



UNIVERSITÉ ABDERRAHMANE MIRA

Faculté des sciences exactes

Département de Mathématique

Mémoire de fin de cycle

En vue de l'obtention du diplôme de Master en Mathématiques

Option :

Probabilités Statistique et Applications

Intitulé :

Champernowne Transformation Dans L'estimation à Noyau de la Densité

Présenté par :

Moussaoui Anis

Dirigé par :

Pr.Lagha eps Cherfi

Devant le jury :

M^{me} Barache.B

Présidente

M^{me} Bouraine.L

Examinatrice

Promotion : **2024/2025**

Remerciements

Je tiens à exprimer de tout mon coeur ma profonde gratitude à toutes les personnes qui ont accompagné mon parcours. À mon encadreuse, merci pour votre expertise, votre disponibilité constante et vos conseils précieux qui ont guidé chaque étape de ce travail. Aux membres du jury, un immense merci pour d'avoir accepté de juger mon travail et pour le temps que vous y avez consacré. Je remercie chaleureusement tous mes enseignants depuis mon entrée dans cette formation pour leurs enseignements inspirants, leur rigueur et leur soutien. Merci également à l'ensemble du département pour son appui logistique et sa disponibilité envers nous. Enfin, à toutes les personnes qui m'ont encouragée par leurs mots, leur présence ou leur confiance.

Moussaoui Anis

Table des matières

Introduction générale	1
1 Théorie des valeurs extrêmes	4
1.1 Distribution à queue lourde	6
1.1.1 Fonctions à variations régulières	9
1.2 Le Modèle GEV (Generalized Extreme Value Distribution)	10
1.2.1 Statistique d'ordre	10
1.2.2 Domaines d'attraction	12
1.2.3 Distribution des valeurs extrêmes généralisée GEVD	15
1.3 Le Modèle POT (Peaks-Over-Threshold)	16
1.3.1 Loi des excès	17
1.3.2 Le nombre d'excès au-dessus du seuil u	18
1.3.3 Distribution de Pareto généralisée (GPD)	18
1.3.4 Modélisation de la loi des Excès	19
2 Estimation à noyau de la densité de probabilité	21
2.1 Estimation par Histogramme	22
2.1.1 Construction de L'Histogramme	22
2.1.2 Estimateur de densité par Histogramme	26
2.1.3 Valeur de l'estimateur \hat{f} dans chaque classe	27
2.1.4 Théorie asymptotique de l'estimateur par Histogramme	27
2.2 Estimation par Histogramme Mobile	29
2.2.1 Estimateur de densité par Histogramme Mobile	30
2.2.2 Théorie asymptotique de l'estimateur par Histogramme Mobile	31
2.3 Estimateur à noyau de densité de probabilité	32
2.3.1 Estimateur de densité de Parzen-Rosenblatt	33
2.3.2 Théorie asymptotique de l'estimateur à noyau symétrique	36
2.3.3 Propriétés de l'estimateur de Parzen-Rosenblatt	37
2.3.4 Problème de Biais aux Limites dans l'Estimation par Noyau de Densité.	42
2.3.5 Noyaux asymétriques	43
2.3.6 Noyau Gamma	44
2.3.7 Théorie asymptotique de l'estimateur à noyau Gamma	45

2.3.8	Noyau Bêta	47
2.3.9	Théorie asymptotique de l'estimateur à noyau Beta	49
2.3.10	Transformation des données en estimation de densité de probabilité	51
3	Estimateur KMCE (Kernel Modified Champernowne Estimator)	57
3.1	Distribution de Champernowne	58
3.1.1	Distribution de Champernowne originale	58
3.1.2	Distribution de Champernowne modifiée	58
3.1.3	Estimation des paramètres de la distribution de Champernowne modifiée	60
3.2	Estimation semi-paramétrique de densité de probabilité	61
3.2.1	Transformation des données avec la distribution de Champernowne	61
3.2.2	Calcul de l'estimateur de densité par noyau pour les données transformées	63
3.2.3	Estimateur KMCE (Kernel Modified Champernowne Estimator)	66
3.2.4	Théorie asymptotique de l'estimateur KMCE	68
4	Simulation	73
4.1	Données utilisées dans la simulation	73
4.2	Méthodes d'estimation et critères d'erreur retenus	76
4.3	Algorithme du programme de simulation	76
4.4	Résultats de Simulation	78
4.4.1	Interprétation des Résultats de Simulation	81
	Conclusion générale	83
	Bibliographie	84

Introduction générale

L'analyse statistique des distributions à queues lourdes constitue un enjeu critique dans des domaines où les événements extrêmes, bien que rares, engendrent des conséquences économiques, sociales ou environnementales majeures. En finance, ces distributions modélisent les krachs boursiers et les mouvements erratiques des marchés ; en assurance, elles décrivent les sinistres catastrophiques qui peuvent menacer la solvabilité des compagnies ; en ingénierie, elles quantifient les défaillances structurelles aux conséquences potentiellement dramatiques ; en hydrologie, elles prédisent les crues centennales ; en épidémiologie, elles modélisent l'émergence de pandémies dévastatrices. La modélisation précise de ces phénomènes exige des méthodes capables de capturer à la fois la dynamique centrale des données, là où se concentre la majorité des observations et le comportement des queues de distribution, où résident les risques les plus extrêmes et les plus coûteux.

L'analyse traditionnelle a toutefois été critiquée sur plusieurs plans. Premièrement, l'estimation paramétrique classique impose une forme fonctionnelle restrictive, souvent inadaptée à la diversité et à la complexité des distributions rencontrées en pratique. Une mauvaise spécification du modèle peut conduire à une sous-estimation dangereuse des risques extrêmes. Deuxièmement, les méthodes non paramétriques traditionnelles, comme l'estimation par noyaux symétriques, souffrent d'un biais systématique aux frontières pour les données à support borné ou semi-borné (telles que les montants de sinistres, nécessairement positifs). Ce biais se traduit par une sous-estimation persistante de la densité près des limites, faussant ainsi l'analyse des zones où de nombreuses observations peuvent pourtant se concentrer. Parmi les travaux qui ont été faits dans ce domaine on trouve : Économétrie non paramétrique [36], Estimation de l'Expected Shortfall basée sur les quantiles à noyau [64], estimation paramétrique et non paramétrique de la VaR [40], modélisation d'un indice de sécheresse à l'aide d'une approche non paramétrique [24]. Troisièmement, la théorie des valeurs extrêmes (EVT), bien que théoriquement puissante pour modéliser les maxima via les lois GEV ou l'approche POT,

présente des limitations significatives : elle néglige l'information contenue dans les grandes valeurs non maximales, et son efficacité dépend cruciallement du choix souvent subjectif et difficile à justifier d'un seuil au-delà duquel le modèle POT s'applique, parmi les travaux qui ont été faits dans ce domaine on trouve : La VaR et les événements de marché extrêmes [16], la Caractérisation des distributions à queue lourde pour l'analyse des crues [19], Application de la théorie des valeurs extrêmes en assurance automobile [3], Estimation des queues des distributions de sévérité des pertes [39], la gestion des risques du marché pétrolier [45].

Face à ces limitations, ce mémoire présente l'estimateur KMCE (Kernel Modified Champernowne Estimator), une approche semi-paramétrique novatrice qui synthétise les forces des paradigmes paramétrique et non paramétrique, introduit pour la première fois par Buch-Larsen et al., 2005 [11]. Son architecture repose sur trois étapes : (1) une transformation paramétrique des données via la distribution de Champernowne modifiée, ajustée aux données ; (2) une estimation non paramétrique par noyaux (symétriques ou asymétriques) appliquée aux données ainsi transformées, qui résident désormais dans l'intervalle unitaire $[0, 1]$; et (3) une rétro-transformation préservant la cohérence avec l'échelle originale, via l'application du jacobien de la transformation. Cette approche permet de réduire le biais aux bords, tout en capturant fidèlement la forme des queues de la distribution, y compris dans les zones les plus extrêmes. D'autres travaux sur l'estimation de densité par transformation peuvent être cités : Estimation par noyau de la densité des fonctions de perte actuarielles [6], l'estimateur à noyau basé sur une transformation améliorée pour les densités sur l'intervalle unitaire [62], l'approche non paramétrique basée sur le noyau pour l'anamorphose gaussienne des données [43].

Les simulations présentées dans ce travail, couvrant un spectre de distributions (Exponentielle à queue légère, Gaussienne inverse, mélange Weibull-Fréchet, et Burr à queue très lourde), attestent de la supériorité de performance du KMCE. Les résultats quantitatifs montrent une réduction du biais aux frontières, ainsi qu'une amélioration significative de la précision globale par rapport aux estimateurs à noyaux standards. Cette robustesse, alliée à sa flexibilité, positionne le KMCE comme un outil de choix pour l'estimation des risques extrêmes dans des contextes applicatifs réels, où la fiabilité et la justesse des modèles sont des impératifs non négociables.

Ce mémoire s'articule autour d'un parcours en quatre chapitres, conçu pour guider le lecteur des fondements théoriques aux validations pratiques.

Chapitre 1 : Nous y établissons les fondements mathématiques de la théorie des valeurs extrêmes, qui est indispensable pour appréhender les événements rares. Nous détaillons ses deux piliers : le modèle GEV (Generalized Extreme Value), qui caractérise la loi limite des maxima de séries de données indépendantes (Gnedenko, 1943 [25]), et l'approche POT (Peaks-Over-Threshold), fondée sur la distribution GPD (Generalized Pareto Distribution) pour modéliser les excès au-delà d'un seuil élevé. Les concepts clés de domaines d'attraction et de fonctions à variation régulière y sont expliqués.

Chapitre 2 : Nous y explorons l'évolution de l'estimation de densité par noyau, en analysant ses forces et, surtout, ses faiblesses. Après un examen des estimateurs par histogramme et histogramme mobile, nous nous concentrons sur l'estimateur à noyau de Parzen-Rosenblatt (Parzen, 1962 [42] ; Rosenblatt, 1956 [17]). Une analyse approfondie est consacrée au problème du biais aux frontières, et nous passons en revue les principales stratégies de correction proposées dans la littérature, notamment l'emploi de noyau bêta de Chen, S. X. (1999) [13], noyau gamma [15] et les transformations de données (Box et Cox (1964) [9], Yeo et Johnson (2000) [63]).

Chapitre 3 : Coeur théorique et méthodologique de ce mémoire, ce chapitre est entièrement dédié à la construction de l'estimateur KMCE. Nous y présentons la distribution de Champernowne modifiée, décrivons la procédure d'estimation de ses paramètres par maximum de vraisemblance, et présentons pas à pas l'estimateur KMCE dans ses deux versions : avec noyau symétrique (muni d'une correction de bord) [11] et avec noyau bêta de Chen, S. X. (1999) [13]. Les propriétés asymptotiques (biais, variance, MISE) de l'estimateur sont établies.

Chapitre 4 : Dans ce chapitre de validation par l'expérience, nous évaluons et comparons par simulation les performances du KMCE sur des distributions aux propriétés variées. L'analyse des résultats, fondée sur plusieurs mesures d'erreur (ISE, distances L1 et WISE), met en lumière sa polyvalence, sa robustesse et sa supériorité pour modéliser simultanément des données à queues légères et à queues lourdes. Cela vaut particulièrement pour le KMCE avec noyau bêta.

En synthèse, ce mémoire présente et valide une méthodologie unifiée d'estimation de densité de probabilité, particulièrement adaptée aux défis posés par les distributions à queues lourdes. En transcendant les limitations des approches purement paramétriques ou non paramétriques, le KMCE émerge comme un outil fiable et puissant, notamment avec le noyau bêta. Son applicabilité s'étend à de nombreux champs appliqués (finance, actuariat, ingénierie, sciences environnementales), offrant aux chercheurs et aux praticiens un outil pour mieux quantifier, anticiper et gérer l'incertitude inhérente aux événements extrêmes.

Chapitre 1

Théorie des valeurs extrêmes

La prise en compte des risques majeurs dans notre société technologique est devenue cruciale en raison des impacts économiques, environnementaux et humains des catastrophes industrielles. L'une des approches standards pour étudier ces risques repose sur la théorie des valeurs extrêmes, une branche des statistiques consacrée à l'analyse des événements rares situés dans les queues de distribution.

L'objectif principal de cette théorie est d'estimer les probabilités d'événements extrêmes non observés ou très rares. Initialement développée pour étudier les niveaux de crues, son champ d'application s'étend désormais à d'autres phénomènes météorologiques (précipitations, vitesse du vent), à l'industrie (défaillances majeures), à la finance (crises boursières), à l'assurance (sinistres catastrophiques) et aux sciences environnementales (concentrations d'ozone atmosphérique), entre autres.

Bien que son application pratique nécessite toujours un examen critique, la théorie des valeurs extrêmes demeure un outil privilégié pour la prédiction des événements exceptionnels. Dans ce mémoire, nous proposons une synthèse des fondements mathématiques et statistiques de cette théorie.

Définition 1.1 (Inverse généralisée d'une fonction monotone) [20] (p.130) , [46] (p.13)
Pour tout $x \in [0, 1]$ l'inverse généralisée d'une fonction monotone est définie par :

$$F^{\leftarrow}(x) = \inf \{y \in \mathbb{R} : F(y) \geq x\} \quad , \quad 0 < x < 1 \quad .$$

Lorsque F est inversible (c'est-à-dire continue et strictement croissante), la fonction inverse généralisée coïncide avec l'inverse classique : $F^{\leftarrow} = F^{-1}$. La fonction de répartition F est toujours non décroissante, mais présente des discontinuités (sauts) aux points x où $P(X = x) > 0$ et des intervalles de constance sur $[\alpha, \beta]$ où $P(X \in [\alpha, \beta]) = 0$. Ces deux situations apparaissent généralement pour les variables aléatoires discrètes ou continues avec des supports disjoints. Dans ces cas, la fonction inverse classique n'est pas définie, ce qui justifie l'usage de l'inverse généralisée.

Définition 1.2 (Fonction de quantile) [20] (p.130)

Soit X une variable aléatoire réelle continue, de fonction de répartition F .

L'inverse généralisée de la fonction de répartition F (ou fonction de quantile) est définie par :

$$q(t) = F^{\leftarrow}(t) = \inf \{y \in \mathbb{R} : F(y) \geq t\} \quad , \quad 0 < t < 1 \quad .$$

La quantité $q(t)$ définit le quantile d'ordre t de F .

t : est la probabilité cumulée pour laquelle on cherche la valeur seuil $q(t)$ dans la distribution. C'est un paramètre fondamental pour caractériser les positions dans une distribution statistique.

Exemple

Supposons une loi normale standard (moyenne 0, écart-type 1) :

Si $t = 0.95$ (probabilité cumulée = 95%), alors $q(0.95) \simeq 1.645$.

Autrement dit, 95% des observations sont inférieures ou égales à 1.645.

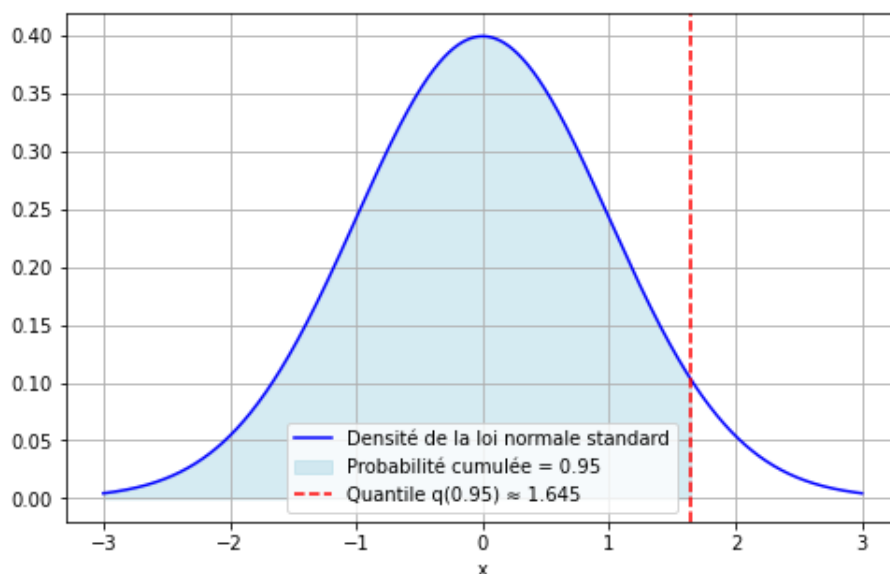


FIGURE 1.1 – Interprétation graphique du quantile.

Définition 1.3 (Fonction de quantile de queue)

La fonction de quantile de queue, notée Q , est définie comme suit :

$$\begin{aligned}
 Q(u) &= q\left(1 - \frac{1}{u}\right) = \inf \{y \in \mathbb{R} : F(y) \geq 1 - \frac{1}{u}\} \\
 &= \inf \{y \in \mathbb{R} : 1 - F(y) \leq \frac{1}{u}\} \\
 &= \inf \{y \in \mathbb{R} : \frac{1}{1 - F(y)} \geq u\} \\
 &= \left(\frac{1}{\bar{F}}\right)^{\leftarrow}(u) \quad , \quad u \in (1, \infty) \quad .
 \end{aligned} \tag{1.1}$$

où $\bar{F}(x) = (1 - F(x) = P(X > x))$: Fonction de queue (ou fonction de survie).

1.1 Distribution à queue lourde

Dans la théorie des probabilités et les statistiques, une distribution F est dite à queue lourde si la décroissance de sa fonction de survie dans les queues (les régions éloignées de la moyenne) est asymptotiquement plus lente que toute décroissance exponentielle. Cela signifie que les événements extrêmes ont une probabilité non négligeable de se produire, contrairement aux distributions à queue légère (comme la distribution normale) où de tels événements sont exponentiellement rares.

Cette propriété a des implications profondes sur le comportement des moments et la prévalence des valeurs extrêmes. [21]

Les réclamations d'assurance, en particulier dans des secteurs comme l'assurance automobile, l'assurance habitation, et l'assurance responsabilité civile, présentent souvent des distributions où la majorité des réclamations sont de faible montant, mais quelques réclamations sont extrêmement élevées. Ces réclamations élevées, bien que rares, ont un impact significatif sur les pertes totales. Dans l'assurance automobile, la majorité des réclamations concernent des dommages mineurs, tandis que les accidents graves peuvent entraîner des réclamations très élevées, surtout si des blessures corporelles sont impliquées.

les réclamations d'assurance sont souvent modélisées par des distributions à queue lourde pour capturer la présence significative des événements extrêmes. Les distributions de Pareto et log-normale sont couramment utilisées à cette fin. Ces modèles permettent une meilleure estimation des risques et des pertes potentielles, aidant les assureurs à gérer les sinistres de manière plus efficace et à se préparer aux événements rares mais coûteux.

Définition 1.4 [21] (p.02). Une distribution F est dite à queue lourde (à droite) ssi

$$\int_{-\infty}^{+\infty} e^{\lambda x} F(dx) = +\infty \quad \forall \lambda > 0 \quad . \tag{1.2}$$

C'est-à-dire si et seulement si F ne possède aucun moment exponentiel positif. Sinon, F est dite à queue légère.

où $F(dx)$ désigne la probabilité que la variable aléatoire X appartienne à un intervalle infinitésimal autour de x ($P(X \in dx)$), si X est une variable continue (avec une fonction de densité $f(x)$), alors $F(dx)$ est équivalent à $f(x)dx$.

Définition 1.5 [21] (p.08). Une distribution F est dite à queue légère si et seulement si :

$$\exists \lambda > 0, \quad \int_{-\infty}^{+\infty} e^{\lambda x} F(dx) < +\infty \quad .$$

C'est-à-dire si et seulement si elle n'est pas à queue lourde.

Pour toute distribution à queue légère F sur \mathbb{R}^+ , tous les moments sont finis, c'est-à-dire $\int_0^{+\infty} x^k F(dx) < \infty$ pour tout $k > 0$.

Nous dirons qu'une fonction non négative (généralement tendant vers zéro) est à queue lourde si elle n'est pas majorée par une fonction exponentielle décroissante.

Définition 1.6 [21]. Une distribution F sur \mathbb{R} est dite de support non borné à droite si

$$\bar{F}(x) > 0 \quad \forall x \in \mathbb{R} \quad .$$

Définition 1.7 [21]. On dit qu'une fonction $f \geq 0$ est à queue lourde si et seulement si :

$$\limsup_{x \rightarrow \infty} f(x)e^{\lambda x} = \infty \quad \forall \lambda > 0 \quad .$$

Définition 1.8 [21] (p.08). Pour toute distribution F , la fonction $R(x) = -\log F(x)$ est appelée la fonction de risque de la distribution. Si la fonction de risque est dérivable, alors sa dérivée $r(x) = R'(x)$ est appelée le taux de risque.

Théorème 1.1 Pour toute distribution F , les assertions suivantes sont équivalentes :

- (i) F est une distribution à queue lourde.
- (ii) La fonction \bar{F} est à queue lourde.
- (iii) La fonction de risque R correspondante satisfait : $\lim_{x \rightarrow +\infty} \inf \frac{R(x)}{x} = 0$.
- (iv) La fonction $F(x, x + T]$ est à queue lourde pour un (ou tout) $T > 0$ fixé.

La démonstration de ce théorème est disponible dans [21] (p.08).

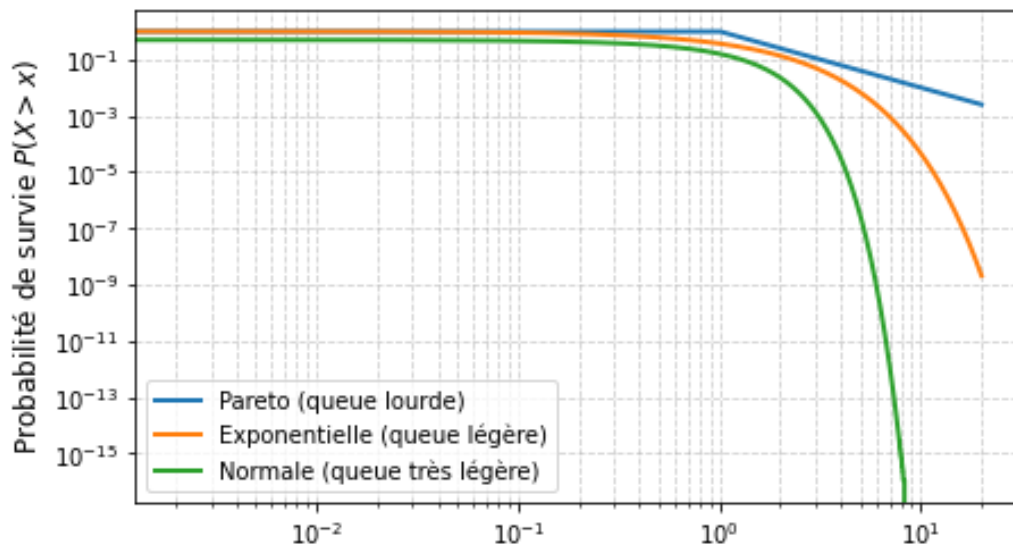


FIGURE 1.2 – Comparaison des queues de distribution (côté droit) pour trois distributions.

Proposition 1.1 . Soit F une distribution absolument continue de densité f . Si la distribution F est à queue lourde, alors sa fonction de densité $f(x)$ est aussi à queue lourde.

La preuve est disponible dans : [21] (p.09).

Le tableau suivant résume quelques distributions à queue lourde utilisées en finance et en assurance.[20] (p.35).

Distribution	Support	Paramètres
log-gamma	$(1, \infty)$	α (forme), β (échelle) > 0
log-Cauchy	$(0, \infty)$	$x_0 \in \mathbb{R}$ (position), $\gamma > 0$ (échelle).
Benktander (type I)	$(0, \infty)$	$\alpha, \beta > 0$.
Benktander (type II)	$(0, \infty)$	$\alpha > 0$, $0 < \beta < 1$.
Burr (Type XII)	$(0, \infty)$	$c > 0, k > 0$, Plus $c \cdot k$ est petit, plus la queue devient lourde.
Cauchy	\mathbb{R}	$x_0 \in \mathbb{R}$ (position), $\gamma > 0$ (échelle).
Distribution t de Student	\mathbb{R}	$k > 0$ (degrés de liberté).
Distribution de Lévy	$[\mu, +\infty)$	μ (position), c (échelle) > 0
Fréchet (Type II)	$(0, \infty)$	$\xi > 0$ (paramètre de forme).
Pareto Généralisée (GPD)	$[\mu, +\infty)$	μ (position), $\sigma > 0$ (échelle), $\xi > 0$ (forme), Plus ξ est grand, plus la queue devient lourde.

TABLE 1.1 – Distributions à queue lourde utilisées en finance et en assurance.

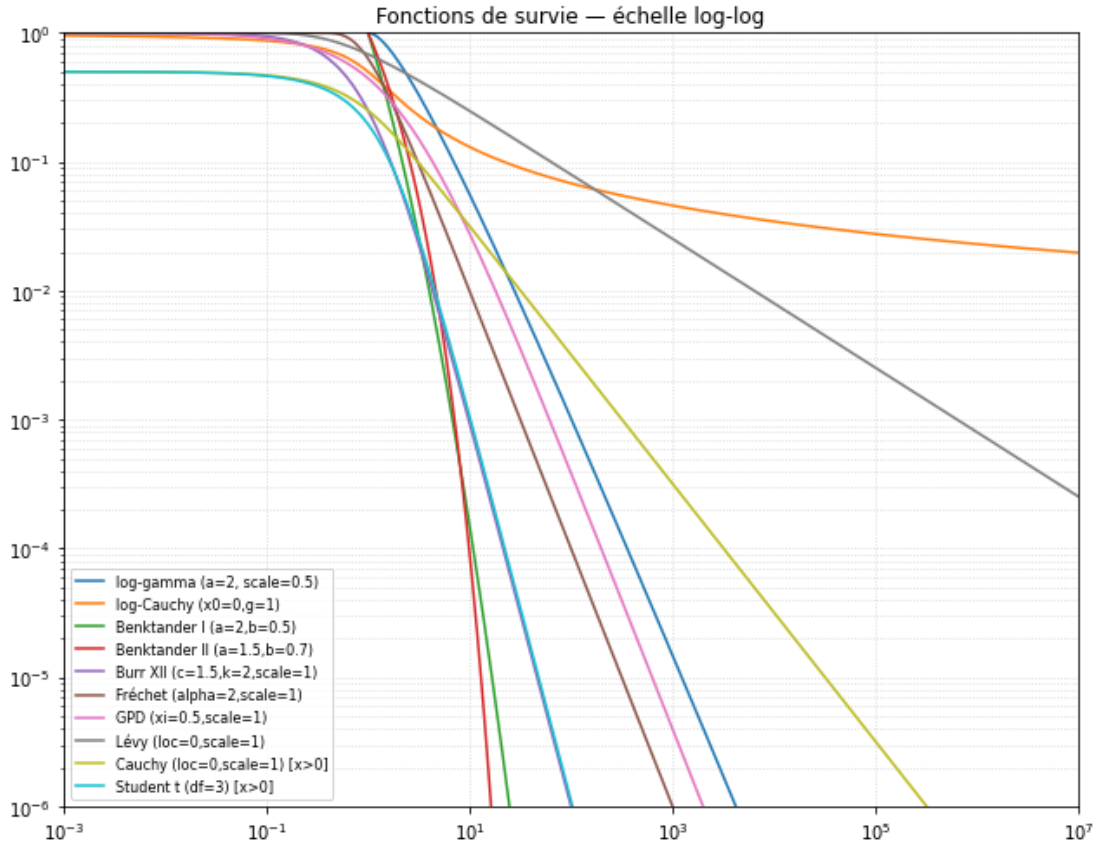


FIGURE 1.3 – Comparaison des queues (à droite) des distributions mentionnées dans le tableau 1.1

1.1.1 Fonctions à variations régulières

La Théorie de la Variation Régulière (Regular Variation, notée \mathcal{RV}) est un concept fondamental en analyse mathématique, introduit par Jovan Karamata en 1930. Elle fournit le cadre théorique rigoureux pour étudier le comportement des fonctions pour de grandes valeurs de leur argument ($x \rightarrow \infty$).

Définition 1.9 [4] (p.1). On dit qu'une fonction mesurable positive f , définie sur un voisinage de l'infini de la forme $[y, \infty)$, est à variations régulières d'indice $\alpha \in \mathbb{R}$, ce que l'on note $f \in \mathcal{RV}_\alpha$ si et seulement si :

$$\forall \lambda > 0, \quad \lim_{x \rightarrow \infty} \frac{f(\lambda x)}{f(x)} = \lambda^\alpha \quad .$$

Si $f \in \mathcal{RV}_0$ ($\alpha = 0$), f est dite à variation lente à l'infini, c'est-à-dire :

$$\forall \lambda > 0, \quad \lim_{x \rightarrow \infty} \frac{f(\lambda x)}{f(x)} = 1 \quad .$$

Remarque 1.1 : Pour que l'intervalle $[y, \infty)$ soit un voisinage de l'infini, y doit être un nombre réel (généralement supposé positif et grand, $y > 0$).

Proposition 1.2 : Soient $\alpha \in \mathbb{R}$ et $f \in \mathcal{RV}_\alpha$. Alors il existe une fonction à variation lente ℓ à l'infini telle que

$$\forall x > 0, \quad f(x) = x^\alpha \ell(x) \quad .$$

Proposition 1.3 : [4](p.26)

(i) Si $f \in \mathcal{RV}_\alpha$, alors $f^\rho \in \mathcal{RV}_{\alpha\rho}$.

(ii) Si $f_i \in \mathcal{RV}_{\alpha_i}$ ($i = 1, 2$), $f_2 \rightarrow \infty$ lorsque $x \rightarrow \infty$, alors $f_1 \circ f_2 \in \mathcal{RV}_{\alpha_1\alpha_2}$.

(iii) Si $f_i \in \mathcal{RV}_{\alpha_i}$ ($i = 1, 2$), alors la fonction $f_1 + f_2$, définie par :

$$(f_1 + f_2)(x) = f_1(x) + f_2(x), \quad x \in \mathbb{R}_+, \text{ appartient à la classe } \mathcal{RV}_\alpha, \text{ avec } \alpha = \max(\alpha_1, \alpha_2).$$

1.2 Le Modèle GEV (Generalized Extreme Value Distribution)

Dans ce modèle-ci, on s'intéresse seulement à la valeur la plus grande, contrairement au modèle POT que l'on va examiner par la suite.

1.2.1 Statistique d'ordre

Définition 1.10 [53] (p.07)

Supposons que les données soient des réalisations d'un échantillon de n variables aléatoires (X_1, X_2, \dots, X_n) indépendantes et identiquement distribuées (i.i.d). Les données ordonnées seront alors désignées par $X_{(1,n)} \leq X_{(2,n)} \leq \dots \leq X_{(n,n)}$ telle que $\{X_{(i,n)}, i \in [1, n]\}$ appelée la i^{eme} statistique d'ordre d'un l'échantillon de taille n , on note :

$$\hat{X}_n = X_{(n,n)} = \max(X_{(1,1)}, \dots, X_{(n,n)}) \quad .$$

$$\check{X}_n = X_{(1,n)} = \min(X_{(1,1)}, \dots, X_{(n,n)}) \quad .$$

La loi du maximum et du minimum de ces variables aléatoires peut être facilement déterminée, comme le montre la proposition suivante.

Proposition 1.3 [46] (p.20) , [53] (p.09) : $\forall x \in \mathbb{R}$

$$P(\hat{X}_n \leq x) = (F_X(x))^n \quad , \quad P(\check{X}_n \leq x) = 1 - (1 - F_X(x))^n \quad .$$

Preuve : Comme les (X_i) sont i.i.d, alors $\forall x \in \mathbb{R}$:

$$P(\hat{X}_n \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = \prod_{i=1}^n P(X_i \leq x) = (F_X(x))^n ,$$

pour la seconde partie on a $\check{X}_n = -\max(-X_1, \dots, -X_n)$

$$\Rightarrow P(\check{X}_n > x) = P(\max(-X_1, \dots, -X_n) < -x) ,$$

$$= \prod_{i=1}^n P(-X_i < -x) = (1 - F_X(x))^n .$$

Fonction de répartition empirique (E.D.F) :

Pour étudier la fonction de répartition F qui est inconnue, on utilise en pratique la fonction de répartition empirique obtenue à partir des observations $X_{(1,n)}, \dots, X_{(n,n)}$, qui est définie comme suit :

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x\}} = \begin{cases} 0 & \text{si } x < X_{(1,n)} \\ \frac{i-1}{n} & \text{si } X_{(i-1,n)} \leq x < X_{(i,n)} , \text{ pour } 2 \leq i \leq n \\ 1 & \text{si } x \geq X_{(n,n)} \end{cases} \quad (1.3)$$

Définition 1.11 (Point extrême) [20] (p.114)

On note par x_F (resp. x_F^*) le point extrême supérieur (resp. inférieur) de la distribution F , définie par :

$$x_F := \sup \{x \in \mathbb{R} : F(x) < 1\} , \quad (1.4)$$

respectivement

$$x_F^* := \inf \{x \in \mathbb{R} : F(x) > 0\} . \quad (1.5)$$

Proposition 1.4 [20] (p.115)

Soit (X_1, X_2, \dots, X_n) un échantillon de n variables aléatoires i.i.d. de fonction de répartition F .

- Pour tout $x < x_F$,

$$P(X_{(n,n)} \leq x) = F^n(x) \xrightarrow[n \rightarrow \infty]{} 0 .$$

- Si $x_F < +\infty$, alors pour tout $x \geq x_F$,

$$P(X_{(n,n)} \leq x) = F^n(x) = 1 \quad .$$

Ainsi,

$$X_{(n,n)} \xrightarrow{\text{p.s.}} x_F \quad \text{lorsque } n \rightarrow \infty \quad .$$

1.2.2 Domaines d'attraction

Un des résultats fondamentaux de la TVE est le théorème suivant qui a été énoncé pour la première fois par Fisher et Tippett en (1928), puis démontré par Gnedenko en (1943) ensuite a été repris et simplifié par Haan en (1970).

Théorème 1.2 [Fisher-Tippett(1928)-Gnedenko(1943)] [25],[47] (p.09),[20] (p.121)

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d) de fonction de répartition F . S'il existe deux suites normalisantes réelles $(a_n)_{n \geq 1} > 0$ et $(b_n)_{n \geq 1} \in \mathbb{R}$ et une loi non-dégénérée \mathcal{H}_ξ telles que :

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{\hat{X}_n - b_n}{a_n} \leq x \right) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = \mathcal{H}_\xi(x), \quad \xi \in \mathbb{R} \quad . \quad (1.6)$$

\mathcal{H}_ξ est appelée fonction de répartition de la loi des valeurs extrêmes, elle dépend d'un seul paramètre de forme ξ appelé indice des valeurs extrêmes ou indice de queue. Si F vérifie le théorème de Fisher-Tippett-Gnedenko on dit que F appartient au domaine d'attraction de \mathcal{H}_ξ on le note $F \in \mathcal{D}(\mathcal{H}_\xi)$.

On distingue trois domaines d'attraction selon le signe de ξ :

1. si $\xi < 0$ on dit que F appartient au domaine d'attraction de Weibull. Il contient des lois dont la fonction de survie n'a pas de queue de distribution.

La fonction de distribution standard de Weibull est définie comme suit :

$$\mathcal{H}_\xi(x) = \Psi_\xi(x) = \begin{cases} \exp(-[-x]^{-\xi}) & x \leq 0 \\ 1 & x > 0 \end{cases} \quad (1.7)$$

2. si $\xi = 0$ on dit que F appartient au domaine d'attraction de Gumbel. Il contient les lois dont la fonction de survie est à décroissance exponentielle, i.e. les lois à queues légères.

La fonction de distribution standard de Gumbel est définie comme suit :

$$\mathcal{H}_0(x) = \Lambda(x) = \exp(-\exp(-x)), \quad x \in \mathbb{R} \quad (1.8)$$

3. si $\xi > 0$ on dit que F appartient au domaine d'attraction de Fréchet. Il contient les lois dont la fonction de survie est à décroissance polynomiale, i.e. les lois à queues lourdes.

La fonction de distribution standard de Fréchet est définie comme suit :

$$\mathcal{H}_\xi(x) = \Phi_\xi(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-(x)^{-1/\xi}) & x > 0 \end{cases} \quad (1.9)$$

Le tableau suivant résume l'ensemble des lois avec leur apparition dans différents domaines d'attraction.[12] (p.75-76) .

Weibull ($\xi < 0$)	Gumbel ($\xi = 0$)	Fréchet ($\xi > 0$)
Uniforme	Normale	Pareto
Beta	Exponentielle	Student
ReverseBurr	Log-normale	Burr
	Gamma	Chi-deux
	Weibull	Fréchet
	Logistique	Log-gamma
	Gumbell	Log-logistique
		Cauchy

TABLE 1.2 – Domaines d'attraction de certaines distributions de probabilité.

Domaine d'attraction de Fréchet $\mathcal{D}(\Phi_\xi)$:

Les lois appartenant à ce domaine d'attraction sont caractérisées par une queue à décroissance lente (polynomiale) à l'infini (lois à queues lourdes), et un point terminal $x_F = \infty$

Théorème 1.3 [25] ,[47](p.54) :

Pour qu'une fonction de distribution F appartienne au domaine d'attraction de la loi de Fréchet ($\mathcal{D}(\Phi_\xi)$) il faut et il suffit qu'on ait

$$\lim_{t \rightarrow \infty} \frac{1 - F(t)}{1 - F(kt)} = k^\xi \quad , \quad \forall k > 0 \quad .$$

c'est-à-dire si et seulement si $\bar{F} \in \mathcal{RV}_{-\xi}$.

Domaine d'attraction de Gumbel $\mathcal{D}(\Lambda)$

Théorème 1.4 [25] : Pour qu'une fonction de distribution F appartienne au domaine d'attraction de la loi de Gumbel ($\mathcal{D}(\Lambda)$), il faut et il suffit qu'il existe une fonction continue et monotone $g(t)$ telle que, $g \rightarrow 0$ pour $t \rightarrow x_F$ (par valeurs inférieures), et que pour toutes les valeurs de x

$$\lim_{t \rightarrow x_F} \frac{1 - F(t + txg(t))}{1 - F(t)} = e^{-x} \quad , \quad (1.10)$$

avec $x_F \leq \infty$ est le point terminal (extême supérieur (1.4)) de F .

Théorème 1.5 [25] : Supposons que la fonction de distribution $F(x)$ soit telle que $F(x) < 1$ pour toute valeur de x (c'est-à-dire que le support de la distribution s'étend jusqu'à l'infini). Pour que la fonction $F(x)$ appartienne au domaine d'attraction de la loi de Gumbel $\mathcal{D}(\Lambda)$, il faut que :

$$\frac{1 - F(kt)}{1 - F(t)} \rightarrow 0 \quad \text{pour } t \rightarrow \infty \quad .$$

pour tout $k > 0$.

Domaine d'attraction de Weibull $\mathcal{D}(\Psi_\xi)$:

Les lois de ce domaine sont bornées à droite, et par conséquent, le point terminal x_F est fini.

Théorème 1.6 [25], [47](p.59) : Pour qu'une fonction de distribution F appartienne au domaine d'attraction de la loi de Weibull ($\mathcal{D}(\Psi_\xi)$) il faut et il suffit que

1. Il existe un point terminal x_F tel que

$$F(x_F) = 1 \quad \text{et} \quad F(x_F - \epsilon) < 1 \quad \forall \epsilon > 0 \quad ,$$

2. pour tout $\epsilon > 0$.

$$\lim_{t \rightarrow -0} \frac{1 - F(kt + x_F)}{1 - F(t + x_F)} = k^\xi \quad , \quad \forall k > 0 \quad .$$

c'est-à-dire si et seulement si $x_F < \infty$ et $\bar{F}(x_F - t^{-1}) \in \mathcal{RV}_{-\xi}$, $t \rightarrow \infty$.

Conditions de Von Mises (1936) [35](p.16), [47](p.62-66)

Certaines conditions suffisantes pour qu'une distribution appartienne à un domaine d'attraction ont été initialement formulées par Von Mises (1936) et s'avèrent souvent plus pratiques à vérifier que certaines des conditions présentées jusqu'à présent.

Théorème 1.7 : Supposons que la fonction de répartition F des variables aléatoires de la suite i.i.d. $\{X_n\}$ soit absolument continue avec une densité f . Alors, les conditions suffisantes pour que F appartienne à chacun des trois domaines d'attraction possibles sont :

Type 1 (Gumbel) : f a une dérivée négative f' pour tout x dans un certain intervalle (x_0, x_F) , $x_F \leq \infty$, $f(x) = 0$ pour $x \geq x_F$, et

$$\lim_{t \rightarrow x_F} \frac{f'(t)(1 - F(t))}{f^2(t)} = -1 \quad ;$$

Type 2 (Fréchet) : $f(x) > 0$ pour tout $x \geq x_0$ fini, et

$$\lim_{t \rightarrow \infty} \frac{tf(t)}{1 - F(t)} = \alpha > 0 \quad ;$$

Type 3 (Weibull) : $f(x) > 0$ pour tout x dans un certain intervalle fini (x_0, x_F) , et $f(x) = 0$ pour $x > x_F$, et

$$\lim_{t \rightarrow x_F} \frac{(x_F - t)f(t)}{1 - F(t)} = \alpha > 0 \quad ;$$

Proposition 1.5 : Soit X une variable aléatoire positive alors :

$$X \in \mathcal{D}(\Phi_\xi) \iff \log(X^\xi) \in \mathcal{D}(\Lambda) \iff -X^{-1} \in \mathcal{D}(\Psi_\xi) \quad .$$

1.2.3 Distribution des valeurs extrêmes généralisée GEVD

Le résultat théorique établi par von Mises et Jenkinson permet de généraliser les lois des valeurs extrêmes en unifiant les trois types de lois (Fréchet, Weibull, Gumbel) au sein d'une famille unique : la loi GEV, paramétrée par μ (localisation), σ (dispersion) et ξ de (forme).

Définition 1.12 (GEVD) [20](p.152, p.158) , [2](p.132)

la distribution généralisée des valeurs extrêmes a la forme suivante :

$$\mathcal{H}_{\mu, \sigma, \xi}(x) = \begin{cases} \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} & \text{si } 1 + \frac{\xi}{\sigma}(x - \mu) > 0, \xi \neq 0 \\ \exp \left\{ - \exp \left\{ - \frac{x - \mu}{\sigma} \right\} \right\} & \text{si } \xi = 0, x \in \mathbb{R} \end{cases}$$

avec $\mu \in \mathbb{R}$ et $\sigma > 0$.

En remplaçant $(\frac{x-\mu}{\sigma})$ par x on obtient la forme standard de la GEVD :

$$\mathcal{H}_\xi(x) = \begin{cases} \exp \{-(1 + \xi x)^{-1/\xi}\} & \text{si } \xi \neq 0, 1 + \xi x > 0 \\ \exp \{-\exp(-x)\} & \text{si } \xi = 0, x \in \mathbb{R} \end{cases}$$

La fonction de densité standard correspondante h_ξ est :

$$h_\xi(x) = \begin{cases} \mathcal{H}_\xi(x)(1 + \xi x)^{-\frac{1+\xi}{\xi}} & \text{si } \xi \neq 0, 1 + \xi x > 0 \\ \exp \{x - e^{-x}\} & \text{si } \xi = 0, x \in \mathbb{R} \end{cases}$$

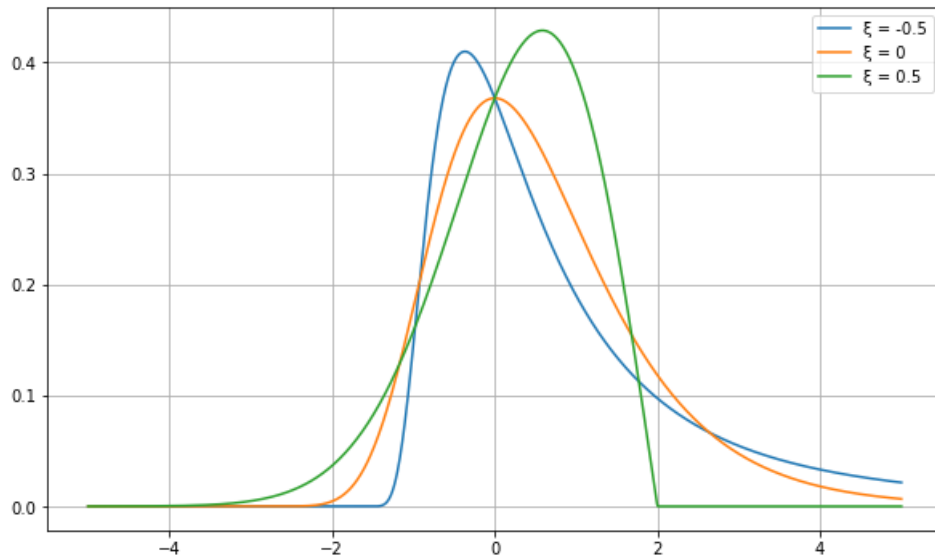
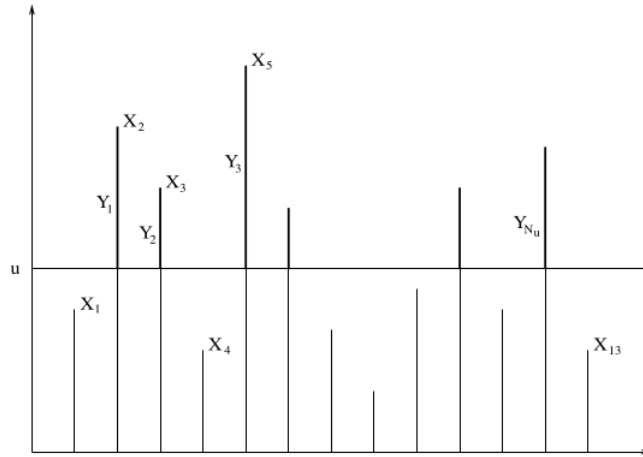


FIGURE 1.4 – La fonction de densité standard de la GEVD pour $(\mu=0$ et $\sigma=1)$

1.3 Le Modèle POT (Peaks-Over-Threshold)

L'approche par la loi des valeurs extrêmes généralisée (GEV) présente une limitation importante : en ne considérant que le maximum de chaque période, elle néglige l'information contenue dans les autres valeurs extrêmes de l'échantillon. Pour contourner cette difficulté, Pickands (1975) a introduit la méthode POT (Peaks-Over-Threshold), dite (des excès de seuil), qui s'appuie sur la distribution de Pareto généralisée (GPD).

FIGURE 1.5 – Observations X_i , et excès Y_j au-dessus du seuil u

Définition 1.13 (Excès) [44](p.86) : On appelle excès de la variable aléatoire X au-dessus du seuil u avec $(u < x_F)$ la variable aléatoire Y qui prend ses valeurs dans $(0, x_F - u)$ définie par :

$$Y = X - u | X > u \quad .$$

Pour n variables aléatoires i.i.d. X_1, \dots, X_n , nous pouvons définir les variables d'excès $Y_j = X_i - u | X_i > u$, où i est l'indice correspondant au j^{me} excès. pour $j=1, \dots, N_u$. Ici, $N_u \leq n$ représente le nombre d'excès au-dessus du seuil u , $N_u = \text{card}\{i : i = 1, \dots, n, X_i > u\}$.

1.3.1 Loi des excès

Soit une X une variable aléatoire de fonction de répartition F et de point terminale x_F . Pour tout seuil $u < x_F$, on définit la fonction de répartition des excès au-dessus du seuil u par :

$$\begin{aligned} F_u(x) &= P(X - u \leq x | X > u) \\ &= P(X \leq x + u | X > u) \\ &= \frac{P(\{X \leq x + u\} \cap \{X > u\})}{P(X > u)} \\ &= \frac{P(u < X \leq x + u)}{P(X > u)} \\ &= \frac{F(x + u) - F(u)}{1 - F(u)} \end{aligned}$$

Alors

$$F_u(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \frac{F(x+u)-F(u)}{1-F(u)} & \text{si } 0 \leq x \leq x_F - u \\ 1 & \text{si } x \geq x_F - u \end{cases} \quad (1.11)$$

Remarque 1.2 Il est clair que cette méthode nécessite de déterminer un seuil, noté u , qui soit ni trop faible (pour éviter de prendre en compte des valeurs non extrêmes), ni trop élevé (afin de disposer d'un nombre suffisant d'observations).

1.3.2 Le nombre d'excès au-dessus du seuil u

Dans ce théorème, on s'intéresse au nombre de fois où, dans un échantillon de taille n , le seuil u est dépassé.

Théorème 1.8 [34](p.1), [27](p.58)

Soit X_1, \dots, X_n un échantillon de variables aléatoires indépendantes et identiquement distribuées de fonction de répartition F . Si le seuil u est choisi de sorte que :

$$\lim_{n \rightarrow +\infty} n(1 - F(u_n)) = \lambda \quad \text{avec} \quad 0 < \lambda < \infty \quad .$$

Alors le nombre de dépassement du seuil u_n noté N_{u_n} suit asymptotiquement une loi de Poisson de paramètre λ .

Preuve [34](p.1), [27]

Ce résultat est la conséquence du fait que N_{u_n} est distribué selon une loi binomiale.

$$N_{u_n} \rightsquigarrow \beta(n, P_n) \quad \text{avec} \quad P_n = 1 - F(u_n) \quad ,$$

et

$$\beta\left(n, \frac{\lambda}{n}\right) \xrightarrow{d} \mathcal{P}(\lambda) \quad \text{quand} \quad n \rightarrow +\infty \quad (\text{converge en loi}) \quad .$$

1.3.3 Distribution de Pareto généralisée (GPD)

Définition 1.14 : La GPD (Generalized Pareto Distribution) est une distribution à trois paramètres où deux paramètres supplémentaires, $\sigma > 0$ et $\mu \geq 0$, ont été introduits. Le support de $G_{\xi, \sigma, \mu}(x)$ est $[\mu, +\infty)$ pour $\xi \geq 0$ et $[\mu, \mu - \frac{\sigma}{\xi}]$ pour $\xi < 0$. La distribution est à queue lourde lorsque $\xi > 0$.

Pour $\sigma > 0$, $\xi \in \mathbb{R}$ et $\mu \geq 0$, la GPD (Generalized Pareto Distribution) est définie par la fonction de répartition :

$$G_{\xi, \sigma, \mu}(x) := \begin{cases} 1 - (1 + \xi(\frac{x-\mu}{\sigma}))^{-1/\xi} & \text{si} \quad \xi \neq 0 \\ 1 - \exp(-\frac{x-\mu}{\sigma}) & \text{si} \quad \xi = 0 \end{cases}$$

où

$$\begin{cases} x \geq 0 & \text{si} \quad \xi \geq 0 \\ x \in [\mu, \mu - \frac{\sigma}{\xi}] & \text{si} \quad \xi < 0 \end{cases}$$

La GPD standard correspond au cas : $\mu = 0$, $\sigma = 1$

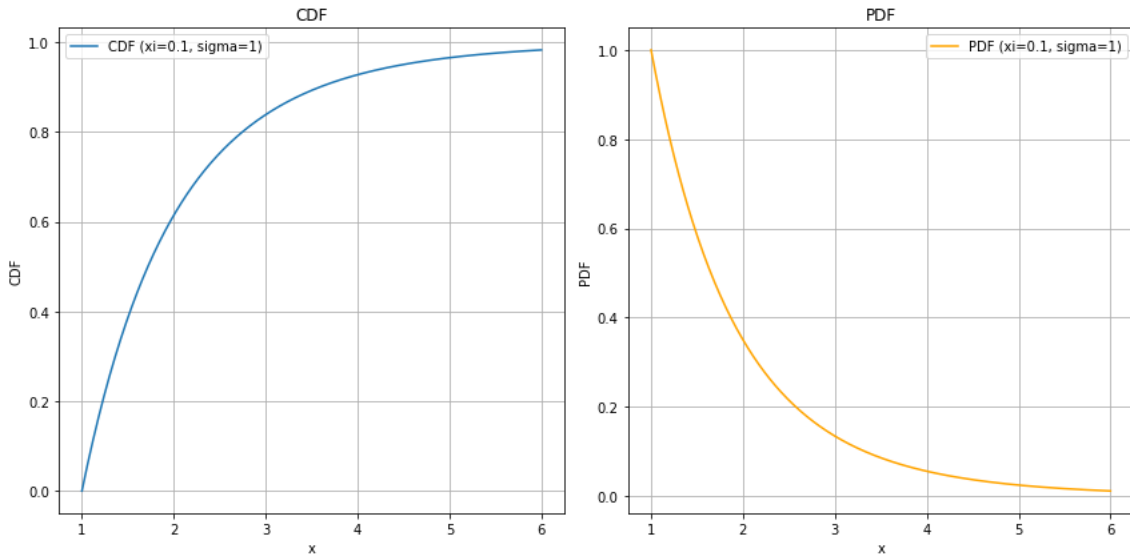


FIGURE 1.6 – La fonction de répartition cumulative (CDF) et La fonction de densité de probabilité (PDF) de la GPD pour ($\xi = 0.1, \sigma = 1, \mu = 0$)

1.3.4 Modélisation de la loi des Excès

La loi des excès F_u est approchée par la loi de Pareto généralisée (GPD) grâce à un résultat théorique fondamental en statistique des valeurs extrêmes : le théorème de Pickands-Balkema-de Haan. Ce théorème établit le lien entre le comportement asymptotique de la distribution des excès et la loi de Pareto généralisée.

Théorème 1.9 [Balkema-de Haan-Pickands] [44](p.86)

Si F appartient à l'un des trois domaines d'attraction de la loi limite des extrêmes (Fréchet, Gumbel ou Weibull) (1.6), alors il existe une fonction de répartition des excès au-delà de u , noté F_u qui peut être uniformément approchée par une loi de Pareto généralisée (GPD) telle que :

$$\lim_{u \rightarrow x_F} \sup_{0 < y < x_F - u} |F_u(y) - G_{\xi, \sigma(u)}(y)| = 0$$

ou σ est une fonction strictement positive.

Ce théorème indique que la distribution des excès F_u peut être remplacée par la distribution GPD $G_{\xi, \sigma(u)}$ lorsque u est suffisamment élevé. Pour comprendre son utilisation, remarquons que d'après l'équation (1.11) ci-dessus, nous pouvons écrire :

$$\bar{F}(x) = \bar{F}(u) \bar{F}_u(x - u)$$

pour $x > u$. En supposant que u est suffisamment grand, nous pouvons alors approximer F_u par $G_{\xi, \sigma(u)}$ et utiliser l'estimateur empirique pour $\bar{F}(u)$:

$$\widehat{\bar{F}}(u) = \frac{N_u}{n}, \quad \text{où } N_u = \sum_{i=1}^n \mathbf{1}_{\{X_i > u\}}$$

et où n est le nombre total d'observations. La queue supérieure de $F(x)$ peut alors être estimée par :

$$\widehat{F}(x) = 1 - \widehat{\bar{F}} = 1 - \frac{N_u}{n} \left(1 + \hat{\xi} \frac{x - u}{\hat{\sigma}} \right)^{-1/\hat{\xi}} \quad \text{pour tout } x > u.$$

où $\hat{\sigma}$ et $\hat{\xi}$ représentent les estimateurs de σ et ξ respectivement.

Conclusion

Nous avons exploré le domaine de la théorie des valeurs extrêmes, qui nous a permis de comprendre et de prévoir les événements rares mais aux conséquences significatives. À travers l'étude des distributions à queue lourde et des fonctions à variation régulière, nous avons découvert que les événements extrêmes ne sont pas totalement imprévisibles, mais suivent des lois mathématiques précises.

Les deux piliers de cette théorie - le modèle GEV pour les maxima et le modèle POT pour les dépassements de seuil - nous offrent des outils complémentaires pour appréhender l'extrême. Le modèle GEV nous permet de comprendre le comportement des pires scénarios, tandis que le modèle POT, avec sa distribution de Pareto généralisée, nous aide à analyser tout ce qui dépasse un certain seuil de risque.

Ces approches ne sont pas que théoriques : elles trouvent des applications concrètes dans la finance pour gérer les crises, dans l'environnement pour anticiper les catastrophes naturelles, et dans l'ingénierie pour construire des infrastructures plus résistantes. La maîtrise de ces concepts nous fait passer d'une position de vulnérabilité face aux événements extrêmes à une capacité active de gestion des risques.

Chapitre 2

Estimation à noyau de la densité de probabilité

L'analyse des distributions de probabilité constitue un outil fondamental en modélisation statistique, particulièrement dans le domaine économique où les variables aléatoires positives apparaissent fréquemment. Il existe de nombreuses applications économétriques où l'estimation précise des densités de variables aléatoires positives s'avère essentielle, que ce soit comme objet d'étude principal ou comme élément clé dans la spécification de modèles plus complexes.

Parmi les exemples les plus significatifs, on peut citer l'analyse des distributions de revenus, où la forme de la densité informe sur les inégalités économiques ; le traitement des données de transactions financières, crucial pour la gestion des risques ; la modélisation de la volatilité sur les marchés financiers ; ainsi que l'étude des données de durée et des temps de survie en assurance et en épidémiologie.

Dans la majorité de ces situations pratiques, les formes fonctionnelles appropriées restent inconnues ou font l'objet de vives controverses théoriques, rendant nécessaire le recours à des méthodes d'estimation non paramétriques. L'estimation à noyau de la densité émerge alors comme une approche particulièrement flexible, permettant de capturer les caractéristiques de la distribution sans imposer de restrictions paramétriques préalables.

Ce chapitre se propose donc d'examiner en détail la théorie et la pratique de l'estimation non paramétrique par noyau, en accordant une attention particulière aux enjeux liés à l'estimation près des frontières et aux méthodes développées pour améliorer la précision dans ces régions critiques.

2.1 Estimation par Histogramme

L'histogramme, bien au-delà d'une simple représentation graphique, constitue l'une des méthodes non paramétriques les plus fondamentales pour estimer la densité de probabilité d'une variable aléatoire continue. Lorsque la forme analytique de la distribution sous-jacente n'est pas connue, l'histogramme émerge comme un outil d'exploration indispensable permettant d'approcher la fonction de densité directement à partir des données observées.

Le principe fondamental de l'estimation par histogramme repose sur une idée à la fois simple et puissante : partitionner l'espace des observations en classes disjointes et estimer la densité dans chaque intervalle par la proportion relative des données qu'il contient.

Dans cette section, nous allons transcender la vision descriptive de l'histogramme pour aborder ses propriétés en tant qu'estimateur : son biais, sa variance, et sa convergence vers la vraie densité lorsque la taille de l'échantillon augmente. Nous examinerons également les règles de sélection optimales pour les paramètres de l'histogramme, qui permettent de trouver le compromis idéal entre la fidélité aux données et le lissage nécessaire pour capturer la structure fondamentale de la distribution.

Cette méthode, bien qu'ancienne, reste d'une actualité remarquable en science des données, en apprentissage automatique, et dans tous les domaines où l'exploration visuelle et quantitative des distributions est essentielle à la compréhension des phénomènes étudiés.

2.1.1 Construction de L'Histogramme

Lorsqu'on considère une distribution statistique à variable continue, on partitionne en général l'intervalle de référence $\Omega = [e_{min}, e_{max}]$ en $s \in \mathbb{N}^*$ classes (ou intervalles) C_1, \dots, C_s . Ces classes sont définies comme des intervalles disjoints, généralement semi-ouverts, pour garantir une partition unique :

$$C_j = [e_j, e_{j+1}), \quad j \in \{1, \dots, s\},$$

avec $e_1 = e_{min}$, $e_{s+1} = e_{max}$ et $e_{j+1} > e_j$ pour tout j .

La construction d'un histogramme à partir d'un échantillon x_1, \dots, x_n se déroule en trois étapes clés.

1. Détermination des Classes

Le choix des classes est l'étape la plus critique. On note $b_j = e_{j+1} - e_j$, $b_j \in \mathbb{R}_+$ la largeur (ou l'amplitude) de la $j^{\text{ème}}$ classe.

• **Classes d'amplitude égale** C'est le cas le plus simple, on fixe un nombre de classes s , et la largeur des classes est obtenue par la relation :

$$b = \frac{e_{max} - e_{min}}{s} . \tag{2.1}$$

. **Classes d'amplitude inégale** Pour les distributions asymétriques ou à queue lourde, il est préférable d'utiliser des largeurs inégales (b_j), des classes étroites là où la densité est forte (le pic) et des classes larges là où la densité est faible (les queues).

2. Comptage des Effectifs

Il s'agit de déterminer l'effectif de chaque classe C_j , noté n_j . C'est le nombre d'observations de l'échantillon qui appartiennent à cette classe, et est défini par :

$$n_j = \sum_{i=1}^n \mathbb{1}_{C_j}(x_i) \quad ,$$

où $\mathbb{1}_{C_j}(x_i)$ est la fonction indicatrice de la classe C_j , définie comme suit :

$$\mathbb{1}_{C_j}(x_i) = \begin{cases} 1 & \text{si } x_i \in C_j \\ 0 & \text{sinon} \end{cases}$$

3. Détermination de la Hauteur des Barres (Normalisation)

L'objectif est d'estimer la fonction de densité sous-jacente. Pour ce faire, l'aire de la barre ($h_j \times b_j$) doit être égale à la fréquence relative ($f_j = n_j/n$). La hauteur h_j de la barre C_j est donc calculée par :

$$h_j = \frac{f_j}{b_j} = \frac{n_j}{n \cdot b_j} \quad .$$

Cette normalisation garantit que l'aire totale de l'histogramme est égale à 1 ($\sum_{j=1}^s h_j b_j = \sum_{j=1}^s f_j = 1$). Et l'histogramme reste une estimation correcte de la densité, même si les classes sont d'amplitude inégale (cas b_j variable).

Dans l'analyse qui va suivre, nous travaillerons sous l'hypothèse que toutes les classes sont de même amplitude.

Choix de la largeur des classes d'un Histogramme

Le choix de la largeur des classes (b) constitue la décision la plus déterminante dans la construction d'un histogramme, un paramètre dont dépendent entièrement de la qualité et l'interprétabilité de l'estimation de densité résultante. L'apparence visuelle de l'histogramme en tant qu'estimateur de la fonction de densité de probabilité est entièrement déterminée soit par le nombre de classes (s), soit par l'amplitude (largeur) des classes (b).

Ce choix peut être formulé comme un problème de compromis entre biais et variance. Une largeur de classe trop grande (correspondant à un nombre faible de classes) a pour effet de lisser excessivement la distribution, masquant ainsi les caractéristiques fines de la densité (biais élevé). Inversement, une largeur de classe trop petite (nombre de classes élevé) introduit une variabilité excessive et du bruit dans l'estimation, rendant l'histogramme trop irrégulier (variance élevée).

Afin de minimiser l'erreur entre la vraie densité et la densité estimée, des méthodologies ont été développées pour déterminer la largeur b de manière adaptative. Parmi les approches modernes les plus reconnues, qui s'appuient sur les propriétés de dispersion de l'échantillon, on trouve : la règle de Scott (1979) [51], utilisant l'écart-type (σ) comme mesure de dispersion ; la règle de Freedman-Diaconis (1981) [22], basée sur l'écart interquartile (IQR), et l'approche la plus ancienne celle de Sturges (1926) [56], fondée sur la taille de l'échantillon (n).

Règle de Freedman-Diaconis (1981)

La règle de Freedman-Diaconis est une méthode spécifique pour déterminer la largeur de classe optimale (b) d'un histogramme directement à partir des données, en se basant sur deux facteurs : le nombre d'observations (n) et la dispersion (l'étalement) des données. Le paramètre de lissage optimal est donné par la formule suivante :

$$b = \frac{2 \cdot \text{IQR}(x)}{n^{1/3}} ,$$

où $\text{IQR} = Q_3 - Q_1$ est l'écart interquartile (Q_1 et Q_3 sont le premier et le troisième quartile respectivement).

Le nombre de classes s est donné par :

$$s = \frac{e_{max} - e_{min}}{b} .$$

comme s doit être un nombre entier, on arrondit généralement à l'entier supérieur.

Règle de Scott (1979)

La Règle de Scott est une méthode très proche de celle de Freedman-Diaconis qui sert à déterminer la largeur de classe optimale (b) d'un histogramme, qui minimise l'erreur entre la vraie densité et la densité estimée. Le paramètre de lissage optimal est donné par la formule suivante :

$$b = \frac{3.491 \cdot \sigma}{n^{1/3}} . \quad (2.2)$$

Règle de Sturges (1926)

La règle de Sturges est la méthode la plus ancienne pour déterminer le nombre de classes d'un histogramme. Contrairement aux règles modernes (Scott, Freedman-Diaconis) qui calculent d'abord la largeur (b), la règle de Sturges calcule d'abord le nombre de classes (s). Bien qu'elle soit simple, la règle de Sturges est fortement déconseillée dans les statistiques modernes car elle a été conçue pour des données qui suivent une distribution normale (courbe en cloche, symétrique et mésokurtique). De plus, pour les grandes tailles de données, Sturges donne des classes beaucoup trop larges (paramètre de lissage trop fort), ce qui efface tous les détails de la distribution. Le nombre de classes optimal est donné par la formule suivante :

$$s = 1 + 3.322 \log_{10}(n) \quad ,$$

où n est la taille de l'échantillon et $\log_{10}(n)$ est le logarithme en base 10 de n .

comme s doit être un nombre entier, on arrondit généralement à l'entier supérieur. Une fois s obtenu, on calcule la largeur b par la relation (2.1) .

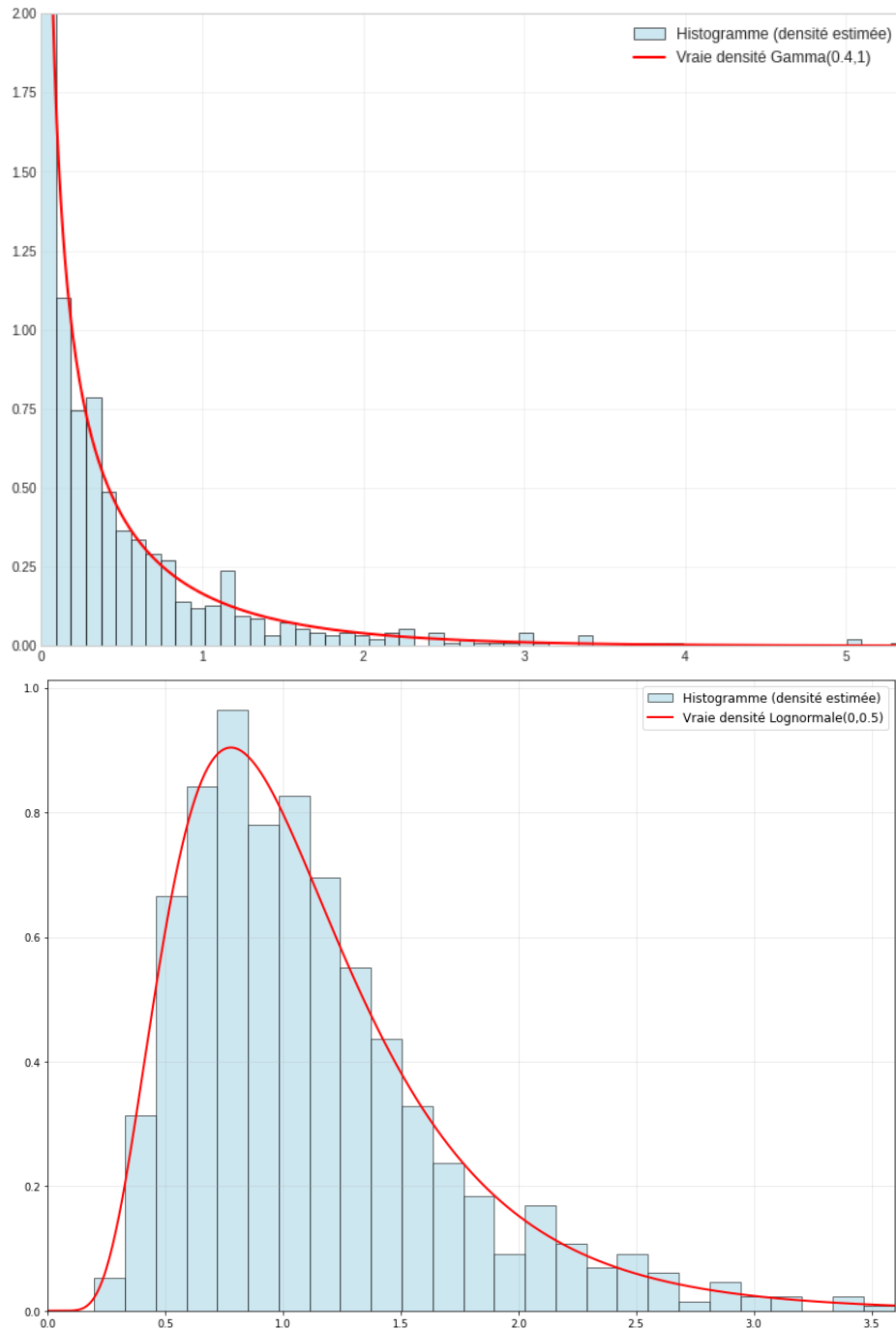


FIGURE 2.1 – Représentation des deux histogrammes issus de deux échantillons de taille $n = 1000$ chacun, provenant d'une loi Gamma($\alpha = 0.4, \beta = 1$) (en haut) et Lognormale($\mu = 0, \sigma = 0.5$) (en bas) et de leur véritable densité.

Les paramètres de lissage ont été calculés par la règle de Freedman-Diaconis (1981). Les résultats obtenus sont respectivement $b = 0,0933$ pour la loi $\text{Gamma}(0.4, 1)$ et $b = 0,1318$ pour la loi $\text{Lognormale}(0, 0.5)$. Les nombres de classes associés sont $s = 58$ et $s = 51$.

2.1.2 Estimateur de densité par Histogramme

Soit (x_1, \dots, x_n) un ensemble d'observations provenant de n variables aléatoires (X_1, \dots, X_n) indépendantes et identiquement distribuées (i.i.d) de densité commune f . L'estimation de la densité f en un point x par l'histogramme est donnée par :

$$\begin{aligned} \hat{f}(x) &= \frac{\text{Nombre d'observations dans la classe contenant } x}{n \cdot b} \\ &= \frac{1}{n \cdot b} \sum_{j=1}^s n_j \mathbb{1}_{C_j(x)} \quad . \end{aligned} \quad (2.3)$$

Une méthode intuitive pour convertir un histogramme en une fonction continue consiste à relier les milieux des sommets des barres, formant ainsi ce qu'on appelle un polygone des fréquences (illustré dans la figure 2.2). De plus, il est crucial de s'assurer que la fonction résultante est normalisée, c'est-à-dire que son intégrale est égale à 1, afin de représenter correctement une densité [52] (p.100).

Pour une lecture approfondie sur l'estimation de la densité par l'histogramme en général, vous pouvez consulter ce qui suit : [26](p.07), [52](p.101), [29] (p.21)

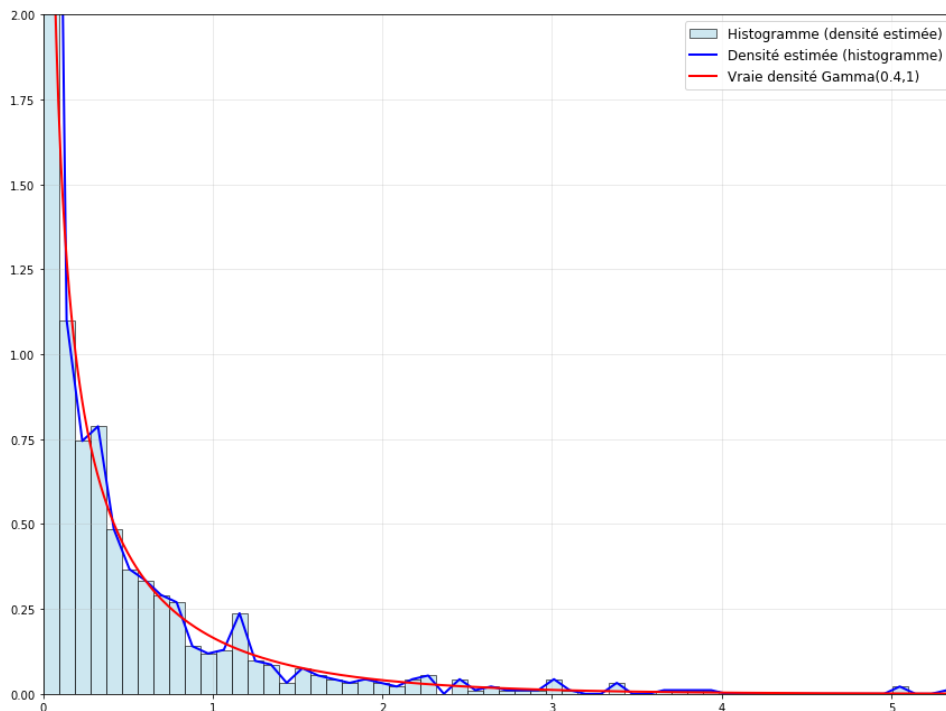


FIGURE 2.2 – Comparaison de la densité théorique et de la densité estimée par histogramme pour le même échantillon issu de la loi $\text{gamma}(0.4, 1)$ (figure 2.1 en haut).

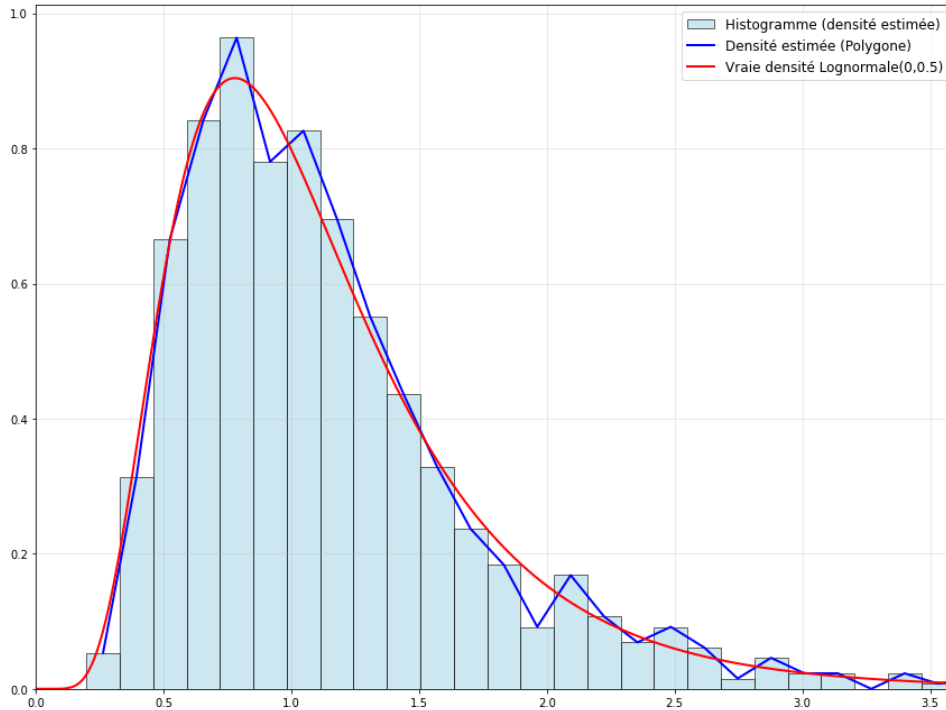


FIGURE 2.3 – Comparaison de la densité théorique et de la densité estimée par histogramme pour le même échantillon issu de la loi Lognormale(0, 0.5) (figure 2.1 en bas).

2.1.3 Valeur de l'estimateur \hat{f} dans chaque classe

Considérons une classe (intervalle) C_j de largeur b . Pour chaque observation X_i de l'échantillon, il n'y a que deux possibilités :

1. Succès : X_i tombe dans l'intervalle C_j .
2. Échec : X_i tombe ailleurs.

La probabilité de succès (p_j) est l'intégrale de la vraie densité $f(x)$ sur cet intervalle :

$$p_j = \int_{C_j} f(u) du \quad .$$

Par conséquent, le nombre de points n_j qui tombent dans cet intervalle suit une Loi Binomiale de paramètres n (taille totale) et p_j :

$$n_j \sim \mathcal{B}(n, p_j) \quad .$$

2.1.4 Théorie asymptotique de l'estimateur par Histogramme

En statistiques, il est essentiel de mesurer la qualité d'un estimateur. Dans cette partie, nous allons donc présenter une méthode pour évaluer $\hat{f}(x)$ en tant qu'estimateur de la densité théorique $f(x)$, en utilisant des mesures mathématiques simples à manipuler. Le fait que

l'effectif de chaque classe suivie une distribution binomiale est un point fondamental pour mesurer la qualité de l'estimateur.

Biais [54](p.16)

Le biais de l'estimateur \hat{f} est défini comme suit : pour tout $x \in [e_j, e_{j+1})$ et $j \in \{1, \dots, s\}$

$$\begin{aligned} \text{Biais} [\hat{f}(x)] &= E [\hat{f}(x)] - f(x) \\ &= \frac{1}{2} f'(x) [b - 2(x - e_j)] + \mathcal{O}(b^2) \quad . \end{aligned} \quad (2.4)$$

où f' est une fonction absolument continue et de carré intégrable ($f' \in L^2(\Omega)$).

Variance [54](p.16)

La variance de l'estimateur \hat{f} est défini pour tout $x \in \Omega$ par :

$$\begin{aligned} V [\hat{f}(x)] &= \mathbb{E} [(\hat{f}(x))^2] - (\mathbb{E} [\hat{f}(x)])^2 \\ &= \frac{f(x)}{nb} + \mathcal{O}(n^{-1}) \quad . \end{aligned} \quad (2.5)$$

L'erreur quadratique moyenne (MSE) [54](p.16)

Le MSE mesure l'erreur quadratique moyenne en un point particulier x . Il évalue la différence entre la valeur estimée par l'estimateur et la vraie valeur en ce point spécifique. Pour tout $x \in [e_j, e_{j+1})$:

$$\begin{aligned} \text{MSE} [\hat{f}(x)] &= V [\hat{f}(x)] + \text{Biais}^2 [\hat{f}(x)] \\ &= \frac{f(x)}{nb} + \frac{(f'(x))^2}{4} [b - 2(x - e_j)]^2 + \mathcal{O}(n^{-1}) + \mathcal{O}(b^3) \quad . \end{aligned} \quad (2.6)$$

L'erreur quadratique moyenne intégrée (MISE) [54](p.17)

Le MISE quantifie l'erreur quadratique moyenne intégrée entre la densité estimée et la densité réelle sur tout l'intervalle Ω .

En intégrant le MSE sur tout l'espace Ω on aura :

$$\text{MISE} [\hat{f}] = \frac{1}{nb} + \frac{b^2 R(f')}{12} + \mathcal{O}(n^{-1}) + \mathcal{O}(b^3) \quad . \quad (2.7)$$

où $R(f) = \int_{\Omega} [f(x)]^2 dx$, pour toute fonction f de carré intégrable.

Approximation du MISE (AMISE)

L'AMISE est une approximation du MISE lorsque la taille de l'échantillon n tend vers l'infini. Il est souvent utilisé car il permet de simplifier le calcul du MISE en utilisant des

approximations asymptotiques.

$$\text{AMISE} [\hat{f}] = \frac{1}{nb} + \frac{b^2 R(f')}{12} . \quad (2.8)$$

Le paramètre b optimale [54](p.17)

Le paramètre optimal (b_0) qui minimise l'AMISE est donné comme suite :

$$b_0 = \left[\frac{6}{R(f')} \right]^{1/3} \cdot n^{-1/3} . \quad (2.9)$$

Cette formule est l'origine et le fondement de la Règle de Scott (1979) (2.2). En pratique, l'intégrale $R(f')$ est inconnue car elle dépend de la vraie densité f . Pour rendre cette règle utilisable sans connaître f , David Scott a supposé que la densité sous-jacente est une distribution normale ($\mathcal{N}(\mu, \sigma)$). Pour la distribution normale, $R(f')$ est connu et égal à $\frac{1}{4\sigma^3\sqrt{\pi}}$. En substituant cette valeur dans la formule (2.9), on obtient la Règle pratique de Scott :

$$b_{\text{Scott}} \approx 3.49 \cdot \hat{\sigma} \cdot n^{-1/3} .$$

Malgré sa simplicité, l'histogramme présente deux inconvénients majeurs en tant qu'estimateur de densité : il dépend fortement du choix arbitraire de l'origine (voir [29] (p.30)) et de la largeur des classes (b), et son taux de convergence du biais est relativement lent ($\mathcal{O}(b)$), le rendant moins efficace que l'estimateur à noyau pour obtenir une approximation lisse et stable de la densité théorique.

Une solution simple, pour utiliser l'idée de l'histogramme comme estimation de la fonction de densité, c'est d'utiliser l'estimateur par histogramme Mobile.

2.2 Estimation par Histogramme Mobile

Définition 2.1 : L'estimateur de densité par l'histogramme mobile (ou l'estimateur simple) est une méthode d'estimation non-paramétrique qui généralise l'histogramme classique en faisant glisser une fenêtre sur tout le domaine des données. Cet estimateur, analysé rigoureusement pour la première fois par Rosenblatt (1956) [48], constitue la base des méthodes non-paramétriques d'estimation de densité. L'estimation de la densité en un point $x \in \Omega$ par cette méthode consiste à construire autour de x un intervalle (ou fenêtre) de largeur b , défini par :

$$\left[x - \frac{b}{2}, \quad x + \frac{b}{2} \right] .$$

2.2.1 Estimateur de densité par Histogramme Mobile

L'estimateur de densité par histogramme mobile, également connu sous le nom d'estimateur naïf, est la méthode la plus simple d'estimation non-paramétrique par fenêtre glissante (méthode des fenêtres), consistant à estimer la densité en un point donné x en comptant la proportion d'observations tombant dans un intervalle de largeur b centré sur ce point.

Partons du fait que :

$$\forall x \in \Omega, F(x) = \int_{-\infty}^x f(u) du \quad ,$$

on peut écrire :

$$\begin{aligned} f(x) &= \lim_{b \rightarrow 0} \frac{P(x - \frac{b}{2} \leq X \leq x + \frac{b}{2})}{b} \\ &= \lim_{b \rightarrow 0} \frac{F(x + \frac{b}{2}) - F(x - \frac{b}{2})}{b} \quad , \end{aligned}$$

L'estimateur simple de f , noté \hat{f}_b , peut alors être défini comme suit, $\forall x \in \Omega$, par :

$$\begin{aligned} \hat{f}_b &= \frac{\#\{i : x - \frac{b}{2} \leq x_i \leq x + \frac{b}{2}\}}{n \cdot b} \\ &= \frac{1}{nb} \sum_{i=1}^n \mathbb{1}_{[x-\frac{b}{2}, x+\frac{b}{2}]}(x_i) \\ &= \frac{1}{nb} \sum_{i=1}^n \mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]} \left(\frac{x - x_i}{b} \right) \quad , \end{aligned}$$

où $\#E$ est le cardinal de l'ensemble E , et $\mathbb{1}_{[x-\frac{b}{2}, x+\frac{b}{2}]}(x_i)$ est la fonction indicatrice, définie comme suit :

$$\mathbb{1}_{[x-\frac{b}{2}, x+\frac{b}{2}]}(x_i) = \begin{cases} 1 & \text{si } x_i \in [x - \frac{b}{2}, x + \frac{b}{2}] \\ 0 & \text{sinon} \end{cases}$$

L'estimateur \hat{f}_b peut aussi s'écrire :

$$\hat{f}_b(x) = \frac{1}{nb} \sum_{i=1}^n \mathcal{W} \left(\frac{x - x_i}{b} \right) \quad , \quad (2.10)$$

où \mathcal{W} est une fonction de poids définie par :

$$\mathcal{W}(x) = \begin{cases} 1 & \text{si } x \in [-\frac{1}{2}, \frac{1}{2}] \\ 0 & \text{sinon} \end{cases} \quad (2.11)$$

Cette fonction de poids n'est autre que la densité de probabilité de la loi uniforme sur l'intervalle $[-\frac{1}{2}, \frac{1}{2}]$.

2.2.2 Théorie asymptotique de l'estimateur par Histogramme Mobile

Biais [48](p.06)

Le biais de l'estimateur \hat{f}_b (2.10) est calculé comme suit, pour tout $x \in \Omega$:

$$\begin{aligned} \text{Biais} \left[\hat{f}_b(x) \right] &= \mathbb{E} \left[\hat{f}_b(x) \right] - f(x) \\ &= \frac{1}{nb} \cdot n \cdot E \left[\mathcal{W} \left(\frac{x-X}{b} \right) \right] - f(x) \\ &= \frac{1}{b} \int_{\Omega} \mathcal{W} \left(\frac{x-u}{b} \right) f(u) du - f(x) \quad , \end{aligned}$$

Posons $y = \frac{u-x}{b}$, ce qui implique $u = x + b \cdot y$ et $du = b \cdot dy$. Le biais de \hat{f}_b devient :

$$\text{Biais} \left[\hat{f}_b(x) \right] = \int_{\Omega} \mathcal{W}(-y) f(x + b \cdot y) dy - f(x) \quad ,$$

Comme la fonction de poids \mathcal{W} est symétrique ($\mathcal{W}(-y) = \mathcal{W}(y)$), on a :

$$\text{Biais} \left[\hat{f}_b(x) \right] = \int_{-1/2}^{1/2} f(x + b \cdot y) dy - f(x) \quad ,$$

En supposant que f est trois fois continûment dérivable, on effectue un développement de Taylor de $f(x + by)$ au voisinage de x :

$$f(x + by) = f(x) + byf'(x) + \frac{b^2 y^2}{2} f''(x) + \frac{b^3 y^3}{6} f^{(3)}(x) + o(b^3) \quad ,$$

Alors :

$$\begin{aligned} \text{Biais} \left[\hat{f}_b(x) \right] &= \int_{-1/2}^{1/2} \left[f(x) + b \cdot y f'(x) + \frac{b^2 y^2}{2} f''(x) + \frac{b^3 y^3}{6} f^{(3)}(x) + o(b^3) \right] dy - f(x) \\ &= \frac{b^2}{24} f''(x) + o(b^2) \quad . \end{aligned} \tag{2.12}$$

Variance [48](p.06)

La variance de \hat{f}_b est calculé comme suit, pour tout $x \in \Omega$:

$$\begin{aligned} V \left[\hat{f}_b(x) \right] &= \frac{1}{n^2 b^2} \cdot n \cdot V \left[\mathcal{W} \left(\frac{x-X}{b} \right) \right] = \frac{1}{nb^2} \left(E \left[\mathcal{W}^2 \left(\frac{x-X}{b} \right) \right] - \left(E \left[\mathcal{W} \left(\frac{x-X}{b} \right) \right] \right)^2 \right) \\ &= \frac{1}{nb^2} \left(E \left[\mathcal{W} \left(\frac{x-X}{b} \right) \right] - \left(E \left[\mathcal{W} \left(\frac{x-X}{b} \right) \right] \right)^2 \right) \quad , \end{aligned}$$

Pour la fonction de poids, on a $\forall y \in \mathbb{R}$, $\mathcal{W}(y) = \mathcal{W}^2(y)$. Comme $E \left[\mathcal{W} \left(\frac{x-X}{b} \right) \right]$ a déjà été

calculé dans (2.12), alors $V[\hat{f}_b(x)]$ devient :

$$V[\hat{f}_b(x)] = \frac{f(x)}{nb} + \frac{b}{24n}f''(x) - \frac{f^2(x)}{n} - \frac{b^2}{12n}f(x)f''(x) + o\left(\frac{b^2}{n}\right) . \quad (2.13)$$

L'erreur quadratique moyenne (MSE)

$$\begin{aligned} \text{MSE}[\hat{f}_b(x)] &= \left(\text{Bias}[\hat{f}_b(x)]\right)^2 + V[\hat{f}_b(x)] \\ &= \frac{b^4}{576}(f''(x))^2 + \frac{f(x)}{nb} + o\left(\frac{1}{nb} + b^4\right) . \end{aligned} \quad (2.14)$$

L'erreur quadratique moyenne intégrée (MISE)

$$\begin{aligned} \text{MISE}[\hat{f}_b] &= \int_{\Omega} \text{MSE}[\hat{f}_b(x)] dx \\ &= \frac{1}{nb} + \frac{b^4}{576}R(f'') + o\left(\frac{1}{nb} + b^4\right) . \end{aligned} \quad (2.15)$$

où $R(f'') = \int (f''(x))^2 dx$.

L'erreur quadratique moyenne intégrée asymptotique (AMISE)

$$\text{AMISE}[\hat{f}_b] = \frac{1}{nb} + \frac{b^4}{576}R(f'') . \quad (2.16)$$

Largeur de fenêtre optimale

Pour trouver la bande passante optimale (b^*), on minimise l'AMISE par rapport à b . Cela conduit à l'expression suivante :

$$b^* = \left(\frac{144}{nR(f'')}\right)^{1/5} .$$

2.3 Estimateur à noyau de densité de probabilité

Nous avons vu que l'estimateur naïf (histogramme mobile) permet de s'affranchir du choix arbitraire de l'origine des classes. Cependant, il conserve un défaut majeur : la discontinuité de la fonction estimée, due à l'utilisation d'une fenêtre rectangulaire (poids uniforme). Pour obtenir une densité estimée lisse et régulière, nécessaire à une analyse fine, il convient de généraliser cette approche. Dans cette section, nous introduisons l'estimateur à noyau, qui remplace la fenêtre rectangulaire par une fonction de poids plus douce, appelée noyau [54].

L'estimation de la densité de probabilité par noyau est une technique indispensable qui offre une analyse plus fine de la distribution de probabilité qu'un histogramme traditionnel.

Deux concepts jouent un rôle fondamental dans l'estimation par noyau : la forme de la fonction noyau elle-même et le paramètre de lissage (bande passante). Ce dernier est crucial, car c'est lui qui détermine le lissage de l'estimation finale. Un choix approprié de ce paramètre est donc essentiel pour la qualité de l'estimation [61].

Définition 2.2 (Noyau) [37] : Un noyau sommatif (ou simplement noyau) est une fonction $K : \mathbb{D} \rightarrow \mathbb{R}$ (avec $\mathbb{D} \subseteq \mathbb{R}$), qui vérifie la propriété de sommativité, c'est-à-dire :

$$\int_{\mathbb{D}} K(u) du = 1 \quad .$$

Le noyau sommatif K peut être vu comme une densité de probabilité induisant une mesure de probabilité, notée P_K , définie par :

$$\forall A \subset \mathbb{D}, P_K(A) = \int_A K(u) du \quad .$$

L'estimateur simple de densité défini par l'expression (2.10) peut être généralisé en remplaçant la fonction de poids \mathcal{W} (2.11) (qui correspond à une densité de probabilité uniforme) par une fonction plus générale, notée K , appelée noyau (en anglais, kernel). Cette fonction K doit également être une densité de probabilité. Cependant, Wand et Jones (1995) [59] mentionnent que des noyaux qui ne sont pas des densités sont parfois aussi utilisés.

2.3.1 Estimateur de densité de Parzen-Rosenblatt

L'estimateur à noyau classique, appelé aussi estimateur de densité de Parzen-Rosenblatt a été introduit par M.Rosenblatt [17] et développé par E.Parzen [42].

Définition 2.3 (Noyau de Parzen-Rosenblatt) [42], [58]

Soit (X_1, \dots, X_n) un échantillon i.i.d. de taille n issu d'une variable aléatoire X de densité de probabilité inconnue f . L'estimateur de densité à noyau de Parzen-Rosenblatt, noté \hat{f}_X , est défini pour tout $x \in \mathbb{R}$ par :

$$\hat{f}_X(x) = \frac{1}{nb} \sum_{i=1}^n K\left(\frac{x - X_i}{b}\right) \quad , \quad (2.17)$$

Où :

- n est la taille de l'échantillon.
- b est le paramètre de lissage (fenêtre ou bandwidth). C'est une suite de réels strictement positifs ($b_n > 0$) vérifiant les conditions de convergence suivantes quand $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} b_n = 0 \quad \text{et} \quad \lim_{n \rightarrow \infty} nb_n = \infty \quad ,$$

- K est la fonction Noyau (Kernel), une fonction intégrable vérifiant les propriétés suivantes :

1. Normalisation : $\int_{\mathbb{D}} K(u) du = 1$.
2. Moment d'ordre 1 est nul : $\int_{\mathbb{D}} uK(u) du = 0$.
3. Symétrie : $K(-u) = K(u)$.
4. Positivité : $K(u) \geq 0$.
5. Moment d'ordre 2 fini : $\int_{\mathbb{D}} u^2 K(u) du = \mu_2(K) < \infty$.

Le Tableau suivant résume quelques fonctions de noyau symétriques [26] (p.27), [52] (p.152).

Noyau	Définition
Epanechnikov(parabolique)	$\frac{3}{4}(1 - u^2)\mathbb{1}_{[-1,1]}(u)$
Biweight	$\frac{15}{16}(1 - u^2)^2\mathbb{1}_{[-1,1]}(u)$
Triweight	$\frac{35}{32}(1 - x^2)^3\mathbb{1}_{[-1,1]}(u)$
cosine arch (cosinus)	$\frac{\pi}{4} \cos\left(\frac{\pi}{2}u\right)\mathbb{1}_{[-1,1]}(u)$
Gaussien	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$
Uniforme(Rectangulaire)	$\frac{1}{2}\mathbb{1}_{[-1,1]}(u)$
triangulaire	$(1 - u)\mathbb{1}_{[-1,1]}(u)$

TABLE 2.1 – Exemples de fonctions noyaux symétriques

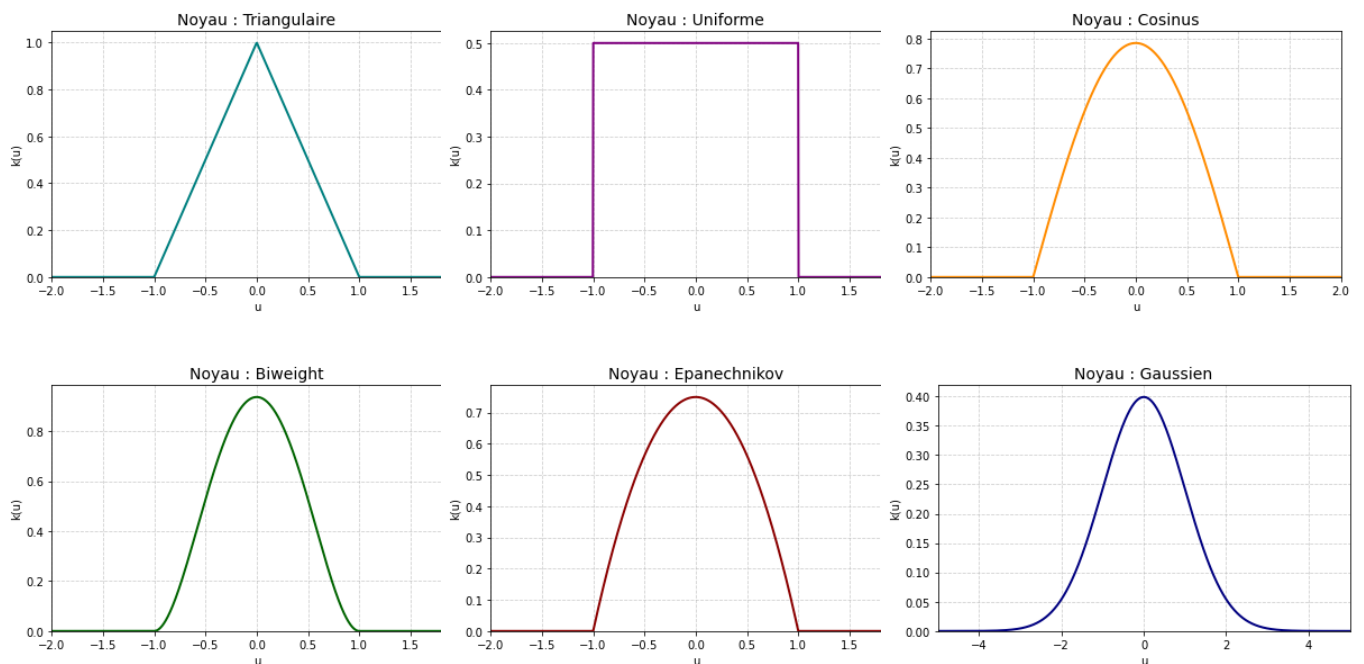


FIGURE 2.4 – Ensemble de six noyaux sommatifs symétriques

L'estimation de densité par noyau (Kernel Density Estimation, KDE) est une méthode non-paramétrique permettant d'estimer la fonction de densité de probabilité inconnue d'une variable aléatoire à partir d'un échantillon de données observées. Il s'agit d'une estimation ponctuelle, où la densité estimée en un point x est obtenue en calculant la moyenne des contributions (ou ordonnées) des fonctions noyau centrées sur chaque observation.

Le principe est le suivant : on choisit une fonction noyau, puis on centre une copie de cette fonction sur chaque point de données de l'échantillon, comme le montre la figure 2.5.

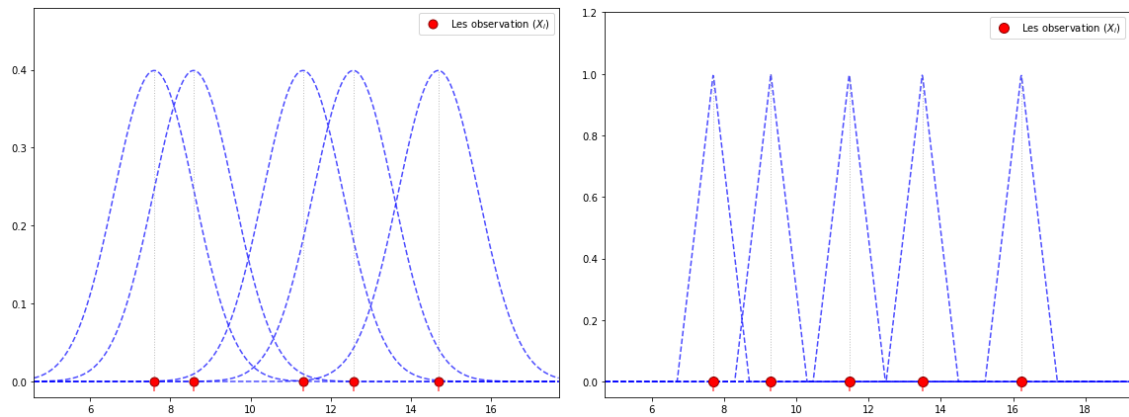


FIGURE 2.5 – la 1^{er} étape de la KDE, avec le noyau gaussien à gauche et le noyau triangulaire à droite.

L'estimateur final (KDE) est obtenu en sommant toutes ces fonctions noyau ainsi positionnées, puis en divisant cette somme par le nombre total de points de données. Le résultat est une estimation de la densité de probabilité. [59] (p.11).

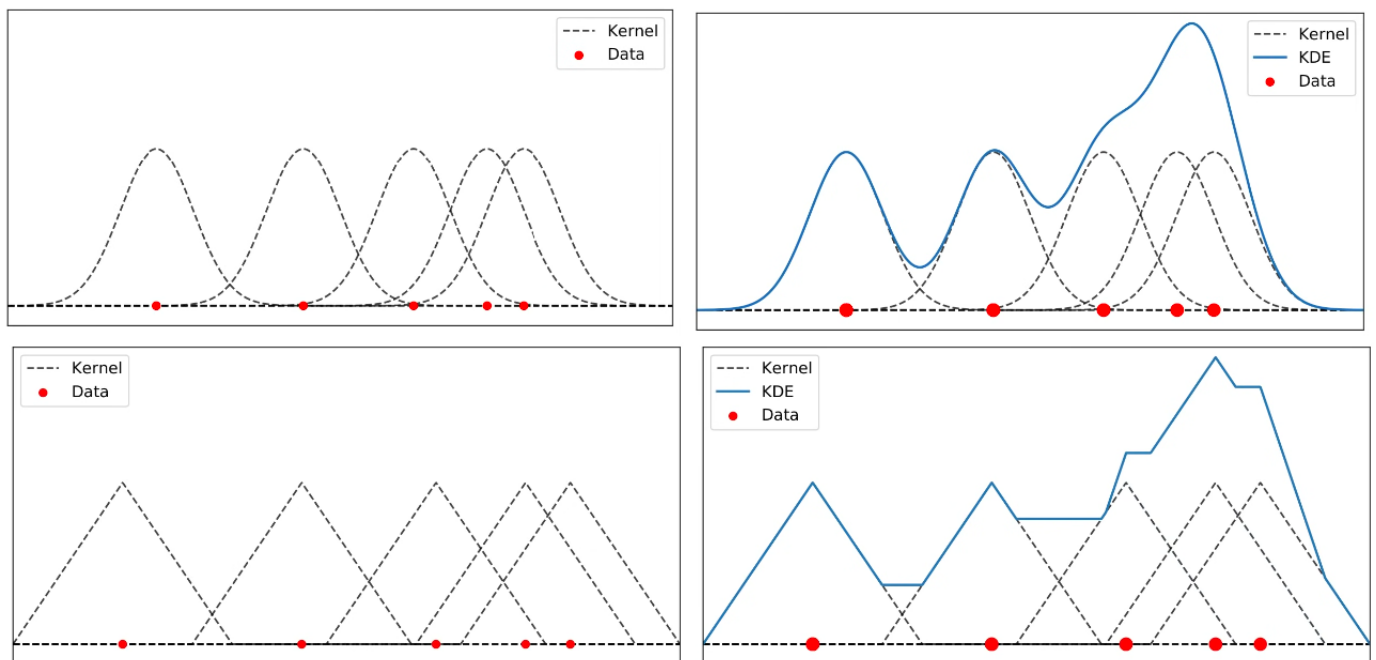


FIGURE 2.6 – Estimation point par point, avec le noyau gaussien en haut et le noyau triangulaire en bas.

2.3.2 Théorie asymptotique de l'estimateur à noyau symétrique

Biais [59] (p.20)

Pour tout $x \in \mathbb{D}$, l'espérance de \hat{f}_X est donnée comme suit :

$$E \left[\hat{f}_X(x) \right] = f(x) + \frac{b^2}{2} f''(x) \int t^2 K(t) dt + o(b^2) \quad ,$$

Alors le biais de l'estimateur \hat{f}_X (2.17) est définie comme suit : $\forall x \in \mathbb{D}$

$$\text{Biais} \left[\hat{f}_X(x) \right] = \frac{b^2}{2} \mu_2(K) f''(x) + o(b^2) \quad . \quad (2.18)$$

où

f'' est la dérivée seconde de la fonction de densité f , et elle doit être une fonction absolument continue, et $\mu_2(K) = \int t^2 K(t) dt$.

E. Parzen (1962) [42] a établi que les estimateurs définis en (2.17) sont asymptotiquement non biaisés en tout point x où f est continue, à condition que la suite b_n satisfasse $\lim_{n \rightarrow \infty} b_n = 0$ et que le noyau K soit une fonction borélienne (soit $\mathbb{D} \subseteq \mathbb{R}$, $K(u)$ et une fonction borélienne si $K^{-1}(B)$ est un ensemble borélien pour tout ensemble borélien $B \subset \mathbb{R}$ [1]) satisfait :

$$\sup_{-\infty < u < \infty} |K(u)| < \infty \quad , \quad (2.19)$$

$$\int_{-\infty}^{\infty} |K(u)| du < \infty \quad , \quad (2.20)$$

$$\lim_{u \rightarrow \infty} |K(u)| = 0 \quad , \quad (2.21)$$

$$\int_{-\infty}^{\infty} K(u) du = 1 \quad . \quad (2.22)$$

Variance [59] (p.20)

La variance de l'estimateur \hat{f}_X (2.17) est définie comme suit : $\forall x \in \mathbb{D}$

$$V \left[\hat{f}_X(x) \right] = (nb)^{-1} f(x) R(K) + o\left(\frac{1}{nb}\right) \quad , \quad (2.23)$$

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

D'après l'expression (2.23) on a :

$$V \left[\hat{f}_X(x) \right] \longrightarrow 0 \quad \text{quand} \quad n \cdot b \rightarrow +\infty \quad ,$$

L'erreur quadratique moyenne (MSE) [59] (p.21)

L'erreur quadratique moyenne de \hat{f}_X (2.17) en un point $x \in \mathbb{D}$ est définie comme suit :

$$\text{MSE} \left[\hat{f}_X(x) \right] = (nb)^{-1} f(x) R(K) + \frac{b^4}{4} (\mu_2(K))^2 (f''(x))^2 + o \left(\frac{1}{nb} + b^4 \right) . \quad (2.24)$$

L'erreur quadratique moyenne intégrée (MISE) [59] (p.21)

L'erreur quadratique moyenne intégrée de \hat{f}_X (2.17) est obtenue par l'intégration de (2.24) par rapport à x sur tout l'intervalle \mathbb{D} , elle est définie comme suit :

$$\text{MISE} \left[\hat{f}_X \right] = (nb)^{-1} R(K) + \frac{b^4}{4} (\mu_2(K))^2 R(f'') + o \left(\frac{1}{nb} + b^4 \right) , \quad (2.25)$$

où $R(f'') = \int_{\mathbb{D}} f''(u)^2 du$.

L'erreur quadratique moyenne intégrée asymptotique (AMISE) [59] (p.21)

L'erreur quadratique moyenne intégrée asymptotique de \hat{f}_X (2.17) est définie comme suit :

$$\text{AMISE} \left[\hat{f}_X \right] = (nb)^{-1} R(K) + \frac{b^4}{4} (\mu_2(K))^2 R(f'') . \quad (2.26)$$

Largeur de fenêtre optimale [59] (p.22)

La Largeur de fenêtre optimale (b^*) qui minimise l'AMISE est donnée comme suit :

$$b^* = \left(\frac{R(K)}{n(\mu_2(K))^2 R(f'')} \right)^{\frac{1}{5}} . \quad (2.27)$$

2.3.3 Propriétés de l'estimateur de Parzen-Rosenblatt

L'objectif de l'estimateur de Parzen-Rosenblatt est d'estimer la densité de probabilité f en chaque point x de l'espace des données \mathbb{D} . Pour cela, il repose sur l'idée de centrer une fonction noyau K autour de chaque observation X_i de l'échantillon. La contribution de chaque observation X_i à l'estimation de la densité au point x est pondérée par $\frac{1}{b} K\left(\frac{x-X_i}{b}\right)$, où le noyau K détermine la forme du voisinage local tandis que la largeur de bande b contrôle le degré de lissage. Afin d'obtenir un estimateur raisonnable, ces deux paramètres, le noyau et la largeur de fenêtre, doivent être choisis avec soin [59].

Choix de la Largeur de Bande (b)

La largeur de bande b contrôle le degré de lissage de l'estimation. Une petite valeur de b donne une estimation plus détaillée mais plus bruitée, tandis qu'une grande valeur de b donne une estimation plus lisse mais moins détaillée.

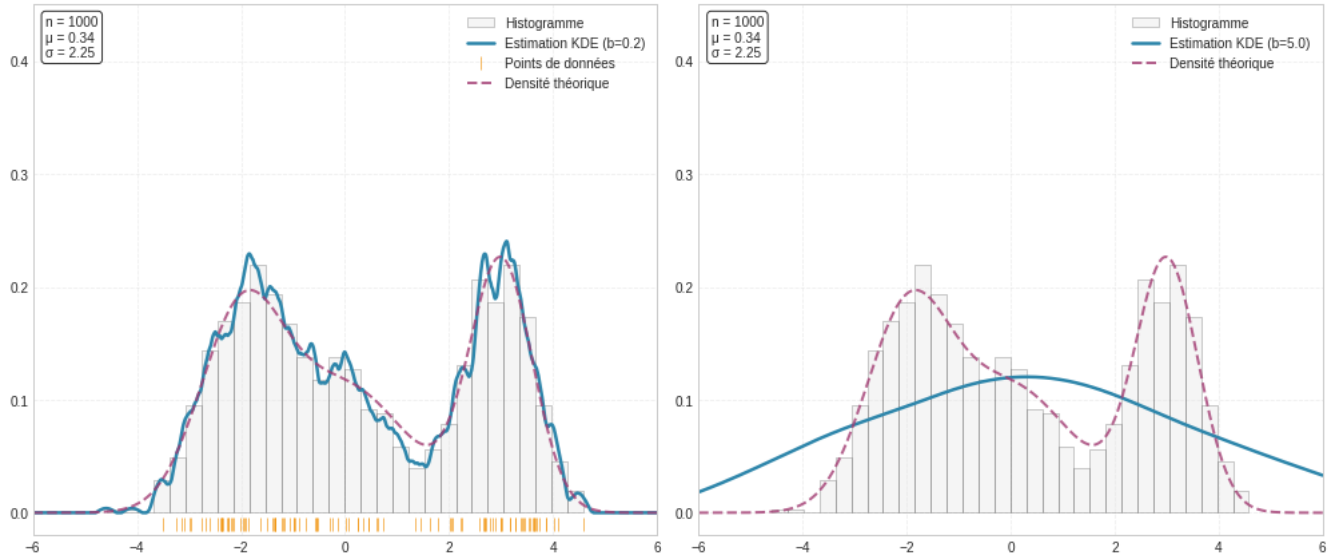


FIGURE 2.7 – L’estimateur de densité de Parzen-Rosenblatt construit à partir de 1000 observations en utilisant le noyau cosinus avec une largeur de bande $b = 5$ (à droite) et $b = 0.2$ (à gauche), pour μ et σ , la moyenne et l’écart-type des données.

Le problème du choix du paramètre de lissage (bandwidth) constitue une question cruciale qui se pose fréquemment dans le cadre de l’estimation par noyau (KDE). Cette partie présente une synthèse des méthodes existantes pour la sélection optimale de ce paramètre.

Plusieurs méthodologies ont été développées pour choisir le paramètre de lissage b de manière automatisée et optimale, chacune cherchant à minimiser une mesure d’erreur, généralement l’erreur quadratique moyenne intégrée (MISE). Dans ce mémoire, on se limite uniquement à deux méthodes :

• Méthode empirique (Rule-of-Thumb (RoT))

• Règle de Silverman (1986) [26] (p.65)

La règle de Silverman (1986) est une méthode paramétrique de sélection de la bande passante (b) pour l’estimation de densité par noyau, qui suppose que la densité sous-jacente des données appartient à la famille des distributions normales avec une moyenne μ et une variance σ^2 .

$$b^* = \left(\frac{8\sqrt{\pi} \cdot R(K)}{3 \cdot \mu_2(K)^2} \right)^{1/5} \cdot \hat{\sigma} \cdot n^{-1/5} \quad (2.28)$$

$$\approx 1.06 \cdot \hat{\sigma} \cdot n^{-1/5} .$$

Où

$\hat{\sigma}$ est l’estimation de σ (l’écart-type des données).

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

$$\mu_2(K) = \int_{\mathbb{D}} u^2 K(u) du \text{ (moment d’ordre 2 du noyau).}$$

Bien que cette règle empirique soit facile à calculer, elle doit être utilisée avec prudence, car elle peut produire des estimations très inexactes lorsque la densité n'est pas proche d'une normale. En pratique, la vraie distribution de X est généralement inconnue et, dans un tel cas, (2.28) donne une largeur de fenêtre proche de l'optimale seulement si la distribution de X n'est pas significativement différente de la distribution normale.

La version améliorée de la règle empirique est donné comme suite :

$$b^* = 1.06 \cdot \min \left\{ \hat{\sigma}, \frac{\text{IQR}}{1.34} \right\} \cdot n^{-1/5} \quad .$$

• Méthodes de Validation Croisée

• Validation croisée par moindres carrés (Least Squares Cross Validation)

[8], [26] (p.67), [18]

La méthode de Validation Croisée des Moindres Carrés (Least Squares Cross-Validation ou LSCV) est une technique statistique puissante pour choisir automatiquement la largeur de fenêtre optimale b sans faire d'hypothèse sur la forme de la vraie densité f . C'est une méthode purement basée sur les données.

L'objectif idéal est de minimiser l'erreur quadratique intégrée (ISE), qui mesure la distance globale entre notre estimation \hat{f}_X et la vraie densité f .

$$\begin{aligned} \text{ISE}(b) &= \int (\hat{f}_X(x) - f(x))^2 dx \\ &= \int (\hat{f}_X(x))^2 dx + \int (f(x))^2 dx - 2 \int \hat{f}_X(x) f(x) dx \quad . \end{aligned}$$

Notez que le second terme de l'équation ci-dessus ne dépend pas de b , donc minimiser l'ISE en b revient à minimiser :

$$\text{LSCV}(b) = \int (\hat{f}_X(x))^2 dx - 2 \int \hat{f}_X(x) f(x) dx \quad .$$

L'idée de la validation croisée par moindre carrés (LSCV) est de remplacer ces termes inconnus par des estimateurs basés sur les données :

$$\text{LSCV}(b) = \int \hat{f}_X^2(x) dx - \frac{2}{n} \sum_{i=1}^n \hat{f}_{X,-i}(X_i) \quad ,$$

où $\hat{f}_{X,-i}$, appelé l'estimateur jack-knife, est l'estimateur obtenu sans l'observation i . Comme le nom l'indique, l'observation X_i n'est pas utilisée dans les calculs. Le terme validation croisée fait référence à l'utilisation d'un sous-ensemble de données pour réaliser l'analyse sur un autre sous-ensemble. Cela garantit que les observations utilisées pour calculer $\hat{f}_{X,-i}$ sont indépendantes de X_i .

En substituant la définition de $\hat{f}_X(x)$ (2.17) et en simplifiant, la fonction LSCV finale (que l'on minimise) est :

$$\text{LSCV}(b) = \frac{1}{n^2 b} \sum_{i=1}^n \sum_{j=1}^n K^* \left(\frac{X_i - X_j}{b} \right) - \frac{2}{n(n-1)b} \sum_{i=1}^n \sum_{j=1, j \neq i}^n K \left(\frac{X_i - X_j}{b} \right) . \quad (2.29)$$

Où : $K^*(u) = \int K(x)K(u-x)dx$ est le noyau de convolution.

Nous avons effectué une petite simulation où nous avons généré deux échantillons (de tailles 50 et 1000) d'une loi de Weibull(1.8, 1) et comparé, sur 1000 répliques, l'estimation à noyau lorsque le paramètre de lissage est sélectionné par LSCV (à partir de l'échantillon) ou bien par la minimisation directe de l'ISE (à partir de la vraie densité). Les résultats de l'estimation et les paramètres de lissage sont calculés par moyenne sur les répliques (Lors de chaque réplique, un nouvel échantillon a été généré. De nouveaux paramètres de lissage et une estimation de la densité sur l'ensemble du domaine ont ensuite été calculés. Finalement, on calcule la moyenne des paramètres de lissage ainsi que celle des estimations de la vraie densité en chaque point) et sont présentés ci-dessous.

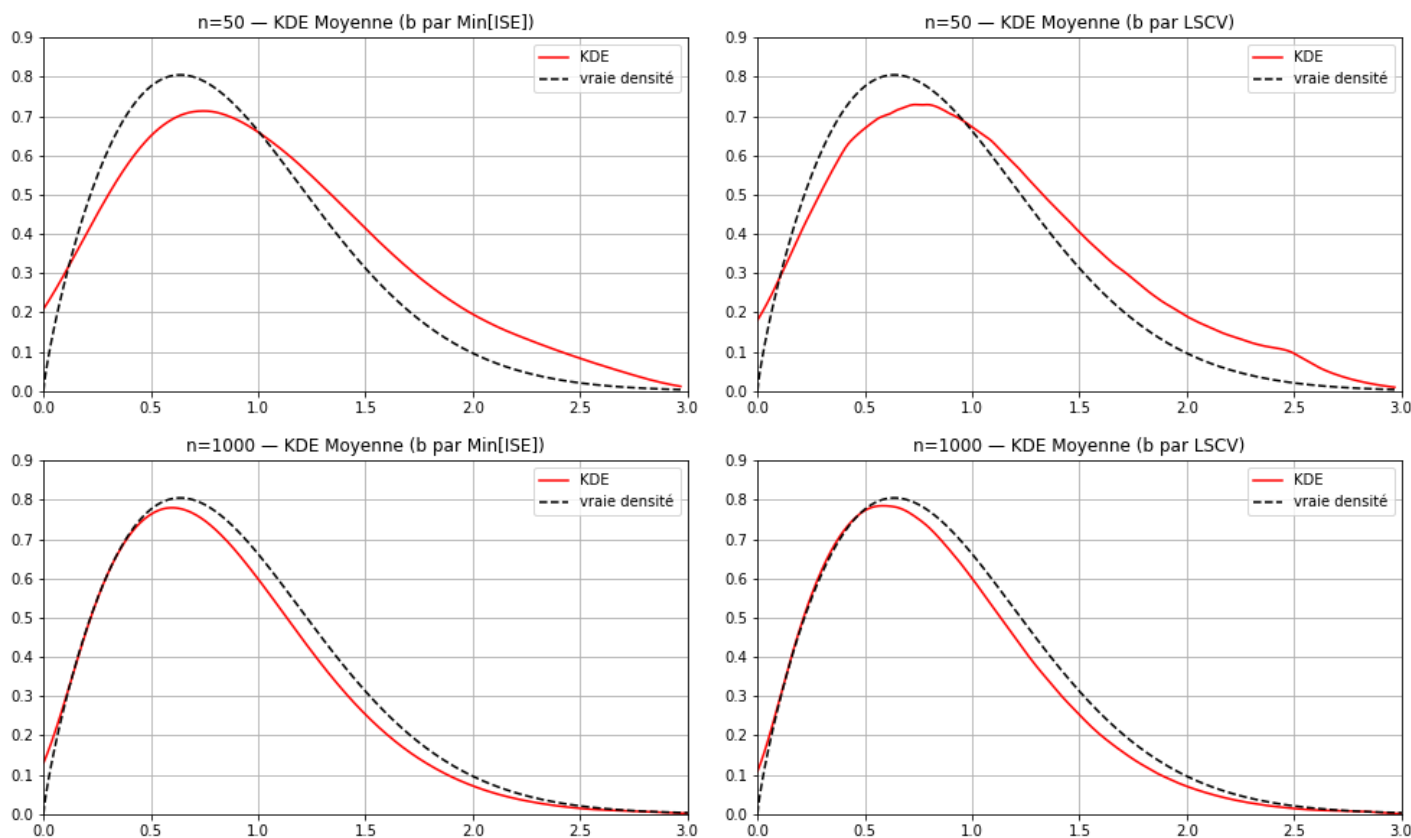


FIGURE 2.8 – L'estimation de la densité de Weibull(1.8, 1) par noyau d'Epanechnikov en fonction de la taille de l'échantillon ($n = 50$ en haut et $n = 1000$ en bas) et de la méthode de sélection de la bande passante (min(ISE) à gauche et LSCV à droite.).

n	Min(ISE)	LSCV
50	0.5171	0.4586
1000	0.2611	0.2104

TABLE 2.2 – Paramètres de lissage obtenus

Les résultats montrent que malgré l'absence de connaissance a priori de la vraie densité, la méthode LSCV parvient à sélectionner une fenêtre qui produit une estimation dont l'écart avec l'optimum théorique (Min ISE) est négligeable, Cela valide l'usage de la LSCV comme méthode pratique fiable pour l'analyse exploratoire de données réelles où la vraie distribution est inconnue.

Choix de la Fonction Noyau (K)

Wand et Jones (1995) [59], ainsi que Węglarczyk (2018) [61], mentionnent que pour les fonctions noyau symétriques, le choix de la forme du noyau $K(\cdot)$ a relativement peu d'influence sur l'estimateur comparativement au paramètre de lissage b . Cependant, Węglarczyk (2018) [61] indique que les noyaux triangulaire et rectangulaire (uniforme), en particulier ce dernier, tendent à produire de nombreux maxima locaux et ne sont donc pas recommandés en pratique.

Le noyau Biweight présente une décroissance plus rapide que le noyau d'Epanechnikov, ce qui peut réduire l'influence des points éloignés et potentiellement mieux révéler les modes de la distribution. Quant au noyau gaussien, défini sur l'ensemble des réels (support infini), il produit l'estimation la plus régulière (C^∞).

Dans les études plus récentes, Skorski (2019) [55] souligne que la précision de l'estimation à noyau de la fonction de densité dépend de la nécessité de maintenir un certain équilibre entre la décroissance du noyau (kernel decay) et la largeur de fenêtre ; sinon, quelles que soient la régularité du noyau et ses moments, les estimations ne sont pas bornées. Il a démontré à travers son étude que le problème lié à la formule de Taylor peut être résolu sans hypothèses plus complexes sur la densité f , mais qu'il peut être traité uniquement par la décroissance du noyau. Autrement dit, cette décroissance constitue un mécanisme suffisant pour contrôler le biais d'estimation, même lorsque la densité cible f ne vérifie pas les hypothèses fortes de régularité globale ou de support compact généralement requises. Contrairement à l'analyse traditionnelle comme celle de Wand et Jones (1995) [59], qui repose sur le développement de Taylor de f sur tout son domaine et nécessite précisément de telles hypothèses pour garantir la validité des approximations.

Les noyaux à décroissance rapide, tels que le noyau gaussien (décroissance exponentielle), présentent des avantages théoriques et pratiques substantiels par rapport aux noyaux à décroissance lente (par exemple polynômiale). Ceci justifie la popularité de noyaux comme le noyau gaussien, dont les propriétés de décroissance et de régularité infinie en font un choix

privilegié dans de nombreuses applications, notamment en grande dimension où le contrôle des queues de distribution est critique.

2.3.4 Problème de Biais aux Limites dans l'Estimation par Noyau de Densité.

L'estimation par noyau de densité (KDE) est une méthode non paramétrique fondamentale pour estimer une densité de probabilité. Cependant, sa formulation standard souffre d'un défaut important, connu sous le nom de problème de bord ou, plus précisément, de biais aux limites. Ce biais systématique apparaît lorsque le support des données est borné ou semi-borné (par exemple, limité à \mathbb{R}^+ comme l'âge d'une personne ou la durée d'un processus technique), tandis que le noyau utilisé possède un support non borné ou symétrique.

Ce phénomène, également appelé effet de bordure, constitue un sérieux inconvénient lorsque de nombreuses observations se situent près des frontières du support. Plusieurs méthodes ont été développées pour corriger ce problème. Parmi elles, on trouve la méthode de réflexion [26](p.54),[33], l'utilisation d'un noyau asymétrique [13],[32], ou encore les techniques de transformation des données [38],[23]. Ce mémoire présente une analyse de ce phénomène et des principales méthodes de correction.

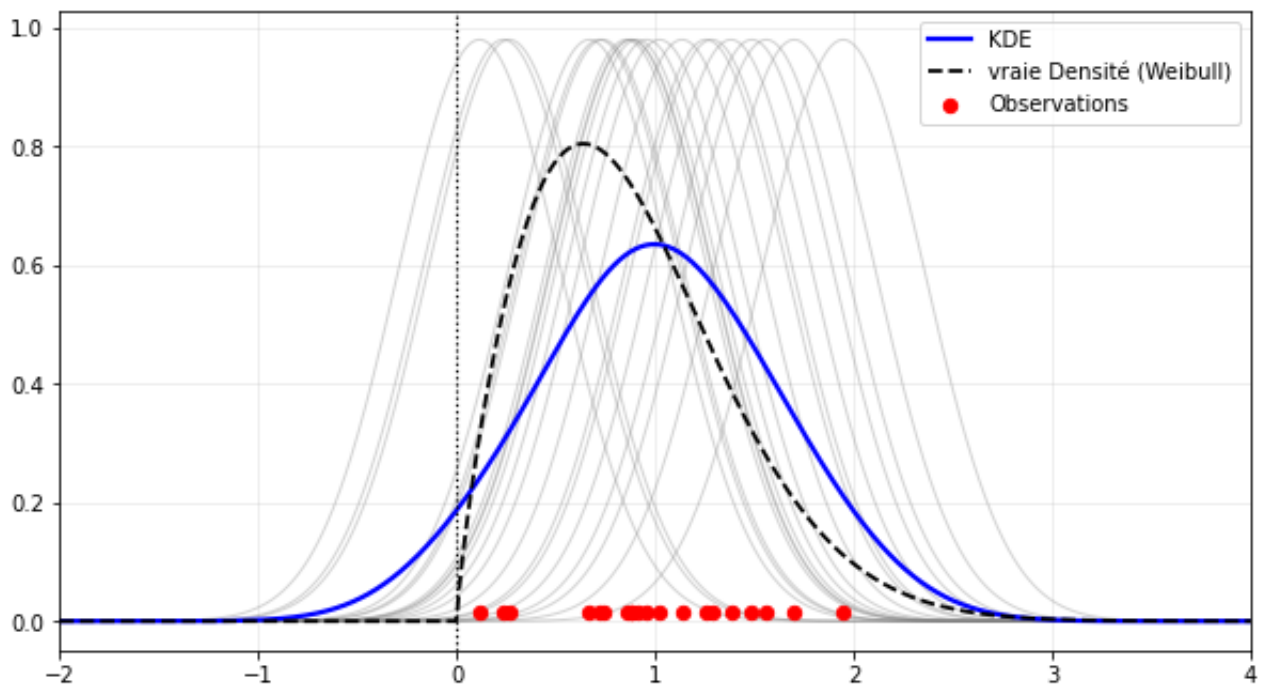


FIGURE 2.9 – Visualisation du problème de biais au bord dans l'estimation par noyau gaussien d'un échantillon de taille $n = 20$ issu d'une loi de Weibull(1.8,1), avec le paramètre de lissage ($b = 0.4069$) calculé par la méthode LSCV.

Définition 2.4 Le problème de bord (ou biais aux limites) survient lorsque l'on souhaite estimer $f(x)$ pour un x proche d'une frontière du support de la distribution (par exemple,

$x \rightarrow 0^+$ pour des données strictement positives). Dans ce contexte, le noyau symétrique attribue une partie de sa masse à une région en dehors du support physique des données, là où par définition $f = 0$. Cela induit une sous-estimation systématique de la densité réelle à proximité de la limite, ce qui constitue un biais d'estimation non négligeable et produit des estimations faussées. (Rappelons qu'un estimateur biaisé ne représente pas fidèlement la population étudiée et peut conduire à des conclusions erronées.)

2.3.5 Noyaux asymétriques

L'estimateur à noyau standard (2.17) a été principalement développé pour des densités ayant des supports non bornés. Cependant, lorsque la fonction de densité des données a un support borné, l'utilisation du noyau classique (symétrique) conduit à un estimateur avec un biais important.

Ce n'est qu'à la fin des années 1990 que les recherches sur les noyaux asymétriques ont commencé. Les premiers travaux se sont concentrés sur la correction des bords pour les estimations de régression et de densité à support compact. Brown et Chen (1999) [10] ont approché les courbes de régression à support dans l'intervalle unitaire $[0,1]$ en utilisant des polynômes de Bernstein lissés par une famille de densités bêta. Dans une extension du lissage par polynômes de Bernstein, Chen (1999, 2000a) [13],[14] propose d'utiliser une famille de densités bêta comme noyaux pour l'estimation de la densité.

L'estimateur de densité de Parzen-Rosenblatt [48],[42] constitue le fondement théorique de l'estimation non paramétrique par noyaux. Sa formulation générale établit un cadre méthodologique unifié pour l'estimation de densités sur différents types de supports, sous réserve d'un choix approprié de la fonction noyau $K(\cdot)$.

Les noyaux asymétriques représentent des réalisations spécifiques de ce cadre théorique, développés pour répondre à des contraintes de support particulières, notamment pour corriger les biais au bord. Ainsi, le cadre de Parzen-Rosenblatt offre une structure formelle flexible, permettant d'intégrer diverses fonctions noyau, chacune optimisée pour des caractéristiques de support spécifiques. Cette adaptabilité méthodologique permet de traiter rigoureusement des problèmes d'estimation sur des supports contraints, où les noyaux symétriques présenteraient des biais systématiques.

Plusieurs noyaux adaptés aux données à support borné ou semi-borné ont été proposés dans la littérature, notamment le noyau Gamma [15], le noyau Gamma généralisé [30], le noyau inverse Gamma [41], le noyau Bêta [13], le noyau inverse Gaussien [50] et le noyau log-normal [31]. Dans le cadre de ce mémoire, notre analyse s'est limitée à l'utilisation du noyau Bêta, de support $[0, 1]$, et du noyau Gamma, de support $[0, \infty)$. Ce choix est justifié par la nature de nos données et les objectifs spécifiques de l'étude, exposés au Chapitre 3.

2.3.6 Noyau Gamma

Chen, Song Xi (2000) [15] étudie l'estimation de fonctions de densité à support $[0, \infty)$ en utilisant des densités gamma comme noyaux. Cette approche vise à remplacer le noyau fixe et symétrique de l'estimateur à noyau classique. L'auteur a mentionné que les noyaux gamma sont non négatifs et présentent une forme naturellement variable. De plus, les estimateurs à noyau gamma sont exempts de biais au bord (boundary bias) et atteignent le taux de convergence optimal pour l'erreur quadratique moyenne intégrée. Enfin, la variance de ces estimateurs, pour un point situé à une distance x de l'origine, est en $O(n^{-4/5}x^{-1/2})$, ce qui indique une décroissance de la variance à mesure que x augmente.

Soient X_1, \dots, X_n un échantillon aléatoire issu d'une distribution admettant une fonction de densité de probabilité inconnue f , à support dans $[0, \infty)$. Nous supposons que f est de classe \mathcal{C}^2 et que $\int_0^\infty f'^2(x)dx$ et $\int_0^\infty [xf''(x)]^2dx$ sont finis. Soit $K_{p,q}$ la fonction de densité d'une variable aléatoire suivant une loi $\text{Gamma}(p, q)$.

Chen (2000) [15] considère deux estimateurs à noyau Gamma pour f . Le premier estimateur utilise $K_{x/b+1,b}$ comme noyau au point $x \in [0, \infty)$, où b est un paramètre de lissage satisfaisant la condition que $b \rightarrow 0$ et $nb \rightarrow \infty$ lorsque $n \rightarrow \infty$, et est défini comme suit, pour $x \in [0, \infty)$:

$$\hat{f}_1(x) = n^{-1} \sum_{i=1}^n K_{x/b+1,b}(X_i) \quad , \quad (2.30)$$

où

$$K_{x/b+1,b}(t) = \frac{t^{x/b} e^{-t/b}}{b^{x/b+1} \Gamma(x/b + 1)} \quad .$$

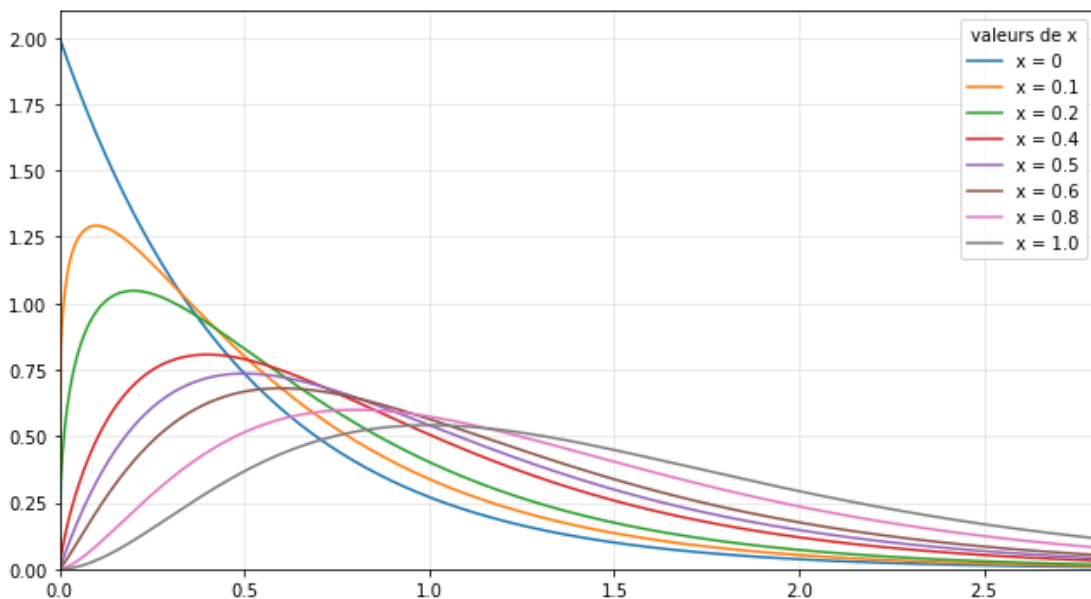


FIGURE 2.10 – Variation de la forme du noyau Gamma ($K_{x/b+1,b}(t)$) en fonction de x , ($b=0.5$).

Le deuxième estimateur à noyau Gamma, tel que défini dans [15], est le suivant :
pour $x \in [0, \infty)$

$$\hat{f}_2(x) = n^{-1} \sum_{i=1}^n K_{\rho_b(x), b}(X_i) \quad , \quad (2.31)$$

où

$$\rho_b(x) = \begin{cases} x/b & \text{si } x \geq 2b; \\ \frac{1}{4}(x/b)^2 + 1 & \text{si } x \in [0, 2b). \end{cases}$$

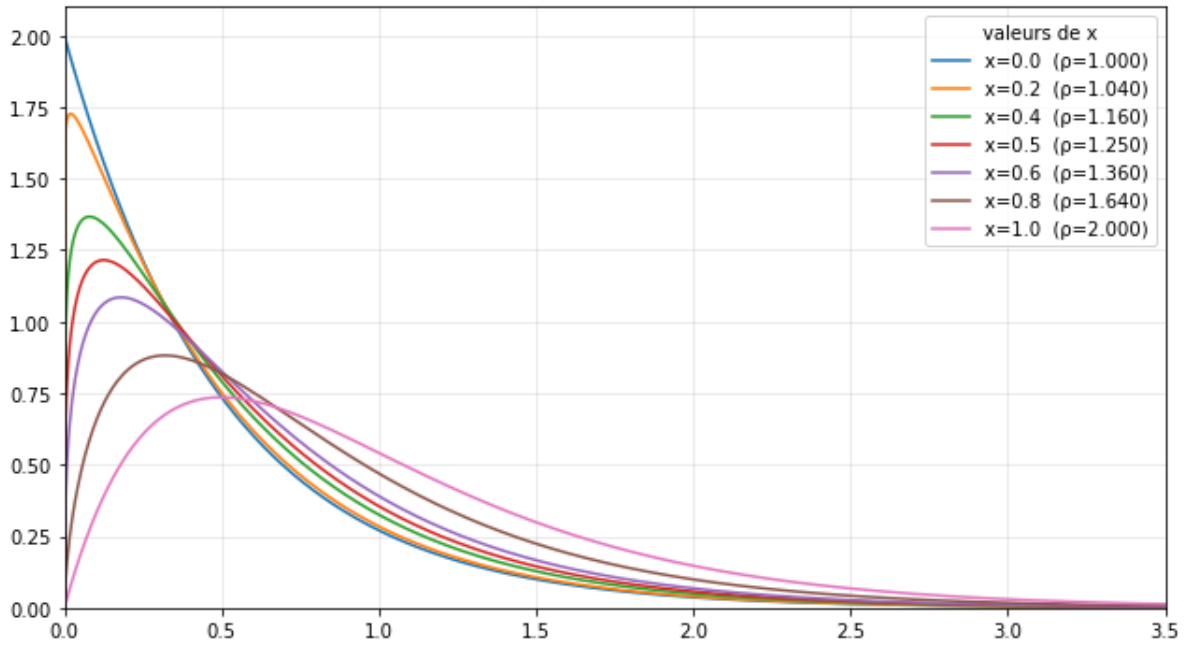


FIGURE 2.11 – Variation de la forme du noyau Gamma ($K_{\rho_b(x), b}(t)$) en fonction de x , ($b=0.5$).

2.3.7 Théorie asymptotique de l'estimateur à noyau Gamma

Biais [15]

Le biais de l'estimateur \hat{f}_1 (2.30) est donné comme suit :

$$\text{Biais} [\hat{f}_1(x)] = b \left(f'(x) + \frac{1}{2} x f''(x) \right) + o(b) \quad . \quad (2.32)$$

et le biais de l'estimateur \hat{f}_2 (2.31) est donné comme suit :

$$\text{Biais} [\hat{f}_2(x)] = \begin{cases} \frac{1}{2} b x f''(x) + o(b) & \text{si } x \geq 2b; \\ b \xi_b(x) f'(x) + o(b) & \text{si } x \in [0, 2b). \end{cases} \quad (2.33)$$

où $\xi_b(x) = (1 - x) \frac{\rho_b(x) - x/b}{1 + b\rho_b(x) - x}$.

Variance [15]

La variance de l'estimateur \hat{f}_1 (2.30) est donné comme suit :

$$V \left[\hat{f}_1(x) \right] \sim \begin{cases} \frac{1}{2\sqrt{\pi}} n^{-1} (bx)^{-1/2} f(x) & \text{si } x/b \rightarrow \infty; \\ \frac{\Gamma(2k+1)}{2^{2k+1} \Gamma^2(k+1)} (nb)^{-1} f(x) & \text{si } x/b \rightarrow k. \end{cases} \quad (2.34)$$

La variance de \hat{f}_2 est similaire, à l'exception du multiplicateur devant $n^{-1}b^{-1}$ dans le cas où $x/b \rightarrow \kappa$, qui présente une forme légèrement différente.

L'erreur quadratique moyenne intégrée (MISE) [15]

L'erreur quadratique moyenne intégrée de $\hat{f}_1(x)$ (2.30) est définie comme suit :

$$\text{MISE} \left[\hat{f}_1 \right] = b^2 \int_0^\infty \left\{ x f'(x) + \frac{1}{2} x f''(x) \right\}^2 dx + \frac{1}{2\sqrt{\pi}} n^{-1} b^{-1/2} \int_0^\infty x^{-1/2} f(x) dx + o(n^{-1} b^{-1/2} + b^2) \quad , \quad (2.35)$$

L'erreur quadratique moyenne intégrée de $\hat{f}_2(x)$ (2.31) est définie comme suit :

$$\text{MISE} \left[\hat{f}_2 \right] = \frac{1}{4} b^2 \int_0^\infty \{ x f''(x) \}^2 dx + \frac{1}{2\sqrt{\pi}} n^{-1} b^{-1/2} \int_0^\infty x^{-1/2} f(x) dx + o(n^{-1} b^{-1/2} + b^2) \quad . \quad (2.36)$$

Paramètre de lissage optimale [15]

Les largeurs de fenêtre optimales (bandwidths) qui minimisent les termes principaux dans (2.35) et (2.36) sont respectivement :

$$b_1^* = \left[\frac{\int_0^\infty x^{-1/2} f(x) dx}{8\sqrt{\pi} \int_0^\infty (x f'(x) + \frac{1}{2} x f''(x))^2 dx} \right]^{2/5} \cdot n^{-2/5} \quad ,$$

et

$$b_2^* = \frac{\left[\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-1/2} f(x) dx \right]^{2/5}}{\left[\int_0^\infty (x f''(x))^2 dx \right]^{2/5}} \cdot n^{-2/5} \quad .$$

En substituant les largeurs de fenêtre optimales ci-dessus, les erreurs quadratiques moyennes intégrées optimales sont :

$$\text{MISE}^* \left[\hat{f}_1 \right] = \frac{5}{4^{4/5}} \left[\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-1/2} f(x) dx \right]^{4/5} \cdot \left[\int_0^\infty \left\{ x f'(x) + \frac{1}{2} x f''(x) \right\}^2 dx \right]^{1/5} \cdot n^{-4/5} \quad ,$$

et

$$\text{MISE}^* \left[\hat{f}_2 \right] = \frac{5}{4^{4/5}} \left[\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-1/2} f(x) dx \right]^{4/5} \cdot \left[\int_0^\infty \{ x f''(x) \}^2 dx \right]^{1/5} \cdot n^{-4/5} \quad .$$

Il peut être démontré que pour toute fonction de densité f , si $\int_0^\infty (f'(x))^2 dx$ et $\int_0^\infty (f''(x))^2 dx$ sont toutes deux finies, alors $\text{MISE}^*[\hat{f}_1] \geq \text{MISE}^*[\hat{f}_2]$ et $b_1^* \geq b_2^*$. Par conséquent, \hat{f}_2 devrait avoir une meilleure performance globale et utiliser une largeur de fenêtre plus petite que \hat{f}_1 .

2.3.8 Noyau Bêta

L'idée du lissage par le noyau bêta a été considérée pour la première fois par Brown et Chen (1999) [10] afin de remédier au biais de bord dans la régression non paramétrique, sans diviser le support compact en zones intérieures, et sans appliquer explicitement des traitements de lissage distincts de manière séparée.

Plus tard, Chen (1999) [13] a proposé un estimateur de densité à noyau bêta pour des densités définies sur l'intervalle $[0, 1]$. Cet estimateur présente deux avantages spécifiques. Premièrement, son support correspond exactement à celui de la densité à estimer, ce qui permet de conserver toute l'information de l'échantillon. Deuxièmement, la forme du noyau varie automatiquement en fonction de la position x où la densité est estimée, garantissant un bon comportement aux frontières du support.

Contrairement aux noyaux symétriques classiques (gaussien, épanechnikov, etc.) dont la forme reste fixe sur tout le domaine, le noyau bêta de Chen (1999) [13] adapte sa forme en fonction du point d'estimation x . Plus précisément, à l'intérieur de l'intervalle $(0, 1)$, le noyau reste à peu près symétrique, tandis qu'aux voisinages des bornes 0 ou 1, il devient progressivement asymétrique, éliminant ainsi le biais aux frontières. Cette variation de la forme du noyau modifie effectivement le degré de lissage appliqué par les estimateurs à noyau bêta. Ce schéma d'adaptation se distingue du schéma adaptatif classique, où un degré de lissage différent est obtenu en modifiant uniquement la valeur de la largeur de bande b [13].

L'estimateur de densité à noyau bêta est exempt de biais de bord, toujours non négatif, et atteint le taux de convergence optimal ($n^{-\frac{4}{5}}$) pour l'erreur quadratique moyenne intégrée (MISE) au sein de la classe des estimateurs de densité à noyau non négatifs [13].

Soit X_1, \dots, X_n un échantillon aléatoire issu d'une distribution avec une fonction de densité de probabilité inconnue f , qui a un support compact $[0, 1]$. Nous supposons que f possède une dérivée seconde continue (C^2). Soit $K_{p,q}$ la fonction de densité d'une variable aléatoire suivant une loi Beta(p, q).

Song Xi Chen. (1999) [13] considère deux estimateurs à noyau bêta pour f . Le premier estimateur à noyau bêta utilise $K_{\frac{x}{b}+1, \frac{1-x}{b}+1}$ comme noyau au point $x \in [0, 1]$, où b est un paramètre de lissage satisfaisant la condition que $b \rightarrow 0$ lorsque $n \rightarrow \infty$, et est défini comme suit, pour $x \in [0, 1]$:

$$\hat{f}_1(x) = n^{-1} \sum_{i=1}^n K_{x,b}(X_i) \quad , \quad (2.37)$$

où

$$K_{x,b}(t) = K_{\frac{x}{b}+1, \frac{1-x}{b}+1}(t) = \frac{t^{\frac{x}{b}}(1-t)^{\frac{1-x}{b}}}{B(x/b+1, (1-x)/b+1)} \cdot \mathbb{I}_{0 < t < 1} \quad . \quad (2.38)$$

La fonction $K_{\frac{x}{b}+1, \frac{1-x}{b}+1}(t)$ (figure 2.12) est symétrique en $x = 0,5$ et devient asymétrique lorsque x se rapproche des bornes 0 ou 1.

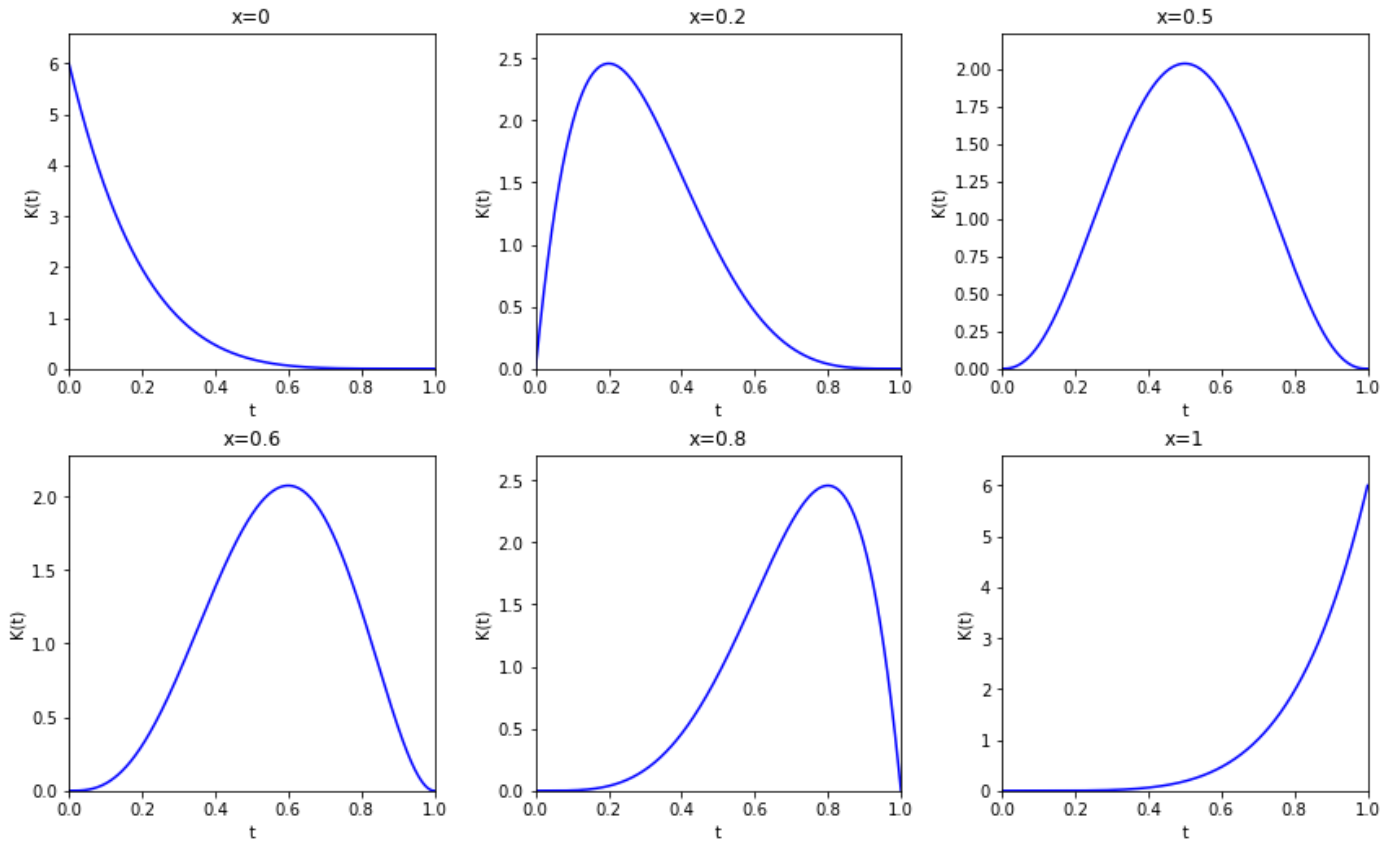


FIGURE 2.12 – Variation de la forme du noyau Beta ($K_{x,b}(t)$) en fonction de x , ($b=0.2$)

Le deuxième estimateur à noyau bêta, tel que défini dans [13], est le suivant : pour $x \in [0, 1]$

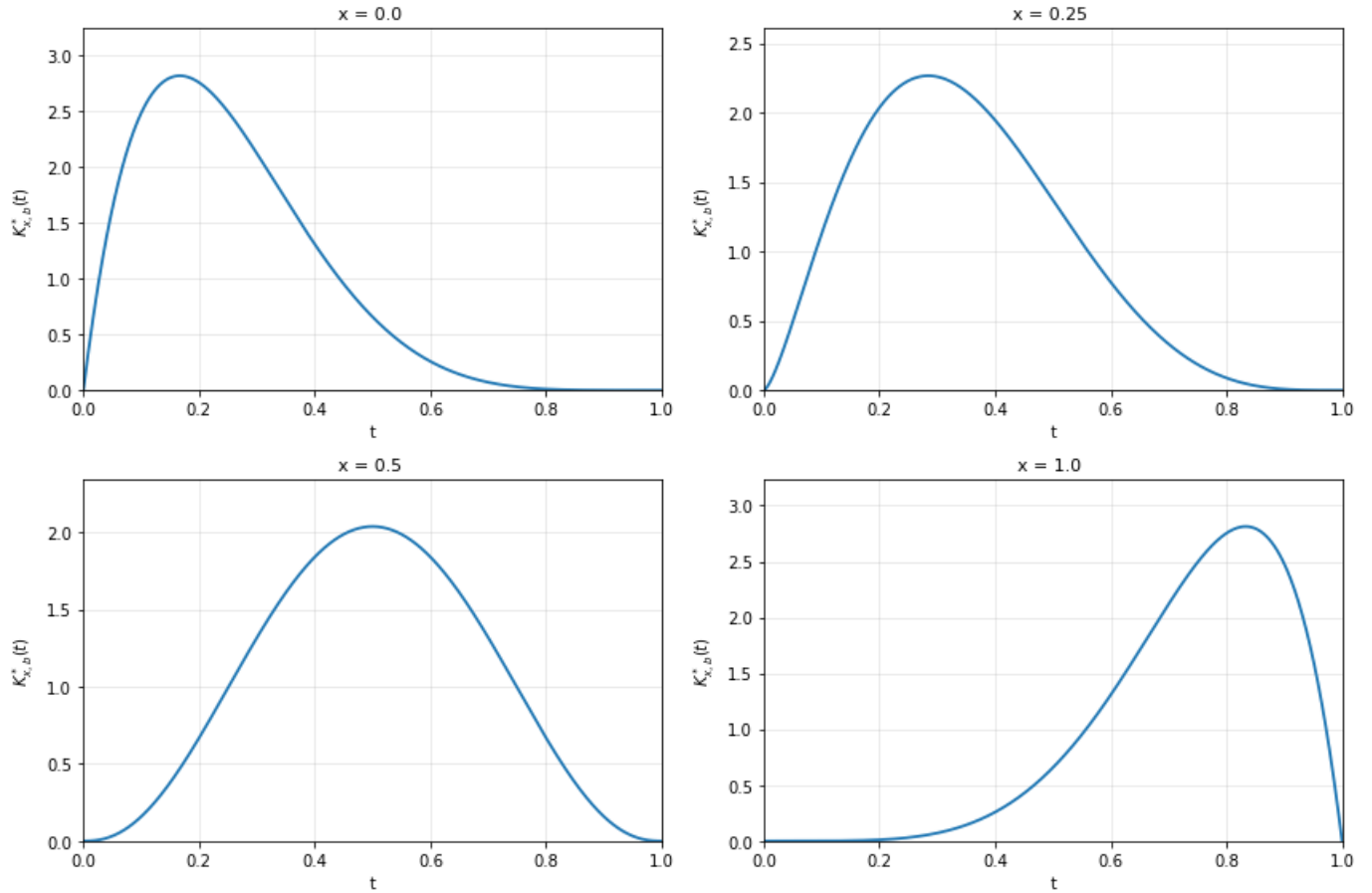
$$\hat{f}_2(x) = n^{-1} \sum_{i=1}^n K_{x,b}^*(X_i) \quad , \quad (2.39)$$

où $K_{x,b}^*$ sont des noyaux bêta de bord, défini comme suit :

$$K_{x,b}^*(u) = \begin{cases} K_{x/b, (1-x)/b}(u) & \text{si } x \in [2b, 1 - 2b], \\ K_{\rho(x), (1-x)/b}(u) & \text{si } x \in [0, 2b], \\ K_{x/b, \rho(1-x)}(u) & \text{si } x \in (1 - 2b, 1], \end{cases}$$

où $\rho(x, b) = 2b^2 + 2.5 - \sqrt{4b^4 + 6b^2 + 2.25 - x^2 - x/b}$.

Song Xi Chen (1999) [13] a remarqué que, pour tout b fixé, la fonction $\rho(x, b)$ est strictement croissante par rapport à x sur l'intervalle $[0, 2b]$. De plus, il souligne que la forme du noyau varie en fonction de la valeur de x , comme illustré par l'ensemble des noyaux bêta présentés dans la figure 2.13, Cette variation de la forme du noyau modifie en fait le niveau de lissage appliqué par les estimateurs à noyau bêta.

FIGURE 2.13 – Variation de la forme du noyau Beta ($K_{x,b}^*$) en fonction de x , ($b=0.2$)

2.3.9 Théorie asymptotique de l'estimateur à noyau Beta

Biais [13]

On a :

$$\begin{aligned}
 \mathbb{E}[\hat{f}_1(x)] &= \mathbb{E}\left[n^{-1} \sum_{i=1}^n K_{\frac{x}{b}+1, \frac{(1-x)}{b}+1}(X_i)\right] \\
 &= n^{-1} \sum_{i=1}^n \mathbb{E}\left[K_{\frac{x}{b}+1, \frac{(1-x)}{b}+1}(X_i)\right] \\
 &= \int_0^1 K_{\frac{x}{b}+1, \frac{(1-x)}{b}+1}(y) f(y) dy \quad .
 \end{aligned}$$

Alors le biais de l'estimateur \hat{f}_1 (2.37) est donné comme suit :

$$\text{Biais}[\hat{f}_1(x)] = \left((1-2x)f'(x) + \frac{1}{2}x(1-x)f''(x) \right) b + o(b) \quad . \quad (2.40)$$

Ce biais est de $\mathcal{O}(b)$ sur le tout l'intervalle $[0,1]$, et cela indique que \hat{f}_1 n'a pas de biais au limites [13].

De la même méthode, le biais de l'estimateur \hat{f}_2 (2.39) est donné comme suit :

$$\text{Biais} \left[\hat{f}_2(x) \right] = \begin{cases} \frac{1}{2}x(1-x)f''(x)b + \mathcal{O}(b^2) & \text{si } x \in [2b, 1-2b], \\ \xi(x)bf'(x) + o(b) & \text{si } x \in [0, 2b), \\ -\xi(1-x)bf'(x) + o(b) & \text{si } x \in (1-2b, 1], \end{cases} \quad (2.41)$$

$$\text{Où } \xi(x) = \frac{(1-x)(\rho(x) - \frac{x}{b})}{(1+b\rho(x) - x)} .$$

Variance [13]

La variance de l'estimateur \hat{f}_1 (2.37) est donné comme suit :

$$V \left[\hat{f}_1(x) \right] = \begin{cases} \frac{n^{-1}b^{-1/2}}{2\sqrt{\pi x(1-x)}} [f(x) + \mathcal{O}(n^{-1})] & \text{si } \frac{x}{b} \rightarrow \infty \text{ et } \frac{1-x}{b} \rightarrow \infty, \\ \frac{\Gamma(2c+1)}{2^{2c+1}\Gamma^2(c+1)} n^{-1}b^{-1} [f(x) + \mathcal{O}(n^{-1})] & \text{si } \frac{x}{b} \rightarrow c \text{ ou } \frac{1-x}{b} \rightarrow c. \end{cases}$$

Où c est une constante $\in \mathbb{R}^+$, et $\Gamma(\cdot)$ est la fonction gamma définie comme suit :

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

La variance de l'estimateur \hat{f}_2 (2.39) est donné comme suit :

$$V \left[\hat{f}_2(x) \right] = \begin{cases} \frac{n^{-1}b^{-1/2}}{2\sqrt{\pi x(1-x)}} [f(x) + \mathcal{O}(n^{-1})] & \text{si } \frac{x}{b} \rightarrow \infty \text{ et } \frac{1-x}{b} \rightarrow \infty \\ g(c) \cdot n^{-1}b^{-1} [f(x) + \mathcal{O}(n^{-1})] & \text{si } \frac{x}{b} \rightarrow c \text{ ou } \frac{1-x}{b} \rightarrow c \end{cases}$$

Où $g(c)$ est une constante positive dépend de c , ($g(c) \in \mathbb{R}^+$).

L'erreur quadratique moyenne intégrée (MISE) [13]

L'erreur quadratique moyenne intégrée de \hat{f}_1 (2.37) est définie comme suit :

$$\begin{aligned} \text{MISE} \left[\hat{f}_1 \right] &= b^2 \int_0^1 \left((1-2x)f'(x) + \frac{1}{2}x(1-x)f''(x) \right)^2 dx \\ &+ \frac{1}{2\sqrt{\pi}} n^{-1}b^{-1/2} \int_0^1 [x(1-x)]^{-1/2} f(x) dx + o(n^{-1}b^{-1/2} + b^2) . \end{aligned} \quad (2.42)$$

L'erreur quadratique moyenne intégrée de $\hat{f}_2(x)$ (2.39) est définie comme suit :

$$\begin{aligned} \text{MISE} \left[\hat{f}_2 \right] &= \frac{1}{4} b^2 \int_0^1 (x(1-x)f''(x))^2 dx \\ &+ \frac{1}{2\sqrt{\pi}} n^{-1} b^{-1/2} \int_0^1 [x(1-x)]^{-1/2} f(x) dx + o(n^{-1} b^{-1/2} + b^2) \quad . \end{aligned} \quad (2.43)$$

Paramètre de lissage optimale [13]

On note par b_1^* et b_2^* les paramètres de lissage qui minimisent (2.42) et (2.43) respectivement et on donne comme suit :

$$b_1^* = \left[\frac{\frac{1}{2\sqrt{\pi}} \int_0^1 [x(1-x)]^{-1/2} f(x) dx}{4 \int_0^1 \left[(1-2x)f'(x) + \frac{1}{2}x(1-x)f''(x) \right]^2 dx} \right]^{2/5} \cdot n^{-2/5} \quad , \quad (2.44)$$

et

$$b_2^* = \left[\frac{\int_0^1 (x(1-x))^{-1/2} f(x) dx}{2\sqrt{\pi} \int_0^1 [x(1-x)f''(x)]^2 dx} \right]^{2/5} \cdot n^{-2/5} \quad . \quad (2.45)$$

En substituant les largeurs de fenêtre optimales ci-dessus, les erreurs quadratiques moyennes intégrées optimales sont :

$$\begin{aligned} \text{MISE}^* \left[\hat{f}_1 \right] &= \frac{5}{4^{4/5}} \left(\frac{1}{2\sqrt{\pi}} \int_0^1 \frac{f(x)}{\{x(1-x)\}^{1/2}} dx \right)^{4/5} \\ &\times \left[\int_0^1 \left\{ (1-2x)f'(x) + \frac{1}{2}x(1-x)f''(x) \right\}^2 dx \right]^{1/5} n^{-4/5} \end{aligned}$$

$$\begin{aligned} \text{MISE}^* \left[\hat{f}_2 \right] &= \frac{5}{4^{3/5}} \left(\frac{1}{2\sqrt{\pi}} \int_0^1 \{x(1-x)\}^{-1/2} f(x) dx \right)^{4/5} \\ &\times \left[\int_0^1 \{x(1-x)f''(x)\}^2 dx \right]^{1/5} n^{-4/5} . \end{aligned}$$

Il peut être démontré que pour toute fonction de densité f , si à la fois $\int (f'(x))^2 dx$ et $\int (f''(x))^2 dx$ sont finis, alors : $\text{MISE}^* \left[\hat{f}_1 \right] \geq \text{MISE}^* \left[\hat{f}_2 \right]$ et $b_1^* \geq b_2^*$. Par conséquent, \hat{f}_2 devrait offrir une meilleure performance globale et utiliser une bande passante plus petite que \hat{f}_1 , et est donc recommandée.

2.3.10 Transformation des données en estimation de densité de probabilité

En inférence statistique, et plus particulièrement en estimation non paramétrique de la densité, les transformations de données constituent une approche préalable fondamentale

visant à adapter le support et la structure des observations aux hypothèses sous-jacentes des méthodes d'estimation. L'objectif principal est de remédier aux limitations induites par des supports bornés ou semi-bornés (e.g., $[0, 1]$, \mathbb{R}^+), une asymétrie (skewness) marquée ou une forte multimodalité (plusieurs maxima locaux de la fonction de densité), qui peuvent dégrader significativement la performance des estimateurs à noyau.

La technique de transformation repose sur le principe du changement de variable aléatoire. Son objectif est de projeter les données d'un support \mathcal{S}_X où le biais aux limites pose problème vers un support \mathcal{S}_Y sur lequel on peut appliquer un estimateur non paramétrique (par exemple à noyau) exempt de biais de frontière. Pour retrouver l'estimation de la densité originale $\hat{f}_X(x)$, on applique ensuite la transformation inverse. D'après le théorème de changement de variable en théorie des probabilités, la densité de X s'exprime à partir de celle de Y par la valeur absolue de la dérivée de la transformation (le terme jacobien).

Formellement, soit X une variable aléatoire de densité f_X inconnue, de support $\mathcal{S}_X \subseteq \mathbb{R}$, et $\{X_i\}_{i=1}^n$ un échantillon i.i.d. L'idée est de construire une application bijective et différentiable $T : \mathcal{S}_X \rightarrow \mathcal{S}_Y$, où \mathcal{S}_Y est un support plus régulier (typiquement \mathbb{R} tout entier ou $[0, 1]$), afin d'estimer plus efficacement la densité f_Y de $Y = T(X)$. L'estimateur de la densité originale est ensuite obtenu par la formule du changement de variable :

$$\hat{f}_X(x) = \hat{f}_Y(T(x)) \cdot |T'(x)| \quad ,$$

où \hat{f}_Y est un estimateur non paramétrique (par exemple à noyau) construit sur les données transformées $\{Y_i = T(X_i)\}$, et $|T'(x)|$ représente le jacobien de la transformation.

Cette section présente deux transformations classiques utilisées en statistique ainsi que leurs domaines d'application privilégiés dans le cadre de l'estimation de densité.

● Transformations par fonctions de puissance

Cette famille de transformations, monotones et non linéaires, vise principalement à atténuer l'asymétrie des distributions et à stabiliser la variance.

● Transformations de Box–Cox

Introduite par Box et Cox (1964) [9], cette famille généralise les transformations de puissance par un paramètre λ , conçue pour stabiliser la variance et rendre la distribution des données plus proche d'une loi normale :

$$T_\lambda(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \log(x), & \lambda = 0, \end{cases} \quad x > 0.$$

Le paramètre λ est estimé par maximisation de la vraisemblance (sous l'hypothèse de normalité des données transformées) ou par validation croisée.

Pour étendre le domaine d'application aux réels quelconques, Yeo et Johnson (2000) [63] ont proposé une famille définie pour $x \in \mathbb{R}$ et paramétrée par λ . Elle préserve la continuité et la différentiabilité en $x = 0$ et permet de traiter simultanément des données positives et négatives :

$$T_\lambda(x) = \begin{cases} \frac{(x+1)^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0, x \geq 0 \\ \log(x+1) & \text{si } \lambda = 0, x \geq 0 \\ -\frac{(1-x)^{2-\lambda} - 1}{2-\lambda} & \text{si } \lambda \neq 2, x < 0 \\ -\log(1-x) & \text{si } \lambda = 2, x < 0 \end{cases}$$

Formellement, la transformation $T_\lambda(x)$ est définie par une fonction définie par morceaux qui assure la continuité et la convexité à l'origine.

Si $x \geq 0$, la transformation ressemble à une Box-Cox en $(x+1)$.

• Transformations par fonction de répartition paramétrique

La famille des transformations par fonction de répartition paramétrique s'inscrit dans un cadre explicitement probabiliste. Elle consiste à transformer les données à l'aide de la fonction de répartition (F_θ) d'une loi paramétrique donnée, de sorte que la transformation ($T(x) = F_\theta(x)$) projette les observations sur l'intervalle borné $[0, 1]$. Cette approche repose sur un résultat fondamental de la théorie des probabilités selon lequel, si $(X \sim F_\theta)$, alors la variable transformée ($Y = F_\theta(X)$) suit une loi uniforme sur $[0, 1]$. Cette propriété confère à ces transformations un rôle central dans la correction du biais aux limites en estimation de densité, en particulier lorsque le support des données est borné ou semi-borné.

L'efficacité de cette transformation dépend crucialement de l'adéquation du modèle paramétrique choisi, c'est-à-dire de l'estimation précise des paramètres θ qui caractérisent la fonction de répartition F_θ . Cette étape d'estimation constitue le fondement pratique de la méthode. Elle s'appuie généralement sur les données observées (x_1, x_2, \dots, x_n) . L'estimateur du maximum de vraisemblance est fréquemment employé, où $\hat{\theta}$ est choisi pour maximiser la vraisemblance $\prod_{i=1}^n f_\theta(x_i)$, avec f_θ la densité de probabilité associée à F_θ . D'autres méthodes, comme la méthode des moments ou des approches bayésiennes, peuvent également être mobilisées, en particulier si la maximisation de la vraisemblance s'avère complexe.

Une fois le paramètre θ estimé par $\hat{\theta}$, la transformation appliquée aux données devient $T(x) = F_{\hat{\theta}}(x)$. Il est important de noter que si la famille paramétrique choisie est mal spécifiée, la transformation ne produira pas une uniformité stricte (la variable transformée ne suivra pas exactement une loi uniforme), ce qui peut affaiblir son utilité pour la correction de biais. Le choix de la loi initiale (log-normale, GPD, Champernowne, etc.) doit donc être guidé par la nature du support des données et la forme de leur distribution empirique.

Plusieurs transformations ont été proposées dans la littérature. Parmi les approches notables, la transformation par la fonction de répartition (cdf) de Champernowne par Buch-Larsen et al. (2005) [11] (chapitre 3) et la transformation inverse Beta-Champernowne de Bolancé, Guillén et Nielsen (2008) [7]. Ils ont démontré, via des études de simulation, la capacité de leurs estimateurs à réduire l'erreur entre la vraie densité et la densité estimée.

Pour plus de documentation sur l'estimation à noyau par transformation, on peut se référer à Gustafsson et al. (2009) [28] qui ont traité l'estimation des distributions de pertes actuarielles à noyau asymétrique par transformation de Champernowne généralisée. De même, Marron et Ruppert (1994) [38] abordent les transformations pour réduire le biais de frontière. D'autres contributions notables incluent la transformation optimale inverse Beta(3,3) de Bolancé (2010) [5], l'estimateur amélioré pour les densités sur l'intervalle unitaire de Wen et Wu (2015) [62], ainsi que la réduction du biais par transformations empiriques lissées de Ruppert et Cline (1994) [49].

Algorithme d'estimation à noyau par transformation

L'estimation de densité par noyau avec transformation se déroule en quatre étapes :

1. Estimer les paramètres de la fonction de transformation ($T_\theta(x)$) pour ajuster au mieux les données à cette dernière.
2. Transformer les données par la fonction $T_{\hat{\theta}}(x)$.
3. Estimer la densité des données transformées par la méthode du noyau.
4. Effectuer une rétrotransformation après l'estimation à noyau pour revenir à l'échelle originale des données.

Exemple d'estimation à noyau par transformation

Nous avons considéré deux échantillons de taille 1000 chacun, tirés respectivement d'une loi de Burr(1,1,1) et d'une loi inverse-gaussienne(2,1). Deux approches ont été comparées :

- Approche (A) :
Estimation par noyau gamma (asymétrique).
- Approche (B) :
Transformation de Yeo et Johnson, suivie de l'utilisation d'un noyau gamma .

Les paramètres de lissage optimaux ont été sélectionnés via la méthode de validation croisée par moindres carrés (LSCV). Les simulations ont été réalisées sur 100 répliques, et nous avons retenu la moyenne des paramètres de lissage ainsi que celle des densités estimées. Ces dernières ont ensuite été tracées et comparées graphiquement à la densité théorique.

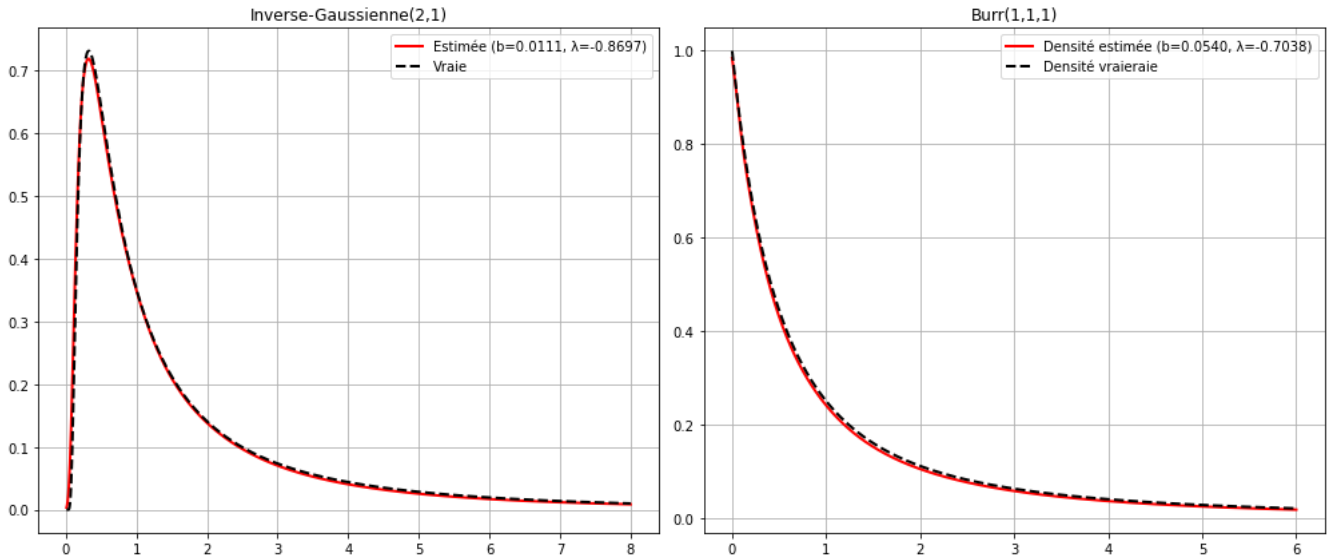


FIGURE 2.14 – Estimation des densités des lois de Burr(1,1,1) et d'inverse-gaussienne(2,1) par l'approche (B).

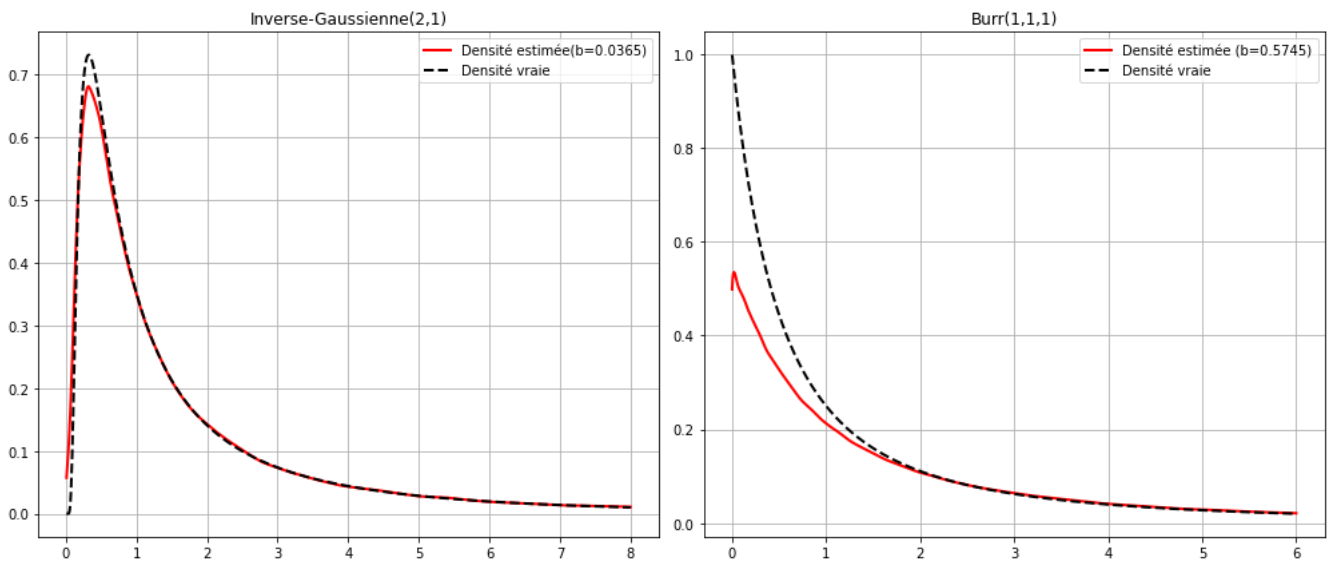


FIGURE 2.15 – Estimation des densités des lois de Burr(1,1,1) et d'inverse-gaussienne(2,1) par l'approche (A).

L'analyse comparative des performances d'estimation, présentées dans les figures 2.14 et 2.15, met en évidence la supériorité empirique de l'approche intégrant la transformation de Yeo-Johnson (Approche B) par rapport à l'utilisation directe du noyau gamma asymétrique (Approche A). Alors que la figure 2.14 démontre une convergence quasi parfaite entre les densités estimées et théoriques pour les lois de Burr et Inverse-Gaussienne, la figure 2.15 révèle les limites structurelles de l'estimateur direct par noyau gamma. En effet, bien que l'Approche A fournisse une approximation acceptable pour la loi Inverse-Gaussienne(2,1) qui est à queue légère, elle présente une sous-estimation marquée de la densité au voisinage de l'origine dans le cas de la loi de Burr(1,1,1) à queue lourde.

Ces observations valident ainsi la capacité de la transformation à stabiliser les données et à reproduire fidèlement la forme des distributions.

Conclusion

Dans ce chapitre, nous avons présenté et comparé les principales méthodes d'estimation non paramétrique de la densité : l'histogramme (classique et mobile), l'estimateur de Parzen-Rosenblatt, ainsi que les estimateurs à noyaux spécifiquement conçus pour des supports bornés ou semi-bornés (noyaux bêta et noyaux gamma). Nous avons exposé à la fois les aspects pratiques, tels que l'implémentation et le choix de la largeur de bande, et les propriétés asymptotiques fondamentales (biais, variance et erreur quadratique moyenne intégrée (MISE)) qui justifient les performances observées en pratique.

Nous avons accordé une attention particulière au problème du biais aux frontières, un défi récurrent lorsque le support des données est limité (par exemple $[0, 1]$ ou \mathbb{R}^+). Deux approches majeures ont été étudiées : (i) l'utilisation de noyaux adaptés au bord, tels que les noyaux bêta et gamma, et (ii) la transformation préalable des données (Box-Cox, Yeo et Johnson, transformées par CDF paramétrique), suivie d'une rétro-transformation incluant le Jacobien pour garantir la validité de la densité finale.

En conclusion, l'estimation à noyau constitue un outil flexible et puissant pour approcher une densité inconnue sans imposer d'hypothèses paramétriques fortes. Son efficacité repose essentiellement sur un diagnostic précis du support, un choix judicieux du paramètre de lissage, et une correction adéquate du biais aux frontières dans le cas de domaines bornés. Les développements récents, notamment les transformations par CDF et les noyaux asymétriques, ouvrent des perspectives prometteuses pour améliorer la précision et la robustesse des estimateurs dans des contextes applicatifs réels tels que l'assurance, la finance ou l'hydrologie.

Chapitre 3

Estimateur KMCE (Kernel Modified Champernowne Estimator)

Il est bien connu dans la littérature que, lorsque le domaine de la distribution étudiée n'est pas $(-\infty, +\infty)$, l'estimateur de densité par noyau classique, malgré ses propriétés avantageuses, souffre d'un problème de biais aux frontières (boundary bias).

Dans ce chapitre, nous recourons à un estimateur semi-paramétrique de densité par transformation pour remédier à ce défaut. La première étude fondatrice de cette approche est attribuée à Wand, Marron et Ruppert (1991) [60], qui ont étudié l'estimation de densité de probabilité via la transformation des données initiales, ainsi que l'effet du paramètre de lissage variable et global sur l'estimation finale.

Buch-Larsen et al. (2005) [11] proposent une approche unifiée pour l'estimation des distributions de pertes en assurance, adaptée aux situations à queues lourdes. L'estimateur proposé est obtenu en transformant les données à l'aide d'un modèle paramétrique, en l'occurrence la distribution de Champernowne modifiée (généralisée). Ensuite, la densité des données transformées est estimée à l'aide de l'estimateur de densité par noyau classique :

$$\hat{f}_X(x) = \frac{1}{nb} \sum_{i=1}^n K\left(\frac{x - X_i}{b}\right) . \quad (3.1)$$

L'estimateur de densité des données originales est ensuite obtenu par rétro-transformation de $\hat{f}_X(x)$. Cette méthode constitue une procédure d'estimation semi-paramétrique, car elle combine une transformation paramétrique et une estimation non paramétrique appliquées à un seul échantillon de données. Les auteurs ont démontré, à travers des études de simulation,

que cette approche est préférable aux méthodes d'estimation de densité existantes pour les distributions à queues lourdes (comme c'est typiquement le cas pour les revenus ou les montants des sinistres en assurance).

Buch-Larsen et al. (2005) [11] mentionnent qu'ils n'ont pas choisi de travailler avec les extensions classiques de la distribution de Pareto telles que la distribution Pareto généralisée (GPD). La raison en est que la GPD estime souvent que des distributions à support infini ont un support fini, et ne peut donc pas être utilisée comme transformation. De plus, la GPD nécessite un seuil (difficile à déterminer) à partir duquel la distribution commence, ce qui pose également des problèmes.

3.1 Distribution de Champernowne

3.1.1 Distribution de Champernowne originale

La distribution de Champernowne originale a une densité [11] :

$$f(x) = \frac{c}{x((1/2)(x/\beta)^{-\alpha} + \lambda + (1/2)(x/\beta)^{-\alpha})} , \quad x \geq 0 \quad (3.2)$$

où c est une constante de normalisation et α, λ et β sont des paramètres.

Lorsque λ est égal à 1 et que la constante de normalisation c est égale à $(1/2)\alpha$, la densité de la distribution originale est simplement appelée la distribution de Champernowne.

$$f(x) = \frac{\alpha\beta^\alpha x^{\alpha-1}}{(x^\alpha + \beta^\alpha)^2} , \quad x \geq 0$$

et sa fonction de distribution cumulative (cdf) est donnée comme suit :

$$F(x) = \frac{x^\alpha}{x^\alpha + \beta^\alpha} , \quad x \geq 0$$

3.1.2 Distribution de Champernowne modifiée

d'après Buch-Larsen et al. (2005) [11], la transformation des données avec la fonction de répartition de Champernowne standard entraîne des problèmes de bordure en raison de sa forme rigide près de 0. Pour résoudre cet inconvénient, ils ont utilisé la distribution en y ajoutant un paramètre supplémentaire c .

La fonction de distribution de Champernowne modifiée (cdf) est donnée comme suit :

$$F_{\alpha,\beta,c}(x) = \frac{(x+c)^\alpha - c^\alpha}{(x+c)^\alpha + (\beta+c)^\alpha - 2c^\alpha} \quad \forall x \in \mathbb{R}_+ , \text{ avec } \alpha > 0, \beta > 0 \text{ et } c \geq 0 ,$$

et sa fonction de densité (pdf) est donnée comme suit :

$$f_{\alpha,\beta,c}(x) = \frac{\alpha(x+c)^{\alpha-1}((\beta+c)^\alpha - c^\alpha)}{((x+c)^\alpha + (\beta+c)^\alpha - 2c^\alpha)^2} \quad \forall x \in \mathbb{R}_+ \quad (3.3)$$

Propriété. 3.1

la distribution de Champernowne modifiée converge vers une distribution de Pareto dans la partie extrême ,ce qui fait que cette distribution est un choix intéressant pour modéliser les distributions à queues lourdes.

$$\lim_{x \rightarrow \infty} f_{\alpha,\beta,c} = \frac{(((\beta+c)^\alpha - c^\alpha)^{1/\alpha})^\alpha}{x^{\alpha+1}} .$$

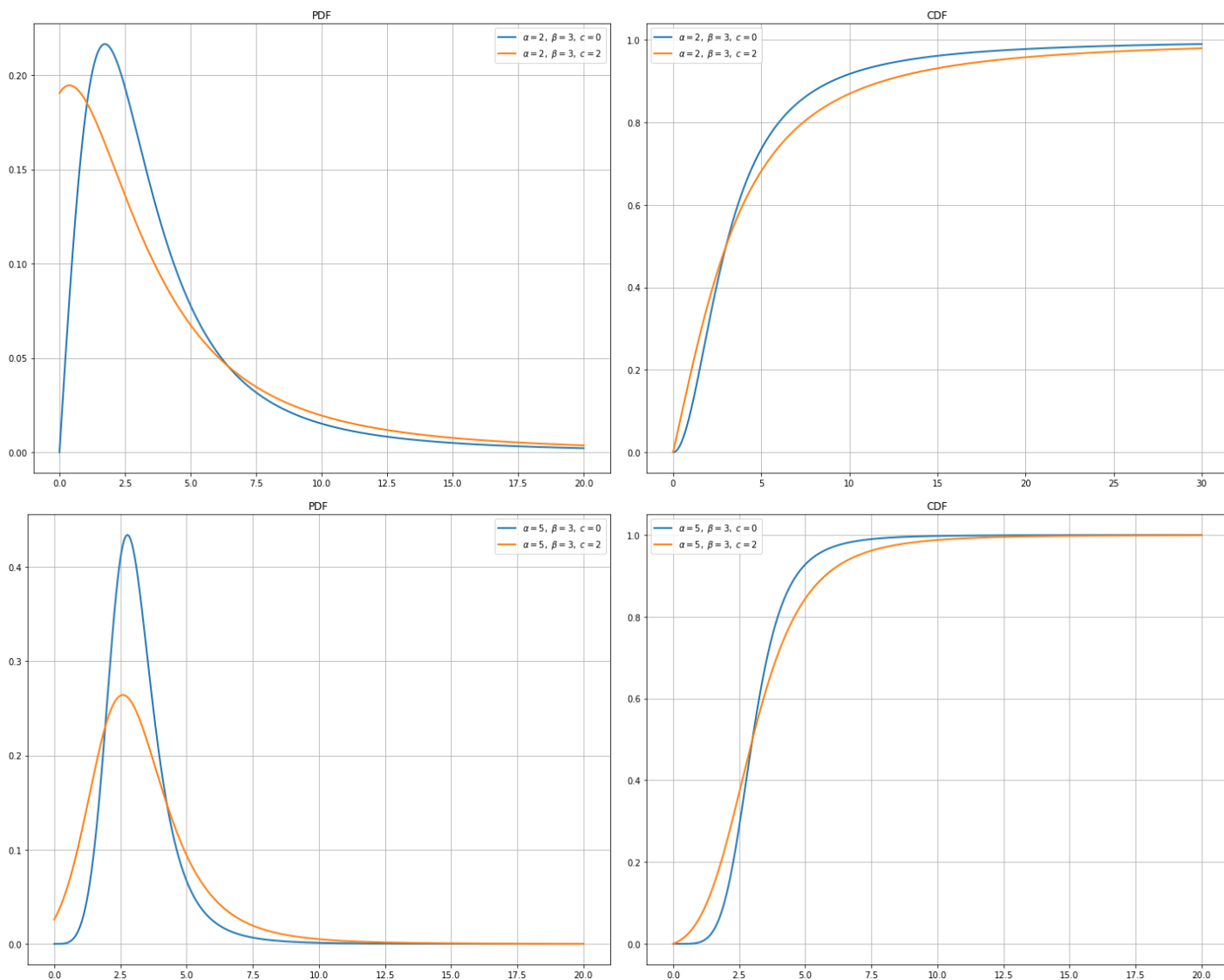


FIGURE 3.1 – Fonctions de répartition (CDF) et de densité (PDF) de la distribution de Champernowne en fonction des paramètres

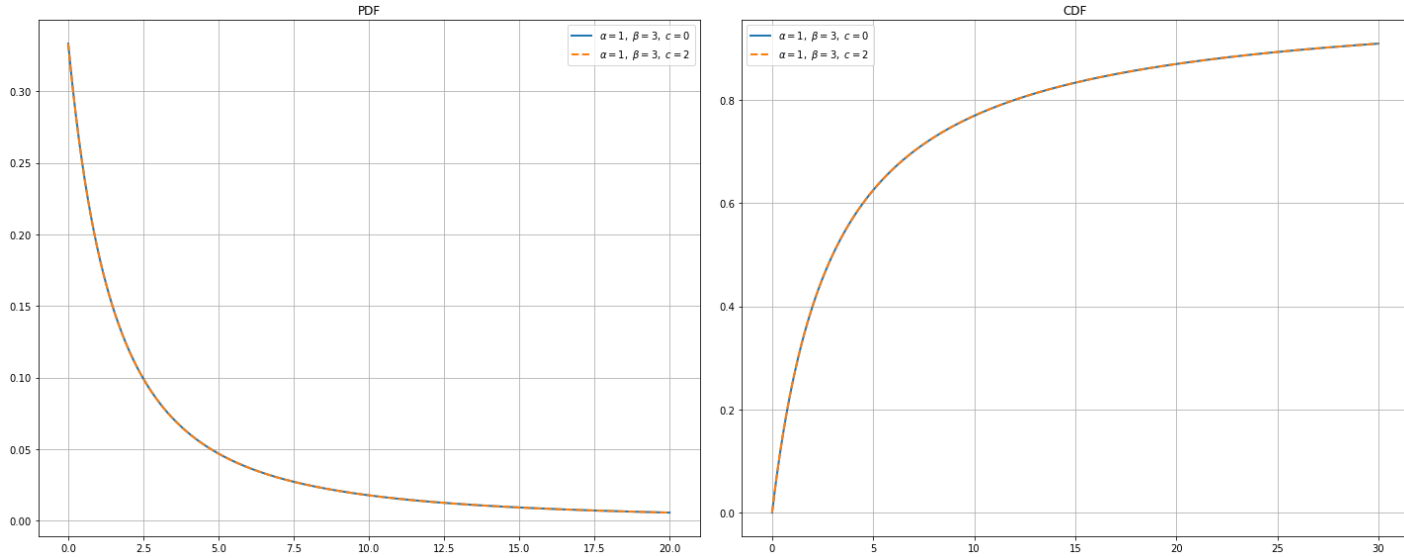


FIGURE 3.2 – Fonctions de répartition (CDF) et de densité (PDF) de la distribution de Champernowne en fonction des paramètres

3.1.3 Estimation des paramètres de la distribution de Champernowne modifiée

Buch-Larsen et al. (2005) [11] recommandent de commencer par l'estimation du paramètre β avant de traiter les autres paramètres (α et c). Cette approche simplifie le calcul global. Ils ont observé que, dans la distribution de Champernowne, $F_{\alpha,\beta,0}(\beta) = 0,5$, une propriété également valable pour la distribution de Champernowne modifiée, où $F_{\alpha,\beta,c}(\beta) = 0,5$. Cela suggère que β peut être estimé en tant que médiane empirique des données. Une fois β estimé, l'étape suivante consiste à estimer le couple (α, c) qui maximise la fonction de log-vraisemblance.

Soient X_1, \dots, X_N des variables aléatoires positives, supposons qu'elles soient indépendantes et identiquement distribuées (i.i.d) de densité de probabilité inconnue f_X , le but est de trouver la manière optimale d'ajuster la distribution champernownen généralisée aux données observées (X_i). Pour cela on va estimer les paramètres (α, c) de la distribution de champernowne généralisée par la méthode MLE (Maximum likelihood estimator).

Pour simplifier les calculs, on maximise la fonction de log-vraisemblance $\ell(\alpha, c | X_1, \dots, X_n) = \log(L(\alpha, c | X_1, \dots, X_n))$ au lieu de la fonction de vraisemblance $L(\alpha, c | X_1, \dots, X_n)$, car le logarithme naturel est une fonction monotone croissante. Par conséquent, les valeurs des paramètres (α, c) qui maximisent la fonction de log-vraisemblance sont exactement les mêmes valeurs qui maximisent la fonction de vraisemblance, c'est-à-dire :

$$\arg \max_{\alpha, c} \ell(\alpha, c | X_1, \dots, X_n) = \arg \max_{\alpha, c} L(\alpha, c | X_1, \dots, X_n) \quad .$$

$$\begin{aligned}
\text{Alors } \ell(\alpha, c | X_1, \dots, X_n) &= \ell(\alpha, c) = \sum_{i=1}^N \log f_{\alpha, \beta, c}(X_i | \alpha, c) \\
&= \sum_{i=1}^N \log \left(\frac{\alpha (X_i + c)^{\alpha-1} ((\beta + c)^\alpha - c^\alpha)}{((X_i + c)^\alpha + (\beta + c)^\alpha - 2c^\alpha)^2} \right) \\
&= \sum_{i=1}^N \log (\alpha (X_i + c)^{\alpha-1} ((\beta + c)^\alpha - c^\alpha)) \\
&\quad - 2 \log ((X_i + c)^\alpha + (\beta + c)^\alpha - 2c^\alpha)
\end{aligned}$$

$$\begin{aligned}
\text{D'où } \ell(\alpha, c) &= N \cdot \log(\alpha) + N \cdot \log((\beta + c)^\alpha - c^\alpha) + (\alpha - 1) \sum_{i=1}^N \log(X_i + c) \\
&\quad - 2 \sum_{i=1}^N \log((X_i + c)^\alpha + (\beta + c)^\alpha - 2c^\alpha) \quad . \tag{3.4}
\end{aligned}$$

Si β est une constante fixée, la fonction de vraisemblance (3.4) est concave et possède un maximum.

3.2 Estimation semi-paramétrique de densité de probabilité

Pour obtenir l'estimateur de densité à noyau par transformation avec la distribution de Champernowne modifiée (KMCE), il convient de suivre les quatre étapes suivantes :

1. Estimer les paramètres de la distribution de Champernowne généralisée.
2. Transformer les données à l'aide de la fonction de transformation $T_{\hat{\alpha}, \hat{\beta}, \hat{c}}$.

où

$$T_{\hat{\alpha}, \hat{\beta}, \hat{c}} = F_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x) = \frac{(x + \hat{c})^{\hat{\alpha}} - \hat{c}^{\hat{\alpha}}}{(x + \hat{c})^{\hat{\alpha}} + (\hat{\beta} + \hat{c})^{\hat{\alpha}} - 2\hat{c}^{\hat{\alpha}}} \forall x \in \mathbb{R}_+ \quad , \quad \text{avec } \alpha > 0, \beta > 0 \quad \text{et } c \geq 0 \quad . \tag{3.5}$$

3. Calculer l'estimateur de densité par noyau pour les données transformées $(\{Y_i, i = 1 \dots N\})$.
4. Rétro-transformer les données, c'est-à-dire appliquer l'inverse de la transformation initiale afin d'obtenir une estimation de densité cohérente avec l'échelle originale des données.

3.2.1 Transformation des données avec la distribution de Champernowne

Après avoir estimé les paramètres de la fonction de distribution cumulative (CDF) de Champernowne, l'étape suivante consiste à transformer les données originales à l'aide de cette fonction qui transforme les données positives en données dans $[0,1]$.

Soient $X_i, i = 1, \dots, N$, des variables aléatoires positives ayant une fonction de répartition F et une densité f inconnues. L'ensemble des données transformées $\{Y_i, i = 1, \dots, N\}$, est obtenu par la relation :

$$Y_i = T_{\hat{\alpha}, \hat{\beta}, \hat{c}}(X_i) \quad i = 1, \dots, N \quad . \quad (3.6)$$

Tels que $\hat{\alpha}$, $\hat{\beta}$ et \hat{c} sont des estimations de α , β et c respectivement .

L'estimation des paramètres est conçue pour rapprocher autant que possible les données transformées d'une distribution uniforme. En effet, si, après transformation, les données ne suivent pas une distribution uniforme sur $[0,1]$, cela pourrait indiquer que la forme de la fonction de distribution (CDF) $T_{\hat{\alpha}, \hat{\beta}, \hat{c}}$ ou les paramètres choisis ne sont pas adéquats pour les données en question. En revanche, si les données transformées suivent bien une distribution uniforme sur $[0,1]$, Cela signifie que la transformation a correctement capturé la structure des données originales en termes de leur distribution, car la distribution de Champernowne modifiée est ajustée à cet ensemble de données.

On a considéré un échantillon de taille 1000 provenant d'une loi gamma de paramètres de forme et d'échelle 0.9,1 respectivement ($\Gamma(0.9, 1)$). Ensuite, on a estimé les paramètres de la fonction cumulative de Champernowne généralisée comme on l'a mentionné déjà dans 3.1.3, et les résultats de l'estimation des paramètres sont les suivants :

$$\hat{\alpha} = 86.3536 \quad , \quad \hat{\beta} = 0.5821 \quad , \quad \hat{c} = 95.8027 \quad .$$

Ensuite, on a transformé les données initiales par la fonction $T_{\hat{\alpha}, \hat{\beta}, \hat{c}}$. La figure 3.3 représente l'histogramme des données transformées.

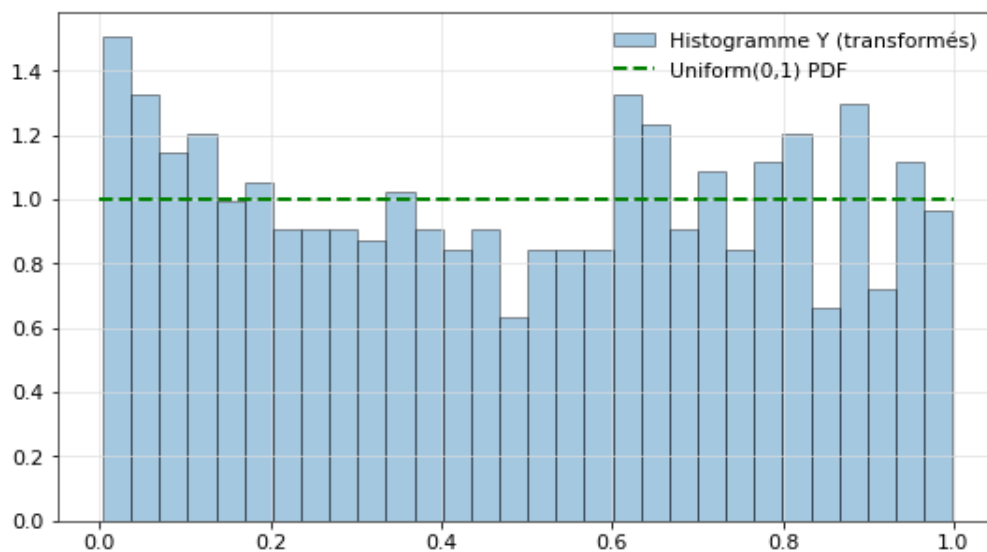


FIGURE 3.3 – Histogramme des données transformées (Y_i).

Pour évaluer la précision d'ajustement de la distribution de Champernowne modifiée ($T_{\hat{\alpha}, \hat{\beta}, \hat{c}}$) aux données considérées (Gamma(0.9,1)), on a tracé les deux densités dans le même

graphe. Le résultat est le suivant :

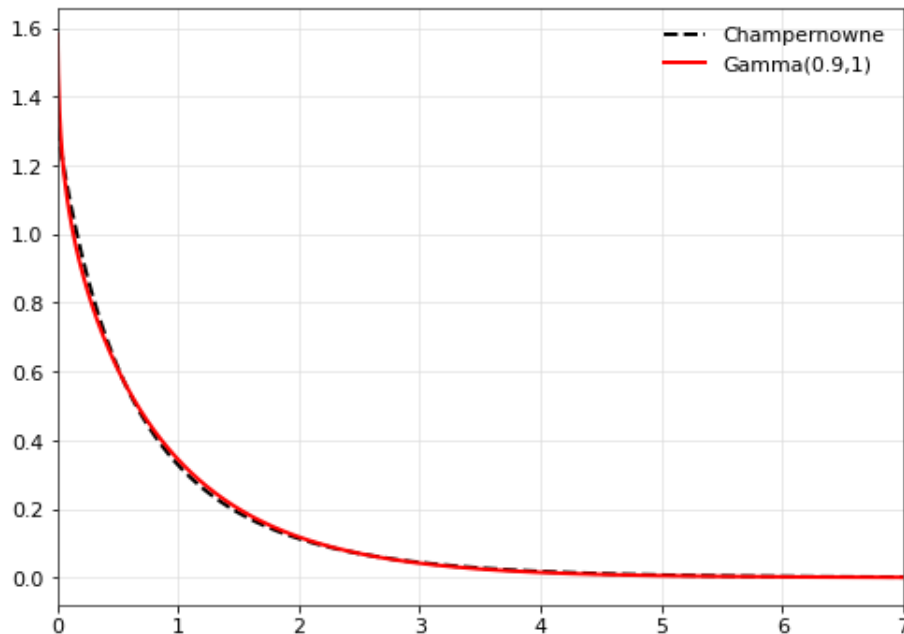


FIGURE 3.4 – Comparaison entre la densité de probabilité de la distribution de Champernowne modifiée après l'estimation des paramètres (α, β, c) , et la densité des données originales Gamma(0.9,1).

3.2.2 Calcul de l'estimateur de densité par noyau pour les données transformées

Après avoir transformé les données, l'étape suivante consiste à calculer l'estimateur de densité par noyau pour les données transformées $(Y_i, i = 1, \dots, N)$.

Buch-Larsen et al. (2005) proposent, dans leur étude [11], d'utiliser le noyau classique comme fonction de pondération pour le calcul de l'estimateur. Dans notre travail, nous étendrons cette approche en considérant trois noyaux différents : un noyau symétrique, couramment utilisé dans la littérature, et deux noyaux asymétriques que nous introduirons afin d'étudier leur impact sur les performances de l'estimateur.

• Cas d'un noyau symétrique

Soit $Y_i, i = 1, \dots, N$ l'ensemble de données transformées par la fonction de distribution de Champernowne $(T_{\hat{\alpha}, \hat{\beta}, \hat{c}})$, alors l'estimateur de densité à noyau classique des données transformées est donné par :

$$\hat{f}_Y(y) = \frac{1}{N b k_y} \sum_{i=1}^N K\left(\frac{y - Y_i}{b}\right) . \quad (3.7)$$

- $\hat{f}_Y(y)$: L'estimation de la densité de probabilité au point y pour les données transformées.

- Y_i : représente la i -ème observation dans l'ensemble de données transformées.
- N : La taille de l'échantillon, c'est-à-dire le nombre de points de données Y_i .
- k_y : Facteur de correction de bord, (qui sera expliqué plus en détail ci-dessous).
- b : Largeur de bande (bandwidth).
- K : La fonction noyau .

L'idée principale de cette formule est d'adapter l'estimateur de densité par noyau standard aux données bornées dans l'intervalle $[0, 1]$. La correction de bordure k_y permet de normaliser l'estimateur afin de tenir compte de la partie du noyau qui sort de cet intervalle. Cela permet d'obtenir une estimation plus précise de la densité près des bornes. Sans cette correction, si les points Y_i se trouvent proches des limites 0 ou 1, une partie du noyau pourrait tomber en dehors de l'intervalle admissible, ce qui fausserait l'estimation de la densité.

La correction est donnée par :

$$k_y = \int_{\max(-1, \frac{-y}{b})}^{\min(1, \frac{1-y}{b})} K(u) du \quad .$$

$\max(-1, \frac{-y}{b})$ et $\min(1, \frac{1-y}{b})$ définissent l'intervalle où le noyau reste dans $[0, 1]$.

Cette correction signifie que lorsqu'on est proche des bornes, on ne considère que la portion du noyau qui se trouve dans les bornes autorisées ($[0, 1]$), et on normalise l'estimation en conséquence pour que l'intégrale sous le noyau dans $[0, 1]$ soit égale à 1.

Reprenons l'exemple précédent, où un échantillon de taille 1000 a été généré selon une loi Gamma(0.9,1). Dans cette partie, la densité des données transformées a été estimée par une méthode à noyau avec la correction k_y . Le noyau d'Epanechnikov a été utilisé comme fonction noyau $K(u)$ et le paramètre de lissage ($b = 0.4999$) a été sélectionné par minimisation de ISE définie par : $ISE = \int_0^\infty [\hat{T}_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x) - T_{\alpha, \beta, c}(x)]^2 dx$. Le résultat de cette estimation est présenté dans la figure suivante.

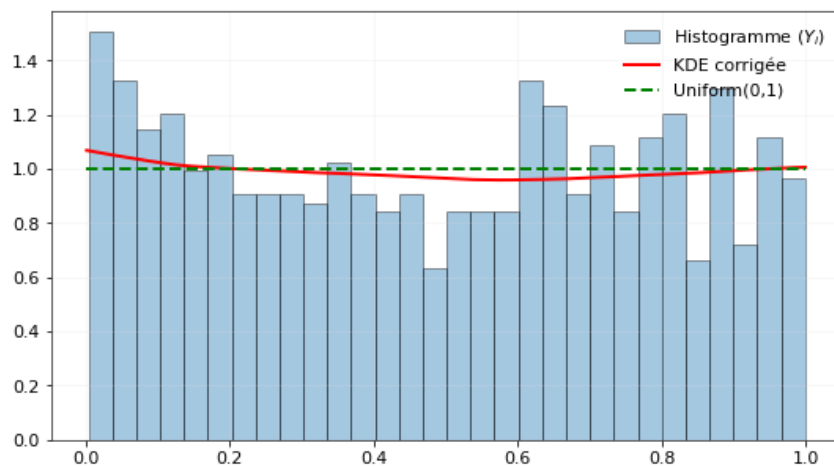


FIGURE 3.5 – Estimation à noyau d'Epanechnikov de densité des données transformées.

• Cas d'un noyau bêta

Comme nous l'avons vu dans la partie précédente, dans le cas d'estimation de densité de données transformées par un noyau symétrique, qui cause un problème de biais au niveau des bornes 0 et 1, plusieurs méthodes ont déjà été proposées, dans notre travail, nous avons utilisé une correction (k_y) proposé par Buch-Larsen et al. (2005) [11] afin d'obtenir une estimation plus précise.

Afin d'éviter les effets de bord, dans cette partie, on a fait appel à un autre noyau : le noyau bêta, défini sur l'intervalle $[0,1]$ qui est proposé et étudié par : Song Xi Chen (1999) [13] .

Contrairement aux autres noyaux qui nécessitent une correction explicite ,ce noyau n'attribue aucun poids en dehors de l'intervalle $[0,1]$ [14].

Soit $Y_i, i = 1, \dots, N$ l'ensemble de données transformées par la fonction de distribution de Champernowne ($T_{\hat{\alpha}, \hat{\beta}, \hat{c}}$), alors l'estimateur de densité à noyau beta (2.39) des données transformées est donné comme suit :

$$\hat{f}_Y(y) = \frac{1}{N} \sum_{i=1}^N K_{y,b^*}^*(Y_i) \quad . \quad (3.8)$$

où K_{y,b^*}^* et la fonction noyau définie dans (2.3.8).
et $b^* = b_2^*$ est le paramètre de lissage optimale (2.45).

Reprenons le même exemple où nous avons généré un échantillon de taille 1000, distribué selon une loi Gamma de paramètres (0,9,1) .Dans cette partie, la densité des données transformées est estimée à l'aide du noyau bêta, qui remplace ici le noyau symétrique classique. Le paramètre de lissage optimal, obtenu par minimisation de ISE, est $b^* = 0.2479$. La figure suivante présente le résultat de cette estimation.

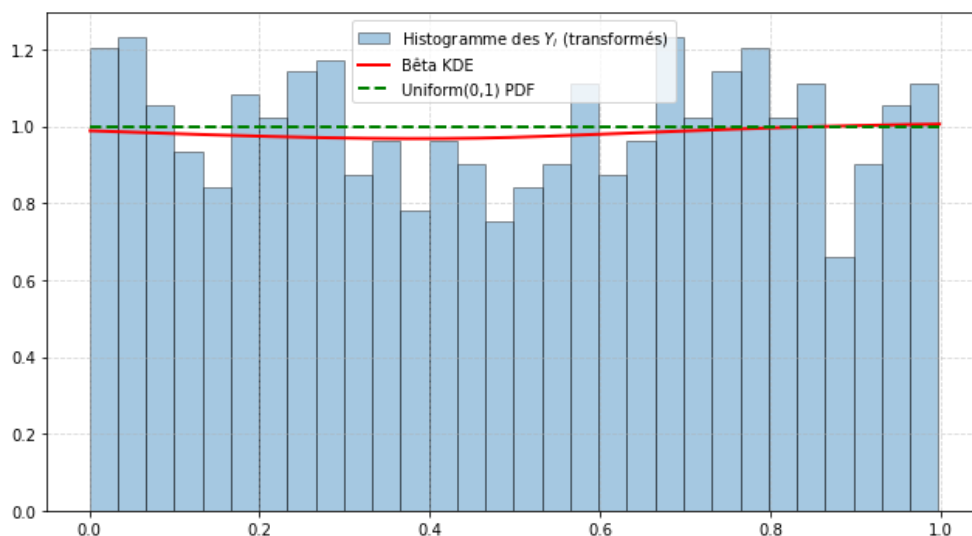


FIGURE 3.6 – Estimation a noyau bêta de densité des donnés transformées.

3.2.3 Estimateur KMCE (Kernel Modified Champernowne Estimator)

Après avoir estimé les données transformées par la méthode du noyau, la dernière étape consiste à rétrotransformer les données Y_i tel que $Y_i \in [0, 1]$, $\forall i \in [1, N]$ afin de revenir à l'échelle originale.

À partir de (3.6) on a :

$$Y_i = T_{\hat{\alpha}, \hat{\beta}, \hat{c}}(X_i) \Leftrightarrow X_i = T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(Y_i) \quad , \quad Y_i \sim \mathcal{U}(0, 1) \quad , \quad \forall i \in [1, N] \quad .$$

On note par F_Y et \hat{F}_X les fonctions de répartitions des variables aléatoires Y_i et X_i respectivement, et par \hat{f}_X et f_Y leurs densités respectivement (F_Y et la fonctions de répartition originale de la variables aléatoire Y_i , et f_Y et sa densité originale).

Selon les paramètre $\hat{\alpha}, \hat{\beta}, \hat{c}$ estimées, la fonction de transformation $T_{\hat{\alpha}, \hat{\beta}, \hat{c}}$ (3.5) peut être croissante ($\Leftrightarrow T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}$ croissante) ou bien décroissante ($\Leftrightarrow T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}$ décroissante). Donc on va considérer deux cas :

1^{er} cas : ($T_{\hat{\alpha}, \hat{\beta}, \hat{c}}$ croissante)

On a :

$$\begin{aligned} F_Y(y) &= P(Y_i < y) = P(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}(X_i) < y) \\ &= P(X_i < T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y)) \\ &= F_X \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y) \right) \quad . \end{aligned}$$

D'autre part, on a :

$$\begin{aligned} \hat{f}_Y(y) \simeq f_Y(y) &= \frac{dF_Y(y)}{dy} = \frac{dF_X \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y) \right)}{dy} \\ &= \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y) \right)' \frac{dF_X(y)}{dy} [T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y)] \\ &= \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y) \right)' f_X \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y) \right) \\ &\simeq \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y) \right)' \hat{f}_X \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y) \right) \\ \Leftrightarrow \hat{f}_X \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y) \right) &= \frac{\hat{f}_Y(y)}{\left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y) \right)'} = \frac{\hat{f}_Y(y)}{\left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1} \right)'(y)} \quad . \end{aligned} \quad (3.9)$$

On prend $y = T_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x)$, tel que $x \in [0, \infty)$, $y \in [0, 1]$, on aura :

$$\hat{f}_X(x) = \frac{\hat{f}_Y(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x))}{\left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1} \right)' [T_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x)]} = T'_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x) \cdot \hat{f}_Y \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x) \right) \quad , \quad (3.10)$$

tel que \hat{f}_Y c'est l'estimateur de densité à noyau symétrique (3.7) ou bêta (3.8) des données transformées.

Remarque 3.1 :

Le résultat (3.10) vient du fait que : $1 = \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1} [T_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x)] \right)' = T'_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x) \cdot \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1} \right)' [T_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x)]$

2^{eme} cas : ($T_{\hat{\alpha}, \hat{\beta}, \hat{c}}$ décroissante).

On a :

$$\begin{aligned} F_Y(y) &= P(Y_i < y) = P(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}(X_i) < y) \\ &= P(X_i > T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y)) \\ &= 1 - F_X \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y) \right) \quad . \end{aligned}$$

D'autre part, on a :

$$\begin{aligned} \hat{f}_Y(y) \simeq f_Y(y) &= \frac{dF_Y(y)}{dy} = - \frac{dF_X \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y) \right)}{dy} \\ &= - \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y) \right)' \frac{dF_X(y)}{dy} [T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y)] \\ &= - \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y) \right)' f_X \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y) \right) \\ &\simeq - \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y) \right)' \hat{f}_X \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y) \right) \\ \Leftrightarrow \hat{f}_X \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y) \right) &= - \frac{\hat{f}_Y(y)}{\left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1}(y) \right)'} = - \frac{\hat{f}_Y(y)}{\left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1} \right)'(y)} \quad . \end{aligned} \quad (3.11)$$

On prend $y = T_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x)$, tel que $x \in [0, \infty)$, $y \in [0, 1]$, on aura :

$$\hat{f}_X(x) = - \frac{\hat{f}_Y \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x) \right)}{\left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}^{-1} \right)' \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x) \right)} = - T'_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x) \cdot \hat{f}_Y \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x) \right) \quad . \quad (3.12)$$

À partir de (3.10) et (3.11) on aura l'estimateur de densité de l'ensemble de données originale $\{X_i, i = 1, \dots, N\}$ est donné comme suit :

$$\hat{f}_X(x) = \left| T'_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x) \right| \cdot \hat{f}_Y \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x) \right) \quad . \quad (3.13)$$

• Cas d'un noyau symétrique

Pour obtenir l'estimateur KMCE de densité de probabilité de Buch-Larsen et al. (2005) [11] dans le cas symétrique il suffit de remplacer \hat{f}_Y (3.7) dans (3.13), et on obtient :

$$\hat{f}_X(x) = \frac{|T'_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x)|}{N \cdot b \cdot k_{T_{\hat{\alpha}, \hat{\beta}, \hat{c}}}(x)} \sum_{i=1}^N K \left(\frac{T_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x) - T_{\hat{\alpha}, \hat{\beta}, \hat{c}}(X_i)}{b} \right) . \quad (3.14)$$

• Cas d'un noyau bêta

Pour obtenir l'estimateur KMCE dans le cas où le noyau utilisé appartient à la famille des noyaux bêta, il suffit de remplacer \hat{f}_Y (3.8) dans (3.13), et on obtient :

$$\hat{f}_X(x) = \frac{|T'_{\hat{\alpha}, \hat{\beta}, c}(x)|}{N} \sum_{i=1}^N K_{T_{\hat{\alpha}, \hat{\beta}, c}(x), b^*}^* \left(T_{\hat{\alpha}, \hat{\beta}, c}(X_i) \right) . \quad (3.15)$$

3.2.4 Théorie asymptotique de l'estimateur KMCE

• Cas d'un noyau symétrique [11]

Théorème 3.1 : Soient X_1, \dots, X_N des variables aléatoires indépendantes et identiquement distribuées (i.i.d) de densité f_X de support \mathbb{D} . Soit \hat{f}_X l'estimation à noyau symétrique par transformation de f_X , est alors :

$$\hat{f}_X(x) = \frac{|T'(x)|}{N \cdot b} \sum_{i=1}^N K [T(x) - T(X_i)] ,$$

tel que T est la fonction de transformation.

D'après l'équation (3.13), on a :

$$E \left[\hat{f}_X(x) \right] = |T'(x)| \cdot E \left[\hat{f}_Y(T(x)) \right] = \begin{cases} T'(x) \cdot E \left[\hat{f}_Y(T(x)) \right] & \text{si } T'(x) > 0 \\ -T'(x) \cdot E \left[\hat{f}_Y(T(x)) \right] & \text{si } T'(x) < 0 \end{cases}$$

Alors le biais et la variance de \hat{f}_X sont donnés comme suit :

$$\text{Biais} \left[\hat{f}_X(x) \right] = \mathbb{E} \left[\hat{f}_X(x) \right] - f_X(x) = \frac{1}{2} \mu(K) b^2 \left(\left(\frac{f_X(x)}{T'(x)} \right)' \frac{1}{T'(x)} \right)' \cdot \frac{|T'(x)|}{T'(x)} + o(b^2) ,$$

tel que $\mu(K) = \int u^2 K(u) du$.

$$V \left[\hat{f}_X(x) \right] = \frac{1}{Nb} R(K) |T'(x)| f_X(x) + o \left(\frac{1}{Nb} \right) ,$$

tel que $R(K) = \int K^2(u) du$.

Pour une preuve détaillée, voir :[11]

L'erreur quadratique moyenne (MSE)

L'erreur quadratique moyenne de $\hat{f}_X(x)$ est définie comme suit :

$$\begin{aligned} \text{MSE} \left[\hat{f}_X(x) \right] &= \text{Biais} \left[\hat{f}_X(x) \right]^2 + V \left[\hat{f}_X(x) \right] \\ &= \frac{1}{4} \mu^2(K) b^4 \left[\left(\left(\frac{f_X(x)}{T'(x)} \right)' \frac{1}{T'(x)} \right)' \right]^2 + \frac{1}{Nb} R(K) |T'(x)| f_X(x) + o \left(\frac{1}{Nb} + b^4 \right) . \end{aligned}$$

L'erreur quadratique moyenne intégrée (MISE) :

L'erreur quadratique moyenne intégrée de $\hat{f}_X(x)$ est définie comme suit :

$$\begin{aligned} \text{MISE} \left[\hat{f}_X \right] &= \int_{\mathbb{D}} \text{MSE} \left[\hat{f}_X(x) \right] dx \\ &= \frac{1}{4} \mu^2(K) b^4 \int_{\mathbb{D}} \left[\left(\left(\frac{f_X(x)}{T'(x)} \right)' \frac{1}{T'(x)} \right)' \right]^2 dx \\ &\quad + \frac{1}{Nb} R(K) \int_{\mathbb{D}} |T'(x)| f_X(x) dx + o \left(\frac{1}{Nb} + b^4 \right) . \end{aligned}$$

L'erreur quadratique moyenne intégrée asymptotique (AMISE) :

L'erreur quadratique moyenne intégrée asymptotique de $\hat{f}_X(x)$ est définie comme suit :

$$\text{AMISE} \left[\hat{f}_X \right] = \frac{1}{4} \mu^2(K) b^4 \int_{\mathbb{D}} \left[\left(\left(\frac{f_X(x)}{T'(x)} \right)' \frac{1}{T'(x)} \right)' \right]^2 dx + \frac{1}{Nb} R(K) \int_{\mathbb{D}} |T'(x)| f_X(x) dx .$$

Paramètre de lissage optimal :

Le paramètre de lissage optimal, noté b^* , est la largeur de bande qui minimise l'erreur quadratique moyenne intégrée (AMISE), et il est donnée comme suit :

$$b^* = \left[\frac{R(K) \int_{\mathbb{D}} |T'(x)| f_X(x) dx}{\mu^2(K) \int_{\mathbb{D}} \left[\left(\left(\frac{f_X(x)}{T'(x)} \right)' \frac{1}{T'(x)} \right)' \right]^2 dx} \right]^{1/5} \cdot N^{-1/5} \quad . \quad (3.16)$$

• Cas d'un noyau bêta

L'estimateur KMCE (noyau bêta) de f_X est donnée comme suit :

$$\hat{f}_X(x) = \frac{|T'(x)|}{N} \sum_{i=1}^N K_{T(x), b^*}^*(T(X_i)) \quad .$$

Biais

D'après (3.13) on a :

$$\text{Biais} \left[\hat{f}_X(x) \right] = |T'(x)| \text{Biais} \left[\hat{f}_Y \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x) \right) \right] \quad ,$$

Alors le biais de \hat{f}_X est définie comme suit :

$$\text{Biais} \left[\hat{f}_X(x) \right] = \frac{|T'(x)|}{T'(x)} \cdot \begin{cases} \frac{b}{2} T(x)(1-T(x)) \left(\left(\frac{f_X(x)}{T'(x)} \right)' \frac{1}{T'(x)} \right)' + \mathcal{O}(b^2) & \text{si } T(x) \in [2b, 1-2b] \\ b \cdot \xi(T(x)) \left(\frac{f_X(x)}{T'(x)} \right)' + o(b) & \text{si } T(x) \in [0, 2b] \\ -b \cdot \xi(1-T(x)) \left(\frac{f_X(x)}{T'(x)} \right)' + o(b) & \text{si } T(x) \in (1-2b, 1] \end{cases}$$

$$\text{où } \xi(x) = \frac{(1-x)(\rho(x) - \frac{x}{b})}{(1+b\rho(x)-x)} \quad \text{et} \quad \rho(x) = 2b^2 + 2.5 - \sqrt{4b^4 + 6b^2 + 2.25 - x^2 - x/b} \quad .$$

Variance

D'après (3.13) on a :

$$\text{V} \left[\hat{f}_X(x) \right] = [T'(x)]^2 \cdot \text{V} \left[\hat{f}_Y \left(T_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x) \right) \right] \quad ,$$

Alors la variance de \hat{f}_X est définie comme suit :

$$\text{V} \left[\hat{f}_X(x) \right] = \begin{cases} \frac{N^{-1} b^{-1/2} |T'(x)|}{2\sqrt{\pi T(x)(1-T(x))}} [f_X(x) + \mathcal{O}(N^{-1})] & \text{si } \frac{T(x)}{b} \rightarrow \infty \text{ et } \frac{1-T(x)}{b} \rightarrow \infty \\ |T'(x)| \cdot g(c) \cdot N^{-1} b^{-1} [f_X(x) + \mathcal{O}(N^{-1})] & \text{si } \frac{T(x)}{b} \rightarrow c \text{ ou } \frac{1-T(x)}{b} \rightarrow c \end{cases}$$

Où $g(c)$ est une constante positive, ($g(c) \in \mathbb{R}$).

L'erreur quadratique moyenne intégrée (MISE)

on a :

$$\text{MISE} [\hat{f}_X(x)] = \int_0^\infty (T'(x))^2 \cdot \text{MSE} [\hat{f}_Y(T(x))] dx \quad ,$$

Alors l'erreur quadratique moyenne intégrée de $\hat{f}_X(x)$ est définie comme suit :

$$\begin{aligned} \text{MISE} [\hat{f}_X(x)] &= \frac{1}{4} b^2 \int_0^\infty \left(T(x)(1 - T(x)) \cdot \left(\left[\frac{f_X}{T'(x)} \right]' \cdot \frac{1}{T'(x)} \right)' \right)^2 dx \\ &\quad + \frac{1}{2\sqrt{\pi}} N^{-1} b^{-1/2} \int_0^\infty [T(x)(1 - T(x))]^{-1/2} f_X(x) \cdot |T'(x)| dx + o(N^{-1} b^{-1/2} + b^2) \quad . \end{aligned}$$

L'erreur quadratique moyenne intégrée asymptotique (AMISE)

$$\begin{aligned} \text{AMISE} [\hat{f}_X(x)] &= \frac{1}{4} b^2 \int_0^\infty \left(T(x)(1 - T(x)) \cdot \left(\left[\frac{f_X}{T'(x)} \right]' \cdot \frac{1}{T'(x)} \right)' \right)^2 dx \\ &\quad + \frac{1}{2\sqrt{\pi}} N^{-1} b^{-1/2} \int_0^\infty |T'(x)| \cdot [T(x)(1 - T(x))]^{-1/2} f_X(x) dx \quad . \end{aligned}$$

Paramètre de lissage optimal

Le paramètre de lissage optimal qui minimise l'erreur quadratique moyenne intégrée asymptotique (AMISE) est donné comme suit :

$$b^* = \left(\frac{\int_0^\infty |T'(x)| \cdot [T(x)(1 - T(x))]^{-1/2} f_X(x) dx}{2\sqrt{\pi} \int_0^\infty \left(T(x)(1 - T(x)) \cdot \left(\left[\frac{f_X}{T'(x)} \right]' \cdot \frac{1}{T'(x)} \right)' \right)^2 dx} \right)^{2/5} \cdot N^{-2/5} \quad .$$

Conclusion

Ce chapitre a présenté l'estimateur KMCE (Kernel Modified Champernowne Estimator), une méthode semi-paramétrique pour l'estimation de densité, particulièrement adaptée aux données à support positif présentant une asymétrie prononcée et des queues de distribution épaisses.

Nous avons d'abord introduit la distribution de Champernowne dans ses formulations originale et modifiée. Cette dernière, caractérisée par le nouveau paramètre c , s'est révélée particulièrement flexible pour modéliser divers comportements de données continues à support positif. Les méthodes d'estimation des paramètres de la distribution de Champernowne, notamment par maximisation de la vraisemblance, ont été explicitées, garantissant ainsi une adéquation optimale aux caractéristiques des échantillons observés.

Le coeur de ce chapitre a consisté en la construction méthodologique de l'estimateur KMCE. Cette approche combine une transformation préalable des données via la fonction de répartition de la distribution de Champernowne modifiée avec une estimation par noyau sur l'intervalle transformé $[0,1]$. Cette double structure confère à l'estimateur KMCE les avantages complémentaires d'une modélisation paramétrique qui guide la transformation et d'une estimation non paramétrique qui préserve la flexibilité.

En synthèse, l'estimateur KMCE émerge comme une alternative pertinente aux approches purement paramétriques ou non paramétriques traditionnelles, notamment pour modéliser des distributions asymétriques aux comportements complexes en queue de distribution. Son architecture lui permet de capturer à la fois la structure globale des données et leurs spécificités locales, tout en atténuant les biais typiques des méthodes conventionnelles.

Chapitre 4

Simulation

Dans ce chapitre, nous avons étudié les performances de l'estimateur semi-paramétrique de densité présenté précédemment, l'estimateur KMCE (Kernel Modified Champernowne Estimator). L'objectif était d'évaluer sa capacité à modéliser des distributions à queues lourdes et légères simultanément, à atténuer les biais aux limites du support et à réduire l'erreur globale, en comparaison avec certaines méthodes existantes.

Pour ce faire, nous avons simulé des données issues de quatre distributions distinctes présentant différentes formes de queues, puis estimé leurs densités à l'aide de l'estimateur KMCE, d'une part avec un noyau classique muni d'une correction, d'autre part avec un noyau bêta. Enfin, nous avons quantifié l'écart entre les densités vraies et les densités estimées au moyen de trois mesures d'erreur différentes.

4.1 Données utilisées dans la simulation

Nous avons simulé quatre échantillons de données avec des tailles respectives de $N = 50$, $N = 100$, $N = 500$ et $N = 1000$. Ces simulations ont été réalisées pour chacune des distributions suivantes : exponentielle, inverse gaussienne (qui appartient au domaine d'attraction de Weibull), mixte Weibull-Fréchet, et Burr (qui appartient au domaine d'attraction de Fréchet).

La distribution Mixte Weibull-Fréchet [57] est un mélange de la loi de Weibull et de la loi de Fréchet. Cette approche est particulièrement pertinente pour modéliser les réclamations d'assurance, notamment les montants des sinistres. En effet, ces montants ne proviennent généralement pas d'une cause ou d'un type unique, permettant de distinguer :

- Sinistres fréquents : Ce sont des sinistres de faible à moyenne ampleur, qui surviennent

régulièrement (ex : petits accrochages en auto, bris de glace, dégâts des eaux mineurs). Leur distribution peut être bien capturée par une loi comme la Weibull.

- Sinistres extrêmes (catastrophiques) : Ce sont des événements rares mais qui entraînent des coûts très élevés (ex : accidents corporels graves, catastrophes naturelles, incendies majeurs, très gros vols). Ces événements se situent dans la "queue" de la distribution et sont très bien modélisés par la loi de Fréchet, qui est par nature une distribution à "queue lourde" (heavy-tailed).

Ce mélange de distributions, de densité $p \cdot f_W + (1 - p) \cdot f_R$, où p est le paramètre de poids ($p \in [0, 1]$) et f_W, f_R sont respectivement les densités de Weibull et de Fréchet, forme un modèle plus flexible, capable de s'adapter aux différentes composantes de la distribution.

Les fonctions de densité des distributions utilisées et leurs paramètres sont donnés dans le tableau suivant :

Distribution	Densité	Paramètres
Exponentielle (λ)	$\lambda e^{-\lambda x}$	$\lambda = 1$
Inverse Gaussienne (μ, λ)	$\sqrt{\frac{\lambda}{2\pi x^3}} \exp\left(-\frac{\lambda(x - \mu)^2}{2\mu^2 x}\right)$	$\mu = 2, \lambda = 1$
0.5 Weibull(γ)+0.5 fréchet(ξ, σ)	$0.5\gamma x^{\gamma-1}e^{-x^\gamma} + 0.5\frac{\xi}{\sigma}\left(\frac{x}{\sigma}\right)^{-1-\xi} \exp\left(-\left[\frac{x}{\sigma}\right]^{-\xi}\right)$	$\gamma = 0.9, \sigma = 0.5, \xi = 1.5$
Burr(c, k, λ)	$\frac{kc \cdot x^{c-1}}{\lambda^c(1 + x^c)^{k+1}}$	$c = 1, k = 1, \lambda = 1$

TABLE 4.1 – Densités de probabilité des distributions utilisées dans la simulation.

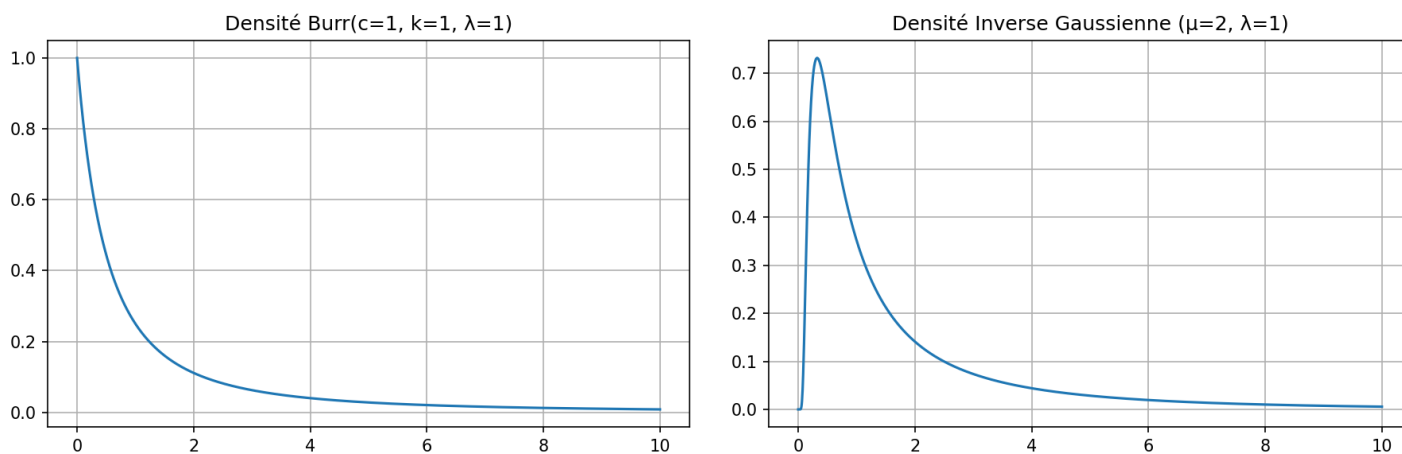


FIGURE 4.1 – Densités de probabilité : Burr (à gauche), inverse-gaussienne (à droite)

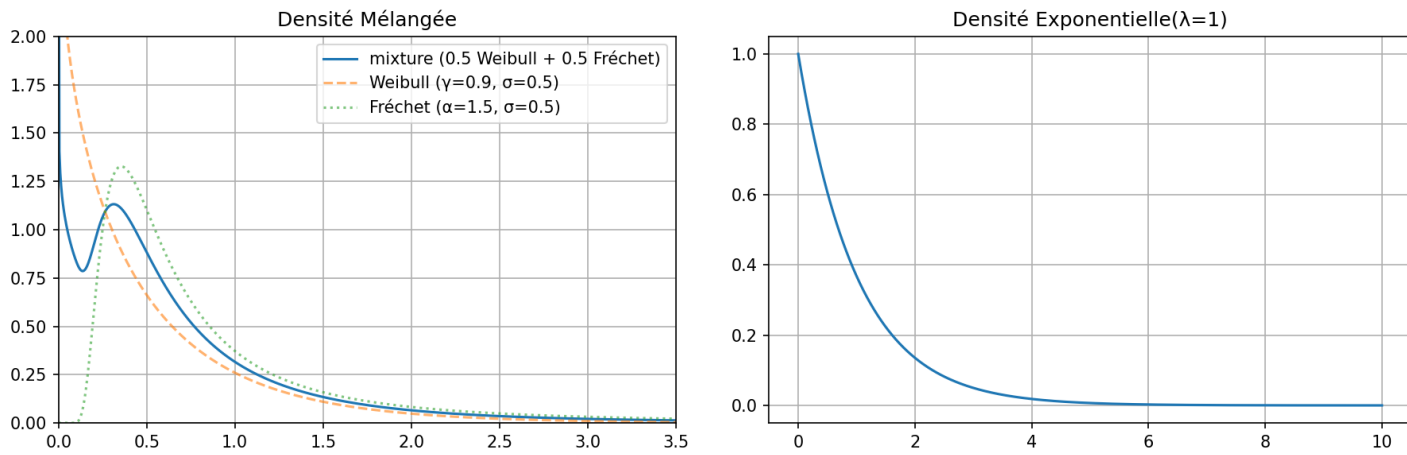


FIGURE 4.2 – Densités de probabilité : Mixte Weibull-Fréchet (à gauche), Exponentielle (à droite)

Les données simulées proviennent de quatre distributions aux comportements de queue variés :

1. La distribution Exponentielle (1), avec une queue légère.
2. La distribution Inverse Gaussienne (2,1), avec une queue légère .
3. La distribution $0.5 \times \text{Weibull}(0.9) + 0.5 \times \text{Fréchet}(1.5, 0.5)$, avec une queue lourde.
4. La distribution Burr(1,1,1), avec une queue lourde.

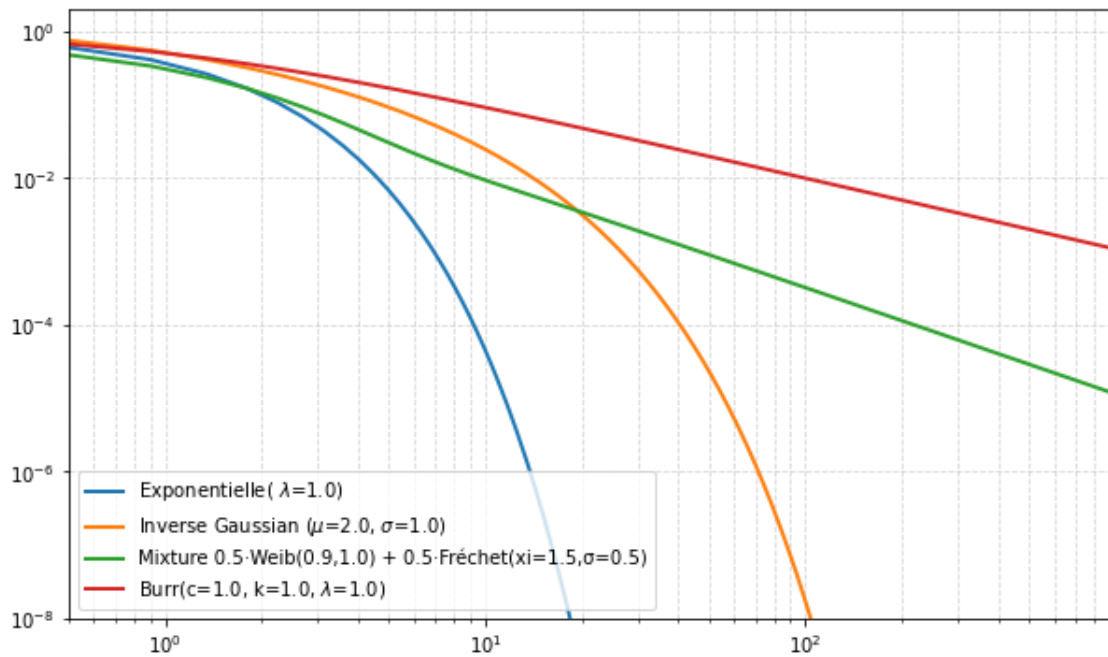


FIGURE 4.3 – Comparaison des queues des quatre distributions (Fonctions de survie)

4.2 Méthodes d'estimation et critères d'erreur retenus

Nous avons comparé les performances de quatre estimateurs à noyau : l'estimateur à noyau d'Epanechnikov avec transformation de Champernowne modifiée (noté EKMC), l'estimateur à noyau Beta avec transformation de Champernowne modifiée (noté BKMC), l'estimateur à noyau d'Epanechnikov classique (noté EKDE), et l'estimateur à noyau gaussien classique (noté GKDE).

Pour la sélection des paramètres de lissage optimaux, nous avons choisi la minimisation de ISE (erreur quadratique intégrée) plutôt que la validation croisée par moindres carrés (LSCV), car elle fournit des résultats plus précis et un temps d'exécution réduit. Toutefois, cette méthode n'est applicable qu'en simulation, lorsque la vraie densité est connue. Avec des données réelles, il convient d'utiliser la LSCV ou d'autres méthodes de sélection de la bande passante.

La comparaison est basée sur des données simulées à partir des quatre distributions décrites dans le tableau 4.1, et pour quatre tailles d'échantillon : $N_1 = 50$, $N_2 = 100$, $N_3 = 500$ et $N_4 = 1000$, avec les mesures suivantes : ISE, L1 et WISE, qui se définissent comme suit :

$$L_1 = \int_0^\infty |\hat{f}(x) - f(x)| dx \quad , \quad \text{WISE} = \left(\int_0^\infty (\hat{f}(x) - f(x))^2 x^2 dx \right)^{1/2} \quad ,$$

$$\text{ISE} = \int_0^\infty (\hat{f}_X(x) - f(x))^2 dx \quad .$$

4.3 Algorithme du programme de simulation

Comme nous l'avons déjà vu au chapitre 3 dans la section sur l'estimation semi-paramétrique de densité de probabilité 3.2, l'estimation de la densité d'un échantillon de données par l'estimateur KMCE, que le noyau utilisé soit symétrique ou asymétrique, implique quatre étapes :

1. Estimer les paramètres de la fonction de distribution (cdf) de Champernowne pour ajuster au mieux les données à cette dernière.
2. Transformer les données par la fonction $T_{\hat{\alpha}, \hat{\beta}, \hat{c}}(x)$ (3.5) .
3. Estimer la densité des données transformées par la méthode du noyau.
4. Effectuer une rétrotransformation après l'estimation à noyau pour revenir à l'échelle originale des données.

Dans notre programme d'estimation, nous avons calculé les estimations et mesuré les erreurs pour $R = 2000$ répliques. Nous avons ensuite calculé, pour chaque taille d'échantillon N , la moyenne des estimations, la moyenne des erreurs et la moyenne des paramètres de lissage sur l'ensemble des répliques.

L'algorithme suivant correspond aux itérations fondamentales du programme de simulation.

Début programme

- $T = [50, 100, 500, 1000]$ (Tailles d'échantillons);
 - $R = 2000$; (Nombre de réplifications)
 - $G_b = [b_{min}, 1]$; (Intervalle d'évaluation de b)
 - $G_x = [0.001, 7]$; (Intervalle de x)
- I. Pour chaque taille d'échantillon $N \in T$:

I.J. Pour chaque réplication $r \in \{1, \dots, R\}$:

I.J.1. Génération des données

- Simuler $D_N = X_1, \dots, X_N \sim Burr(1, 1, 1)$; (Exemple avec la distribution de Burr)

I.J.2. Estimation paramétrique préliminaire

- Estimer $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{c})$ par maximum de vraisemblance où $\hat{\beta} = \text{médiane}(D_N)$ et $(\hat{\alpha}, \hat{c}) = \text{argmax } L(\alpha, c | D_N, \hat{\beta})$

I.J.3. Estimation semi-paramétrique et non paramétrique.

I.J.3.K. Pour chaque estimateur $E \in \{EKMCE, BKMCE, EKDE, GKDE\}$:

I.J.3.K.1. Recherche de bande passante optimale.

I.J.3.K.1.L. Pour chaque valeur b dans l'intervalle G_b (avec un pas p_b) :

- Estimer la densité;
- Calculer le ISE de chaque estimation;

Fin-Pour (boucle L);

- Déterminer la valeur optimale b^* qui minimise le ISE ($b^* = \text{argmin}_{(b \in G_b)} ISE[E, f]$);

I.J.3.K.2. Calcul des estimations

I.J.3.K.2.M Pour chaque $x \in G_x$ (avec un pas p_x) :

- $\hat{f}(n, r, e) = E[x, D_N, \hat{\theta}, b_{opt}]$;
- (n : indice d'échantillon, r : indice de réplification, e : indice d'estimateur).

Fin-Pour (boucle M);

I.J.3.K.3. Mesure d'erreur des estimations

- Calculer le ISE(n,r,e), $L_1(n,r,e)$, WISE(n,r,e);

Fin-pour (boucle K);

Fin-pour (boucle J);

- $ISE(n,e) = \frac{1}{R} \sum_{r=1}^R ISE(n,r,e)$;
- $L_1(n,e) = \frac{1}{R} \sum_{r=1}^R L_1(n,r,e)$;
- $WISE(n,e) = \frac{1}{R} \sum_{r=1}^R WISE(n,r,e)$;
- $b^*(n,e) = \frac{1}{R} \sum_{r=1}^R b^*(n,r,e)$;

Fin-pour (boucle I);

- Stocker les résultats dans l'ordinateur sous forme de fichier Excel (Matrice 4×16).

Fin programme.

4.4 Résultats de Simulation

N	Noyaux	Erreur	Exponentielle	Inverse Gaussienne	Weibull-Fréchet	Burr
50	EKMCE	ISE (b^*)	0.0313 (0.2593)	0.0225 (0.2228)	0.0284 (0.3033)	0.0211 (0.2735)
		L_1	0.2001	0.2274	0.2035	0.1931
		WISE	0.1324	0.1803	0.1290	0.1594
	BKMCE	ISE (b^*)	0.0209 (0.0974)	0.0186 (0.0691)	0.0289 (0.0829)	0.0169 (0.09373)
		L_1	0.1558	0.1868	0.1821	0.1505
		WISE	0.0956	0.1303	0.1005	0.1093
	EKDE	ISE (b^*)	0.0420 (0.3955)	0.0331 (0.4624)	0.0307 (0.4155)	0.0474 (0.4117)
		L_1	0.2621	0.3617	0.2653	0.4746
		WISE	0.2256	0.5269	0.3273	1.1073
	GKDE	ISE b^*	0.0423 (0.1982)	0.0328 (0.2297)	0.0320 (0.2051)	0.0473 (0.2113)
		L_1	0.2584	0.3535	0.2646	0.4574
		WISE	0.2199	0.5094	0.3220	1.0572
100	EKMCE	ISE (b^*)	0.0206 (0.2088)	0.0150 (0.1725)	0.0197 (0.2618)	0.0167 (0.2188)
		L_1	0.1628	0.1884	0.1686	0.1590
		WISE	0.1078	0.1501	0.1091	0.1322
	BKMCE	ISE (b^*)	0.0099 (0.0980)	0.0117 (0.0513)	0.0170 (0.0696)	0.0081 (0.09369)
		L_1	0.1107	0.1503	0.1397	0.1067
		WISE	0.0685	0.1041	0.0763	0.0790
	EKDE	ISE (b^*)	0.0308 (0.2958)	0.0226 (0.3433)	0.0211 (0.3498)	0.0359 (0.2995)
		L_1	0.2267	0.3105	0.2222	0.4158
		WISE	0.1924	0.4463	0.2773	0.9426
	GKDE	ISE (b^*)	0.0308 (0.1416)	0.0230 (0.1741)	0.0223 (0.1698)	0.0360 (0.1432)
		L_1	0.2268	0.3060	0.2243	0.4125
		WISE	0.1939	0.4389	0.2796	0.9526

TABLE 4.2 – Comparaison des erreurs d'estimation pour quatre distributions, pour des échantillons de taille $N=50$ et $N=100$.

N	Estimateur	Erreur	Exponentielle	Inverse Gaussienne	Weibull-Fréchet	Burr
500	EKMCE	ISE (b^*)	0.0086 (0.1079)	0.0057 (0.0953)	0.0135 (0.1463)	0.0108 (0.1138)
		L_1	0.1026	0.1175	0.1165	0.1052
		WISE	0.0669	0.0956	0.0707	0.0819
	BKMCE	ISE (b^*)	0.0019 (0.0978)	0.0040 (0.0180)	0.0063 (0.0247)	0.0016 (0.0929)
		L_1	0.0502	0.0890	0.0837	0.0484
		WISE	0.0312	0.0633	0.0447	0.0356
	EKDE	ISE (b^*)	0.0139 (0.1397)	0.0081 (0.2048)	0.0122 (0.2312)	0.0193 (0.1591)
		L_1	0.1559	0.1932	0.1514	0.2849
		WISE	0.1334	0.2620	0.1721	0.5929
	GKDE	ISE (b^*)	0.0151 (0.1002)	0.0081 (0.1004)	0.0124 (0.1064)	0.0199 (0.1001)
		L_1	0.1360	0.1904	0.1496	0.2515
		WISE	0.1021	0.2557	0.1672	0.4905
1000	EKMCE	ISE (b^*)	0.0059 (0.0786)	0.0036 (0.0728)	0.0130 (0.0826)	0.0090 (0.0790)
		L_1	0.0831	0.0936	0.1030	0.0918
		WISE	0.0534	0.0756	0.0560	0.0682
	BKMCE	ISE (b^*)	0.0009 (0.0979)	0.0024 (0.0134)	0.0040 (0.0154)	0.0007 (0.0938)
		L_1	0.0355	0.0694	0.0675	0.0330
		WISE	0.0222	0.0501	0.0368	0.0243
	EKDE	ISE (b^*)	0.0094 (0.1049)	0.0053 (0.1677)	0.0112 (0.1574)	0.0153 (0.1126)
		L_1	0.1257	0.1569	0.1365	0.2400
		WISE	0.1062	0.2121	0.1605	0.4830
	GKDE	ISE (b^*)	0.0127 (0.1000)	0.0056 (0.1000)	0.0119 (0.1016)	0.0174 (0.1000)
		L_1	0.1069	0.1463	0.1250	0.1941
		WISE	0.0727	0.1814	0.1259	0.3475

TABLE 4.3 – Comparaison des erreurs d'estimation pour quatre distributions, pour des échantillons de taille $N=500$ et $N=1000$.

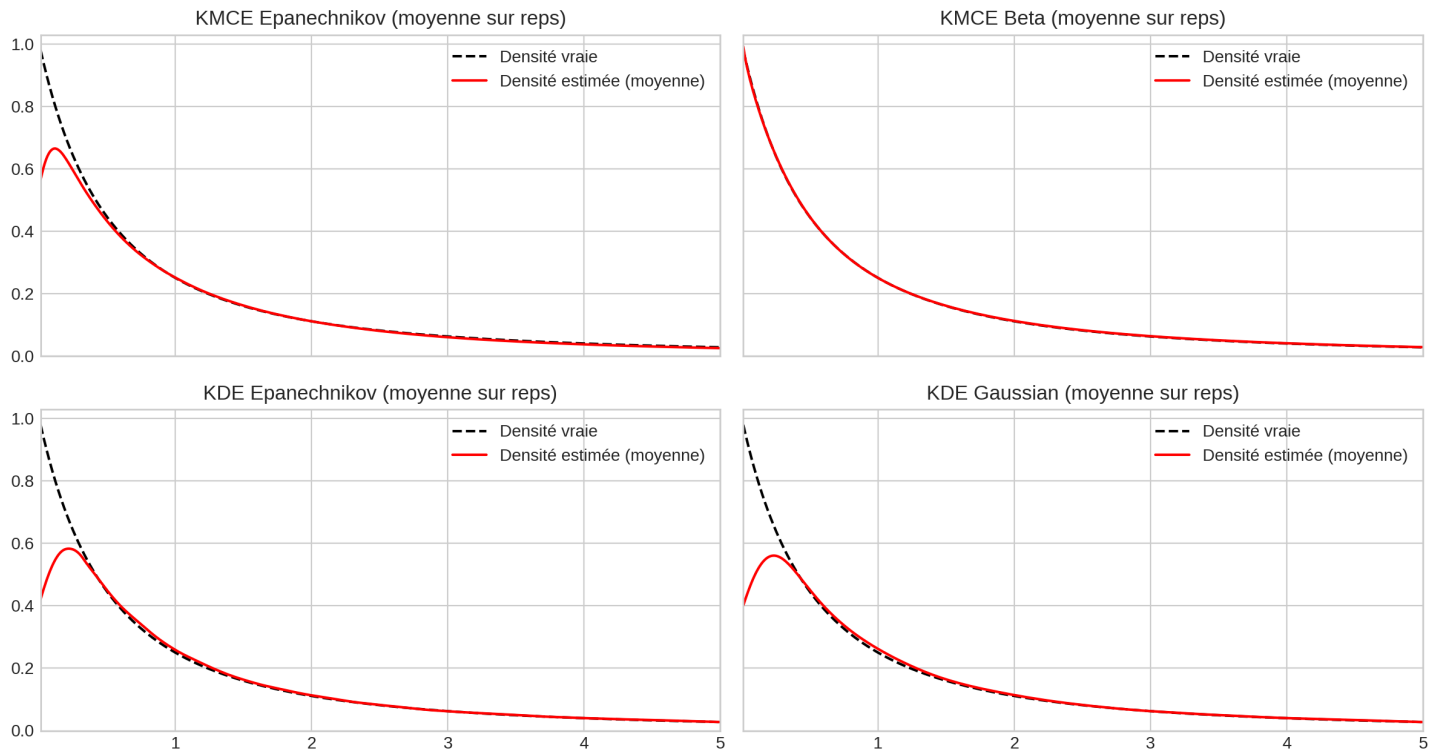


FIGURE 4.4 – Comparaison des estimations de densité d’une loi de Burr pour un échantillon de taille $N = 50$

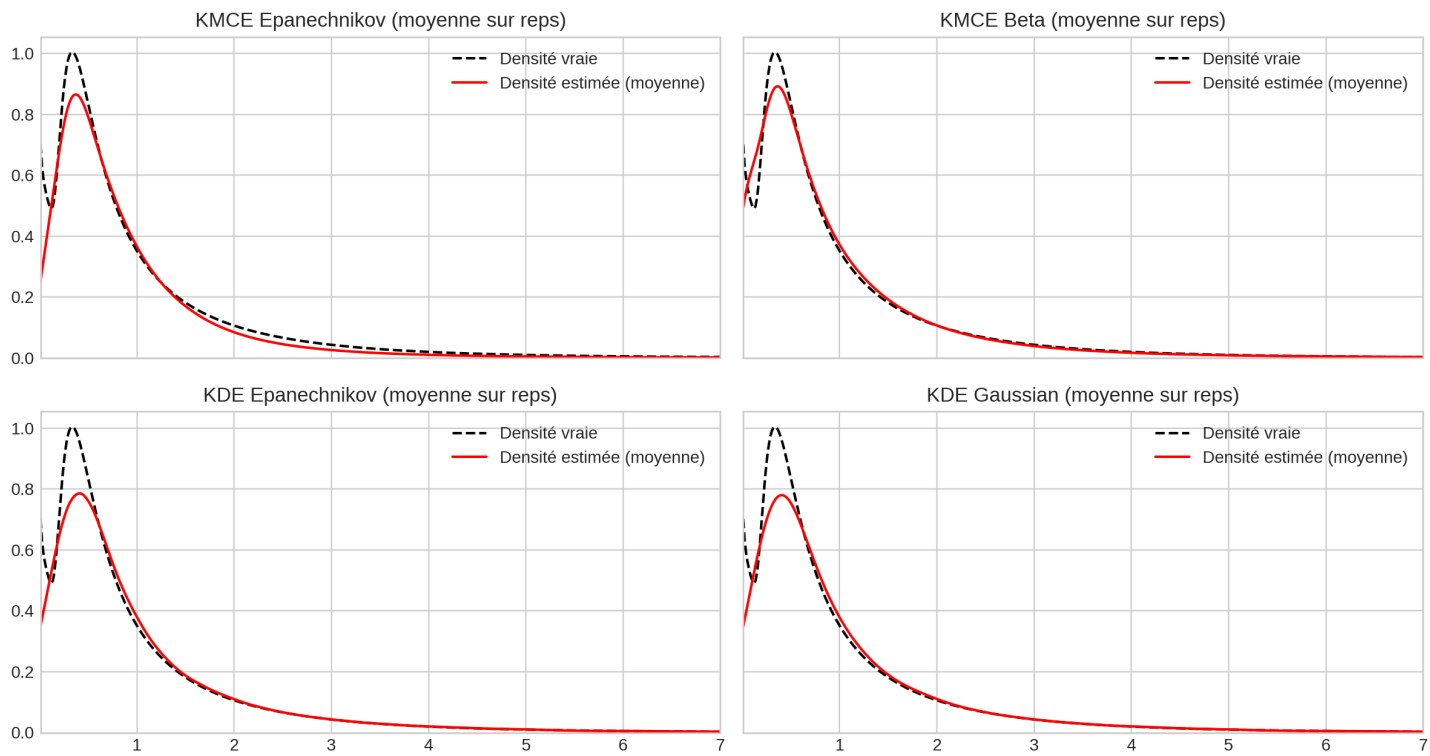


FIGURE 4.5 – Comparaison des estimations de densité d’une loi de Mixte Weibull-Fréchet pour un échantillon de taille $N = 50$

4.4.1 Interprétation des Résultats de Simulation

L'analyse des tableaux de simulation, comparant les performances des estimateurs EKMCE, BKMCE, EKDE et GKDE sur quatre distributions (Exponentielle, Inverse Gaussienne, Weibull-Fréchet et Burr) pour des tailles d'échantillon variant de $N = 50$ à $N = 1000$, permet de dégager plusieurs conclusions majeures :

1. Convergence Asymptotique (Effet de la taille d'échantillon)

Une tendance générale uniforme est observée pour l'ensemble des estimateurs : les métriques d'erreur (ISE, L_1 et WISE) décroissent systématiquement à mesure que la taille de l'échantillon N augmente. Par exemple, pour la distribution Burr avec l'estimateur BKMCE, l'ISE passe de 0.0169 ($N = 50$) à 0.0007 ($N = 1000$).

2. Supériorité de l'Approche par Transformation (KMCE vs KDE)

Les résultats mettent en évidence l'avantage significatif des méthodes basées sur la transformation de Champernowne (EKMCE et BKMCE) par rapport aux méthodes classiques (EKDE et GKDE). Les estimateurs classiques souffrent particulièrement sur les distributions à queue lourde (Burr, Weibull-Fréchet). Pour $N = 1000$ sur la loi Burr, le GKDE affiche un WISE de 0.3475 contre seulement 0.0243 pour le BKMCE. Cela confirme que la transformation permet de mieux capturer la structure des données asymétriques positives, là où les noyaux classiques à fenêtre fixe peinent à s'adapter à la fois au pic de densité et à la queue de distribution.

3. Prééminence de l'Estimateur à Noyau Bêta (Réduction du Biais aux Bords)

Le résultat le plus marquant de cette étude est la surperformance systématique de l'estimateur BKMCE sur tous les autres estimateurs, et ce, quelle que soit la distribution ou la taille de l'échantillon.

Contrairement à l'EKMCE (qui utilise un noyau symétrique Epanechnikov sur des données transformées), le BKMCE utilise un noyau Bêta qui s'adapte naturellement aux bornes de l'intervalle $[0, 1]$ après transformation. Cela se traduit par des erreurs L_1 et ISE nettement plus faibles. Par exemple, pour la loi Exponentielle ($N = 500$), l'ISE du BKMCE (0.0019) est environ 4 fois inférieur à celui de l'EKMCE (0.0086).

4. Analyse de la Robustesse via le WISE

La métrique WISE (Weighted Integrated Squared Error, avec le poids x^2), qui accorde plus d'importance aux erreurs sur les valeurs extrêmes, confirme l'efficacité de l'estimateur BKMCE. Avec la loi de Burr, connue pour avoir beaucoup de valeurs extrêmes, la différence est nette : pour 500 données ($N = 500$), l'erreur du BKMCE est minimale (0.0356), alors que celle des méthodes classiques devient énorme (0.5929). Cela prouve que le BKMCE est le seul capable de bien représenter les événements rares sans perdre en précision sur les valeurs courantes.

Conclusion

L'analyse des résultats démontre de manière convaincante la supériorité des estimateurs par transformation, en particulier l'estimateur BKMCE (l'estimateur à noyau Beta avec transformation de Champernowne modifiée), sur les méthodes KDE classiques. Cette supériorité, qui s'exprime à travers toutes les métriques d'erreur, toutes les distributions testées et toutes les tailles d'échantillon, est particulièrement marquée pour les distributions à queues lourdes et pour les petites tailles d'échantillon.

L'amélioration des performances s'explique par la capacité de la transformation de Champernowne à normaliser le support et à réduire l'asymétrie des données, permettant ainsi une estimation par noyau plus efficace dans l'espace transformé. L'utilisation combinée de cette transformation avec un noyau Beta adaptatif aux bornes constitue une avancée méthodologique significative dans le domaine de l'estimation non paramétrique de densité sur support positif.

Ces résultats justifient l'adoption des estimateurs par transformation dans les applications pratiques nécessitant l'estimation de densités asymétriques sur support positif, particulièrement en finance, en assurance et en sciences environnementales où de telles distributions sont fréquentes.

Conclusion générale

Ce mémoire a exploré des outils statistiques pour analyser des données présentant des événements rares mais coûteux, comme des crises financières ou des catastrophes naturelles. Le but était de mieux modéliser l'ensemble de la distribution de ce type de données, et pas seulement leurs valeurs extrêmes.

La première partie a présenté la théorie des valeurs extrêmes, qui offre des modèles mathématiques spécifiques pour quantifier le risque d'événements exceptionnels. Cette approche est précise pour les queues de distribution, mais elle ne décrit pas toute la forme des données.

Pour obtenir une vision plus complète, la deuxième partie a introduit les méthodes d'estimation non paramétrique de la densité, comme l'histogramme et l'estimateur à noyau. Ces techniques sont flexibles et s'adaptent aux données, mais elles sont souvent imprécises dans l'estimation des valeurs extrêmes lorsque les observations sont peu nombreuses.

La troisième partie a proposé une synthèse pragmatique (l'estimateur KMCE). Cette méthode hybride transforme d'abord les données pour mieux gérer les valeurs extrêmes, puis applique une estimation par noyau classique ou asymétrique. Elle combine ainsi la robustesse des modèles paramétriques pour les queues de distribution avec la flexibilité des méthodes non paramétriques pour le coeur de la distribution. Les simulations ont confirmé l'intérêt de cette approche.

En résumé, ce travail montre qu'une stratégie en deux étapes (transformer puis lisser) est efficace pour estimer la densité de données complexes. L'estimateur KMCE, et en particulier sa version BKMCE (estimateur à noyau Beta avec transformation de Champernowne modifiée), se révèle un outil pertinent pour les domaines où la modélisation des risques extrêmes est cruciale, comme la finance, l'assurance ou la gestion des catastrophes. Il permet de capturer avec précision le comportement des valeurs extrêmes tout en offrant une estimation fidèle de la distribution globale des données, répondant ainsi à l'objectif initial de ce mémoire.

Bibliographie

- [1] Axler, S. (2020). *Measure, integration & real analysis*. Springer Nature.
- [2] Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. L. (2006). *Statistics of extremes : theory and applications*. John Wiley & Sons.
- [3] Benlagha, N., Grun-Réhomme, M., et al. (2007). Application de la théorie des valeurs extrêmes en assurance automobile. *EURO-MEDITERRANEAN ECONOMICS AND FINANCE REVIEW*, page 40.
- [4] Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1989). *Regular variation*, volume 27. Cambridge university press.
- [5] Bolancé, C. (2010). Optimal inverse beta (3, 3) transformation in kernel density estimation. *SORT-Statistics and Operations Research Transactions*, 34(2) :223–238.
- [6] Bolancé, C., Guillen, M., and Nielsen, J. P. (2003). Kernel density estimation of actuarial loss functions. *Insurance : Mathematics and Economics*, 32(1) :19–36.
- [7] Bolancé, C., Guillén, M., and Nielsen, J. P. (2008). Inverse beta transformation in kernel density estimation. *Statistics & Probability Letters*, 78(13) :1757–1764.
- [8] Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2) :353–360.
- [9] Box, G. E. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B : Statistical Methodology*, 26(2) :211–243.
- [10] Brown, B. M. and Chen, S. X. (1999). Beta-bernstein smoothing for regression curves with compact support. *Scandinavian Journal of Statistics*, 26(1) :47–59.

- [11] Buch-Larsen, T., Nielsen, J. P., Guillén, M., and Bolancé, C. (2005). Kernel density estimation for heavy-tailed distributions using the champernowne transformation. *Statistics*, 39(6) :503–516.
- [12] Charras-Garrido, M. and Lezaud, P. (2013). Extreme value analysis : an introduction. *Journal de la société française de statistique*, 154(2) :66–97.
- [13] Chen, S. X. (1999). Beta kernel estimators for density functions. *Computational Statistics & Data Analysis*, 31(2) :131–145.
- [14] Chen, S. X. (2000a). Beta kernel smoothers for regression curves. *Statistica Sinica*, pages 73–91.
- [15] Chen, S. X. (2000b). Probability density function estimation using gamma kernels. *Annals of the institute of statistical mathematics*, 52(3) :471–480.
- [16] Danielsson, J. and De Vries, C. G. (2000). Value-at-risk and extreme returns. *Annales d’Economie et de Statistique*, pages 239–270.
- [17] Davis, R. A., Lii, K.-S., and Politis, D. N. (2011). Remarks on some nonparametric estimates of a density function. In *Selected Works of Murray Rosenblatt*, pages 95–100. Springer.
- [18] Demir, S. (2019). Adaptive kernel density estimation with generalized least square cross-validation. *Hacettepe Journal of Mathematics and Statistics*, 48(2) :616–625.
- [19] El Adlouni, S.-E., Bobée, B., and Ouarda, T. B. (2007). Caractérisation des distributions à queue lourde pour l’analyse des crues.
- [20] Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling extremal events : for insurance and finance*, volume 33. Springer Science & Business Media.
- [21] Foss, S., Korshunov, D., and Zachary, S. S. (2013). *An introduction to heavy-tailed and subexponential distributions*. Springer.
- [22] Freedman, D. and Diaconis, P. (1981). On the histogram as a density estimator : L 2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(4) :453–476.
- [23] Geenens, G. (2014). Probit transformation for kernel density estimation on the unit interval. *Journal of the American Statistical Association*, 109(505) :346–358.
- [24] Ghamghami, M., Hejabi, S., Rahimi, J., Bazrafshan, J., and Olya, H. (2017). Modeling a drought index using a nonparametric approach. *Global Nest Journal*, 19.
- [25] Gnedenko, B. (1943). Sur la distribution limite du terme maximum d’une serie aleatoire. *Annals of mathematics*, 44(3) :423–453.

- [26] Gramacki, A. (2018). *Nonparametric kernel density estimation and its computational aspects*, volume 37. Springer.
- [27] Gumbel, E. J. (2004). *Statistics of extremes*. Courier Corporation.
- [28] Gustafsson, J., Haggmann, M., Nielsen, J., and Scaillet, O. (2009). Local transformation kernel density estimation of loss distributions. *Journal of Business & Economic Statistics*, 27(2) :161–175.
- [29] Härdle, W. (2004). *Nonparametric and semiparametric models*. Springer Science & Business Media.
- [30] Hirukawa, M. and Sakudo, M. (2015). Family of the generalised gamma kernels : a generator of asymmetric kernels for nonnegative data. *Journal of Nonparametric Statistics*, 27(1) :41–63.
- [31] Igarashi, G. (2016). Weighted log-normal kernel density estimation. *Communications in Statistics-Theory and Methods*, 45(22) :6670–6687.
- [32] Igarashi, G. and Kakizawa, Y. (2017). Amoroso kernel density estimation for nonnegative data and its bias reduction. *Department of Policy and Planning Sciences Discussion Paper Series*, (1345).
- [33] Karunamuni, R. J. and Alberts, T. (2005). A generalized reflection method of boundary correction in kernel density estimation. *Canadian Journal of Statistics*, 33(4) :497–509.
- [34] Leadbetter, M. R. (1991). On a basis for ‘peaks over threshold’ modeling. *Statistics & Probability Letters*, 12(4) :357–362.
- [35] Leadbetter, M. R., Lindgren, G., and Rootzén, H. (2012). *Extremes and related properties of random sequences and processes*. Springer Science & Business Media.
- [36] Li, Q. and Racine, J. S. (2007). *Nonparametric econometrics : theory and practice*. Princeton University Press.
- [37] Loquin, K. and Strauss, O. (2008). On the granularity of summative kernels. *Fuzzy sets and systems*, 159(15) :1952–1972.
- [38] Marron, J. S. and Ruppert, D. (1994). Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society : Series B (Methodological)*, 56(4) :653–671.
- [39] McNeil, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin : The Journal of the IAA*, 27(1) :117–137.

- [40] Mentel, G. (2013). Parametric or non-parametric estimation of value-at-risk. *International Journal of Business and Management*, 8 :103–103.
- [41] Mousa, A., Hassan, M. K., and Fathi, A. (2016). A new non parametric estimator for pdf based on inverse gamma distribution. *Communications in Statistics-Theory and Methods*, 45(23) :7002–7010.
- [42] Parzen, E. (1962). On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3) :1065–1076.
- [43] Pavlides, A., Koltsidopoulou, M. D., Chrysanthi, M., and Varouchakis, E. A. (2025). A kernel-based non-parametric approach for data gaussian anamorphosis. *Mathematical Geosciences*, pages 1–16.
- [44] Rachev, S. T. (2003). *Handbook of heavy tailed distributions in finance : Handbooks in finance, Book 1*, volume 1. Elsevier.
- [45] Raggad, B. (2009). Fondements de la théorie des valeurs extrêmes, ses principales applications et son apport à la gestion des risques du marché pétrolier. *Mathématiques et sciences humaines. Mathematics and social sciences*, (186) :29–63.
- [46] Reiss, R.-D. (2012). *Approximate distributions of order statistics : with applications to nonparametric statistics*. Springer science & business media.
- [47] Resnick, S. I. (2008). *Extreme values, regular variation, and point processes*, volume 4. Springer Science & Business Media.
- [48] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27 :832–837.
- [49] Ruppert, D. and Cline, D. B. (1994). Bias reduction in kernel density estimation by smoothed empirical transformations. *The Annals of Statistics*, pages 185–210.
- [50] Scaillet, O. (2004). Density estimation using inverse and reciprocal inverse gaussian kernels. *Nonparametric statistics*, 16(1-2) :217–226.
- [51] Scott, D. W. (1979). On optimal and data-based histograms. *Biometrika*, 66(3) :605–610.
- [52] Scott, D. W. (2015). *Multivariate density estimation : theory, practice, and visualization*. John Wiley & Sons.
- [53] Shahbaz, M. Q., Ahsanullah, M., Shahbaz, S. H., and Al-Zahrani, B. M. (2016). Ordered random variables : Theory and applications.
- [54] Simonoff, J. (2012). *Smoothing Methods in Statistics*. Springer Series in Statistics. Springer New York.

-
- [55] Skorski, M. (2019). Kernel density estimation bias under minimal assumptions. *arXiv preprint arXiv :1901.00331*.
- [56] Sturges, H. A. (1926). The choice of a class interval. *Journal of the american statistical association*, 21(153) :65–66.
- [57] Teamah, A., Elbanna, A. A., and Gemeay, A. M. (2020). Frechet-weibull mixture distribution : properties and applications. *Applied Mathematical Sciences*, 14(2) :75–86.
- [58] Terrell, G. R. and Scott, D. W. (1992). Variable kernel density estimation. *The Annals of Statistics*, pages 1236–1265.
- [59] Wand, M. P. and Jones, M. C. (1994). *Kernel smoothing*. CRC press.
- [60] Wand, M. P., Marron, J. S., and Ruppert, D. (1991). Transformations in density estimation. *Journal of the American Statistical Association*, 86(414) :343–353.
- [61] Węglarczyk, S. (2018). Kernel density estimation and its application. In *ITM web of conferences*, volume 23, page 00037. EDP Sciences.
- [62] Wen, K. and Wu, X. (2015). An improved transformation-based kernel estimator of densities on the unit interval. *Journal of the American Statistical Association*, 110(510) :773–783.
- [63] Yeo, I.-K. and Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4) :954–959.
- [64] Yu, K., Ally, A., Yang, S., and Hand, D. (2010). Kernel quantile-based estimation of expected shortfall. *Journal of Risk*, 12.

Résumé

Ce mémoire a pour objectif de présenter une méthode unifiée d'estimation de densité de probabilité, capable de modéliser simultanément les événements rares et coûteux ainsi que les événements fréquents au sein d'une même distribution à queue lourde. En effet, les méthodes traditionnelles peinent souvent à concilier ces deux aspects : elles privilégient généralement soit la modélisation de la partie extrême, soit celle de la partie centrale, mais rarement les deux conjointement. Ce document propose une solution via l'estimation par transformation. Cette approche semi-paramétrique repose sur trois étapes : d'abord, une transformation des données via la distribution de Champernowne modifiée, ajustée à l'échantillon ; ensuite, une estimation par noyau sur ces données transformées ; enfin, une rétro-transformation préservant la cohérence avec l'échelle originale. Nous étudions ici l'utilisation de noyaux issus des familles symétriques et des noyaux bêta. Les simulations démontrent que l'estimateur résultant, le KMCE, réduit significativement l'écart entre la vraie densité et la densité estimée, particulièrement avec l'usage du noyau bêta. Cette robustesse se vérifie sur divers types de distributions, confirmant l'efficacité de la méthode pour modéliser des scénarios réels complexes.

Mots-clés : Estimation semi-paramétrique – transformation de Champernowne – estimation de la densité – valeurs extrêmes.

Abstract

The objective of this thesis is to present a unified method for probability density estimation, capable of simultaneously modeling rare and costly events as well as frequent occurrences within a single heavy-tailed distribution. Indeed, traditional methods often struggle to reconcile these two aspects ; they generally prioritize modeling either the extreme tail or the central body of the distribution, but rarely handle both jointly with equal accuracy. This document proposes a solution via transformation-based estimation. This semi-parametric approach relies on three steps : first, a data transformation using the modified Champernowne distribution, fitted to the sample ; next, kernel estimation applied to the transformed data ; and finally, a back-transformation preserving consistency with the original scale. We investigate the use of both symmetric kernels and Beta kernels. Simulations demonstrate that the resulting estimator, the KMCE, significantly reduces the discrepancy between the true density and the estimated density, particularly when using the Beta kernel. This robustness is observed across various types of distributions, confirming the method's effectiveness in modeling complex real-world scenarios.

Keywords : Semi-parametric estimation – Champernowne transformation – density estimation – extreme values.