

# RÉDUCTION DU BIAIS DANS L'ESTIMATION NON PARAMÉTRIQUE DES DENSITÉS HEAVY TAILED PAR LA MÉTHODE DU NOYAU

Smail ADJABI<sup>a</sup>, Nabil ZOUGAB<sup>a</sup>

<sup>a</sup>Unité de recherche LaMOS, Université de Béjaia, Algérie

E-mail : adjabi@hotmail.com, nabilzougab@yahoo.fr

## Résumé

On montre que les techniques multiplicatives de correction du biais (MBC) peuvent être appliquées à l'estimateur à noyau général Birnbaum-Saunders (GBS). Les propriétés des estimateurs à noyau MBS-GBS (biais, variance et erreur quadratique moyenne intégrée) sont obtenues. On utilise la méthode de validation croisée pour estimer le paramètre de lissage. Les performances des estimateurs MBC-GBS sont comparées selon les critères du biais quadratique intégré (ISB) et de l'erreur quadratique intégrée (ISE) aux performances de l'estimateur GBS standard, par une étude de simulation sur des densités heavy tailed (queue lourde) connues suivie d'une application sur des données réelles non négatives heavy tailed.

**Mots clés :** Noyau General Birnbaum-Saunders (GBS), Techniques multiplicatives de correction du biais (MBC).

## Abstract

We show that the multiplicative bias correction (MBC) techniques can be applied for generalized Birnbaum-Saunders (GBS) kernel density estimation. Some properties of the MBC-GBS kernel density estimators (bias, variance and mean integrated squared error) are shown. The choice of bandwidth is investigated by adopting the cross-validation technique. The performances of the MBC estimators based on GBS kernel are illustrated by a simulation study, followed by a real application for nonnegative heavy tailed data, in terms of integrated squared bias (ISB) and integrated squared error (ISE).

**Key words :** General Birnbaum-Saunders (GBS) kernel, Multiplicative Bias Correction techniques (MBC).

## Introduction

L'analyse et l'estimation de la densité des données heavy tailed (queue lourde) sont complexes en raison de leurs caractéristiques spécifiques qui sont la décroissance lente vers zéro et présence d'observations rares dans la queue. Dans ce travail, on estime les densités du type heavy tailed par la méthode non paramétrique du noyau introduite par Parzen(1956) et par Chen(1999), en utilisant le noyau asymétrique GBS (Generalized Birnbaum-Saunders), introduit par Marchant et al. (2013), car il est le plus approprié pour estimer les densités heavy tailed non négatives, en raison de sa flexibilité et de ses propriétés. Ce noyau GBS comprend les noyaux : BS-classique, BS-Power-Exponential (BS-PE) et BS-Student-t (BS-t). Cet estimateur à noyau GBS de la densité peut s'appliquer dans divers domaines tel que : l'économie (Hirukawa et Sakudo (2014)), la finance (Jin and Kawczak (2003)), l'actuariat et les sciences de l'environnement (Saulo et al(2013)). Pour corriger le biais de l'estimateur ainsi obtenu, on applique les techniques MBC (Multiplicative Bias Correction) de Terrell et Scott (1980) et Jones, Linton et Nielsen (1995). Nous obtenons deux nouveaux estimateurs de la densité, que nous noterons TS-GBS et JLN-GBS et nous présentons leurs propriétés statistiques : biais, variance et erreur quadratique moyenne intégrée (Zougab et Adjabi (2016)). Le paramètre de lissage qui intervient dans la forme de l'estimateur de la densité sera estimé par la méthode de validation croisée sans biais (UCV). Pour comparer les estimateurs MBC-GBS avec l'estimateur standard GBS, on utilise les critères : erreur quadratique intégrée ISE et le biais quadratique intégré ISB sur des données simulées à partir de densités cibles heavy tailed et sur des données réelles environnementales heavy tailed (Leiva et al(2010)).

# 1 Estimateur à noyau GBS

Soit  $X_1, \dots, X_n$  une séquence de variables aléatoires indépendantes et identiquement distribuées de densité inconnue  $f$ . L'estimateur à noyau asymétrique introduit par Chen(1999) est de la forme :

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_{x,h}(X_i), \quad (1)$$

où  $K_{x,h}(\cdot) = (1/h)K(x - \cdot)/h$ ,  $h > 0$  est le paramètre de lissage, il joue le rôle de paramètre de dispersion autour de la cible  $x$  et  $K$  est une fonction appelée noyau, elle détermine la forme du voisinage autour du point  $x$ . Comme le type de données peut fortement influencer sur le choix du noyau, Marchant et al(2013) ont suggéré que le noyau le plus approprié aux données qui se caractérisent par une queue du type heavy tailed est le noyau GBS (General Birnbaum Saunders) défini par :

$$K_{GBS(h,x;\varphi)}(y) = c\varphi \left( \frac{1}{h} \left( \frac{y}{x} + \frac{x}{y} - 2 \right) \right) \frac{1}{2\sqrt{h}} \left( \frac{1}{\sqrt{xy}} + \sqrt{\frac{x}{y^3}} \right), x > 0, h > 0, y > 0.$$

Différentes formes du noyau GBS :

Distribution	$c$	$\varphi = \varphi(u), u > 0$
Classique	$\frac{1}{\sqrt{2\pi}}$	$\exp(-\frac{1}{2}u)$
PE	$\frac{\nu}{2^{2\nu} \Gamma(\frac{1}{2\nu})}$	$\exp(-\frac{1}{2}u^\nu), \nu > 0$
$t$	$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}$	$(1 + \frac{u}{\nu})^{-\frac{(\nu+1)}{2}}, \nu > 0$

L'estimateur à noyau GBS de  $f$  au point  $x > 0$  est alors de la forme :

$$\hat{f}_{GBS}(x) = \frac{1}{n} \sum_{i=1}^n K_{GBS(h,x;\varphi)}(X_i) = \frac{1}{n} \sum_{i=1}^n c\varphi \left( \frac{1}{h} \left( \frac{X_i}{x} + \frac{x}{X_i} - 2 \right) \right) \frac{1}{2\sqrt{h}} \left( \frac{1}{\sqrt{xX_i}} + \sqrt{\frac{x}{X_i^3}} \right).$$

L'expression du biais, de la variance et de l'erreur quadratique moyenne intégrée de l'estimateur  $\hat{f}_{GBS}(x)$  sont obtenus par Marchant et al(2013) :

$$\text{Biais}(\hat{f}_{GBS}(x)) = \frac{hu_1(\varphi)}{2} (xf'(x) + x^2f''(x)) + o(h), \text{Var}(\hat{f}_{GBS}(x)) = \frac{c^2}{c_\varphi^2 nh^{1/2}x} f(x) + o\left(\frac{1}{nh^{1/2}}\right).$$

$$\begin{aligned} \text{MISE}(\hat{f}_{GBS}) &= \int_0^\infty \text{biais}^2(\hat{f}_{GBS}(x))dx + \int_0^\infty \text{Var}(\hat{f}_{GBS}(x))dx \\ &= \frac{h^2}{4} u_1^2(\varphi) \int (xf'(x) + x^2f''(x))^2 dx + \frac{c^2}{c_\varphi^2 nh^{1/2}} \int \frac{1}{x} f(x)dx + o\left(h^2 + \frac{1}{nh^{1/2}}\right), \end{aligned} \quad (2)$$

où

Distribution	$c$	$c_\varphi^2$	$u_1(\varphi)$	$u_2(\varphi)$
Normal	$\frac{1}{\sqrt{2\pi}}$	$\frac{1}{\sqrt{\pi}}$	1	3
PE	$\frac{\nu}{2^{2\nu} \Gamma(\frac{1}{2\nu})}$	$\frac{\nu}{\Gamma(\frac{1}{2\nu})}$	$\frac{2^{1/\nu} \Gamma(\frac{3}{2\nu})}{\Gamma(\frac{1}{2\nu})}$	$\frac{2^{2/\nu} \Gamma(\frac{5}{2\nu})}{\Gamma(\frac{1}{2\nu})}$
t	$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}$	$\frac{\nu\Gamma(\nu)}{\sqrt{\pi}\Gamma(\frac{1+2\nu}{\nu})}$	$\frac{\nu}{\nu-2}, \nu > 2$	$\frac{3\nu^2}{(\nu-2)(\nu-4)}, \nu > 4$

## Méthode de sélection du paramètre de lissage : Validation croisée non biaisée UCV

On utilise la méthode validation croisée non biaisée qui consiste à sélectionner la paramètre de lissage  $h$  qui minimise un estimateur de l'erreur quadratique intégrée  $ISE(h) = \int (\hat{f}_h(x) - f(x))^2 dx$ . Notons  $UCV(h)$  un estimateur de  $ISE(h)$ . L'optimal paramètre de lissage  $h_{UCV}$  est donné par :

$$h_{UCV} = \arg \min_h UCV(h) \text{ où } UCV(h) = \int \left[ \frac{1}{n} \sum_{i=1}^n K_{x,h}(X_i) \right]^2 - \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n K_{X_i,h}(X_j).$$

## 2 Techniques MBC pour la réduction du biais de l'estimateur

On applique les deux techniques MBC (Multiplicative Bias Correction) proposées par Terrell and Scott (1980) et Jones, Linton and Nielsen (1995) pour l'estimateur à noyau GBS.

### 2.1 Estimateurs

L'estimateur à noyau GBS de la densité basé sur la technique MBC de Terrell et Scott que nous noterons TS-GBS est de la forme :

$$\widehat{f}_{TS-GBS}(x) = \left\{ \widehat{f}_{GBS,h}(x) \right\}^{\frac{1}{1-a}} \left\{ \widehat{f}_{GBS,h/a}(x) \right\}^{-\frac{a}{1-a}}, \quad (3)$$

où  $\widehat{f}_{GBS,h}$ ,  $\widehat{f}_{GBS,h/a}$  sont les estimateurs à noyau GBS de paramètre de lissage  $h$  et  $h/a$ , respectivement, avec  $a \in (0, 1)$  constante qui ne dépend pas de la cible  $x$ . Cette technique consiste à construire un estimateur combinaison de deux estimateurs utilisant le même noyau avec différents paramètres de lissage  $h$  et  $h/a$ ,  $a \in (0, 1)$ .

L'estimateur à noyau GBS de la densité basé sur la technique MBC de Jones, Linton et Nielsen que nous noterons JLN-GBS est de la forme :

$$\widehat{f}_{JLN-GBS}(x) = \widehat{f}_{GBS}(x) \left\{ \frac{1}{n} \sum_{i=1}^n \frac{K_{GBS(h,x;\varphi)}(X_i)}{\widehat{f}_{GBS}(X_i)} \right\}, \quad (4)$$

où  $K_{GBS(h,x;\varphi)}$  est un noyau GBS. Cette technique est basée sur l'expression  $f(x) = \widehat{f}(x) \left\{ f(x)/\widehat{f}(x) \right\}$ , où l'on estime le terme biais correction  $f(x)/\widehat{f}(x)$ .

### 2.2 Propriétés asymptotiques

On suppose que :

- A1. les quatre premières dérivées de  $f$  sont continues et bornées.
- A2. La suite  $h = h(n)$  satisfait  $\lim_{n \rightarrow \infty} h = 0$  et  $\lim_{n \rightarrow \infty} nh^{7/2} = \infty$ .

#### Propriété 1

Soit  $\widehat{f}_{TS-GBS}$  l'estimateur TS-GBS à noyau défini par (3). Pour un  $x > 0$  donné et sous les conditions A1 et A2, alors

(i) Le biais de l'estimateur TS-GBS admet le développement suivant :

$$\text{bias}(\widehat{f}_{TS-GBS}) = \frac{1}{a} \left[ \frac{1}{2} \left\{ \frac{l_1^2(x)}{f(x)} - l_2(x) \right\} \right] h^2 + o(h^2),$$

où  $l_1(x) = \frac{1}{2}xu_1(\varphi)f' + \frac{1}{2}x^2u_1(\varphi)f''$  et  $l_2(x) = \frac{1}{4}x^2u_2(\varphi)f''' + \frac{1}{4}x^3u_2(\varphi)f'''' + \frac{1}{24}x^4u_2(\varphi)f'''''$ .

(ii) La variance l'estimateur TS-GBS est donnée par :

$$\text{Var}(\widehat{f}_{TS-GBS}) = \begin{cases} \frac{\psi(a)c^2}{c_{\varphi_2}nh^{1/2}x}f(x) + o\left(\frac{1}{nh^{1/2}}\right) & \text{si } x/h \rightarrow \infty \\ O\{(nh^{3/2})^{-1}\} & \text{si } x/h \rightarrow \kappa, \end{cases}$$

où  $\psi(a) = \frac{(1+a^{5/2})(1+a)^{1/2}-2\sqrt{2}a^{3/2}}{(1+a)^{1/2}(1-a)^2}$  et  $\kappa > 0$  est une constante.

(iii) L'erreur quadratique moyenne intégrée (MISE) de l'estimateur TS-GBS est de la forme :

$$MISE(\widehat{f}_{TS-GBS}) = \frac{h^4}{4a^2} \int \left\{ \frac{l_1^2(x)}{f(x)} - l_2(x) \right\}^2 dx + \frac{\psi(a)c^2}{c_{\varphi_2}nh^{1/2}} \int \frac{f(x)}{x} dx + o\left(h^4 + \frac{1}{nh^{1/2}}\right), \quad (5)$$

où  $l_1(x) = \frac{1}{2}xu_1(\varphi)f' + \frac{1}{2}x^2u_1(\varphi)f''$  et  $l_2(x) = \frac{1}{4}x^2u_2(\varphi)f''' + \frac{1}{4}x^3u_2(\varphi)f'''' + \frac{1}{24}x^4u_2(\varphi)f'''''$ .

### Propriété 2

Soit  $\widehat{f}_{JLN-GBS}$  l'estimateur JLN-GBS à noyau défini par (4). Pour un  $x > 0$  donné et sous les conditions A1 et A2, alors

(i) Le biais de l'estimateur JLN-GBS admet le développement suivant :

$$\text{biais}(\widehat{f}_{JLN-GBS}) = -f(x) \left[ \frac{1}{2}xu_1(\varphi)q'(x) + \frac{1}{2}x^2u_1(\varphi)q''(x) \right] h^2 + o(h^2),$$

où  $q(x) = l_1(x)/f(x)$  et  $l_1(x) = \frac{1}{2}xu_1(\varphi)f' + \frac{1}{2}x^2u_1(\varphi)f''$ .

(ii) La variance l'estimateur TS-GBS est donnée par :

$$\text{Var}(\widehat{f}_{TS-GBS}) = \begin{cases} \frac{c^2}{c_{\varphi^2}nh^{1/2}x}f(x) + o\left(\frac{1}{nh^{1/2}}\right) & \text{si } x/h \rightarrow \infty \\ O\{(nh^{3/2})^{-1}\} & \text{si } x/h \rightarrow \kappa, \end{cases}$$

où  $\kappa > 0$  est une constante.

L'erreur quadratique moyenne intégrée (MISE) de l'estimateur JLN-GBS est de la forme :

$$MISE(\widehat{f}_{JLN-GBS}) = h^4 \int f^2(x) \left[ \frac{1}{2}xu_1(\varphi)q'(x) + \frac{1}{2}x^2u_1(\varphi)q''(x) \right]^2 dx + \frac{c^2}{c_{\varphi^2}nh^{1/2}} \int \frac{f(x)}{x} dx + o\left(h^4 + \frac{1}{nh^{1/2}}\right), \quad (6)$$

où  $q(x) = l_1(x)/f(x)$  avec  $l_1(x) = \frac{1}{2}xu_1(\varphi)f' + \frac{1}{2}x^2u_1(\varphi)f''$ .

## 2.3 Choix du paramètre de lissage : Méthode UCV

L'optimal paramètre de lissage par la méthode UCV est :

$$h_{ucv} = \arg \min_h UCV(h), \text{ où } UCV(h) = \int \widehat{f}_{MBC-GBS}^2(x) dx - \frac{2}{n} \sum_{i=1}^n \widehat{f}_{MBC-GBS}^{(-i)}(X_i),$$

$\widehat{f}_{MBC-GBS}^{(-i)}(y)$  est obtenu comme  $\widehat{f}_{MBC-GBS}(y)$  en excluant l'observation  $X_i$ .

• Pour l'estimateur TS-GBS, la fonction UCV est :

$$\begin{aligned} UCV_{TS-GBS}(h) &= \int \left\{ \widehat{f}_{GBS,h}(x) \right\}^{\frac{2}{1-a}} \left\{ \widehat{f}_{GBS,h/a}(x) \right\}^{-\frac{2a}{1-a}} dx - \frac{2}{n(n-1)} \\ &\times \sum_i \left[ \left\{ \sum_{j \neq i} K_{GBS(h, X_i; \varphi)}(X_j) \right\}^{\frac{1}{1-a}} \left\{ \sum_{j \neq i} K_{GBS((h/a), X_i; \varphi)}(X_j) \right\}^{-\frac{a}{1-a}} \right] \end{aligned}$$

• Dans le cas de l'estimateur JLN-GBS, l'expression de UCV est

$$\begin{aligned} UCV_{JLN-GBS}(h) &= \frac{1}{n^2} \int \widehat{f}_{GBS}(x)^2 \left\{ \sum_{i=1}^n \frac{K_{GBS(h, x; \varphi)}(X_i)}{\widehat{f}_{GBS}(X_i)} \right\}^2 dx - \frac{2}{n(n-1)} \\ &\times \sum_i \sum_{j \neq i} K_{GBS(h, X_i; \varphi)}(X_j) \frac{\widehat{f}_{GBS}(X_i)}{\widehat{f}_{GBS}(X_j)}. \end{aligned}$$

## 3 Simulations

Pour comparer les performances des estimateurs MBC-GBS avec ceux de l'estimateur GBS, on simule des données à partir de trois densités cibles du type heavy tailed avec  $a = 0.5$  :

(a) Burr(1, 3, 1) de densité :  $f(x) = \frac{3x^2}{(1+x^3)^2}$ ,  $x > 0$ , (b) Pareto(1, 1) de densité :  $f(x) = \frac{1}{(1+x)^2}$ ,  $x > 0$ .

(c) GBS(1/2, 3/2;  $t_{\nu=1}$ ) de densité :  $\frac{2}{3\pi} \left[ 1 + \frac{8}{3}x + \frac{6}{x} + 7 \right]^{-1} \left[ \sqrt{\frac{3}{2x}} + \sqrt{\frac{9}{8x^3}} \right]$ ,  $x > 0$ .

Les critères de comparaison sont  $ISE := \int [\widehat{f}(x) - f(x)]^2 dx$  et  $ISB := \int [\mathbb{E}\widehat{f}(x) - f(x)]^2 dx$ .

Densité	Noyau	Estimateur	$n = 25$	$n = 50$	$n = 100$	$n = 200$
Burr	BS-PE( $\nu = 2$ )	$\hat{f}_{GBS}$	0.06118	0.02378	0.01445	0.00762
		$\hat{f}_{TS-GBS}$	0.05276	0.02170	0.01396	0.00680
		$\hat{f}_{JLN-GBS}$	0.05375	0.01980	0.01558	0.00731
Burr	BS-t( $\nu = 5$ )	$\hat{f}_{GBS}$	0.07593	0.03714	0.01874	0.00721
		$\hat{f}_{TS-GBS}$	0.06524	0.03759	0.01624	0.00603
		$\hat{f}_{JLN-GBS}$	0.06976	0.03809	0.01779	0.00726
Pareto	BS-PE( $\nu = 2$ )	$\hat{f}_{GBS}$	0.04864	0.01612	0.01533	0.00694
		$\hat{f}_{TS-GBS}$	0.05510	0.01905	0.01611	0.00717
		$\hat{f}_{JLN-GBS}$	0.04665	0.01672	0.01560	0.00567
Pareto	BS-t( $\nu = 5$ )	$\hat{f}_{GBS}$	0.07505	0.02071	0.01707	0.00705
		$\hat{f}_{TS-GBS}$	0.04623	0.01917	0.01656	0.00610
		$\hat{f}_{JLN-GBS}$	0.04742	0.01890	0.01671	0.00557
GBS	BS-PE( $\nu = 2$ )	$\hat{f}_{GBS}$	0.03453	0.02149	0.01388	0.01107
		$\hat{f}_{TS-GBS}$	0.03945	0.03363	0.02865	0.02688
		$\hat{f}_{JLN-GBS}$	0.03546	0.02411	0.01272	0.00856
GBS	BS-t( $\nu = 5$ )	$\hat{f}_{GBS}$	0.03338	0.02324	0.01481	0.01165
		$\hat{f}_{TS-GBS}$	0.03463	0.02158	0.01427	0.01085
		$\hat{f}_{JLN-GBS}$	0.03729	0.02297	0.01377	0.00925

Table 1. Valeurs du ISE moyen basées sur 1000 répétitions pour Burr, Pareto et GBS

Densité	Noyau	Estimateur	$n = 25$	$n = 50$	$n = 100$	$n = 200$
Burr	BS-PE( $\nu = 2$ )	$\hat{f}_{GBS}$	0.01799	0.00692	0.00155	0.00101
		$\hat{f}_{TS-GBS}$	0.01510	0.00491	0.00134	0.00088
		$\hat{f}_{JLN-GBS}$	<b>0.01448</b>	<b>0.00410</b>	<b>0.00131</b>	<b>0.00055</b>
Burr	BS-t( $\nu = 5$ )	$\hat{f}_{GBS}$	0.01863	0.00600	0.00218	0.00158
		$\hat{f}_{TS-GBS}$	0.01647	0.00488	0.00182	0.00149
		$\hat{f}_{JLN-GBS}$	<b>0.01554</b>	<b>0.00381</b>	<b>0.00118</b>	<b>0.00049</b>
Pareto	BS-PE( $\nu = 2$ )	$\hat{f}_{GBS}$	0.00607	0.00228	0.00208	0.00118
		$\hat{f}_{TS-GBS}$	0.00733	0.00153	0.00144	0.00092
		$\hat{f}_{JLN-GBS}$	<b>0.00371</b>	<b>0.00104</b>	<b>0.00065</b>	<b>0.00047</b>
Pareto	BS-t( $\nu = 5$ )	$\hat{f}_{GBS}$	0.01183	0.00384	0.00214	0.00141
		$\hat{f}_{TS-GBS}$	0.00877	0.00300	0.00179	0.00095
		$\hat{f}_{JLN-GBS}$	<b>0.00411</b>	<b>0.00220</b>	<b>0.00165</b>	<b>0.00074</b>
GBS	BS-PE( $\nu = 2$ )	$\hat{f}_{GBS}$	0.00817	0.00702	0.00534	0.00322
		$\hat{f}_{TS-GBS}$	0.01591	0.01486	0.01173	0.00915
		$\hat{f}_{JLN-GBS}$	<b>0.00608</b>	<b>0.00518</b>	<b>0.00320</b>	<b>0.00195</b>
GBS	BS-t( $\nu = 5$ )	$\hat{f}_{GBS}$	0.00865	0.00681	0.00476	0.00315
		$\hat{f}_{TS-GBS}$	0.00816	0.00629	0.00445	0.00282
		$\hat{f}_{JLN-GBS}$	<b>0.00554</b>	<b>0.00474</b>	<b>0.00228</b>	<b>0.00105</b>

Table 2. Valeurs de ISB moyen basées sur 1000 répétitions pour Burr, Pareto et GBS (en gras les meilleurs résultats)

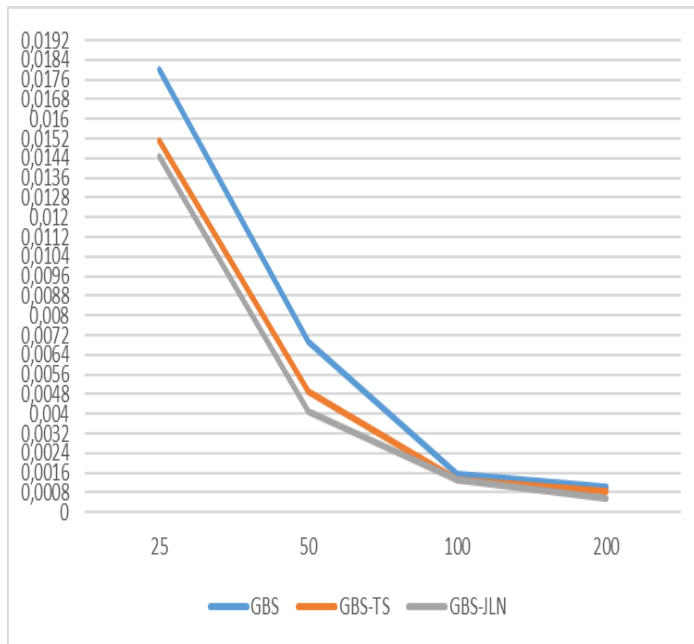


Fig 1. Comparaison du ISB pour la densité de Burr de noyau BS-PE

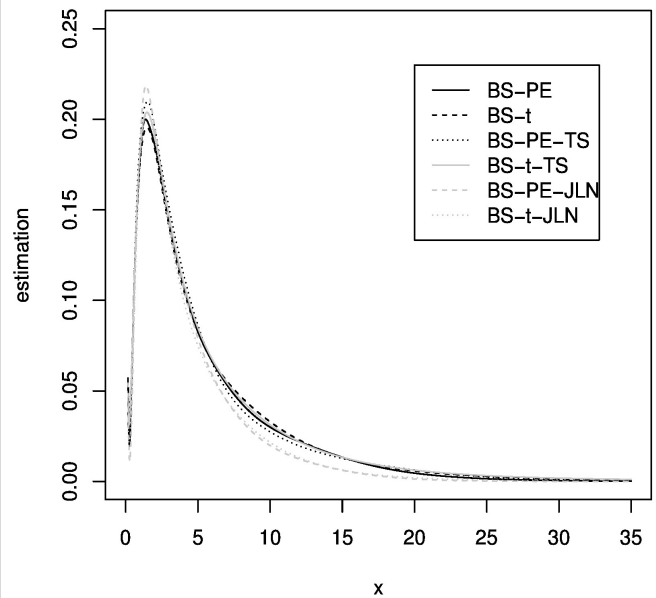


Fig 2. Estimation de la densité des données pollution utilisant le noyau standard BS-PE( $\nu = 2$ ), le standard BS-t( $\nu = 5$ ) le BS-PE( $\nu = 2$ )-TS, le BS-t( $\nu = 5$ )-TS, le BS-PE( $\nu = 2$ )-JLN et le BS-t( $\nu = 5$ )-JLN

A partir des Tables 1, 2 et figure 1 on peut observer que :

- La valeur moyenne de l'erreur quadratique intégrée (ISE) et la valeur moyenne de l'intégrale du biais quadratique (ISB) décroissent quand  $n$  augmente et ceci pour tous les estimateurs.
- Les estimateurs TS-GBS et JLN-GBS sont plus performants que les estimateurs standards GBS au sens du ISE et de l'ISB.
- En général l'estimateur JLN-GBS est meilleur que l'estimateur TS-GBS au sens du biais quadratique (ISB).

## 4 Application aux données de pollution

Ces données concernent l'étude des concentrations journalières ( $n = 116$ ) d'ozone troposphérique  $O_3$  observées à New York entre mai et septembre 1973, fournies par le département de la conservation de l'État de New York (Leiva et al(2010)). La distribution de ces données est une distribution du type heavy tailed. A partir de la Figure 2, on constate que la meilleure qualité de lissage est obtenue avec les estimateurs MBC-GBS. Les estimateurs standards GBS ont tendance à sous-estimer (sous lissage) la densité.

## Conclusion

Dans ce travail, on a appliqué les techniques multiplicatives de correction du biais (MBC) pour l'estimateur à noyaux généralisés de Birnbaum-Saunders (GBS) pour les données non négatives heavy tailed (queue lourde). Nous avons montré que ces deux techniques MBC améliorent l'ordre de grandeur du biais de  $o(h)$  à  $o(h^2)$ . Les performances des estimateurs MBC-GBS (TS-GBS et JLN-GBS) ainsi obtenus, on fait l'objet d'une étude de simulation sur des densités cibles heavy tailed et d'une application réelle sur des données de pollution heavy tailed. Les deux estimateurs MBC-GBS sont plus performants que l'estimateur standard GBS selon les critères de l'erreur quadratique intégrée (ISE) et du biais quadratique intégré (ISB).

## Bibliographie

- [1] Chen, S. X. (1999), *Beta kernels estimators for density functions*, Computational Statistics and Data Analysis, 31, 131-145.
- [3] Hirukawa, M. and Sakudo, M. (2014). *Nonnegative bias reduction methods for density estimation using asymmetric kernels*, Computational Statistics and Data Analysis, 75, 112-123.
- [4] Jin, X. and Kawczak, J. (2003). *Birnbaum-Saunders and lognormal kernel estimators for modelling durations in high frequency financial data*, Annals of Economics and Finance, 4, 103-124.
- [5] Jones, M. C., Linton, O. and Nielsen, J. P. (1995). A simple bias reduction method for density estimation. *Biometrika*, 82, 327-338.
- [6] Leiva, V., Vilca, F., Balakrishnan, N. and Sanhueza, A. (2010), *A skewed sinh-normal distribution and its properties and application to air pollution*, Communications in Statistics-Theory and Methods, 39, 426-443.
- [7] Marchant, C., Bertin, K. and Leiva Vsaulo, H. (2013) *Generalized Birnbaum-Saunders kernel density estimators and an analysis of financial data*, Comput Stat Data Anal 63 :1-15.
- [8] Parzen, E. (1962), *On estimation of a probability density function and mode*, Annals of Mathematical Statistics, 33, 1065-1076.
- [9] Saulo, H., Leiva, V., Ziegelmann, F. A., and Marchant, C. (2013), *A nonparametric method for estimating asymmetric densities based on skewed Birnbaum-Saunders distributions applied to environmental data*, Stochastic Environmental Research and Risk Assessment, 7, 1479-1491.
- [10] Terrell, G. R. and Scott, D. W. (1980), *On improving convergence rates for nonnegative kernel density estimators*, Annals of Statistics, 8, 1160-1163.
- [11] Zougab, N. and Adjabi, S. (2016), *Multiplicative bias correction for generalized Birnbaum-Saunders kernel density estimators and application to nonnegative heavy tailed data*, J Korean Stat Soc 45 :51-63.