de maturité (la vie utile) est négligeable la courbe en baignoire se transforme en courbe de parabole (voir la figure (1.3)).

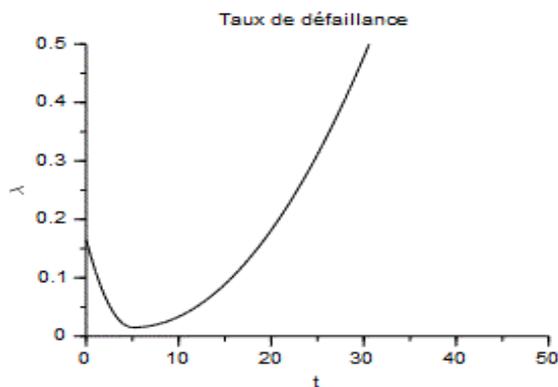


FIG.1.3 Courbe du taux de défaillance en parabole.

1.7 Principales lois de fiabilité

La fiabilité nécessite la connaissance des distributions de durée de vie afin de l'estimer. Dans cette section nous présentons quelques lois de vie qui interviennent le plus fréquem-

ment dans l'analyse de la fiabilité et leur principales propriétés, ainsi que les fonctions de survie (Fiabilité) et les taux de défaillance associées à ces dernières (pour plus de détail voir [20] , [21] , [32]).

1.6.1 Loi exponentielle $\exp(\lambda)$

Cette loi a de nombreuses applications dans plusieurs domaines. C'est une loi simple, très utilisée en fiabilité dont le taux de défaillance est constant. Elle décrit la vie des matériels qui subissent des défaillances brutales.

Une variable aléatoire T est de loi exponentielle de paramètre $\lambda > 0$, notée $\exp(\lambda)$, si et seulement si :

- * La fonction de répartition est : $F(t) = 1 - e^{(-\lambda t)}$, $t \geq 0$.
- * La densité est : $f(t) = \lambda e^{(-\lambda t)}$, $t \geq 0$.
- * La fiabilité est : $R(t) = 1 - F(t) = e^{(-\lambda t)}$, $t \geq 0$.
- * Le taux de défaillance est : $\lambda(t) = \lambda$, $t \geq 0$.

Le taux de défaillance est donc constant, ce qui signifie que la loi exponentielle modélise les durées de vie du systèmes qui ne s'usent pas et qui ne s'améliorent pas.

Remarque 1.7.1 *La loi exponentielle est la loi la plus utilisée en raison de sa simplicité, elle ne permet de modéliser ni l'usure, ni le rajeunissement. En fiabilité, il y a lieu d'utiliser la loi de Weibull.*

1.6.2 Loi de Weibull $\mathcal{W}(\eta, \beta)$

La loi la plus utilisé en fiabilité est la loi Weibull. Elle est défini avec un paramètre d'échelle $\eta > 0$ et un paramètre de forme $\beta > 0$. Elle est souvent utilisée aussi bien en électronique qu'en mécanique, elle caractérise mieux le comportement du produit dans les trois phases de vie selon le paramètre de la forme β : période de jeunesse ($\beta < 1$), période de vie utile ($\beta = 1$) et période d'usure ou vieillissement ($\beta > 1$). Les principales propriétés de cette loi sont données comme suit :

- * La fonction de répartition est : $F(t) = 1 - e^{(-(\frac{t}{\eta})^\beta)}$, $t \geq 0$.
- * La densité est : $f(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} e^{(-(\frac{t}{\eta})^\beta)}$, $x \geq 0$.

* La fiabilité est : $R(t) = e^{-\left(\frac{t}{\eta}\right)^\beta}$, $t \geq 0$.

* Le taux de défaillance est : $\lambda(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1}$, $t \geq 0$.

Dans le cas où $\beta = 1$ on trouve la loi exponentielle de paramètre $\frac{1}{\eta}$.

Remarque 1.7.2 *Le modèle Weibull est un modèle naturel pour des systèmes constitués d'un très grand nombre de composants et dont la panne survient dès qu'un composant est défaillant [34].*

1.6.3 Loi log-normale $Log\mathcal{N}(\mu, \sigma^2)$

Une variable aléatoire continue et positive T est distribuée selon une loi log-normale si son logarithme est distribué suivant une loi normale. Cette distribution est utilisée en fiabilité pour modéliser les défaillances par fatigue. La loi log-normale à deux paramètres $\mu \in \mathbb{R}$ et $\sigma^2 > 0$, vérifie les propriétés suivantes :

* La densité est : $f(t) = \frac{1}{\sigma t \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{\ln(t) - \mu}{\sigma}\right)^2}$, $t \geq 0$.

* La fiabilité est : $R(t) = 1 - \Phi\left(\frac{\ln(t) - \mu}{\sigma}\right)$, $t \geq 0$.

* Le taux de défaillance est : $\lambda(t) = \frac{e^{-\frac{1}{2} \left(\frac{\ln(t) - \mu}{\sigma}\right)^2}}{t \sigma \sqrt{2\pi} \left(1 - \Phi\left(\frac{\ln(t) - \mu}{\sigma}\right)\right)}$, $t \geq 0$.

Où Φ est la fonction de répartition de la loi normale centrée réduite.

1.6.4 Loi Gamma $G(\alpha, \beta)$

La loi Gamma est définie par un paramètre de forme $\alpha > 0$ et un paramètre d'échelle $\beta > 0$. Elle est utilisée pour modéliser le temps de défaillance d'un composant. La densité, la fiabilité et le taux de défaillance de cette dernière sont données par :

* La densité est : $f(t) = \frac{1}{\beta^\alpha \Gamma(\alpha)} t^{\alpha-1} e^{-\frac{t}{\beta}}$, $t \geq 0$.

* La fiabilité est : $R(t) = \int_t^{+\infty} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} dx$, $t \geq 0$.

* Le taux de défaillance est : $\lambda(t) = \frac{t^{\alpha-1} e^{-\frac{t}{\beta}}}{\beta^\alpha \int_t^{+\infty} \Gamma(\alpha) f(x) dx}$, $t \geq 0$.

Où Γ est la fonction gamma d'Euler, définie par: $\Gamma(a) = \int_0^{+\infty} t^{a-1} e^{-t} dt$.

En particulier : $\Gamma(a) = (a-1)!$ pour tout $a \in \mathbb{N}^*$.

La fonction de répartition de la loi Gamma n'a pas d'expression explicite, ce qui est de même pour la fiabilité.

Pour $\alpha = 1$ et $\beta = \frac{1}{\lambda}$ on retrouve la loi exponentielle de paramètre λ .

1.6.6 Loi Béta $\mathcal{B}(\alpha, \beta)$

Cette loi représente, en particulier, la probabilité pour qu'un matériel survive jusqu'à un instant t , quand on essaie n matériels. D'où son intérêt dans l'évaluation de la durée des essais de fiabilité. Les différentes propriétés de la loi Béta à deux paramètres α et β sont données par :

* La densité est : $f(t) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} t^{\alpha-1} (1-t)^{\beta-1}$, $t \geq 0$.

* La fiabilité est : $R(t) = \int_t^{+\infty} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} dx$, $t \geq 0$.

* Le taux de défaillance est : $\lambda(t) = \frac{\Gamma(\alpha + \beta) t^{\alpha-1} (1-t)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta) \int_t^{+\infty} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} dx}$, $t \geq 0$.

Où $\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t} dt$.

Remarque 1.7.3 *Il est important de noter qu'il existe d'autres lois de fiabilité comme la loi Normale, la loi Birnbaum-Saunders, la loi Pareto, la loi binômiale, la loi de Poisson, la loi hypergéométrique,...*

Après avoir donné les propriétés de quelques lois de probabilité qui interviennent dans l'analyse de la fiabilité, nous nous intéressons dans ce qui suit à la fiabilité des systèmes.

1.8 Fiabilité des systèmes

Nous pouvons calculer la fiabilité d'un système par le biais du calcul de la fiabilité par structure (ou architecture). Le principe est de considérer qu'un système est constitué de composants élémentaires, et que sa fiabilité dépend à la fois de la fiabilité de ses composants et de la façon dont le bon fonctionnement ou la panne de chaque composant influe sur le bon fonctionnement ou la panne du système tout entier [20].

1.7.1 Système en série

Un système en série est un système qui ne fonctionne que si tout ses composants fonctionnent (voir la figure (1.4)).

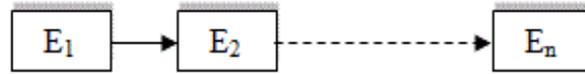


FIG.1.4 Diagramme de fiabilité d'un système en série.

La durée de vie du système est : $T = \min_{i=1,\dots,n} T_i$ où T_i est la durée de vie du composant i .

La fiabilité du système est alors :

$$R_S(t) = P(T > t) = P\left(\min_{i=1,\dots,n} T_i > t\right) = P(\forall i, T_i > t) = \prod_{i=1}^n P(T_i > t) = \prod_{i=1}^n R_i(t). \quad (1.7)$$

1.7.2 Système en parallèle

Un système en parallèle est un système tel qu'il suffit qu'un seul de ses composants fonctionne pour qu'il fonctionne (voir la figure (1.5)).

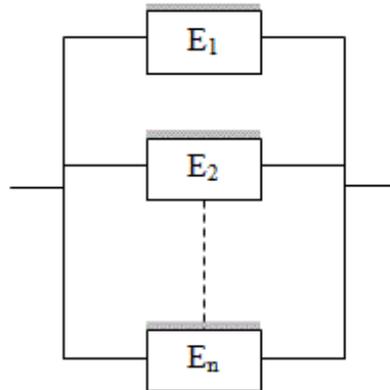


FIG.1.5 Diagramme de fiabilité d'un système en parallèle.

La durée de vie du système sera alors : $T = \max_{i=1,\dots,n} T_i$.

Avec, T_i est la durée de vie du composant i .

La fiabilité du système est alors :

$$\begin{aligned}
R_S(t) &= P(T > t) = P\left(\max_{i=1, \dots, n} T_i > t\right) = 1 - P\left(\max_{i=1, \dots, n} T_i \leq t\right) = 1 - P(\forall i, T_i \leq t) \\
&= 1 - \prod_{i=1}^n P(T_i \leq t) = 1 - \prod_{i=1}^n (1 - P(T_i > t)) = 1 - \prod_{i=1}^n (1 - R_i(t)). \tag{1.9}
\end{aligned}$$

1.7.3 Système k parmi n (k/n)

Un système k/n est un système qui ne fonctionne que si au moins k composants parmi n fonctionnent.

En particuliers

- 1) Le cas $k = 1$ correspond à un système en parallèle.
- 2) Le cas $k = 0$ correspond à un système série.
- 3) Si tous les composants sont indépendantes et identiques, et de même fiabilité $R(t)$,

alors :

$$R_S(t) = \sum_{j=k}^n C_n^j R^j(t) (1 - R(t))^{n-j}, t > 0 \tag{1.10}$$

Remarque 1.8.1 *Il existe d'autres systèmes, tels que les systèmes mixtes qui sont obtenus en combinant des systèmes séries et des systèmes parallèles.*

1.9 Conclusion

Dans cette section, nous avons rappelé quelques notions de base liées à la théorie de la fiabilité ainsi que quelques propriétés de la fonction du taux de défaillance. Nous avons mis l'accent sur les lois de probabilité qui seront utilisées pour l'estimation de la fiabilité et du taux de défaillance. Dans le chapitre suivant, nous donnons quelques méthodes d'estimation de la densité et de la fonction de répartition.

Estimation non paramétrique

2.1 Introduction

Un problème récurrent en statistique est celui de l'estimation non paramétrique d'une densité f et d'une fonction de répartition F d'une *v.a* X à partir d'un échantillon de *v.a.r* X_1, X_2, \dots, X_n *i.i.d* [47]. De nombreux estimateurs de la fonction densité ont été définis (voir [3], [37], [14]), étudiés et comparés. Nous citons les estimateurs : par Histogramme, par la méthode du noyau, par Histogramme modifiés, à base de splines et par des séries orthogonales, ...etc.

Dans ce chapitre, nous présentons les différentes méthodes d'estimation de la densité f et de la fonction de répartition F , en particulier la méthode des séries orthogonales.

2.2 Critères d'erreurs et définitions

Dans ce qui suit, nous donnons une liste des erreurs et leurs définitions mathématiques :

Critères d'erreurs

- L'erreur quadratique intégrée ISE (Integrated Square Error)

$$ISE(\hat{f}(x)) = \int [f(x) - \hat{f}(x)]^2 dx = \int [f(x)^2 + \hat{f}(x)^2 - 2f(x)\hat{f}(x)] dx.$$

- L'erreur quadratique moyenne MSE (Mean Square Error)

$$\begin{aligned} MSE(\hat{f}(x)) &= E[f(x) - \hat{f}(x)]^2 \\ &= [\text{biais}\hat{f}(x)]^2 + Var(\hat{f}(x)). \end{aligned}$$

- L'erreur quadratique moyenne intégrée MISE (Mean Integrated Square Error)

$$MISE(\hat{f}) = \int MSE(\hat{f}(x)) dx = \int [\text{biais}\hat{f}(x)]^2 dx + \int Var(\hat{f}(x)) dx.$$

Quelques définitions

1) On dit qu'un estimateur \hat{f} de f est sans biais si :

$$E(\hat{f}(x)) = f(x).$$

2) On dit qu'un estimateur \hat{f} de f est asymptotiquement sans biais si :

$$\lim_{n \rightarrow \infty} E(\hat{f}(x)) = f(x) \text{ en tout point } x \text{ pour le quel la densité } f \text{ est continue.}$$

3) On dit qu'un estimateur \hat{f} de f est asymptotiquement uniformément sans biais si :

$$\lim_{n \rightarrow \infty} \sup_x |E(\hat{f}(x)) - f(x)| = 0.$$

4) On dit qu'un estimateur \hat{f} de f est ponctuellement consistant en moyenne quadratique si :

$$\lim_{n \rightarrow \infty} MSE(f(x), \hat{f}(x)) = 0.$$

5) On dit qu'un estimateur \hat{f} de f est uniformément consistant en moyenne quadratique intégré si :

$$\lim_{n \rightarrow \infty} MISE(f, \hat{f}) = 0.$$

6) On dit qu'un estimateur \hat{f} de f est asymptotiquement normale si :

$$\hat{f} \xrightarrow{cv\ loi} N(E(\hat{f}(x)), Var(\hat{f}(x))).$$

2.3 Estimation par Histogramme

• Principe de la méthode

Soit X_1, X_2, \dots, X_n des *v.a.i.i.d.*, de densité f inconnue défini sur un intervalle $[a, b[$. Le principe de la méthode d'histogramme est d'approcher f par une fonction en escalier. On découpe l'intervalle $[a, b[$ en k classes $(I_j)_{j=0,1,\dots,k-1}$ avec $(I_j) =]a_{j-1}, a_j]$ de largeur h , h est appelé le paramètre de lissage.

• Estimation de la densité

L'estimateur de la densité au point x sur l'intervalle $]a_{j-1}, a_j]$ est défini par :

$$\hat{f}_h(x) = \frac{n_j}{nh}, \quad x \in]a_{j-1}, a_j]. \quad (2.1)$$

n_j : nombre d'observations dans l'intervalle $]a_{j-1}, a_j]$.

n : nombre d'échantillon.

Les propriétés asymptotiques de cet estimateur (voir [4], [42], [30]) sont données par :

$$MSE(\hat{f}_h) = \frac{f(x)}{nh} + \frac{f'(x)^2}{4} [h - 2(x - a_j)]^2 + o(h^3) + o(n^{-1}). \quad (2.2)$$

$$MISE(\hat{f}_h) = \frac{1}{nh} + \frac{h^2 \int f'(x)^2 dx}{12} + o(h^3) + o(n^{-1}). \quad (2.3)$$

Ces deux quantités tendent vers zéro, lorsque $h \rightarrow 0$ et $nh \rightarrow \infty$.

En pratique, on choisit h en fonction de n . Les règles les plus utilisées sont : règle de Sturges [40], règle de Scott [41] et règle de Freedman-Diaconis [19].

Remarque 2.3.1 *La plus grande qualité de l'histogramme est sa simplicité. Parmi ses inconvénients, citons celui d'être trop peu sensible aux propriétés locales de f . On outre, alors que la plupart des fonctions de densité ne sont pas des fonctions en escalier, l'histogramme est un estimateur toujours de cette forme. L'application de certaines opérations sur l'estimé comme par exemple une dérivée, ou une intégration, devient alors impossible ou très difficile à effectuer.*

2.4 Estimation par la méthode du noyau

• Principe de la méthode

Cette méthode généralise astucieusement la méthode d'estimation par histogramme. L'idée de l'estimation par la méthode du noyau consiste à évaluer la densité $f(x)$ au point x en comptant le nombre d'observations figurant dans un certain voisinage de x sur \mathbb{R} . Par définition, la densité de probabilité f est égale à la dérivée de la fonction de répartition F (si cette dérivée existe).

• Estimation de la densité

L'estimateur à noyau de la densité a été introduit par Parzen [36] et Rosenblatt [37], depuis il est devenu un objet classique, utilisé par des statisticiens et des probabilistes. Cet estimateur est adapté aux variables aléatoires continues. Considérons X_1, X_2, \dots, X_n un échantillon de *v.a.i.i.d* de densité f .

L'estimateur à noyau de la fonction de densité f , appelé estimateur de Parzen-Rosenblatt, notée $\hat{f}(x)$ est défini par :

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x - X_i}{h}\right). \quad (2.4)$$

Où, $K : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction noyau intégrable positive, telle que, $K(u) = K(-u)$, $\int_{-\infty}^{+\infty} K(u) du = 1$ et $\int_{-\infty}^{+\infty} K^2(u) du < \infty$ et h le paramètre de lissage.

Sous les conditions suivantes :

$\int_{\mathbb{R}} xK(x)dx = 0$ et $\sigma_K^2 = \int_{\mathbb{R}} x^2K(x)dx < \infty$ et $h \rightarrow \infty, nh \rightarrow \infty$ quand $n \rightarrow \infty$, on obtient les propriétés suivantes :

- Le biais de l'estimateur

$$\text{Biais} \left(\hat{f}(x) \right) = \frac{h^2}{2} f''(x) \sigma_K^2 + o(h^2). \quad (2.5)$$

Où, f'' est la deuxième dérivée de f .

-La variance de l'estimateur

$$\text{Var} \left(\hat{f}(x) \right) = \frac{1}{n} f(x) R(K) + o\left(\frac{1}{nh}\right). \quad (2.6)$$

Où : $R(g) = \int_{-\infty}^{+\infty} g^2(y)dy$ et f'' est la deuxième dérivée de f .

- L'erreur quadratique moyenne (*MSE*)

$$\text{MSE} \left(\hat{f}_n(x) \right) = \frac{1}{4} h^4 \sigma_K^4 (f''(x))^2 + \frac{1}{nh} f(x) R(K) + o\left(h^4 + \frac{1}{nh}\right). \quad (2.7)$$

- L'erreur quadratique moyenne intégrée (*MISE*)

$$\text{MISE}(\hat{f}_n) = \frac{1}{4} h^4 \sigma_K^4 \int (f''(x))^2 dx + \frac{1}{nh} R(K) + o\left(h^4 + \frac{1}{nh}\right). \quad (2.8)$$

● **Estimation de la fonction de répartition**

On définit l'estimateur \hat{F}_n à noyau de F par :

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n H\left(\frac{x - X_i}{h}\right), \quad H(w) = \int_{-\infty}^w K(u)du \quad \text{où } w = \frac{x - X_i}{h}. \quad (2.9)$$

Comme $K(u)$ a une forme d'une fonction de densité, $H(w)$ est de la forme d'une fonction de répartition.

Pour avoir l'analogie avec la fonction de répartition empirique F_n , rappelons que $F_n(x)$ est la moyenne des indicatrices $1_{]-\infty, x]}(x_i)$.

$$\begin{aligned}
 1_{]-\infty, x]}(x_i) &= 1_{]-\infty, x]}(x_i - x) \\
 &= 1_{]0, +\infty]}(\frac{x_i - x}{h}) \\
 &= \begin{cases} 0 & \text{si } x < x_i \\ 1 & \text{sinon.} \end{cases}
 \end{aligned}$$

\hat{F} est donc de la forme de F_n , avec une fonction $H(w)$ très particulière, donnant le saut brutale de 0 à 1 en $y = 0$. L'apport d'une fonction plus souple répond au même principe général de lissage que celui appliqué à la densité. Puisque H est une primitive de K , donc elle est continue, alors \hat{F} est également continue.

Nous avons ;

$$H\left(\frac{x - X_i}{h}\right) = \begin{cases} 0 & \text{si } \frac{x - X_i}{h} \leq -1 \\ \frac{1}{2}\left(\frac{x - X_i}{h} + 1\right) & \text{si } -1 < \frac{x - X_i}{h} < 1 \\ 1 & \text{si } \frac{x - X_i}{h} \geq 1 \end{cases}$$

Alors,

$$H\left(\frac{x - X_i}{h}\right) = \begin{cases} 0 & \text{si } x + h \geq X_i \\ \frac{1}{2}\left(\frac{x - X_i}{h} + 1\right) & \text{si } x - h < X_i < x + h \\ 1 & \text{si } x - h \geq X_i \end{cases}$$

Soit le noyau K vérifiant les conditions suivantes :

$$\int_{-1}^1 K(u)du = 1, \int_{-1}^1 uK(u)du = \sigma_K^1 = 0, \int_{-1}^1 u^2K(u)du = \sigma_K^2 < \infty \text{ et } \int_{-1}^1 u^3K(u)du = \sigma_K^3 = 0.$$

Nous pouvons montrer que [16]:

$$\int_{-1}^1 u^i H(u)du = \frac{1}{i+1} (1 - \sigma_K^{i+1}), \quad i = 0, 1, 2, 3.$$

Alors

$$\int_{-1}^1 H(u)du = 1, \int_{-1}^1 uH(u)du = \frac{1}{2} (1 - \sigma_K^2) = 0, \int_{-1}^1 u^2 H(u)du = \frac{1}{3}.$$

En utilisant les conditions précédentes, les propriétés de l'estimateur à noyau de la fonction de répartition F sont :

- Biais de l'estimateur

$$\text{Biais}(\hat{F}_n(x)) = E(\hat{F}_n(x) - F(x)) = \frac{h^2}{2} f'(x) \sigma_K^2 + o(h^2). \quad (2.10)$$

- Variance de l'estimateur

$$V(\hat{F}_n(x)) = \frac{1}{n} F(x)(1 - F(x)) + \frac{h}{n} f(x) \left[\int_{-1}^1 H(t) dt - 1 \right] + o(h). \quad (2.11)$$

- Erreur quadratique moyenne (MSE)

$$\begin{aligned} MSE(\hat{F}_n(x)) &= \frac{h^4}{4} f'^2(x) \sigma_K^4 + o(h^2) + \frac{1}{n} F(x)(1 - F(x)) + \frac{h}{n} f(x) \left[\int_{-1}^1 H(t) dt - 1 \right] \\ &+ o(h). \end{aligned} \quad (2.12)$$

Quand $n \rightarrow \infty$ pour $h \rightarrow 0$: $\sigma_K^2 < \infty$ et $\int_{-1}^1 H(t) dt < \infty$ alors $MSE(\hat{F}_n(x)) \rightarrow 0$.

L'estimateur $\hat{F}_n(x)$ est convergent en moyenne quadratique vers $F(x)$.

- Erreur quadratique moyenne intégrée ($MISE$)

$$\begin{aligned} MISE(\hat{F}_n) &= \int_{\mathbb{R}} \left(\frac{h^4}{4} f'^2(x) \sigma_K^4 + \frac{1}{n} F(x)(1 - F(x)) + \frac{h}{n} f(x) \left[\int_{-1}^1 H(t) dt - 1 \right] \right) dx \\ &+ o(h^2) + o(h). \end{aligned} \quad (2.12)$$

Quand $n \rightarrow \infty$ pour $h \rightarrow 0$: $\sigma_K^2 < \infty$ et $\int_{-1}^1 H(t) dt < \infty$ alors $MISE(\hat{F}_n(x)) \rightarrow 0$.

L'estimateur $\hat{F}_n(x)$ est convergent en moyenne quadratique intégrée vers $F(x)$.

• Choix du noyau K

Les premiers noyaux introduit dans l'estimation de la densité sont des noyaux symétriques (voir le tableau (1.1)) :

Noyau	$K(u)$
Uniforme	$\frac{1}{2}I_{[-1,1]}(u)$.
Triangulaire	$(1 - u) I_{[-1,1]}(u)$.
Gaussien	$\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}, u \in \mathbb{R}$.
Epanechnikov	$\frac{3}{4\sqrt{5}} \left(1 - \frac{u^2}{5}\right) I_{[-\sqrt{5},\sqrt{5}]}(u)$.
Biweight	$\frac{15}{16} (1 - u^2)^2 I_{[-1,1]}(u)$.
Triweight	$\frac{35}{32} (1 - u^2)^3 I_{[-1,1]}(u)$.
Cosine	$\frac{\pi}{4} \cos\left(\frac{\pi u}{2}\right) I_{[-1,1]}(u)$.
Silverman	$\frac{1}{2} \exp\left(-\frac{ u }{\sqrt{2}}\right) \sin\left(\frac{ u }{\sqrt{2}} + \frac{\pi}{4}\right), u > 0$.

TAB.1.1 Noyaux usuels.

Lorsque nous déterminons l'estimateur de la densité de probabilité d'une *v.a* à support borné $[0, 1]$ ou semi-borné $[0, +\infty[$, l'inconvénient principal de ce type de noyaux est qu'il assigne un poids à l'extérieur du support quand le lissage est pris en compte près du bord. Cela cause le problème du biais de bordure et donne un estimateur non consistant. La solution consiste alors à utiliser des noyaux asymétriques continus.

Nous citons les noyaux Beta [12], Gamma[13], Gaussien-inverse (*GI*) et Gaussien-inverse-réciproque (*GIR*) [49].

Noyau Gamma :

La première classe des noyaux gamma est :

$$K_{Gam(\frac{x}{h}+1,h)}(t) = \frac{t^{\frac{x}{h}} e^{-\frac{t}{h}}}{\Gamma\left(\frac{x}{h} + 1\right) h^{\frac{x}{h}+1}}, t \in [0, +\infty[.$$

La deuxième classe est défini par :

$$K_{Gam(\rho_h(x),h)}(t) = \frac{x^\alpha (1-x)^\beta}{B(\alpha, \beta)}, x \in [0, 1].$$

Noyau Beta :

Le noyau beta est donné comme suit :

$$K_{Beta}(x, \alpha, \beta) = \frac{x^\alpha e^{-\frac{t}{h}}}{\Gamma(\rho_h(x)) h^{\rho_h(x)}}, \quad t \in [0, +\infty[.$$

Avec, $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ et $\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx, \forall \alpha > 0$.

• **Choix théorique optimal du paramètre de lissage:**

Il existe plusieurs méthodes pour choisir le paramètre de lissage, considérons trois grandes classes :

- La classe des méthodes de validations croisées :
 - La méthode de validation croisée biaisée[43].
 - La méthode de validation croisée non biaisée ([5], [38]).
 - La méthode de validation croisée lissée[24].
- La classe des méthodes plug-in (ré-injection) :
 - La règle du pouce (rule of Thumb)[44].
 - Surlissage (Oversmoothing)[43].
 - Plug-in itéré[41].
 - Estimation optimal du paramètre de lissage[51].
- La classe des méthodes de bootstrap ([52], [8]).

2.5 Estimation par des fonctions orthogonales

2.5.1 Principe de la méthode

Il est souvent possible d'approcher un paramètre fonctionnel g de dimension infinie sur un espace de dimension finie. Si ce nouveau paramètre est facile à estimer on obtient un estimateur de g . Le cas le plus simple est celui où g prend ses valeurs dans un espace de **Hilbert**, on peut alors l'approcher par sa projection orthogonale sur un espace de dimension finie, convenablement choisi.

2.5.2 Estimation de la densité

Soit $(X_n, n \in \mathbb{N})$ une suite de v.a i.i.d définies sur un même espace de probabilité, de densité de probabilité inconnue f par rapport à la mesure de Lebesgue sur \mathbb{R} . Il s'agit d'estimer f à partir de l'échantillon X_1, \dots, X_n .

Pour cela on suppose que:

- L'espace de **Hilbert** L^2 est de dimension infinie muni de sa norme usuelle $\| \cdot \|$ et de son produit scalaire $\langle \cdot, \cdot \rangle$;

- $\{e_k, k \in \mathbb{N}\}$ est une base orthonormée dans L^2 ;
- $f \in L^2$ tel que :

$$f(x) = \sum_{k=0}^{\infty} a_k e_k(x), x \in \mathbb{R}. \quad (2.13)$$

Avec, $a_k, k \in \mathbb{N}$ sont les coefficients de **Fourier** associés à f vérifiant :

$$a_k = \int_{\mathbb{R}} e_k(x) f(x) dx = E[e_k(x)], \quad k = . \quad (2.14)$$

- Le développement à l'ordre d_n de $f(x)$ est donné par :

$$f_{d_n}(x) = \sum_{k=0}^{d_n} a_k e_k(x), x \in \mathbb{R}. \quad (2.15)$$

Avec, $d_n \rightarrow \infty$ lorsque $n \rightarrow \infty$.

$$\hat{a}_k = \frac{1}{n} \sum_{i=1}^n e_k(X_i).$$

Par la méthode des moments, on peut estimer a_k par la moyenne empirique. Ainsi, $f_{d_n}(x)$ peut être estimé par :

$$\hat{f}_{d_n}(x) = \sum_{k=0}^{d_n} \hat{a}_k e_k(x), x \in \mathbb{R} = \frac{1}{n} \sum_{k=0}^{d_n} \sum_{i=1}^n e_k(X_i) e_k(x) = \frac{1}{n} \sum_{i=1}^n K_{d_n}(x, X_i). \quad (2.16)$$

Avec, $K_{d_n}(x, X) = \sum_{k=0}^{d_n} e_k(X) e_k(x)$ est le noyau de l'estimateur et d_n le paramètre de lissage.

Remarque 2.5.1 L'expression de l'estimateur $\hat{f}_{d_n}(x)$ montre que l'estimateur par la méthode des fonctions orthogonales appartient à la classe des estimateurs à noyaux et peut donc s'exprimer sous la forme d'une somme pondérée de la variable aléatoire indépendantes. En conséquence, l'étude de l'estimateur $\hat{f}_{d_n}(x)$ peut être faite à partir des théorèmes généraux régissant l'estimateur à noyau.

• **Propriétés des estimateurs**

Les coefficients \hat{a}_k vérifient :

$$E(\hat{a}_k) = E\left(\frac{1}{n} \sum_{i=1}^n e_k(X_i)\right) = \frac{1}{n} \sum_{i=1}^n E(e_k(X_i)) = E(e_k(X)) = a_k, \quad \forall k = 0, \dots, d_n.$$

$$Var(\hat{a}_k) = Var\left(\frac{1}{n} \sum_{i=1}^n e_k(X_i)\right) = \frac{1}{n} Var(e_k(X)), \quad \forall k = 0, \dots, d_n.$$

Théorème 2.5.1 Soient $(\hat{a}_k)_{k \in \mathbb{N}}$ les estimateurs des coefficients de Fourier associés à f , alors :

$$\lim_{n \rightarrow +\infty} Var(\hat{a}_k) = 0, \quad \forall k = 0, \dots, d_n.$$

Théorème 2.5.2 Soient $(\hat{a}_k)_{k \in \mathbb{N}}$ les estimateurs des coefficients de Fourier associés à f , alors :

$$\lim_{n \rightarrow +\infty} E[\hat{a}_k - a_k]^2 = 0, \quad \forall k = 0, \dots, d_n.$$

Les estimateurs (\hat{a}_k) sont alors asymptotiquement convergent et vérifiant la convergence en moyenne quadratique.

L'estimateur $\hat{f}_{d_n}(x)$ est :

- Le biais est donné par

$$Biais(\hat{f}_{d_n}(x)) = E(\hat{f}_{d_n}(x)) - f(x) = \sum_{k=0}^{d_n} \hat{a}_k e_k(x) - f(x) = f_{d_n}(x) - f(x) \quad (2.17)$$

Ce qui implique que $\hat{f}_{d_n}(x)$ est un estimateur biaisé de $f(x)$.

- L'erreur quadratique moyenne intégrée (*MISE*)

Théorème 2.5.3 (*Kronmal-Tarter* [26])

Si $\int f^2(x)dx < \infty$, alors :

$$MISE\left(\hat{f}_{d_n}\right) = \int f^2(x)dx - \sum_{k=0}^{d_n} a_k^2 + \sum_{k=0}^{d_n} Var(\hat{a}_k). \quad (2.18)$$

Les principaux résultats relatifs à la convergence de l'estimateur suivant divers modes stochastiques sont donnés par [33].

Convergence en écart quadratique intégrée

Théorème 2.5.4 (*Mizahmedov-Hasimov*[31])

Supposons que :

- 1) $\sup_k \|e_k\|_\infty < \infty$ et $(d_n)_{n \in \mathbb{N}}$ une suite d'entiers positifs avec $\lim_{n \rightarrow \infty} d_n = \infty$;
 - 2) $\sum_k d_n \exp\left[\frac{-kb}{d_n^2}\right] < \infty$, $\forall b > 0$ alors :
- $$\sum_{k=0}^{\infty} a_k^2 < \infty \implies \int \left[\hat{f}_{d_n}(x) - f(x)\right]^2 dx \rightarrow 0, \text{ quand } n \rightarrow \infty.$$

Convergence uniforme

Théorème 2.5.5 (*Mizahmedov-Hasimov*[31])

Si :

- 1) $\sum_{k=0}^{\infty} |a_k| k^\alpha < \infty$, $\alpha > 0$;
- 2) $\sum_n d_n \exp\left[\frac{-n\beta}{(d_n)^{2+2\alpha}}\right] < \infty$, $\forall \beta > 0$, alors :

$$\lim_{n \rightarrow \infty} d_n^\alpha \sup_x \left[\hat{f}_{d_n}(x) - f(x)\right] = 0. \quad (2.19)$$

Soit d la distance de la convergence uniforme sur \mathbb{R} définie par :

$$d\left(\hat{f}_{d_n}, f\right) = \sup_x \left|\hat{f}_{d_n}(x) - f(x)\right|. \quad (2.20)$$

Convergence uniforme presque complète

Théorème 2.5.6 (Bosq-Lecoutre[4])

Si $M = \sup_{j, x} |e_j(x)| < \infty$ et $d(\hat{f}_{d_n}, f) \rightarrow 0$, alors :

$$\forall \gamma > 0, \sum_n d_n \exp \left[-\gamma \frac{n}{d_n^2} \right] < \infty \implies d(\hat{f}_{d_n}, f) \xrightarrow{p.co} 0, \text{ quand } n \rightarrow \infty.$$

• Choix pratique du paramètre de lissage

Pour une base fixée, le choix du paramètre de lissage d_n est effectué en minimisant l'erreur quadratique moyenne intégrée $MISE(\hat{f}_{d_n})$. Il existe plusieurs méthodes pour le choix du paramètre de lissage. Dans ce qui suit nous nous contentons par donner deux de ces méthodes à savoir la méthode de Kronmal-Tarter et la méthode de Bosq.

Méthode de Kronmal-Tarter [46]

L'emploi de $\hat{f}_{d_n}(x)$ pour estimer $f(x)$ n'est possible qu'après avoir déterminé le nombre optimum de termes d_n de la somme. Il est naturel de choisir d_n de sorte que l'erreur quadratique intégrée $MISE(\hat{f}_{d_n})$ soit minimum. La règle adoptée pour déterminer la valeur optimum d_n repose sur l'algorithme suivant :

À partir de $d_n = 1$ on augmente la valeur de d_n d'une unité jusqu'à ce que $MISE(\hat{f}_{d_n})$ augmente. On donne alors à d_n la valeur qui précède juste l'augmentation du $MISE(\hat{f}_{d_n})$.

On ajoutera donc à la somme le $d_n^{\text{ème}}$ terme si et seulement si :

$$\Delta_{d_n} = MISE(\hat{f}_{d_n}) - MISE(\hat{f}_{d_n-1}) \leq 0. \quad (2.21)$$

En tenant compte des propriétés statistiques de l'estimateur, Δ_{d_n} se met sous la forme:

$$\begin{aligned} \Delta_{d_n} &= MISE(\hat{f}_{d_n}(x)) - MISE(\hat{f}_{d_n-1}(x)) \\ &= \int f^2(x)dx + \sum_{k=0}^{d_n} [Var(\hat{a}_k) - a_k^2] - \left[\int f^2(x)dx + \sum_{k=0}^{d_n-1} [Var(\hat{a}_k) - a_k^2] \right] \\ &= \frac{(n+1)}{n} Var(e_{d_n}(X)) - E(e_{d_n}(X)^2). \end{aligned}$$

Posons alors $\theta_i = e_{d_n}(X_i)$, $i = 1, \dots, n$ et $\bar{\theta} = \frac{1}{n} \sum_{i=1}^n \theta_i$.

On peut alors définir un estimateur symétrique sans biais de Δ_{d_n} donné par :

$$\widehat{\Delta}_{d_n} = \frac{1}{n} \left[\frac{n+1}{n-1} \sum_{i=1}^n (\theta_i - \bar{\theta})^2 - \sum_{i=1}^n \theta_i^2 \right]. \quad (2.22)$$

L'optimum d^* est alors de la forme :

$$d^* = \begin{cases} \inf \{d_n, 1 \leq d_n \leq D\} & \widehat{\Delta}_d > 0 \\ D & \text{sinon.} \end{cases}$$

Méthode de Bosq

Bosq [18] a proposé un estimateur du paramètre de lissage d_n donné par :

$$\widehat{d}_n = \max \{j : 0 \leq j \leq d_n, |\widehat{a}_j| \geq \gamma_n\}, \quad (2.23)$$

Avec

$$\gamma_n = c \sqrt{\frac{\log n}{n}}, c > 0.$$

Proposition [18]

- 1) Si $\frac{d_n}{n} \rightarrow 0$ on a : $MISE(\widehat{f}_{d_n}) \rightarrow 0$.
- 2) Si $\sum_n |a_j| < \infty$ et $\sum_n d_n \exp\left[-\frac{n}{d_n^2} a\right] < \infty, a > 0$,

Alors

$$\sup_{x \in E} \left| \widehat{f}_{d_n}(x) - f(x) \right| \xrightarrow{p.s} 0.$$

2.5.3 Estimation de la fonction de répartition

Soit $(X_n, n \in \mathbb{N})$ une suite de v.a i.i.d définies sur un même espace de probabilité, de densité de probabilité inconnue $f(x)$ et de fonction de répartition $F(x)$ sur $[a, b]$. Il s'agit d'estimer $F(x)$ à partir de l'échantillon X_1, \dots, X_n .

Pour cela on suppose que:

• L'espace de **Hilbert** L^2 est de dimension infinie muni de sa norme usuelle $\| \cdot \|$ et de son produit scalaire $\langle \cdot, \cdot \rangle$;

• $\{e_k, k \in \mathbb{N}\}$ est une base orthonormée dans $L^2([a, b])$;

• $F \in L^2$ tel que :

$$F_{d_n}(x) = \sum_{k=0}^{+\infty} A_k e_k(x), \quad x \in [a, b]. \quad (2.24)$$

Avec, $A_k, k \in \mathbb{N}$ sont les coefficients de **Fourier** associés à f vérifiant :

$$A_k = \int_a^b e_k(x) F(x) dx, \quad k \in \mathbb{N}. \quad (2.25)$$

• Le développement à l'ordre d_n de $F(x)$ est donné par :

$$F_{d_n}(x) = \sum_{k=0}^{d_n} A_k e_k(x), \quad x \in [a, b]. \quad (2.26)$$

Supposons que $\sum_{k=0}^{+\infty} A_k e_k(x)$ uniformément convergente sur $[a, b]$, et chaque fonction $e_k(x)$ est continue. Pour tout $x \in [a, b]$, l'estimateur de est donné par :

$$\hat{F}_{d_n}(x) = \sum_{k=0}^{d_n} \hat{A}_k e_k(x), \quad x \in [a, b]. \quad (2.27)$$

Avec, $d_n \rightarrow \infty$ lorsque $n \rightarrow \infty$, et

$$\text{pour } k = 0, \dots, d_n \quad \hat{A}_k = \frac{1}{n} \sum_{i=1}^n \Psi_k(X_i) \quad \text{et} \quad \Psi_k(y) = \int_a^b e_k(x) 1_{x-y \geq 0} dx.$$

Où, $1_B(x) = 1$ si $x \in B$ et $1_B(x) = 0$ si $x \notin B$.

D'où

$$\hat{A}_k = \frac{1}{n} \sum_{i=1}^n \int_{X_i}^b e_k(x) dx, \quad x \in [0, b] \quad \text{et} \quad k = 0, \dots, d_n. \quad (2.28)$$

• Propriétés des estimateurs

Les coefficients \hat{A}_k vérifient :

$$E(\hat{A}_k) = E\left(\frac{1}{n} \sum_{i=1}^n \int_{X_i}^b e_k(x) dx\right) = \frac{1}{n} \sum_{i=1}^n E\left(\int_{X_i}^b e_k(x) dx\right), \quad \forall k = 0, \dots, d_n.$$

$$Var(\hat{A}_k) = Var\left(\frac{1}{n} \sum_{i=1}^n \int_{X_i}^b e_k(x) dx\right) = \frac{1}{n^2} \sum_{i=1}^n Var\left(\int_{X_i}^b e_k(x) dx\right), \quad \forall k = 0, \dots, d_n.$$

Les deux théorèmes suivants nous donnent le comportement asymptotique de la variance et de l'erreur quadratique de l'estimateur.

Théorème 2.5.7 Soient $(\hat{A}_k)_{k \in \mathbb{N}}$ les estimateurs des coefficients de Fourier associés à F , alors :

$$\lim_{n \rightarrow +\infty} \text{Var}(\hat{A}_k) = 0, \quad \forall k = 0, 1, \dots$$

Théorème 2.5.8 Soient $(\hat{A}_k)_{k \in \mathbb{N}}$ les estimateurs des coefficients de Fourier associés à F , alors :

$$\lim_{n \rightarrow +\infty} E \left[\hat{A}_k - A_k \right]^2 = 0, \quad \forall k = 0, 1, \dots$$

Les estimateurs (\hat{A}_k) sont alors asymptotiquement convergent et vérifiant la convergence en moyenne quadratique.

Propriétés de l'estimateur $\hat{F}_{d_n}(x)$

- Le biais est donné par

$$\text{Biais} \left(\hat{F}_{d_n}(x) \right) = E \left(\sum_{k=0}^{d_n} \hat{A}_k e_k(x) \right) - \sum_{k=0}^{d_n} A_k e_k(x) = - \sum_{k=d_n+1}^{\infty} A_k e_k(x) \quad (2.29)$$

Ce qui implique que $\hat{F}_{d_n}(x)$ est un estimateur biaisé de $F_{d_n}(x)$.

- La variance de $\hat{F}_{d_n}(x)$ est donné par :

$$\text{Var}(\hat{F}_{d_n}(x)) = \text{Var} \left(\sum_{k=0}^{d_n} \hat{A}_k e_k(x) \right) = \sum_{k=0}^{d_n} e_k^2(x) \text{Var}(\hat{A}_k) + \sum_{k \neq j}^{d_n} \text{Cov}(\hat{A}_k, \hat{A}_j) e_k(x) e_j(x) \quad (2.30)$$

- L'erreur quadratique moyenne intégrée (*MISE*)

$$\text{MISE} \left(\hat{F}_{d_n} \right) = \sum_{k=d_n+1}^{d_n} \hat{A}_k^2 + \sum_{k=0}^{d_n} \text{Var}(\hat{A}_k). \quad (2.31)$$

• **Choix pratique du paramètre de lissage**

Méthode Kronmal et Tarter

Il est naturel de choisir d_n de sorte que l'erreur quadratique moyenne intégrée $MISE(\hat{F}_{d_n})$ soit minimum. La règle adoptée pour déterminer la valeur optimum d_n repose sur l'algorithme suivant :

Algorithme

A partir de $d_n = 1$ on augmente la valeur de d_n d'une unité jusqu'à ce que $MISE$ augmente on donne alors à d_n la valeur qui précède juste l'augmentation $MISE$. On ajoutera donc à la somme (2.27) le $d^{i\text{ème}}$ terme si et seulement si

$$\Delta_{d_n} = MISE(\hat{F}_{d_n}) - MISE(\hat{F}_{d_{n-1}}) \leq 0.$$

Notons que l'expression (2.31) est en fonction des coefficient de Fourier associés à la densité de probabilité $f(x)$. Donc le signe de $[MISE(\hat{F}_{d_n}) - MISE(\hat{F}_{d_{n-1}})]$ est égale au signe de $[MISE(\hat{f}_{d_n}) - MISE(\hat{f}_{d_{n-1}})]$. Par conséquent, le paramètre de lissage qui minimise l'erreur quadratique moyenne intégrée associée à la fonction de répartition est donnée par :

$$d^* = \begin{cases} \inf \{d_n, 1 \leq d_n \leq d\} & \hat{\Delta}_d > 0 \\ d & \text{sinon.} \end{cases} \quad (2.32)$$

Où, $\hat{\Delta}_{d_n} = \frac{1}{n} \left[\frac{n+1}{n-1} \sum_{i=1}^n (\theta_i - \bar{\theta})^2 - \sum_{i=1}^n \theta_i^2 \right]$, avec $\theta_i = \int_{X_i}^b e_{d_n}(x) dx$ et $\bar{\theta} = \frac{1}{n} \sum_{i=1}^n \theta_i$.

Méthode de Saadi et Adjabi [45]

La valeur de d_n optimale etant celle qui minimise le $MISE$, alors par intégration par partie on obtient :

$$MISE(\hat{F}_{d_n}) = \int_a^b F(x)^2 dx + \sum_{k=0}^{d_n} \left[var(\hat{A}_k) - A_k^2 \right] = b - 2E(XF(X)) + \sum_{k=0}^{d_n} \left[var(\hat{A}_k) - A_k^2 \right].$$

Le paramètre de lissage optimal d^* est obtenu en minimisant un estimateur du $MISE(\hat{F}_{d_n})$. Il est donné par :

$$d^* = \underset{0 \leq d \leq d_n}{\text{Arg min}} \left[b - \frac{2}{n} \sum_{i=0}^n \psi_{-i,d}(X_i) + \sum_{k=0}^{d_n} \frac{1}{n} \left(\sum_{i=0}^n \frac{n+1}{n-1} (\vartheta_{i,k} - \bar{\vartheta}_k)^2 - \vartheta_{i,k}^2 \right) \right], \quad (2.33)$$

Où, $\vartheta_{i,k} = \int_{X_i}^b e_k(x) dx$, $\bar{\vartheta} = \frac{1}{n} \sum_{i=0}^n \vartheta_{i,k}$ et $\psi_{-i,k}(x) = \frac{1}{n-1} \sum_{j=0}^n \sum_{j \neq i, k=1}^{d_n} \vartheta_{j,k} x e_k(x)$.

2.5.4 Choix pratique de la base

Le choix primaire de la base dépend du support de la fonction à estimer. En général lorsque le support est la droite réelle $]-\infty, +\infty[$ ou la demi droite réelle $[0, +\infty[$, alors les fonctions d'Hermite et Laguerre sont recommandées voir [54]. Dans le cas où nous n'avons aucune information sur le support, on peut utiliser les fonctions d'Hermite. Les fonctions d'Hermite donnent de bons résultats au voisinage de la loi normale réduite pour la densité puisque le premier élément de la base $e_0(x) = \pi^{-\frac{1}{2}} \exp(-\frac{x^2}{2})$ est la densité d'une v.a de loi normale centrée réduite $N(0, 1)$. Au voisinage d'une loi normale quelconque on peut considérer des fonctions d'Hermite modifiées.

$$e_j^1(x) = e_j \left(\frac{x - \bar{X}}{S_n} \right), \quad j \in \mathbb{N}. \quad (2.34)$$

Où : $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, et $S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$.

Si le support de f est un intervalle compact, tel que $[0, 1]$, alors les fonctions trigonométriques et séries de Fourier sont recommandées. En particulier, la base cosinus [17] est aussi convenable puisqu'elle se rapproche bien des densités apériodiques. Une discussion sur la base trigonométrique peut être trouvée dans [9]. Les polynômes orthogonaux, y compris Legendre, Gegenbauer, Jacobi et Chebyshev sont souvent choisis ([54], [10], [7]). Les bases d'ondelettes sont d'une grande opportunité. Une complication technique dans l'utilisation d'ondelettes pour l'estimation de la densité est qu'aucune formule achevée n'est disponible. Enfin, notons que, dans de nombreux cas, le choix entre un support compact contre un support infini peut être compliqué voir ([55], [23]). Wahba suggère que «Dans de nombreuses applications, il pourrait être préférable de présumer que la vraie densité est à support compact et mettre à l'échelle les données à l'intérieur de l'intervalle $[0, 1]$ ».

On constate qu'il n'y a pas de solution évidente qui se dégage. En effet, les systèmes orthonormaux sont très variés, et il n'existe pas de théorème qui permet de conseiller un système particulier.

2.5.5 Exemples d'estimateurs

Nous allons présenter dans cette partie quelques estimateurs particuliers pour la fonction densité.

• Estimateurs associés aux fonctions trigonométriques

Estimateur de Dirichlet [57]

$I = [-\pi, +\pi]$ est un intervalle de \mathbb{R} . La base orthonormale est définie par:

$$e_1 = \frac{1}{\sqrt{2\pi}}; \quad e_{2k} = \frac{\cos(kx)}{\sqrt{\pi}}; \quad e_{2k+1} = \frac{\sin(kx)}{\sqrt{\pi}}; \quad x \in [-\pi, +\pi]; \quad k = 1, 2, \dots$$

L'estimateur de Dirichlet est donné par:

$$\hat{f}_{d_n}(x) = \frac{1}{2n\pi} \sum_{i=1}^n \frac{\sin \left[\left(\frac{X_i - x}{2} \right) \left(2 \left(\frac{d_n}{2} \right) + 1 \right) \right]}{\sin \left(\frac{X_i - x}{2} \right)}, \quad x \in [-\pi, \pi] \quad \text{pour} \quad x \neq X_i.$$

Estimateur de Fejer [46]

Soit f une fonction nulle en dehors de $I = [a, b]$, un intervalle de \mathbb{R} . On considère les systèmes orthogonaux suivants :

$$\begin{aligned} & \left\{ \cos k\pi \left[\frac{x-a}{b-a} \right], k = 0, 1, \dots \right\}, \\ & \left\{ \sin k\pi \left[\frac{x-a}{b-a} \right], k = 0, 1, \dots \right\} \text{ et} \\ & \left\{ \cos k\pi \left[\frac{x-a}{b-a} \right], \sin k\pi \left[\frac{x-a}{b-a} \right], k = 0, 1, \dots \right\}. \end{aligned}$$

Kronmal et Tarter [26] estiment $f(x)$ à partir d'un développement en série de Fourier. Ils proposent respectivement trois estimateurs :

$$\begin{aligned}\hat{f}_{dn}(x) &= \frac{c_0}{2} + \sum_{k=1}^{d_n} \overline{c}_k \cos k\pi \left[\frac{x-a}{b-a} \right], \\ \hat{f}_{dn}^1(x) &= \sum_{k=1}^{d_n} \overline{s}_k \sin k\pi \left[\frac{x-a}{b-a} \right] \\ \hat{f}_{dn}^2(x) &= \frac{\hat{f}_{dn}^1(x) + \hat{f}_{dn}(x)}{2}.\end{aligned}$$

Dans ces formules \overline{c}_k et \overline{s}_k représentent les moments trigonométriques de l'échantillon. C'est - à- dire :

$$\overline{c}_k = \frac{c_0}{n} \sum_{k=1}^{d_n} \cos k\pi \left[\frac{X_i - a}{b-a} \right], \quad \overline{s}_k = \frac{c_0}{n} \sum_{k=1}^{d_n} \sin k\pi \left[\frac{X_i - a}{b-a} \right], \quad c_0 = \frac{2}{b-a}.$$

Où a et b sont deux constantes arbitraire telles que : $a \leq b$.

Remarque 2.5.2 *L'inconvénient essentiel de ce type d'estimateur réside dans le fait que les fonctions $\hat{f}_{dn}(x)$, $\hat{f}_{dn}^1(x)$, $\hat{f}_{dn}^2(x)$ peuvent prendre des valeurs négatives sur $[a, b]$. Cet inconvénient, signalé par **Kronmal** et **Tarter**, se produit très rarement en pratique dès que n est suffisamment grand mais sa probabilité d'apparition augmente, plus ou moins suivant $f(x)$, lorsque n diminue. Cette possible négativité de $\hat{f}_{dn}(x)$ limite pratiquement l'emploi de ces estimateurs à l'étude de très grand échantillons. On peut contourner cette difficulté en remplaçant la série de **Fourier** par celle de **Fejer**. En effet, la somme de Fejer correspondante est toujours positive ou nulle sur l'intervalle $[a, b]$ considéré, quelle que soit la taille de l'échantillon. Dans ce cas, L'estimateur de $f(x)$ est alors défini par :*

$$g_{dn}(x) = \frac{c_0}{n} + \sum_{k=1}^{d_n} \left(1 - \frac{k}{d_n + 1} \right) [\overline{c}_k \cos(kz(x)) + \overline{s}_k \sin(kz(x))], \quad x \in [a, b]$$

$$\text{Où, } z(x) = \frac{2\pi(x-a)}{b-a},$$

$$\begin{aligned} \overline{c_k} &= \frac{c_0}{n} \sum_{i=1}^n \cos(kz(X_i)) \text{ est l'estimateur sans biais de } c_k = c_0 \int_b^a f(x) \cos(kz(x)) dt, \\ \overline{s_k} &= \frac{c_0}{n} \sum_{i=1}^n \sin(kz(X_i)) \text{ est l'estimateur sans biais de } s_k = c_0 \int_b^a f(x) \sin(kz(x)) dt, \\ \text{et } c_0 &= \frac{2}{b-a} \end{aligned}$$

Théorème 2.5.9 (Asselin de Beauville [1])

$g_{d_n}(x)$ est toujours positif ou nul, il peut s'écrire sous la forme :

$$g_{d_n}(x) = \frac{c_0}{2n(d_n + 1)} \sum_{i=1}^n \left[\frac{\sin(d_n + 1) \frac{v_i(x)}{2}}{\sin\left(\frac{v_i(x)}{2}\right)} \right]^2, \quad v_i(x) = \frac{2\pi(X_i - x)}{b - a}, X_i \neq x \text{ et } x \in [a, b].$$

Estimateur de Saadi et Adjabi [45]

La base proposés par Saadi et Adjaibi est donnée par:

$$e_k(x) = \frac{1}{\sqrt{2\pi}} (\cos(kx) + \sin(kx)) 1_{[-\pi, +\pi]}(x), \quad k = 0, 1, \dots$$

L'estimateur de la densité associé à la base trigonométrique est de la forme:

$$\hat{f}_{d_n}(x) = \frac{1}{4\pi n} \sum_{i=1}^n \left[\frac{\sin\left(\frac{(2d_n+1)(X_i-x)}{2}\right)}{\sin\left(\frac{X_i-x}{2}\right)} + \frac{\sin\left(\frac{(2d_n+1)\left(\frac{\pi}{2}-(X_i+x)\right)}{2}\right)}{\sin\left(\frac{\frac{\pi}{2}-(X_i+x)}{2}\right)} \right], X_i \neq x \text{ et } x \in [-\pi, \pi].$$

● **Estimateur associé au fonctions d'Hermite** [25]

Les fonctions d'Hermite sont données par les formules suivantes :

$$e_j(x) = (2^j j! \pi^{\frac{1}{2}})^{-\frac{1}{2}} Q_j(x) \exp\left(-\frac{x^2}{2}\right), \quad x \in \mathbb{R}, \quad j = 0, 1, \dots$$

Où, $Q_j(x)$ est le $j^{\text{ième}}$ polynôme d'Hermite défini par :

$$Q_j(x) = (-1)^j \exp(-x^2) \frac{d^j}{dx^j} \exp(x^2); \quad x \in \mathbb{R}, \quad j = 0, 1, \dots$$

Le noyau d'Hermite est défini par :

$$K_{d_n}(X, x) = \frac{d_n + 1}{2} \left[\frac{Q_{d_n+1}(X)Q_{d_n}(x) - Q_{d_n}(X)Q_{d_n+1}(x)}{X - x} \right], X \neq x$$

L'estimateur de $f(x)$ est de la forme :

$$\hat{f}_{d_n}(x) = \frac{d_n + 1}{2n} \sum_{i=1}^n \left[\frac{Q_{d_n+1}(X_i)Q_{d_n}(x) - Q_{d_n}(X_i)Q_{d_n+1}(x)}{X_i - x} \right], x \in \mathbb{R}.$$

• **Estimateur associé au fonction de Laguerre [46]**

Les fonctions de Laguerre sont définies par :

$$L_i(x) = \left[\frac{\Gamma(d_n + 1)}{\Gamma(d_n + 1 + \alpha)} x^{-\alpha} \exp(x) \right]^{\frac{1}{2}} \frac{1}{i!} \frac{d^i}{dx^i} (x^{i+\alpha} \exp(-x)), i \geq 0, x \geq 0.$$

Où, $\Gamma(\cdot)$ est la fonction gamma d'Euler, définie pour tout réel positif a par :

$$\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx.$$

Le noyau de Laguerre associé est défini par :

$$K_{d_n}(X, x) = \frac{\Gamma(d_n + 1)}{\Gamma(d_n + 1 + \alpha)} \left[\frac{L_{d_n+1}(X)L_{d_n}(x) - L_{d_n}(X)L_{d_n+1}(x)}{X - x} \right], X \neq x.$$

L'estimateur de $f(x)$ est de la forme :

$$\hat{f}_{d_n}(x) = \frac{\Gamma(d_n + 1)}{n\Gamma(d_n + 1 + \alpha)} \sum_{i=1}^n \left[\frac{L_{d_n+1}(X_i)L_{d_n}(x) - L_{d_n}(X_i)L_{d_n+1}(x)}{X_i - x} \right], x \geq 0.$$

• **Estimateur associé aux pôlynomes de Legendre [33]**

Les fonctions de Legendre sont définies par :

$$p_i(x) = \sqrt{\frac{2i + 1}{2}} \frac{1}{2^i i!} \frac{d^i}{dx^i} ((x^2 - 1)^i), x \in [-1, 1], i \geq 0.$$

Le noyau associé est défini par :

$$\begin{aligned}
K_{d_n}(X, x) &= \sum_{i=1}^{d_n} p_i(X)p_i(x) \\
&= \frac{d_n + 1}{\sqrt{2d_n + 1}\sqrt{2d_n + 3}} \left[\frac{p_{d_n}(X)L_{d_n+1}(x) - p_{d_n+1}(X)L_{d_n}(x)}{x - X} \right], X \neq x
\end{aligned}$$

L'estimateur de $f(x)$ est de la forme :

$$\hat{f}_{d_n}(x) = \frac{d_n + 1}{n\sqrt{2d_n + 1}\sqrt{2d_n + 3}} \sum_{i=1}^n \left[\frac{p_{d_n}(X_i)L_{d_n+1}(x) - p_{d_n+1}(X_i)L_{d_n}(x)}{x - X_i} \right], x \in [-1, 1].$$

2.6 Conclusion

Nous avons mis en œuvre dans ce chapitre différentes méthodes d'estimations non paramétriques de la densité ainsi que la fonction de répartition, en particulier la méthode des fonctions orthogonale. Nous avons donné les principales propriétés de cette méthode et quelques estimateurs basés sur des systèmes orthonormaux particuliers.

Estimation du taux de défaillance: application à base de Dirichlet

3.1 Introduction

Dans ce chapitre, nous présentons les estimateurs des fonctions de densité et de répartition par la méthode des fonctions orthogonales, relativement à la base de Dirichlet pour l'estimation du taux de défaillance. Une simulation est effectuée pour connaître le comportement des estimateurs.

Soit X_1, X_2, \dots, X_n une suite de variables aléatoires indépendantes et identiquement distribuées de la densité de probabilité inconnue f sur $[-\pi, \pi]$. Il s'agit d'estimer $f(x)$ à partir des observations x_1, x_2, \dots, x_n .

La base de Dirichlet est définie par :

$$e_1(x) = \frac{1}{\sqrt{2\pi}}, e_{2k}(x) = \frac{\cos(kx)}{\sqrt{\pi}}, e_{2k+1}(x) = \frac{\sin(kx)}{\sqrt{\pi}} \text{ pour } x \in [-\pi, \pi] \text{ et } k \geq 1.$$

Les données de fiabilité étant positive, nous considérons un intervalle $[0, b]$, $b > 0$.

La base de Dirichlet sur $[0, b]$ est alors :

$$e_1(x) = \frac{1}{\sqrt{b}}, e_{2k}(x) = \sqrt{\frac{2}{b}} \cos \left[\frac{2k\pi}{b} \left(x - \frac{b}{2} \right) \right], e_{2k+1}(x) = \sqrt{\frac{2}{b}} \sin \left[\frac{2k\pi}{b} \left(x - \frac{b}{2} \right) \right]. \quad (3.1)$$

Avec $x \in [0, b]$ et $k \geq 1$.

3.2 Estimation de la densité

En utilisant les relations (2.13) à (2.15) et les notations suivantes :

$$\text{Pour, } \beta_k = E \left[\sin \left(\frac{2k\pi}{b} \left(X - \frac{b}{2} \right) \right) \right] \text{ et } \gamma_k = E \left[\cos \left(\frac{2k\pi}{b} \left(X - \frac{b}{2} \right) \right) \right].$$

On obtient :

- Les coefficients de fourier sont donnés par :

$$a_k = E[e_k(X)] , \quad k = 1, \dots, d_n. \quad (3.2)$$

$$\begin{aligned} a_1 &= \frac{1}{\sqrt{b}} , \quad k = 1, \dots, d_n. \\ a_{2k} &= \sqrt{\frac{2}{b}} \gamma_k , \quad k = 1, \dots, d_n. \\ a_{2k+1} &= \sqrt{\frac{2}{b}} \beta_k , \quad k = 1, \dots, d_n. \end{aligned} \quad (3.3)$$

Les estimateurs des coefficients par la méthode des moments sont données par :

$$\hat{a}_k = \frac{1}{n} \sum_{i=1}^n e_k(X_i) , \quad k = 1, \dots, d_n. \quad (3.4)$$

$$\hat{a}_{2k} = \sqrt{\frac{2}{b}} \hat{\gamma}_k. \quad (3.5)$$

$$\hat{a}_{2k+1} = \sqrt{\frac{2}{b}} \hat{\beta}_k. \quad (3.6)$$

$$\text{Où, } \hat{\gamma}_k = \frac{1}{n} \sum_{i=1}^n \cos \left[\frac{2k\pi}{b} \left(x - \frac{b}{2} \right) \right] \text{ et } \hat{\beta}_k = \frac{1}{n} \sum_{i=1}^n \sin \left[\frac{2k\pi}{b} \left(x - \frac{b}{2} \right) \right].$$

$\hat{\gamma}_k$ et $\hat{\beta}_k$ sont des estimateurs sans biais de γ_k et β_k , respectivement.

- L'estimation de la densité est alors donné par :

$$\begin{aligned}\hat{f}_{d_n}(x) &= \sum_{k=1}^{d_n} \hat{a}_k e_k(x) \\ &= \frac{1}{b} + \frac{2}{nb} \sum_{i=1}^n \sum_{k=1}^{M_n} \cos \left[\frac{2k\pi}{b} (X_i - x) \right], \quad x \in [0, b]\end{aligned}$$

Où : $M_n = \left\lfloor \frac{d_n}{2} \right\rfloor$ et $\left\lfloor \frac{d_n}{2} \right\rfloor$ est la partie entière de $\frac{d_n}{2}$.

En utilisant le fait que : $\sum_{k=0}^m \cos(ky) = \frac{\sin \left(\left(\frac{m+1}{2} \right) y \right) \cos \left(\frac{my}{2} \right)}{\sin \left(\frac{y}{2} \right)}$.

On obtient :

$$\begin{aligned}\hat{f}_{d_n}(x) &= \frac{1}{b} + \frac{2}{nb} \sum_{i=1}^n \sum_{k=1}^{M_n} \cos(ky), \quad \text{avec } y = \frac{2k\pi}{b} (X_i - x) \\ &= \frac{1}{b} + \frac{2}{nb} \sum_{i=1}^n \left(\sum_{k=0}^{M_n} \cos(ky) - 1 \right), \quad \text{avec } y = \frac{2k\pi}{b} (X_i - x) \\ &= -\frac{1}{b} + \frac{2}{nb} \sum_{i=1}^n \left[\frac{\sin \left(\frac{M_n+1}{2} y \right) \cos \left(\frac{M_n}{2} y \right)}{\sin \left(\frac{y}{2} \right)} \right], \quad \text{avec } y = \frac{2k\pi}{b} (X_i - x) \\ &= -\frac{1}{b} + \frac{2}{nb} \sum_{i=1}^n \left[\frac{\frac{1}{2} \sin \left(\frac{(2M_n+1)}{2} y \right)}{\sin \left(\frac{y}{2} \right)} + \frac{1}{2} \right], \quad \text{avec } y = \frac{2k\pi}{b} (X_i - x)\end{aligned}$$

D'où,

$$\hat{f}_{d_n}(x) = \frac{1}{nb} \sum_{i=1}^n \left[\frac{\sin \left(\frac{\pi}{b} (2M_n+1) (X_i - x) \right)}{\sin \left(\frac{\pi}{b} (X_i - x) \right)} \right], \quad X_i \neq x, \quad x \in [0, b]. \quad (3.7)$$

3.2.1 Propriétés des $(\hat{a}_k)_{k \in \mathbb{N}}$

Les propriétés statistiques des $(\hat{a}_k)_{k \in \mathbb{N}}$ seront très utiles pour établir les propriétés statistiques de l'estimateur $\hat{f}_{d_n}(x)$ qui seront données dans la section suivantes.

• Propriétés statistique

Théorème 3.2.1 Soient $(\hat{a}_k)_{k \in \mathbb{N}}$ les estimateurs des coefficients de **Fourier** associés à f , alors :

$$1^\circ / \quad E(\hat{a}_k) = a_k \quad , \quad k = 1, \dots, d_n. \quad (3.8)$$

$$2^\circ / \quad V(\hat{a}_{2k}) = \frac{1}{nb} [1 + \gamma_{2k} - 2\gamma_k^2]. \quad (3.9)$$

$$V(\hat{a}_{2k+1}) = \frac{1}{nb} [1 - \gamma_{2k} - 2\beta_k^2]. \quad (3.10)$$

$$3^\circ / \quad Cov(\hat{a}_{2k'}, \hat{a}_{2j'}) = \frac{2}{nb} \left[\frac{1}{2} (\gamma_{k'+j'} + \gamma_{k'-j'}) - \gamma_{k'} \gamma_{j'} \right]. \quad (3.11)$$

$$Cov(\hat{a}_{2k'}, \hat{a}_{2j'+1}) = \frac{2}{nb} \left[\frac{1}{2} (\beta_{k'+j'} - \beta_{k'-j'}) - \gamma_{k'} \beta_{j'} \right]. \quad (3.12)$$

$$Cov(\hat{a}_{2k'+1}, \hat{a}_{2j'+1}) = \frac{2}{nb} \left[\frac{1}{2} (\gamma_{k'-j'} - \gamma_{k'+j'}) - \beta_{k'} \beta_{j'} \right]. \quad (3.13)$$

Preuve

1°/ On a

$$E(\hat{a}_{2k}) = \sqrt{\frac{2}{b}} \gamma_k = a_{2k}.$$

$$E(\hat{a}_{2k+1}) = \sqrt{\frac{2}{b}} \beta_k = a_{2k+1}.$$

D'où, \hat{a}_{2k} et \hat{a}_{2k+1} sont des estimateurs sans biais de a_{2k} et a_{2k+1} , respectivement.

2°/ On a

$$Var(\hat{a}_k) = \frac{1}{n} Var(e_k(X))$$

$$\begin{aligned} Var(\hat{a}_{2k}) &= Var \left(\sqrt{\frac{2}{b}} \hat{\gamma}_k \right) = \frac{2}{b} Var(\hat{\gamma}_k) = \frac{2}{nb} Var \left(\cos \left(\frac{2k\pi}{b} \left(X - \frac{b}{2} \right) \right) \right) \\ &= \frac{2}{nb} \left[E \left(\cos^2 \left(\frac{2k\pi}{b} \left(X - \frac{b}{2} \right) \right) \right) - \gamma_k^2 \right]. \end{aligned}$$

$$\text{Où, } E \left(\cos^2 \left(\frac{2k\pi}{b} \left(X - \frac{b}{2} \right) \right) \right) = E \left[\frac{1 + \cos \frac{4k\pi}{b} \left(X - \frac{b}{2} \right)}{2} \right] = \frac{1}{2} + \frac{1}{2}\gamma_{2k}.$$

$$\text{D'où : } \quad \text{Var}(\hat{a}_{2k}) = \frac{1}{nb} [1 + \gamma_{2k} - 2\gamma_k^2].$$

On applique les mêmes étapes précédentes on trouve

$$\text{Var}(\hat{a}_{2k+1}) = \frac{1}{nb} [1 - \gamma_{2k} - 2\beta_k^2].$$

3) On a

$$\text{Cov}(\hat{a}_k, \hat{a}_j) = E[\hat{a}_k \hat{a}_j] - a_k a_j.$$

On a trois cas possibles :

1^{ème} cas : k pair et j pair, $k = 2k'$, $j = 2j'$

$$\text{Cov}(e_{2k'}(X), e_{2j'}(X)) = E[\hat{a}_{2k'} \hat{a}_{2j'}] - a_{2k'} a_{2j'}.$$

$$\text{Où : } \quad E[\hat{a}_{2k'} \hat{a}_{2j'}] = E \left[\frac{2}{b} \hat{\gamma}_{k'} \hat{\gamma}_{j'} \right] = \frac{2}{b} E[\hat{\gamma}_{k'} \hat{\gamma}_{j'}].$$

$$\begin{aligned} E[\hat{a}_{2k'} \hat{a}_{2j'}] &= \frac{2}{b} \left[E \left(\frac{1}{n^2} \sum_{i=1}^n \cos \left(\frac{2k'\pi}{b} \left(X_i - \frac{b}{2} \right) \right) \sum_{l=1}^n \cos \left(\frac{2j'\pi}{b} \left(X_l - \frac{b}{2} \right) \right) \right) \right] \\ &= \frac{2}{b} \left[\frac{1}{n} E \left(\cos \left(\frac{2k'\pi}{b} \left(X_i - \frac{b}{2} \right) \right) \cos \left(\frac{2j'\pi}{b} \left(X_l - \frac{b}{2} \right) \right) \right) + \frac{n-1}{n} \gamma_{k'} \gamma_{j'} \right] \\ &= \frac{2}{b} \left[\frac{1}{2n} (\gamma_{k'+j'} + \gamma_{k'-j'}) + \frac{n-1}{n} \gamma_{k'} \gamma_{j'} \right]. \end{aligned}$$

$$\text{D'où : } \quad \text{Cov}(\hat{a}_{2k'}, \hat{a}_{2j'}) = \frac{2}{nb} \left[\frac{1}{2} (\gamma_{k'+j'} + \gamma_{k'-j'}) - \gamma_{k'} \gamma_{j'} \right].$$

2^{ème} cas : k pair et j impair, $k = 2k'$, $j = 2j' + 1$

$$\text{Cov}(e_{2k'}(X), e_{2j'+1}(X)) = E[\hat{a}_{2k'} \hat{a}_{2j'+1}] - a_{2k'} a_{2j'+1}.$$

$$\text{Où : } \quad E[\hat{a}_{2k'} \hat{a}_{2j'+1}] = E \left[\frac{2}{b} \hat{\gamma}_{k'} \hat{\beta}_{j'} \right] = \frac{2}{b} E[\hat{\gamma}_{k'} \hat{\beta}_{j'}].$$

$$\begin{aligned}
 E[\hat{a}_{2k'}\hat{a}_{2j'+1}] &= \frac{2}{b} \left[E \left(\frac{1}{n^2} \sum_{i=1}^n \cos \left(\frac{2k'\pi}{b} \left(X_i - \frac{b}{2} \right) \right) \sum_{l=1}^n \sin \left(\frac{2j'\pi}{b} \left(X_l - \frac{b}{2} \right) \right) \right) \right] \\
 &= \frac{2}{b} \left[\frac{1}{n} E \left(\cos \left(\frac{2k'\pi}{b} \left(X_i - \frac{b}{2} \right) \right) \sin \left(\frac{2j'\pi}{b} \left(X_l - \frac{b}{2} \right) \right) \right) + \frac{n-1}{n} \gamma_{k'} \beta_{j'} \right] \\
 E[\hat{a}_{2k'}\hat{a}_{2j'+1}] &= \frac{2}{b} \left[\frac{1}{2n} (\beta_{k'+j'} - \beta_{k'-j'}) + \frac{n-1}{n} \gamma_{k'} \beta_{j'} \right].
 \end{aligned}$$

$$\text{D'où : } \quad \text{Cov}(\hat{a}_{2k'}, \hat{a}_{2j'+1}) = \frac{2}{nb} \left[\frac{1}{2} (\beta_{k'+j'} - \beta_{k'-j'}) - \gamma_{k'} \beta_{j'} \right].$$

3^{ème} cas : k impair et j impair, $k = 2k' + 1$, $j = 2j' + 1$

$$\text{Cov}(e_{2k'+1}(X), e_{2j'+1}(X)) = E[\hat{a}_{2k'+1}\hat{a}_{2j'+1}] - a_{2k'+1}a_{2j'+1}.$$

$$\text{Où : } \quad E[\hat{a}_{2k'+1}\hat{a}_{2j'+1}] = E \left[\frac{2}{b} \hat{\beta}_{k'} \hat{\beta}_{j'} \right] = \frac{2}{b} E[\hat{\beta}_{k'} \hat{\beta}_{j'}].$$

$$\begin{aligned}
 E[\hat{a}_{2k'+1}\hat{a}_{2j'+1}] &= \frac{2}{b} \left[E \left(\frac{1}{n^2} \sum_{i=1}^n \sin \left(\frac{2k'\pi}{b} \left(X_i - \frac{b}{2} \right) \right) \sum_{l=1}^n \sin \left(\frac{2j'\pi}{b} \left(X_l - \frac{b}{2} \right) \right) \right) \right] \\
 &= \frac{2}{b} \left[\frac{1}{n} E \left(\sin \left(\frac{2k'\pi}{b} \left(X_i - \frac{b}{2} \right) \right) \sin \left(\frac{2j'\pi}{b} \left(X_l - \frac{b}{2} \right) \right) \right) + \frac{n-1}{n} \beta_{k'} \beta_{j'} \right] \\
 E[\hat{a}_{2k'+1}\hat{a}_{2j'+1}] &= \frac{2}{b} \left[\frac{1}{2n} (\gamma_{k'-j'} - \gamma_{k'+j'}) + \frac{n-1}{n} \beta_{k'} \beta_{j'} \right].
 \end{aligned}$$

$$\text{D'où : } \quad \text{Cov}(\hat{a}_{2k'+1}, \hat{a}_{2j'+1}) = \frac{2}{nb} \left[\frac{1}{2} (\gamma_{k'-j'} - \gamma_{k'+j'}) - \beta_{k'} \beta_{j'} \right].$$

3.2.2 Propriétés de \hat{f}_{d_n}

• Propriétés Statistiques

Biais de l'estimateur

$$\begin{aligned}
\text{Biais}(\hat{f}_{d_n}(x)) &= - \sum_{k=d_n+1}^{\infty} a_k e_k(x) \\
&= - \sum_{2k \geq d_n+1}^{\infty} a_{2k} e_{2k}(x) - \sum_{2k+1 \geq d_n+1}^{\infty} a_{k+1} e_{k+1}(x) \\
&= - \sum_{k \geq M_n+1}^{\infty} a_{2k} e_{2k}(x) - \sum_{k \geq \theta}^{\infty} a_{2k+1} e_{2k+1}(x) \\
&= - \sum_{k \geq M_n+1}^{\infty} \gamma_k \cos\left(\frac{2k\pi}{b} \left(X_i - \frac{b}{2}\right)\right) - \sum_{k \geq \theta}^{\infty} \beta_k \sin\left(\frac{2k\pi}{b} \left(X_i - \frac{b}{2}\right)\right), \quad x \in [0, b].
\end{aligned}$$

$$\text{Où, } M_n = \left\lfloor \frac{d_n}{2} \right\rfloor \text{ et } \theta = \begin{cases} M_n + 1 & \text{si } d_n \text{ est impair.} \\ M_n & \text{si } d_n \text{ est pair.} \end{cases}$$

D'où, on conclue que l'estimateur \hat{f} de f est un estimateur biaisé.

Variance de l'estimateur

$$\begin{aligned}
\text{Var}(\hat{f}_{d_n}(x)) &= \text{Var}\left(\sum_{k=1}^{d_n} \hat{a}_k e_k(x)\right) = \sum_{k=1}^{d_n} e_k^2(x) \text{Var}(\hat{a}_k) + \sum_{k \neq j}^{d_n} \text{Cov}(\hat{a}_k, \hat{a}_j) e_k(x) e_j(x) \\
&= \frac{1}{n} \left[\sum_{k=1}^{M_n} \frac{2}{nb^2} \cos^2\left(\frac{2\pi k}{b} \left(x - \frac{b}{2}\right)\right) (1 + \gamma_{2k} - 2\gamma_k^2) \right. \\
&\quad + \sum_{k=1}^{M_n} \frac{2}{nb^2} \sin^2\left(\frac{2\pi k}{b} \left(x - \frac{b}{2}\right)\right) (1 + \gamma_{2k} - 2\beta_k^2) \\
&\quad + \sum_{k' \neq j'}^{M_n} \frac{4}{nb^2} \cos\left(\frac{2\pi k'}{b} \left(x - \frac{b}{2}\right)\right) \cos\left(\frac{2\pi j'}{b} \left(x - \frac{b}{2}\right)\right) \left(\frac{1}{2} (\gamma_{k'+j'} + \gamma_{k'-j'}) - \gamma_{k'} \gamma_{j'}\right) \\
&\quad + \sum_{k' \neq j'}^{M_n} \frac{4}{nb^2} \sin\left(\frac{(2k'+1)\pi}{b} \left(x - \frac{b}{2}\right)\right) \sin\left(\frac{(2j'+1)\pi}{b} \left(x - \frac{b}{2}\right)\right) \\
&\quad \left. \left(\frac{1}{2} (\gamma_{k'-j'} - \gamma_{k'+j'}) - \beta_{k'} \beta_{j'}\right) + \sum_{k' \neq j'}^{M_n} \frac{4}{nb^2} \cos\left(\frac{2k'\pi}{b} \left(x - \frac{b}{2}\right)\right) \right. \\
&\quad \left. \sin\left(\frac{(2j'+1)\pi}{b} \left(x - \frac{b}{2}\right)\right) \left(\frac{1}{2} (\beta_{k'+j'} - \beta_{k'-j'}) - \gamma_{k'} \beta_{j'}\right) \right].
\end{aligned}$$

Avec, $M_n = \left\lceil \frac{d_n}{2} \right\rceil$.

Erreur quadratique moyenne

$$MSE(\hat{f}_{d_n}(x)) = Var(\hat{f}_{d_n}(x)) + Biases^2(\hat{f}_{d_n}(x)).$$

Erreur quadratique moyenne intégrée

$$\begin{aligned} MISE(\hat{f}_{d_n}) &= \int_0^b E \left[\hat{f}_{d_n}(X) - f(X) \right]^2 dx = \int_0^b f^2(x) dx + \sum_{k=0}^{d_n} [Var(\hat{a}_k) - a_k^2] \\ &= \int_0^b f^2(x) dx - \frac{1}{b} + \frac{1}{nb} \sum_{k \geq 1}^{M_n} [2 - 2(1+n)(\gamma_k^2 + \beta_k^2)]. \end{aligned}$$

• **Propriétés Asymptotiques**

Biais asymptotique

Théorème 3.2.2 Si $d_n \rightarrow \infty$ quand $n \rightarrow \infty$, l'estimateur $\hat{f}_{d_n}(x)$ est asymptotiquement sans biais.

Variance asymptotique

Théorème 3.2.3 Si d_n est d'ordre inférieur à \sqrt{n} c.à.d. ($d_n = o(\sqrt{n})$), alors :

$$\lim_{n \rightarrow \infty} V(\hat{f}_{d_n}(x)) = 0.$$

Convergence en moyenne quadratique

Théorème 3.2.4 Si d_n est d'ordre inférieur à \sqrt{n} c.à.d. ($d_n = o(\sqrt{n})$), et $d_n \rightarrow \infty$ quand $n \rightarrow \infty$. L'estimateur $\hat{f}_{d_n}(x)$ est alors un estimateur convergent en moyenne quadratique. C'est-à-dire :

$$\lim_{n \rightarrow \infty} MSE(\hat{f}_{d_n}(x)) = \lim_{n \rightarrow \infty} E \left[\hat{f}_{d_n}(x) - f(x) \right]^2 = 0.$$

Convergence en moyenne quadratique intégrée

Théorème 3.2.5 (Bleuez et Bosq [6])

Si $f \in L^2$, pour que $E \left(\int_0^b \left| \hat{f}_{d_n}(x) - f(x) \right|^2 dx \right) \rightarrow 0$, il faut et il suffit que $\frac{d_n}{n} \rightarrow 0$.

3.3 La fonction de répartition

Soit X_1, X_2, \dots, X_n une suite de v.a.i.i.d de la densité inconnue $f(x)$ et d'une fonction de répartition $F(x)$ définie sur $[0, b]$. Il s'agit d'estimer $F(x)$ à partir des observations x_1, x_2, \dots, x_n .

Les coefficients de Fourier :

$$A_k = \int_0^b e_k F(x) dx, \quad k = 1, \dots, d_n. \quad (3.14)$$

Sont donnés par :

$$A_1 = \frac{1}{\sqrt{b}} (b - \mu), \quad \mu = E[X], \quad (3.15)$$

$$A_{2k} = \frac{b}{k\pi} \left[(-1)^{k+1} + \gamma_k \right], \quad \gamma_k = E \left[\cos \frac{2k\pi}{b} \left(X - \frac{b}{2} \right) \right], \quad (3.16)$$

$$\text{et } A_{2k+1} = -\frac{1}{k\pi} \sqrt{\frac{b}{2}} \beta_k, \quad \beta_k = E \left[\sin \frac{2k\pi}{b} \left(X - \frac{b}{2} \right) \right], \quad k \geq 1. \quad (3.17)$$

Les estimateurs des $(A_k)_{k \in \mathbb{N}}$ sont donnés on utilisant la méthode des moments par :

$$\hat{A}_k = \frac{1}{n} \sum_{i=1}^n \int_{X_i}^b e_k(x) dx, \quad x \in [0, b], \quad k = 1, \dots, d_n. \quad (3.18)$$

$$\hat{A}_1 = \frac{1}{\sqrt{b}} (b - \bar{X}), \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i. \quad (3.19)$$

$$\hat{A}_{2k} = \frac{1}{k\pi} \sqrt{\frac{b}{2}} \left[(-1)^{k+1} + \hat{\gamma}_k \right], \quad \hat{\gamma}_k = \frac{1}{n} \sum_{i=1}^n \cos \left(\frac{2k\pi}{b} \left(X_i - \frac{b}{2} \right) \right), \quad (3.20)$$

$$\hat{A}_{2k+1} = -\frac{1}{k\pi} \sqrt{\frac{b}{2}} \hat{\beta}_k, \quad \hat{\beta}_k = \frac{1}{n} \sum_{i=1}^n \sin \left(\frac{2k\pi}{b} \left(X_i - \frac{b}{2} \right) \right). \quad (3.21)$$

Où, $\hat{\beta}_k$ et $\hat{\gamma}_k$ sont les estimateurs de β_k et γ_k respectivement.

En utilisant les même étapes que la fonction de densité, l'estimateur $\hat{F}_{d_n}(x)$ de $F_{d_n}(x)$ est donné par :

$$\hat{F}_{d_n}(x) = \left(1 - \frac{\bar{X}}{b}\right) - \frac{2}{n} \sum_{i=1}^n \sum_{k \geq 1}^{M_n} \frac{1}{k\pi} \left[\sin\left(\frac{k\pi X_i}{2}\right) \cos\left(\frac{k\pi(2x - X_i)}{2}\right) \right], X_i \neq x \text{ et } x \in [0, b]. \quad (3.22)$$

3.3.1 Propriétés de \hat{F}_{d_n}

• Propriétés statistiques

Biais de l'estimateurs

$$\begin{aligned} \text{Biais}\left(\hat{F}_{d_n}(x)\right) &= - \sum_{k \geq M_n+1} \frac{1}{k\pi} \left[(-1)^{k+1} + \hat{\gamma}_k \right] \sin\left[\frac{2k\pi}{b} \left(x - \frac{b}{2}\right)\right] \\ &\quad - \sum_{k \geq \theta} \frac{1}{k\pi} \hat{\beta}_k \cos\left[\frac{2k\pi}{b} \left(x - \frac{b}{2}\right)\right]. \end{aligned}$$

$$\text{Où, } \theta = \begin{cases} M_n + 1 & \text{si } d_n \text{ est impair.} \\ M_n & \text{si } d_n \text{ est pair.} \end{cases} \quad (1)$$

Avec, $\text{Biais}\left(\hat{F}_{d_n}(x)\right) \rightarrow 0, n \rightarrow +\infty$.

Variance de l'estimateur

$$\begin{aligned} \text{Var}\left(\hat{F}_{d_n}(x)\right) &= \frac{\sigma^2}{nb^2} + \sum_{k=1}^{M_n} \left[\frac{1}{2n(k\pi)^2} \left[(1 - \gamma_k^2 - \beta_k^2) + (\gamma_k^2 - \beta_k^2 + \gamma_{2k}) \cos\left(\frac{4k\pi}{b} \left(x - \frac{b}{2}\right)\right) \right] \right] \\ &\quad + \sum_{k=1}^{M_n} \frac{1}{2nk\pi} \left[a_k \sin\left(\frac{2k\pi}{b} \left(x - \frac{b}{2}\right)\right) + b_k \cos\left(\frac{2k\pi}{b} \left(x - \frac{b}{2}\right)\right) \right] \\ &\quad - \sum_{k=1}^{M_n} \frac{1}{2n\pi^2} \left[c_k \sin\left(\frac{2k\pi}{b} \left(x - \frac{b}{2}\right)\right) \cos\left(\frac{2k\pi}{b} \left(x - \frac{b}{2}\right)\right) \right] \\ &\quad - \sum_{\substack{k=1 \\ k \neq j}}^{M_n} \sum_{j=1}^{M_n} \frac{1}{2n\pi^2} \left[d_k \sin\left(\frac{2\pi}{b} (k-j) \left(x - \frac{b}{2}\right)\right) - g_k \cos\left(\frac{2\pi}{b} (k+j) \left(x - \frac{b}{2}\right)\right) \right]. \end{aligned}$$

$$\text{Où : } \begin{cases} a_k = E \left(X \cos \left(\frac{2k\pi}{b} \left(x - \frac{b}{2} \right) \right) \right) + \mu\gamma_k. \\ b_k = E \left(X \sin \left(\frac{2k\pi}{b} \left(x - \frac{b}{2} \right) \right) \right) - \mu\beta_k. \\ c_k = \frac{\beta_{k+j} - \beta_{k-j}}{2} - \beta_k\beta_j. \\ d_k = \gamma_{k-j} - \gamma_k\gamma_j - \beta_k\beta_j. \\ g_k = \gamma_{k+j} - \gamma_k\gamma_j - \beta_k\beta_j. \end{cases}$$

Erreur quadratique moyenne

$$MSE(\hat{F}_{d_n}(x)) = Var(\hat{F}_{d_n}(x)) + BiAIS^2(\hat{F}_{d_n}(x)).$$

Erreur quadratique moyenne intégrée

$$\begin{aligned} MISE(\hat{F}_{d_n}) &= \frac{\sigma^2}{nb} + \sum_{k=1}^{M_n} \frac{b}{2n(k\pi)^2} (1 - \gamma_k^2 - \beta_k^2) \\ &\quad + \sum_{k \geq M_n+1} \frac{b}{2k\pi} [(-1)^{k+1} + \hat{\gamma}_k] + \sum_{k \geq \theta} \frac{1}{k\pi} \hat{\beta}_k^2. \end{aligned}$$

Où, θ vérifie (1).

• Propriétés asymptotiques

Biais asymptotique

Théorème 3.3.1 Si $d_n \rightarrow \infty$ lorsque $n \rightarrow \infty$, l'estimateur $\hat{F}_{d_n}(x)$ est asymptotiquement sans biais.

Variance asymptotique

Théorème 3.3.2 Si $d_n = o(\sqrt{n})$ et $d_n \rightarrow \infty$ lorsque $n \rightarrow \infty$, alors :

$$\lim_{n \rightarrow \infty} Var \left(\hat{F}_{d_n}(x) \right) = 0.$$

Convergence en moyenne quadratique

Théorème 3.3.3 Si $d_n = o(\sqrt{n})$ et $d_n \rightarrow \infty$ lorsque $n \rightarrow \infty$, alors :

$$\lim_{n \rightarrow \infty} MSE \left(\hat{F}_{d_n}(x) \right) = \lim_{n \rightarrow \infty} E \left[\hat{F}_{d_n}(x) - F(x) \right]^2 = 0.$$

Convergence en moyenne quadratique intégrée

Théorème 3.3.4 Si $d_n = o(\sqrt{n})$ et $d_n \rightarrow \infty$ lorsque $n \rightarrow \infty$, alors :

$$\lim_{n \rightarrow \infty} MISE \left(\hat{F}_{d_n}(x) \right) = \lim_{n \rightarrow \infty} E \int_0^b \left[\hat{F}_{d_n}(x) - F(x) \right]^2 dx = 0.$$

3.4 Estimation de la fiabilité et du taux défaillance

L'estimateur de la fiabilité peut être obtenu à partir de l'estimateur de la fonction de répartition associé à la base de Dirichlet défini dans (3.22)

$$\hat{R}_{d_n}(x) = 1 - \hat{F}_{d_n}(x) \quad (3.23)$$

L'estimateur du taux de défaillance est alors donné par :

$$\hat{\lambda}_{d_n}(x) = \frac{\hat{f}_{d_n}(x)}{\hat{R}_{d_n}(x)}, \quad x > 0. \quad (3.24)$$

Où, $\hat{f}_{d_n}(x)$ est l'estimateur associé à la base Dirichlet de la fonction densité définie dans la formule (3.6).

3.5 Simulation

3.5.1 Position du problème

Soit une variable aléatoire X qui représente la durée de vie d'un élément (durée de bon fonctionnement jusqu'à l'apparition de la première panne). Dans la pratique, nous nous intéressons à calculer la fiabilité et le taux de défaillance pour des valeurs de x positifs, on peut donc considérer un intervalle $[0, b]$, $b > 0$ et utiliser les estimateurs définis dans (3.23) et (3.24).

On est alors amené à utiliser les fonctions suivantes :

considérons $x \in [0, b]$, $b > 0$,

$$R^*(x) = 1 - F^*(x) \quad \text{et} \quad \lambda^*(x) = \frac{f^*(x)}{1 - F^*(x)} \quad \text{pour } x \in [0, b].$$

Où

$$F^*(x) = \frac{F(x)}{F(b)} \quad \text{et} \quad f^*(x) = \frac{f(x)}{F(b)} \quad \text{pour } x \in [0, b].$$

D'où, les estimateurs :

$$\hat{R}_{d_n}^*(x) = 1 - \hat{F}_{d_n}^*(x) \quad \text{et} \quad \hat{\lambda}_{d_n}^*(x) = \frac{\hat{f}_{d_n}^*(x)}{1 - \hat{F}_{d_n}^*(x)} \quad \text{pour } x \in [0, b].$$

Où, $\hat{f}_{d_n}^*$ et $\hat{F}_{d_n}^*$ sont les estimateurs de la densité et de la fonction de répartition obtenus par la méthode des fonctions orthogonales.

Pour cela on se fixe les paramètres suivants :

Base orthogonormée : La base de Dirichlet sur l'intervalle $[0, b]$.

Le paramètre de lissage : Le paramètre de lissage d_n obtenu par la méthode de Kronmal-Tarter.

Lois de fiabilités : Nous construisons les échantillons simulés à partir des densités cibles suivantes :

La loi Beta (α, β) .

La loi Gamma (α, λ) .

La loi Weibull (η, β) .

La loi Log-Normale (μ, σ^2) .

Pour les calculs et les représentations graphiques, nous avons utilisé le logiciel de statistique R dans lequel nous avons programmé les différents estimateurs.

3.5.2 Algorithme de simulation

Afin de visualiser l'allure de la fiabilité, nous suivons les étapes suivantes :

1. Fixé la valeur de b et la taille de l'échantillon n .
2. Générer les durées de vie suivant les lois usuelles en fiabilité (Beta, Gamma, Log-normale).
3. Utiliser la base de Dirichlet, et la fenêtre de lissage d_n par la méthode de Kronmal-Tarter.
4. Introduire les points cibles x .
5. Calculer la valeur théorique de la fonction densité $f(x)$ et son estimateur $\hat{f}_{d_n}^*(x)$.

6. Calculer la valeur théorique de la fonction de répartition $F(x)$ et son estimateur $\hat{F}_{d_n}^*(x)$.
7. Calculer la valeur théorique de la fiabilité $R(x)$ et son estimateur $\hat{R}_{d_n}^*(x)$.
8. Calculer la valeur théorique du taux de hasard $\lambda(x)$ et son estimateur $\hat{\lambda}_{d_n}^*(x)$.
9. Calculer le ISE de la fiabilité et du taux de hasard, tel que :

$$ISE = \int \left(\hat{f}_{d_n}^*(x) - f(x) \right)^2 dx$$

10. Tracer les courbes correspondantes à $f(x)$, $R(x)$, $\lambda(x)$ et leurs estimateurs $\hat{f}_{d_n}^*(x)$, $\hat{R}_{d_n}^*(x)$, $\hat{\lambda}_{d_n}^*(x)$.

En se basant sur l'algorithme décrit ci-dessus, nous simulons les lois usuelles de fiabilité citées auparavant pour $n = (100, 500, 5000)$.

Les résultats obtenus par la simulation sont donnés sous forme de tableaux et graphiques. Ce qui nous permet de visualiser le comportement et l'allure des fonctions densité, répartition, fiabilité et le taux de hasard qui diffère selon la loi retenue. Les graphiques permettent de faire une comparaison entre les résultats théoriques et les résultats obtenus par estimation.

3.5.3 Résultats de simulation

Les résultats obtenus pas la simulation sont donnés sous forme de tableaux voir[(3.1), (3.2) et (3.3)] et graphiques voir [(3.1) à (3.12)]. Ce qui nous permet de visualiser le comportement de l'allure des courbes des fonctions densite de défaillance, répartition, fiabilité et celle du taux de hasard. Les courbes correspondantes aux fonctions théoriques sont en couleur noire et celles correspondantes aux fonctions estimés sont en couleur rouge.

• Loi Beta(3,5)

		Densité de défaillance			Fonction de répartition		Fiabilité			Taux de défaillance				
n	x	d	$f_{ih}(x)$	$\hat{f}^*(x)$	MSE	D	$F_{ih}(x)$	$\hat{F}^*(x)$	$R_{ih}(x)$	$\hat{R}^*(x)$	MSE	$\lambda_{ih}(x)$	$\hat{\lambda}^*(x)$	MSE
100	0,228	2	1.631	1,364	$7.13e^{-2}$	2	0,279	0,194	0,720	0,805	$7.22e^{-3}$	2,265	1,694	$3.26e^{-1}$
	0,267		1.721	1,509	$4.49e^{-2}$		0,326	0,198	0,673	0,801	$1.79e^{-2}$	2,557	1,883	$4.54e^{-1}$
	0,341		1.797	1,694	$1.06e^{-2}$		0,447	0,289	0,552	0,710	$2.52e^{-2}$	3,255	2,385	$7.56e^{-1}$
	0,409		1.714	1,735	$4.41e^{-4}$		0,599	0,448	0,400	0,551	$2.24e^{-2}$	4,285	3,148	$9.99e^{-1}$
500	0,228	3	1.631	1,528	$1.68e^{-2}$	3	0,279	0,200	0,720	0,799	$6.26e^{-3}$	2,265	1,912	$1.24e^{-1}$
	0,267		1.721	1,886	$2.62e^{-2}$		0,326	0,229	0,673	0,770	$9.66e^{-3}$	2,557	2,449	$1.16e^{-2}$
	0,341		1.797	1,820	$5.29e^{-3}$		0,447	0,331	0,552	0,668	$1.85e^{-2}$	3,255	2,724	$2.81e^{-1}$
	0,409		1.714	1,805	$8.56e^{-3}$		0,599	0,507	0,400	0,492	$5.81e^{-3}$	4,285	3,668	$3.80e^{-1}$
5000	0,228	4	1.631	1,635	$1.60e^{-5}$	4	0,279	0,229	0,720	0,770	$3.28e^{-3}$	2,265	2,123	$2.02e^{-2}$
	0,267		1.721	1,763	$1.52e^{-3}$		0,326	0,300	0,673	0,699	$5.16e^{-4}$	2,557	1,991	$3.20e^{-1}$
	0,341		1.797	1,828	$9.61e^{-4}$		0,447	0,452	0,552	0,547	$2.46e^{-5}$	3,255	2,681	$3.29e^{-1}$
	0,409		1.714	1,729	$2.50e^{-4}$		0,599	0,601	0,400	0,398	$3.72e^{-6}$	4,285	4,344	$3.48e^{-3}$

TAB.3.1 Résultats de simulation de la loi Beta(3,5).

n	Densité de défaillance	Fiabilité	Taux de défaillance
100	$1.166e^{-2}$	$3.303e^{-2}$	$4.329e^{-2}$
500	$3.491e^{-4}$	$2.817e^{-3}$	$3.478e^{-2}$
5000	$3.614e^{-6}$	$6.845e^{-5}$	$6.034e^{-3}$

TAB.3.2 Résultats du ISE de la loi Beta (3,5).

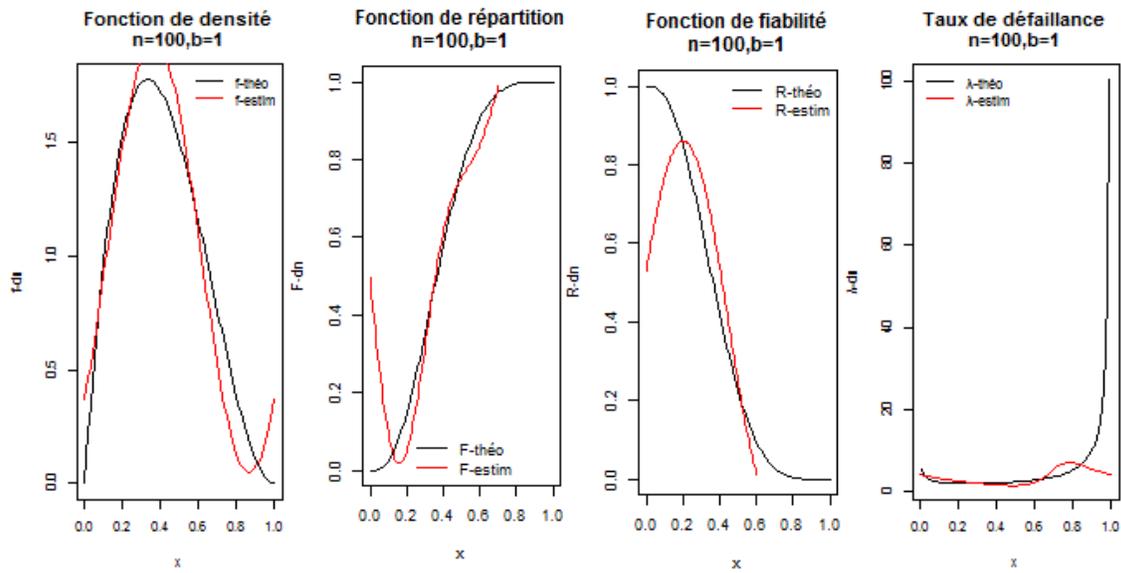


FIG.3.1 Densité, Fonction de répartition, Fiabilité et Taux de hasard associés à la loi Beta (3,5) pour n est à 100.

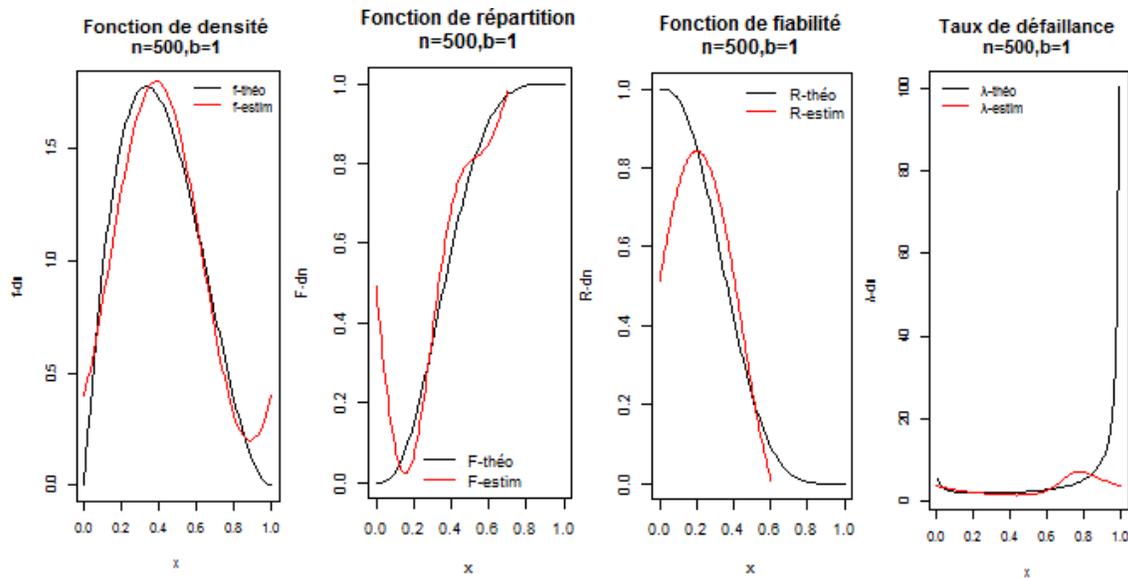


FIG.3.1 Densité, Fonction de répartition, Fiabilité et Taux de hasard associés à la loi Beta (3,5) pour n est à 500.

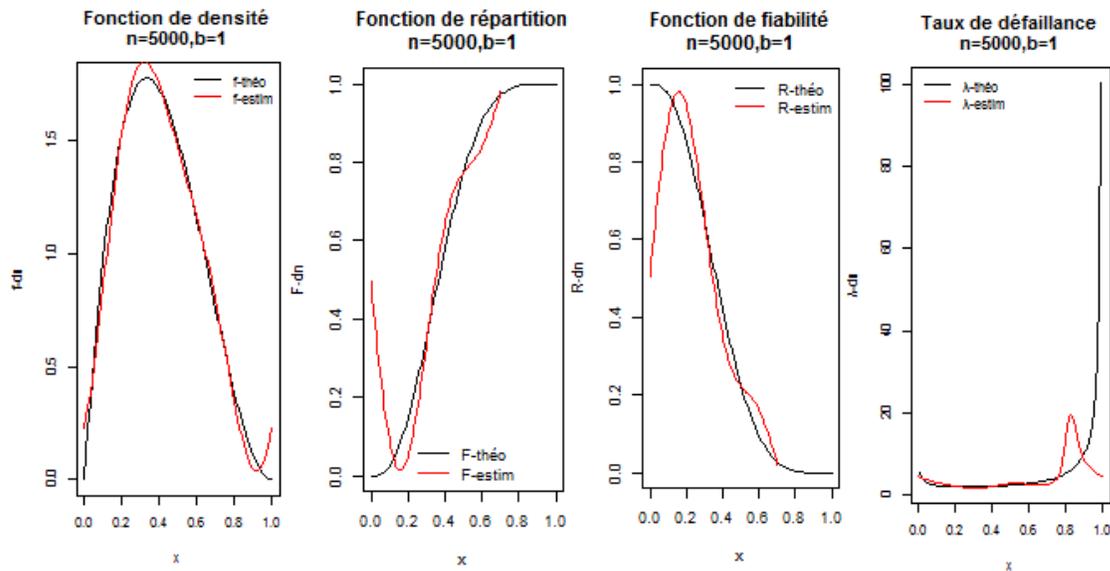


FIG.3.3 Densité, Fonction de répartition, Fiabilité et Taux de hasard associés à la loi Beta (3,5) pour n est à 5000.

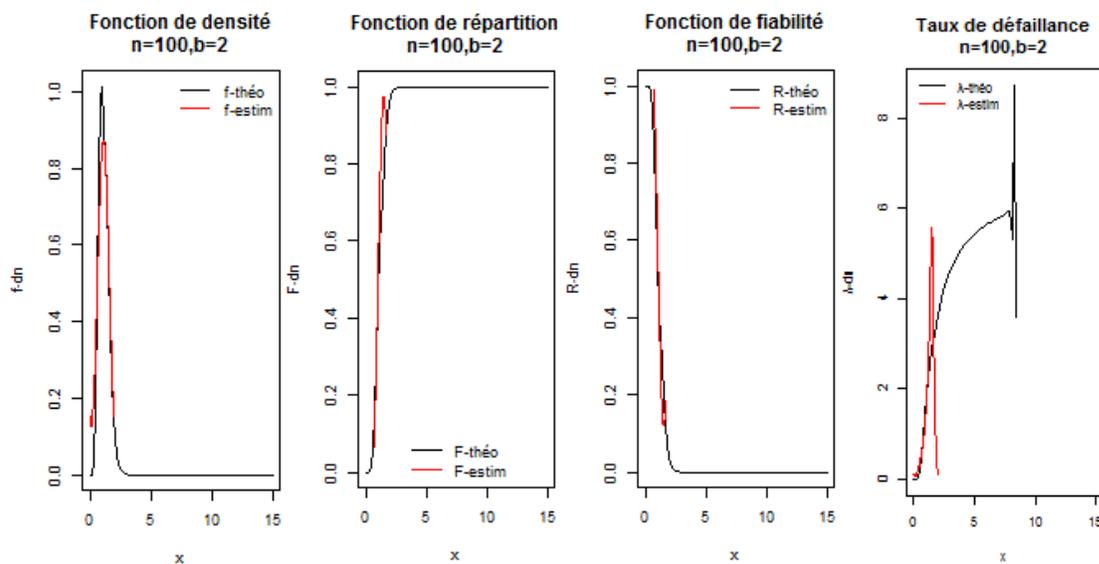
● Loi Gamma (7.8,6.75)

		Densité de défaillance			Fonction de répartition		Fiabilité			Taux de défaillance				
n	x	d	$f_{dh}(x)$	$\hat{f}^*(x)$	MSE	d	$F_{dh}(x)$	$\hat{F}^*(x)$	$R_{dh}(x)$	$\hat{R}^*(x)$	MSE	$\lambda_{dh}(x)$	$\hat{\lambda}^*(x)$	MSE
100	0.906	2	0.270	0.133	$2.33e^{-2}$	2	0.087	0.186	0.912	0.813	$9.34e^{-3}$	0.296	0.163	$1.84e^{-2}$
	1.003		0.914	0.843	$5.69e^{-4}$		0.306	0.350	0.693	0.649	$1.13e^{-3}$	1.319	1.298	$4.12e^{-4}$
	1.225		0.927	0.960	$1.06e^{-3}$		0.512	0.489	0.487	0.510	$5.88e^{-4}$	1.903	1.882	$4.88e^{-4}$
	1.308		0.844	0.908	$8.58e^{-3}$		0.585	0.450	0.414	0.549	$1.59e^{-2}$	2.038	1.654	$1.16e^{-1}$
500	0.906	3	0.270	0.187	$6.84e^{-3}$	3	0.078	0.135	0.912	0.864	$1.32e^{-3}$	0.296	0.216	$5.96e^{-3}$
	1.003		0.914	0.930	$2.46e^{-4}$		0.306	0.403	0.693	0.596	$9.21e^{-3}$	1.319	1.560	$4.08e^{-2}$
	1.225		0.927	0.951	$5.72e^{-4}$		0.512	0.542	0.487	0.457	$2.63e^{-4}$	1.903	2.054	$2.79e^{-2}$
	1.308		0.844	0.837	$4.94e^{-5}$		0.585	0.494	0.414	0.505	$9.93e^{-3}$	2.038	1.657	$1.45e^{-1}$
5000	0.906	4	0.270	0.267	$9.97e^{-6}$	4	0.087	0.091	0.912	0.908	$1.38e^{-5}$	0.296	0.294	$4.36e^{-6}$
	1.003		0.914	0.905	$8.15e^{-5}$		0.306	0.320	0.693	0.679	$1.06e^{-4}$	1.319	1.333	$1.67e^{-4}$
	1.225		0.927	0.919	$6.42e^{-5}$		0.512	0.511	0.487	0.488	$2.29e^{-6}$	1.903	1.889	$1.91e^{-4}$
	1.308		0.844	0.845	$1.55e^{-6}$		0.585	0.583	0.414	0.416	$4.22e^{-6}$	2.038	2.031	$1.24e^{-1}$

TAB.3.3 Résultats de simulation de la loi Gamma (7.8,6.75).

n	Densité de défaillance	Fiabilité	Taux de défaillance
100	$3.648.10^{-2}$	$5.246.10^{-2}$	$6.953.10^{-2}$
500	$1.293.10^{-4}$	$2.869.10^{-3}$	$4.507.10^{-2}$
5000	$7.452.10^{-6}$	$1.348.10^{-5}$	$5.528.10^{-3}$

TAB.3.4 Résultats du ISE de la loi Gamma (7.8,6.75).

FIG.3.4 Densité, Fonction de répartition, Fiabilité et Taux de hasard associés à la loi Gamma (7.8,6.75) pour n est à 100.

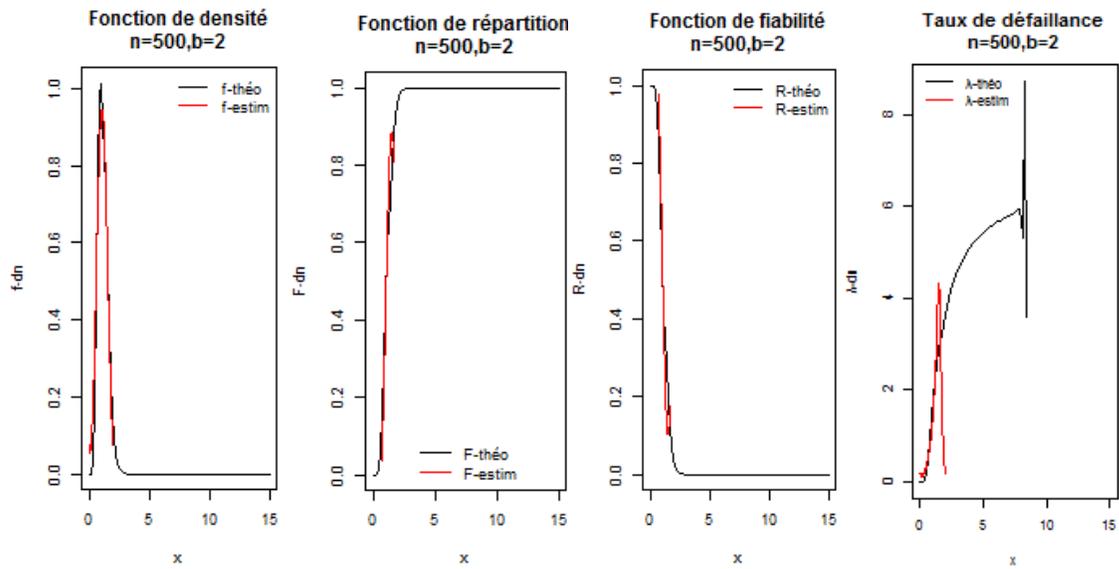


FIG.3.5 Densité, Fonction de répartition, Fiabilité et Taux de hasard associés à la loi Gamma (7.8,6.75) pour n est à 500.

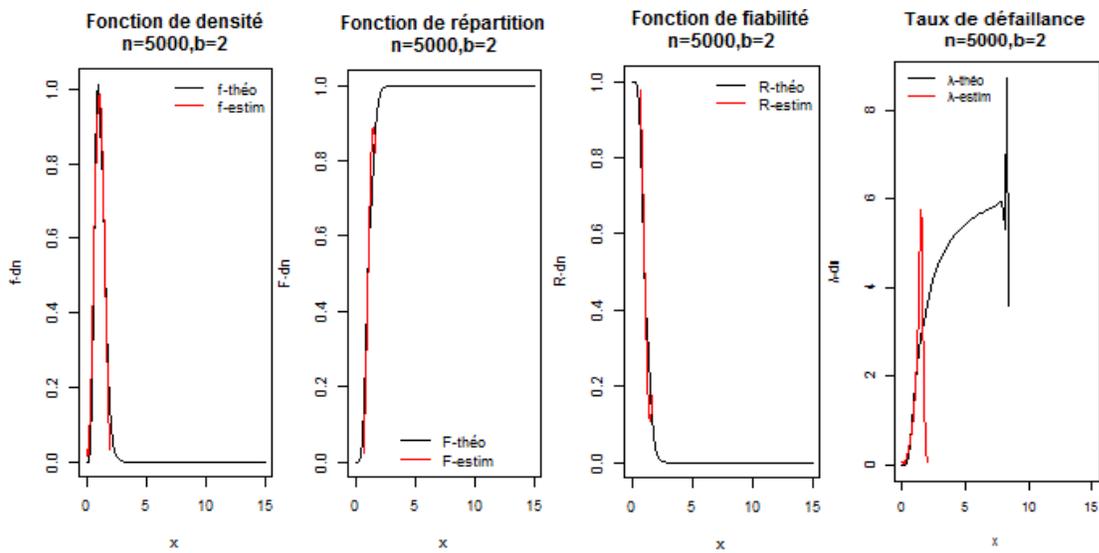


FIG.3.6 Densité, Fonction de répartition, Fiabilité et Taux de hasard associés à la loi Gamma (7.8,6.75) pour n est à 5000.

• Loi Weibull $\mathcal{W}(4.8,3)$

		Densité de défaillance			Fonction de répartition		Fiabilité			Taux de défaillance				
n	x	d	$f_{\hat{\eta}}(x)$	$\hat{f}^*(x)$	MSE	d	$F_{\hat{\eta}}(x)$	$\hat{F}^*(x)$	$R_{\hat{\eta}}(x)$	$\hat{R}^*(x)$	MSE	$\lambda_{\hat{\eta}}(x)$	$\hat{\lambda}^*(x)$	MSE
100	2.158	2	0.372	0.389	$2.89e^{-3}$	2	0.186	0.172	0.814	0.828	$1.04e^{-3}$	0.458	0.470	$1.44e^{-4}$
	2.358		0.468	0.429	$1.52e^{-3}$		0.270	0.292	0.729	0.707	$4.05e^{-3}$	0.641	0.617	$5.76e^{-4}$
	2.616		0.566	0.502	$1.29e^{-3}$		0.404	0.461	0.596	0.539	$2.06e^{-2}$	0.951	0.933	$3.24e^{-4}$
	3.001		0.588	0.471	$1.37e^{-2}$		0.633	0.707	0.367	0.293	$3.30e^{-2}$	1.602	1.615	$2.25e^{-3}$
500	2.158	3	0.372	0.407	$1.22e^{-3}$	3	0.186	0.164	0.814	0.835	$1.24e^{-3}$	0.458	0.487	$1.69e^{-4}$
	2.358		0.468	0.449	$3.61e^{-4}$		0.270	0.286	0.729	0.713	$1.55e^{-4}$	0.641	0.621	$4.00e^{-4}$
	2.616		0.566	0.529	$1.13e^{-3}$		0.404	0.457	0.596	0.542	$1.40e^{-3}$	0.951	0.976	$6.25e^{-4}$
	3.001		0.588	0.472	$1.34e^{-2}$		0.633	0.706	0.367	0.293	$3.48e^{-3}$	1.602	1.610	$6.25e^{-4}$
5000	2.158	3	0.372	0.387	$2.25e^{-4}$	4	0.186	0.170	0.814	0.830	$4.11e^{-4}$	0.458	0.477	$8.74e^{-4}$
	2.358		0.468	0.451	$2.89e^{-4}$		0.270	0.282	0.729	0.718	$3.02e^{-4}$	0.641	0.628	$1.69e^{-4}$
	2.616		0.566	0.551	$2.25e^{-4}$		0.404	0.439	0.596	0.561	$2.95e^{-5}$	0.951	0.979	$7.84e^{-4}$
	3.001		0.588	0.490	$9.60e^{-3}$		0.633	0.691	0.367	0.309	$5.59e^{-5}$	1.602	1.609	$4.90e^{-5}$

TAB.3.5 Résultats de simulation de la loi Weibull (4.8,3).

n	Densité de défaillance	Fiabilité	Taux de défaillance
100	$1.043 \cdot 10^{-3}$	$5.660 \cdot 10^{-4}$	$6.368 \cdot 10^{-4}$
500	$1.60 \cdot 10^{-4}$	$9.210 \cdot 10^{-5}$	$1.398 \cdot 10^{-4}$
5000	$9.65 \cdot 10^{-5}$	$3.201 \cdot 10^{-5}$	$5.781 \cdot 10^{-6}$

TAB.3.6 Résultats du ISE de la loi Weibull (4.8,3).

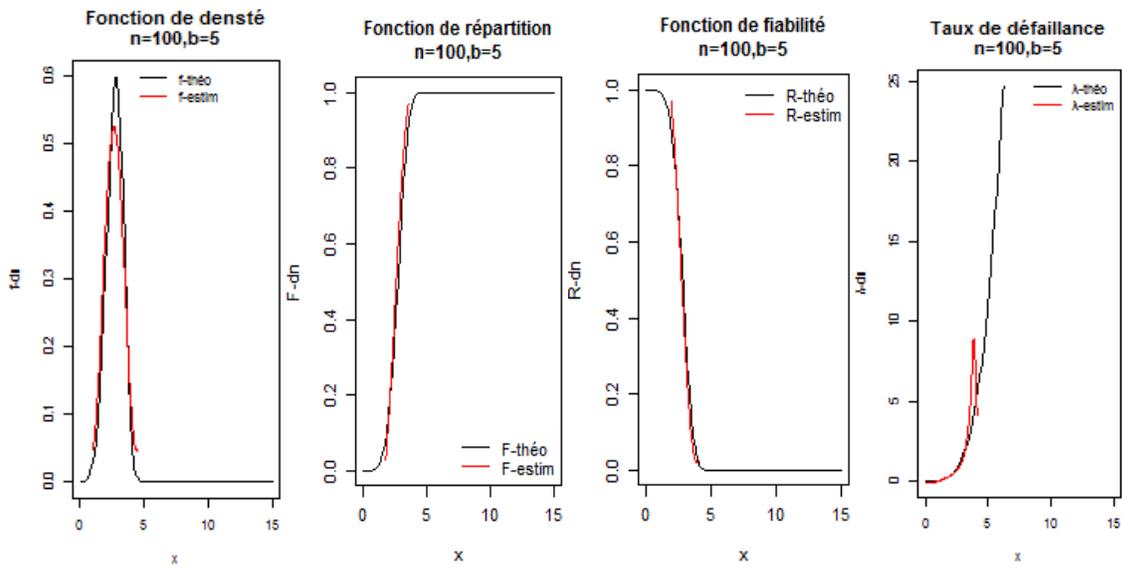


FIG.3.8 Densité, Fonction de répartition, Fiabilité et Taux de hasard associés à la loi Weibull (4.8,3) pour n est à 100.

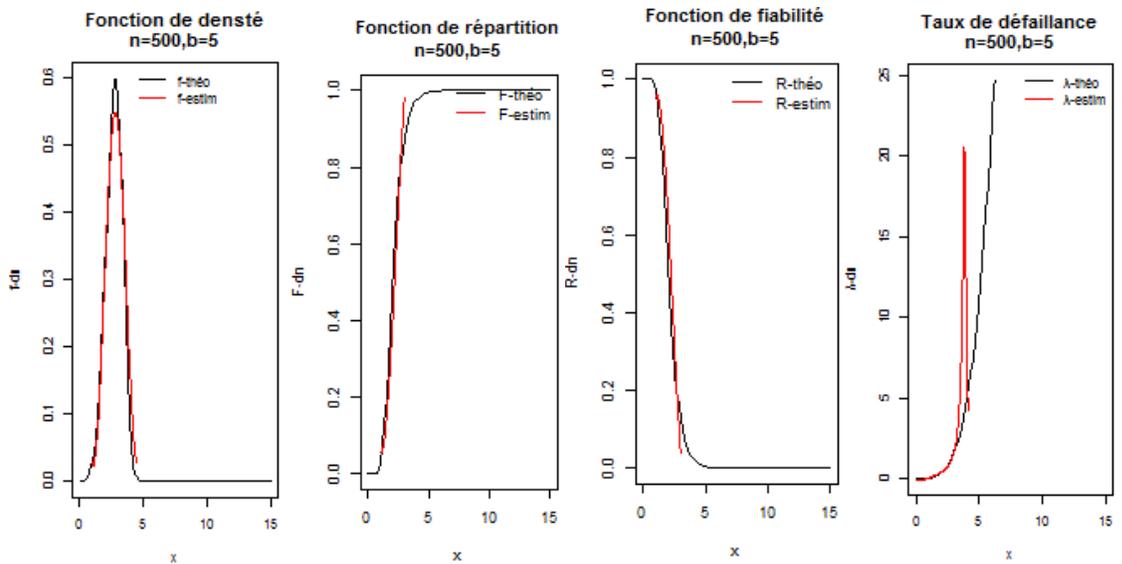


FIG.3.8 Densité, Fonction de répartition, Fiabilité et Taux de hasard associés à la loi Weibull (4.8,3) pour n est à 500.

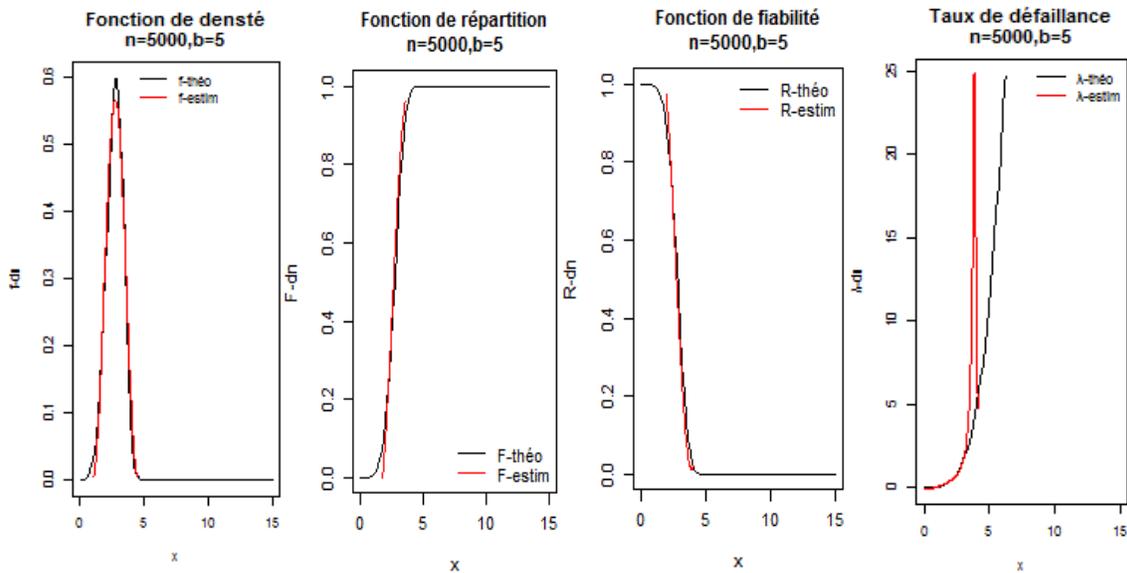


FIG.3.9 Densité, Fonction de répartition, Fiabilité et Taux de hasard associés à la loi Weibull (4.8,3) pour n est à 5000.

● Loi log-normal log- \mathcal{N} (0.7,0.35)

		Densité de défaillance				Fonction de répartition			Fiabilité			Taux de défaillance		
n	x	d	$f_{th}(x)$	$\hat{f}^*(x)$	MSE	d	$F_{th}(x)$	$\hat{F}^*(x)$	$R_{th}(x)$	$\hat{R}^*(x)$	MSE	$\lambda_{th}(x)$	$\hat{\lambda}^*(x)$	MSE
100	1.589	2	0.570	0.555	$2.21e^{-4}$	2	0.249	0.218	0.750	0.781	$9.62e^{-4}$	0.760	0.732	$7.85e^{-4}$
	1.680		0.593	0.592	$1.89e^{-6}$		0.302	0.298	0.708	0.702	$3.61e^{-5}$	0.850	0.842	$6.01e^{-5}$
	1.822		0.601	0.632	$6.69e^{-4}$		0.387	0.420	0.613	0.580	$1.09e^{-3}$	0.980	1.091	$1.24e^{-2}$
	2.258		0.478	0.422	$3.16e^{-3}$		0.628	0.680	0.371	0.321	$2.52e^{-3}$	1.287	1.325	$1.48e^{-3}$
500	1.589	3	0.570	0.560	$1.02e^{-4}$	3	0.249	0.252	0.750	0.748	$4.01e^{-6}$	0.760	0.740	$4.01e^{-4}$
	1.680		0.593	0.590	$9.01e^{-6}$		0.302	0.339	0.708	0.671	$1.37e^{-3}$	0.850	0.879	$8.42e^{-4}$
	1.822		0.601	0.583	$3.24e^{-4}$		0.387	0.444	0.613	0.556	$3.25e^{-3}$	0.980	1.040	$3.61e^{-3}$
	2.258		0.478	0.452	$9.69e^{-4}$		0.628	0.660	0.371	0.351	$4.10e^{-4}$	1.287	1.282	$2.53e^{-5}$
5000	1.589	3	0.570	0.564	$3.63e^{-5}$	4	0.249	0.273	0.750	0.727	$5.29e^{-4}$	0.760	0.775	$2.26e^{-4}$
	1.680		0.593	0.560	$6.45e^{-3}$		0.302	0.350	0.708	0.650	$3.37e^{-3}$	0.850	0.860	$1.01e^{-4}$
	1.822		0.601	0.593	$6.45e^{-5}$		0.387	0.412	0.613	0.588	$6.28e^{-4}$	0.980	1.008	$7.85e^{-4}$
	2.258		0.478	0.469	$8.12e^{-5}$		0.628	0.642	0.371	0.358	$1.69e^{-4}$	1.287	1.310	$5.30e^{-4}$

TAB.3.7 Résultats de simulation de la loi Log-normal(0.7,0.35).

n	Densité de défaillance	Fiabilité	Taux de défaillance
100	$2.052 \cdot 10^{-4}$	$3.480 \cdot 10^{-4}$	$6.153 \cdot 10^{-4}$
500	$6.274 \cdot 10^{-5}$	$1.843 \cdot 10^{-5}$	$5.963 \cdot 10^{-4}$
5000	$6.051 \cdot 10^{-6}$	$2.263 \cdot 10^{-5}$	$3.351 \cdot 10^{-5}$

TAB.3.8 Résultats du ISE de la loi Log-normal(0.7,0.35).

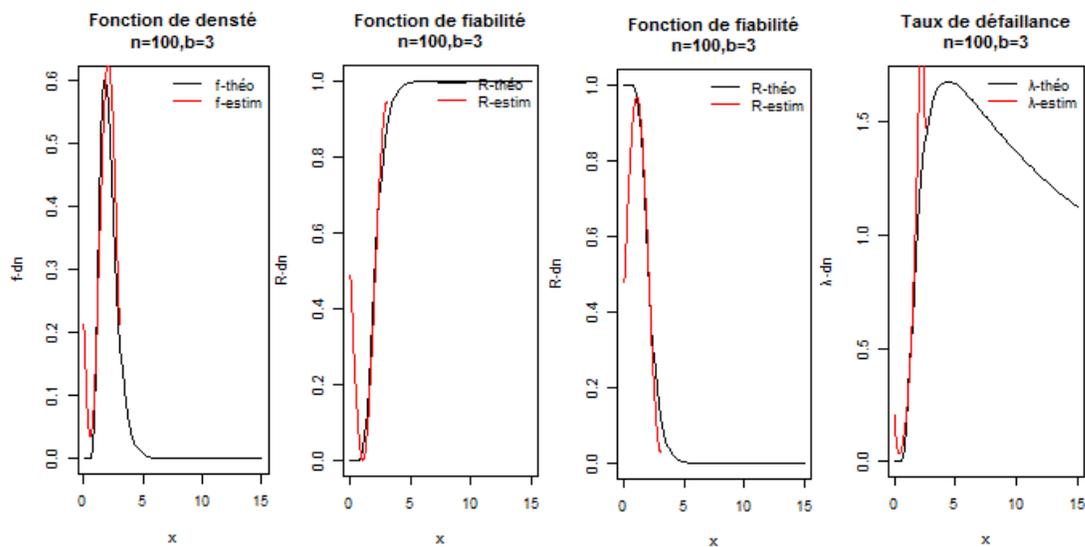


FIG.3.10 Densité, Fonction de répartition, Fiabilité et Taux de hasard associés à la loi Log-normale (0.7,0.35) n est à 100.

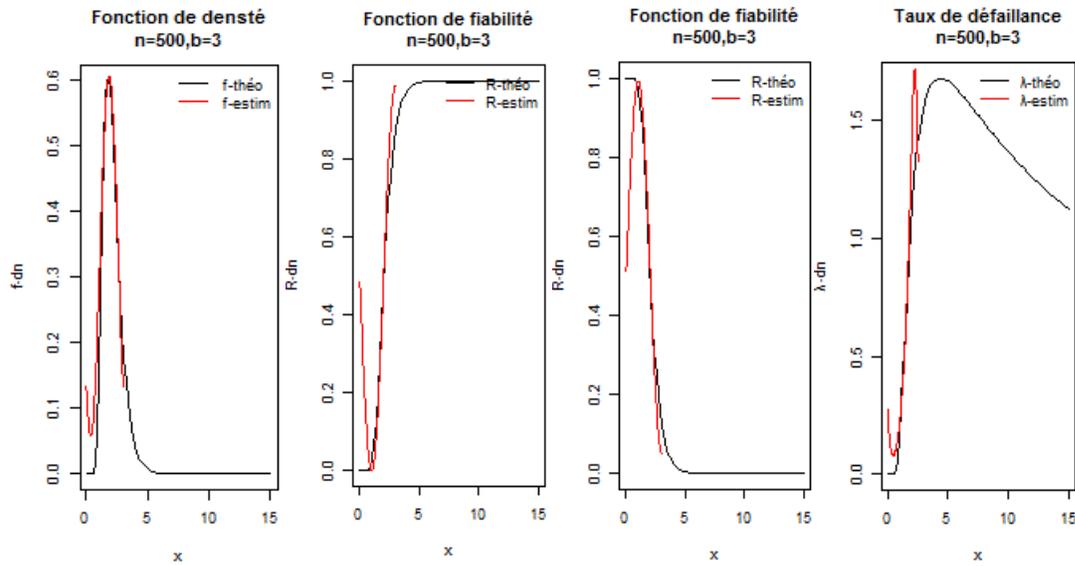


FIG.3.11 Densité, Fonction de répartition, Fiabilité et Taux de hasard associés à la loi Log-normale $(0.7, 0.35)$ n est à 500.

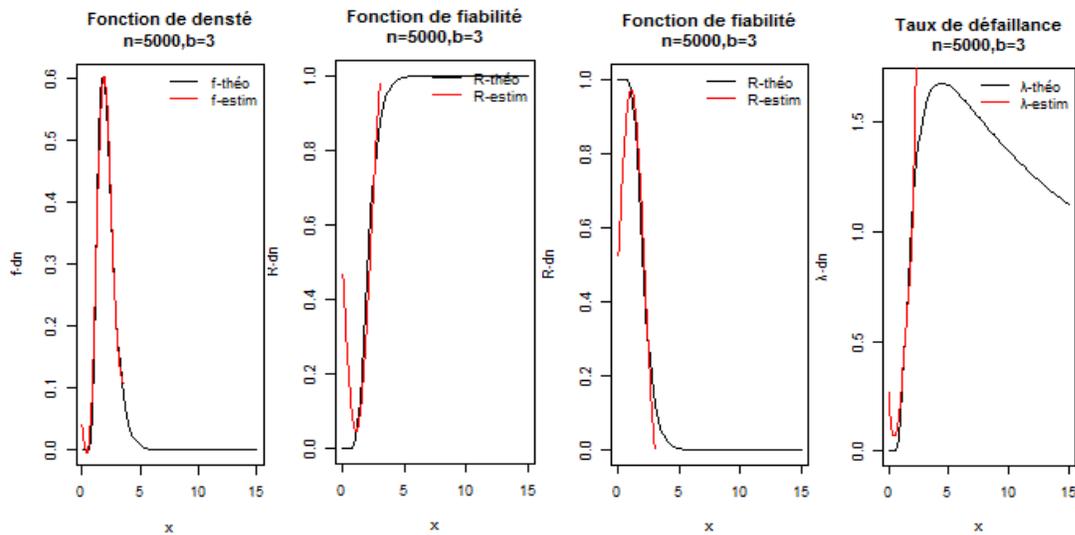


FIG.3.12 Densité, Fonction de répartition, Fiabilité et Taux de hasard associés à la loi Log-normale $(0.7, 0.35)$ n est à 5000 .

3.5.4 Interprétation des résultats

L'interprétation et l'analyse des différents résultats obtenus des différentes lois par la simulation sont résumées ci-dessous :

- **Loi Beta : $b=1$**

Les résultats de simulation sont satisfaisants. D'après les graphes (3.1), (3.2) et (3.3). Nous remarquons que lorsque la taille de l'échantillon augmente, nous obtenons une meilleure estimation pour la densité, fonctions de répartition, fiabilité et taux de défaillance. Ceci est confirmé par les résultats du ISE calculés dans le tableau (3.2) qui diminuent à mesure que la taille d'échantillons augmente.

- **Loi Gamma : $b=2$**

Nous remarquons que l'allure des courbes estimées se rapproche de celles des courbes théoriques (voir (3.4), (3.5), (3.6)) à mesure que nous augmentons la taille de l'échantillon. Ceci est confirmé par les résultats du ISE calculés dans le tableau (3.4) qui diminuent à mesure que la taille de l'échantillon augmente.

- **Loi Weibull : $b=5$**

Les résultats graphiques (voir(3.7), (3.8), (3.9))montrent : Pour la fonction du taux de défaillance, un écart un peu important entre la fonction théorique et estimée, une légère amélioration est constatée avec l'augmentation du de la taille de l'échantillon. Quant aux autres fonctions les courbes des fonctions théoriques et estimées sont presque comparables à mesure que la taille de l'échantillon augmente. Nous confirmons ceci avec les résultats du *ISE* qui diminue avec l'augmentation de la taille de l'échantillon voire le tableau (3.6).

Remarque 3.5.1 *Dans le cas où le paramètre d'échelle ($\beta < 1$) et ($\beta = 1$) les résultats de simulation n'aboutissent pas à un résultat meilleur, en effet, les courbes des fonctions estimés sont très éloignées de celles des courbes théoriques au point ou les courbes des fonctions estimés disparaissent dans l'intervalle $[0, 5]$.*

- **Loi Log-Normale : $b=3$**

L'allure des courbes estimées par la méthode des fonctions orthogonale se rapproche des courbes théoriques graphes (3.10), (3.11), (3.12) à mesure que nous augmentons la taille de l'échantillon, ce qui est confirmé par les résultats numériques obtenus dans le tableau (3.7).

Remarque 3.5.2 *Pour la loi exponentielle, une simulation été effectué mais les résultats obtenus sont non satisfaisant. Cela est dû au fait que l'écart entre la courbe théorique et estimées est très impotrant.*

3.6 Conclusion

Dans ce chapitre, nous avons estimé la densité de probabilité et la fonction de répartition par la méthode des fonctions orthogonales relativement à la base de Dirichlet et déterminé les propriétés de ces estimateurs. Les résultats obtenus sont utilisés pour l'estimation de la fonction de fiabilité et du taux de défaillance.

Les résultats de simulations effectuées pour l'estimation du taux de défaillance par la méthode des fonctions orthogonales sont appréciables pour les lois de fiabilité paramétriques utilisées, à savoir : Beta, Gamma, Log-normale et Weibull.

Les résultats sont meilleurs pour une taille d'échantillon grande.

Conclusion générale

On s'est intéressé dans ce mémoire au problème d'estimation du taux de défaillance par la méthode des fonctions orthogonales.

Nous avons considéré la base de Dirichlet définie sur un support positif $[0, b]$, $b > 0$, et différentes lois de fiabilité à savoir : Beta, Gamma, Log-normale et Weibull.

L'estimation du taux de défaillance est obtenue à partir des estimations de la fonction de densité et de la fonction de répartition. Ces derniers ont été défini dans le chapitre trois. Pour étudier le comportement de ces estimateurs, nous avons effectué des simulations. Les résultats des simulations nous permettent de confirmer que la méthode des fonctions orthogonales peut être appliquée en fiabilité. Ces résultats sont d'autant meilleurs que la taille de l'échantillon est grande.

Comme perspectives, nous pouvons envisager :

1. Utiliser d'autres lois de fiabilité telles que : Pareto, Birnbaum-Saunders, . . .
2. Appliquer d'autres bases : Polynomiales (Laguerre, Legendre, . . .) ou Trigonométrique (Cosine, Fejer, . . .), . . .
3. Effectuer des comparaisons avec d'autres estimateurs.
4. Utiliser la méthode de Saadi et Adjabi pour le choix du paramètre de lissage

Bibliographie

- [1] J.P. Asselindebeauville. Estimation non paramétrique de la densité de probabilité et du mode exemple de la distribution gamma. *Revue de Statistique Appliquée*, (Tome 26,N°3) :47-70, 1978.
- [2] S.Abou-Jouadé. Sur une condition nécessaire et suffisante de L^1 convergence presque complète. *C.R. Acad. Sci. Paris Sér A-B*, 283(16), pages A1107-1110, 1976.
- [3] P.K.Bhattacharya. Estimation of a Probability Density Function and Its Derivatives. *Sankhya : The Indian Journal of Statistics, Series A*, Vol. 29, No. 4, pp.373-382, 1967.
- [4] D.Bosq and J.Lecoutre. *Théorie de l'estimation fonctionnelle*. Economica edition, 1987.
- [5] A.W. Bowman. An alternative method of cross-validation for the smoothing density estimates. *Biometrika*, 71:353-360, 184.
- [6] J.Bleuez and D. Bosq. Condition nécessaire et suffisante de convergence de l'estimateur densité par la méthode des séries orthogonales. *Rev. Roum. Math. Pures et App.*, (24) : 869-886., 1979.
- [7] S.T.Buckland. Fitting density functions with polynomials. *J R Stat Soc [Ser A]* , 41:63–76, 1992.
- [8] M.CHERFAOUI. *Bootstrap dans l'estimation non paramétrique de la densité de probabilité et la courbe de régression de la moyenne* : Université de Béjaïa, 2009.

-
- [9] G.Charlier (1905) – Uber die darstellung willkürlicher funktionne. Arkiv für Matematik Astronomi och Fysik, 9, p.1-35.
- [10] E.Chicken, T.Cai. Block thresholding for density estimation local and global adaptivity. J Multivar Anal , 95:75–106, 2005.
- [11] N.Cencov. Evaluation of unknow distribution density from observation. Sov Maths, (3) :1559-1562, 1962.
- [12] S.X.Chen. Beta kernels estimators or density functions. Computatutational Statistics and Data Analysis, 31:131-145, 1999.
- [13] S.X.Chen. Gamma kernels estimators or density functions. Annals of the Institute of Statistical Mathematics, 52:471-480, 2000.
- [14] L. Devroy and L. Györfi. Non parametric density estimation. Wiley, New York, 1985.
- [15] J.Droesbeke, B.Fichet et Ph.Tassi. Eds, Analyse statistique des durées de vie. Economica, Paris, 1989.
- [16] S.Djouder. Estimation de la fonction de répartition. rapport exposé de méthodes statistiques première année poste-graduation, Faculté des Sciences Exactes département de Recherche Opérationnelle Univ de Béjaia, 2012.
- [17] S.Efromovitch. Distribution estimation for biased data, Journal of Statistical Planning and Inference 124 p. 1-43.0, 2004.
- [18] Edgeworth. The law error. Trans. Camb. Phil. Soc, 20, p.36, p.113, 1904.
- [19] D.Freedman and D.Persi. "On the histogram as a density estimator: L2 theory". Probability Theory and Related Fields(Heidelberg: Springer Berlin) 57 (4): 453-476.(1981).
- [20] O.Gaudoin. Fiabilité des système et logiciels. Note de cours. Ecole d'ingénieurs et informations de docteurs Grenoble, 2010.

-
- [21] F.Guérin. Estimation de la fiabilité pas les essais. ISTIA. Université d'Angers, 2004/2005.
- [22] Geffroy. Sur l'estimation de la densité dans un espace métrique. C.R. Acad. Sci. Paris Sér A-B, 278, pages 1449-1452, 1973.
- [23] I.J.Good, R.A.Gaskins. Density estimation and bumphunting by penalized likelihood method exemplified by scattering and meteorite data, with discussion. J Am Stat Assoc , 75:42–73, 1980.
- [24] J.Hermans, J.D.F. Habbema and K. Vandenbroek. A stepwise discriminat analysis program using density estimation. in Compstat, ed. G. Bruckmann, (101-110), 1974.
- [25] S.Julian. An Assessment of Hemite Function Based Approximations of Mutuel Information Applied to Independant component Analysis. 10.745-756, 2008.
- [26] R.A.Kronmal and M.E.Tarter. The estimation of probababilitydensities and cumulatives by fourier series methods, J. Amer. Statist. Assoc, (63) :925-952, 1968.
- [27] R.A.Kronmal and M.E.Tarter. On multivariate density estimate based on orthogonal expansions. Ann. Math. Statist. 41, No. 2, 718-722, 1970.
- [28] R.A.Kronmal, R.L.Holcomb and M.E.Tarter. A description of new computer methods for estimating the population density. In Proceedinge, Association for Computing Machinery 22, 511-519. Washington, D.C., Thompson Book Company, 1967.
- [29] K.Lagha. Cours de fiabilité et analyse des données de survie. Département de mathématique. Université de Béjaia, 2015.
- [30] J.Lecoutre. Contribution à l'estmation non paramétrique de la regression, PHD thesis, Université de Pierre et Marie curie-Paris VI-France,1982.
- [31] M.Mirzahmedov and S.A.Hasimov. On some properties of density estimation. Progress in Statistics (European Meeting of Statisticians, Budapest,) Colloq. math. Soc. Janos Bolyai, 9:535-546, 1972.

-
- [32] C.Marcovici, J.C.Ligeron. Utilisation des technique de fiabilité en mécanique. Technique et documentation, 1974.
- [33] F.MEKNASSI, A.AMAR. Estimation non paramétrique de la densité de probabilité par des fonctions orthogonales. Appliction au polynôme de Legendre. Université de Béjaia, 2013/2014.
- [34] A.G.Mihalache. Modélisation et évaluation de la fiabilité des systèmes mécatroniques : application sur système embarqué, 2010.
- [35] A.NASSIRI. l'estimation non paramétrique du taux de hasard : application des durées de chômage censurées à droite. Université de Toulouse, 1991.
- [36] E.Parzen. On the estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33 :1065-1027, 1962.
- [37] M.Rosenblatt. Remarks on some nonparametric estimator of a density function. *Annals of Mathematical Statistics*, 27 :832-837,1956.
- [38] M.Rudemo. Emirical choice of Histograms and kernel density estimators. *Scandinavian Jornal of Statistics*, 9:65-78, 1982.
- [39] V.Ryzin. Bayes risk consistency of classification procedures using density estimation *SANKHYA*, Ser.A,28,p. 261-70, 1966.
- [40] G.Saporta. Probabilité, Analyse des données et Statistiques. Technip, Paris, 1990.
- [41] D.W. Scott, R.A. Tapia and J.R Thompson. Kernel density estimation revisited. *Nonlinear Analysis, Theory, Method and Applications*, 1:339-372, 1977.
- [42] J.Simonoff. *Smoothing Methodes in Statistics*. Springer-Verlag New York.Inc, 1996.
- [43] D.W.Scott and G.R.Terrell. Oversmoothed nonparametric density estimates. *Journal of the American Statistical Association*, 80:209-214, 1985.

-
- [44] B.W. Silverman. Weak and strong uniform consistency of the kernel estimate of density function and its derivatives. *Ann.Statist*, 6:177-184, 1978.
- [45] N.Saadi and S. Adjabi. On the estimation of the probability density by trigonometric series. *Communications in Statistics-Theory and Methods*, 38(3583-3595),2009.
- [46] N.Saadi. Estimation de la densité de probabilité par projection orthogonal. Application à la courbe de Régression de la moyenne. Univ-Béjaia, 2016.
- [47] V.Sachs. Estimation non paramétrique d'une fonction de répartition et d'une densité. STAT 2413 chapitre 3, 2002/2003.
- [48] R.Servien. estimation de la fonction de répartition : revue bibliographique. *Journal de la société Française de Statistique, Société Française de Statistique et Société Mathématique de France*, 150(2), pp.84-104, 2009.
- [49] O.Scaillet. Density estimation using inverse and reciprocal inverse gaussian kernels. *Journal of Non parametric Statistics*, 16:217-226, 2004.
- [50] S.Stuart. Estimation of probability density by orthogonal series *Ann.Math.Statist.*,38,p. 1261-5, 1969.
- [51] A.B. Tsybakov. Introduction à l'estimation non -paramétrique. Springer-Verlag Berlin Heidelberg New York, 2004.
- [52] C.Taylor. Bootstrap choice of smoothing parameter in kernel density estimation. *Biometrika*,(76), pages 705-712, 1989.
- [53] O.Tebbi. Estimation des lois de fiabilité en mecanique par les essais accélérés. Autre. Université d'Angers. Français, 2005.
- [54] G.G.Walter. Wavelets and other Orthogonal Systems with Applications. London: CRC Press; 1994.
- [55] G.G.Walter. Properties of Hermite series estimation of probability density. *Ann Stat* , 5:1258-1264, 1977.

- [56] Watson. Density by orthogonal series *Ann.Math.Statist.*,40,p. 1496-8.
- [57] O.Yazourh. estimation fonctionnelle dans les modèles de durée : méthode des fonctions orthogonales. Université des sciences et technologie de Lille, 1992.