Bootstrap dans l'estimation de la densité par la méthode du noyau

Smail Adjabi, Mouloud Cherfaoui

Bootstrap dans l'estimation de la densité par la méthode du noyau

Smail Adjabi, Mouloud Cherfaoui
Laboratoire LAMOS, Université de Béjaia, Algérie
E-mail: adjabi@hotmail.com

Résumé

Dans ce travail, on compare par simulation sur des densités cibles selon le critère de l'erreur quadratique moyenne intégrée, trois méthodes de sélection du paramètre de lissage qui interviennent dans l'estimation de la densité par la méthode du noyau : la méthode plug-in de Sheather and Jones et les deux méthodes de validation croisée : la validation croisée non biaisée et la validation croisée biaisée, en utilisant la technique de ré échantillonnage de bootstrap. On étudie aussi l'influence du noyau selon le même critère sur les performances de l'estimateur de la densité particulièrement pour les densités à support compact.

Mots-clés : Bootstrap, densité, noyau, paramètre de lissage, erreur relative globale, validation croisée, plug-in

Abstract

In this work, we compare by simulation, using the criterion of integrated mean squared error, three methods for selecting the smoothing parameter involved in the estimation of density by the kernel method : the method plug-in Sheather and Jones and two methods of cross validation : the unbiased cross-validation and biased cross-validation, using the technique of bootstrap resampling. It also examines the influence of the kernel using the same test on the performance of the estimator of the density especially for densities with compact support.

Keywords : Bootstrap, density, kernel, parameter of smoothing, total relative error, cross-validation, plug-in

1 Introduction

Étant donné un n-échantillon $X_1, X_2, ..., X_n$ de variables aléatoires indépendantes et de même densité f inconnue. Considérons l'estimateur à noyau de Parzen-Rosenblatt de la densité f donné par :

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

où h est le paramètre de lissage et K une fonction densité définie sur $\mathbb{R}$ appelée noyau.
Pour estimer f, il faut choisir le noyau K et le paramètre h. Si le choix du noyau n'est pas un problème, il n'est pas de même pour le choix de la largeur de la fenêtre h qui dépend essentiellement de la taille n de l'échantillon. En effet, dans l'estimation de la densité par cette méthode, se pose avec acuité le problème du choix du paramètre de lissage. Il existe deux familles de méthodes : la famille des méthodes de validation croisées (Cross-Validation) et la famille des méthodes Plug-in.
Dans ce travail, on étudie l'influence de la technique bootstrap sur le choix du paramètre de lissage dans l'estimation de la densité par rapport aux méthodes de sélection ainsi que dans le cas de changement du noyau. Cette technique due à Efron [5] consiste à tirer B échantillons $X_1^b, X_1^b, ..., X_1^b$ $(b = \overline{1, B})$ de l'échantillon de départ de même taille n, ensuite partant des estimateurs $f_h^b(x)$ on détermine le paramètre h de bootstrap qui minimise l'erreur quadratique intégrée moyenne asymptotique de Bootstrap. On étudie également l'influence de la technique bootstrap sur le choix du noyau particulièrement pour les densités à support compact .

2 Choix du paramètre de lissage

2.1 Méthodes plug-in

2.1.1 Choix optimal

La décision d'un choix optimal pour le paramètre de lissage suppose la spécification d'un critère d'erreur qui puisse être optimisé. Le critère à minimiser est l'Erreur Quadratique Moyenne Intégrée MISE. Dans ce cas [7], on obtient

$$MISE(f,f_h) = \int \mathbb{E}\big[f_h(x) - f(x)\big]^2 dx = \frac{h^4}{4}\sigma_K^4 \int (f''(x))^2 dx \; + \; \frac{\int K^2(y)dy}{nh} \; + \; O(h^5 + \frac{1}{n}). \tag{2}$$

L'Erreur Quadratique Moyenne Intégrée Asymptotique est alors de la forme :

$$AMISE = \frac{h^4}{4}\sigma_K^4 R(f'') + \frac{R(K)}{nh}. \tag{3}$$

où σ_K^2 est la variance du noyau K et $R(g) = \int g^2(x)dx$, pour toute fonction g.

Le paramètre de lissage h^* qui minimise l'Erreur Quadratique Moyenne Intégrée Asymptotique est de la forme :

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

La valeur du $AMISE$ optimale $AMISE^* = AMISE(h^*)$ est alors de forme

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

On constate que la largeur de fenêtre optimale h^* dépend de la densité inconnue f au travers du paramètre $R(f'')$. Cette largeur de fenêtre "idéale" (relativement au critère d'erreur retenu) n'est donc pas directement calculable. Une façon classique de remédier à ce dernier problème consiste à remplacer la quantité $R(f'')$ par un estimateur approprié.

2.1.2 Estimateur de Sheather et Jones

Afin de faire disparaître quelques effets indésirables du terme du biais. Sheather et Jones [10] mettent en place une procédure de type plug-in en trois étapes. Ils choisissent d'estimer $R(f'') = \int_{-\infty}^{+\infty} (f''(x))^2 dx$ par :

$$S(a) = \frac{1}{n(n-1)a^5} \sum_{i=1}^{n}\sum_{j=1}^{n} L^{(4)}\left(\frac{x_i - x_j}{a}\right), \tag{6}$$

où $L^{(4)}$ désigne la dérivée quatrième du noyau suffisamment lisse L et où a est un nouveau paramètre de lissage appelé paramètre pilote.

Pour L noyau gaussien, le paramètre de lissage h_{SJ} est alors la solution du critère $SJ(h)$ défini par :

$$SJ(\hat{h}) = \left(\frac{1}{2\sqrt{\pi}}\right)^{1/5} S(\hat{\alpha}(\hat{h}))\big(f''\big)^{-1/5} n^{-1/5} - \hat{h} \; ,$$

avec

$\hat{\alpha}(h) = 1.357\left(\frac{S(a)}{T(b)}\right)^{1/7} h^{5/7}$, $T(b) = \frac{1}{n(n-1)b^7}\sum_{i=1}^{n}\sum_{j=1}^{n} L^{(6)}\left(\frac{x_i - x_j}{b}\right)$, $a = 0.920\hat{\lambda}n^{-1/7}$, $b = 0.912\hat{\lambda}n^{-1/9}$. $\hat{\lambda}$ représente l'estimateur d'une mesure d'échelle (par exemple l'écart interquartile).

2.2 Méthodes Cross-Validation (Validation Croisée)

2.2.1 Validation croisée non biaisée

Cette méthode appelée Validation Croisée non Biaisée a été proposée par Rudemo [8] en 1982 et Bowman [2] en 1984. Le critère consiste à choisir le paramètre de lissage qui minimise un estimateur convenable de :

$$UCV(h) = \int_{\mathbb{R}} \big[f_h(x) - f(x)\big]^2 dx - \int_{\mathbb{R}} f^2(x)dx = \int_{\mathbb{R}} f_h^2(x)dx - 2\int_{\mathbb{R}} f_h(x)f(x)dx.$$

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

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

A cet effet, Le critère à optimiser est alors :

$$UCV(h) = \int_{\mathbb{R}} f_h^2(x)dx - \frac{2}{n}\sum_{i=1}^{n} f_{h,i}(x_i) \tag{7}$$

La popularité de cette méthode est due à la motivation intuitive et au fait que cet estimateur est asymptotiquement optimal sous de faibles conditions. L'optimalité asymptotique de la validation croisée non biaisée a été obtenue par Stone [11].

2.2.2 Validation croisée biaisée

Le critère de validation croisée biaisée, a été introduit par Scott et Terrell [9] en 1987 pour remédier aux problèmes de validation croisée non biaisée. Il s'agit d'introduire un biais dans le UCV afin de réduire sa variance. La quantité à minimiser est alors

$$BCV(h) = \frac{R(K)}{nh} + h^4\frac{\sigma_K^4}{4n^2}\sum_i\sum_{j,j\neq i} K_h^{(2)}K_h^{(2)}(X_i - X_j). \tag{8}$$

2.3 Bootstrap dans l'estimation globale de la densité de probabilité

Soit h_0 le paramètre qui minimise l'erreur quadratique moyenne intégrée par l'une des méthodes précédentes. Pour calculer la valeur de la fenêtre par la technique de bootstrap on doit rééchantillonner par cette technique à partir de l'échantillon initial et construire ensuite l'estimateur de bootstrap qui s'écrit sous alors sous la forme [6] :

$$f_h^j(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i^j}{h}\right), \quad j = 1, 2, ..., B; \tag{9}$$

où B est le nombre de réplications de bootstrap.
L'estimateur de la variance de $f_h^j(x)$ dans ce cas est donné sous la forme :

$$\frac{1}{B}\sum_{j=1}^{B}\int\left(f_h^j(x) - \bar{f}_h^*(x)\right)^2, \quad \text{avec, } \bar{f}_h^*(x) = \frac{1}{B}\sum_{j=1}^{B} f_h^j(x).$$

Nous construisons un estimateur initial de la densité f_{h_0} ensuite rééchantillonné par la technique de bootstrap à partir de l'échantillon initial pour construire les estimateurs $f_h^j(x)(j = 1, ..., B.)$.
Enfin, nous obtenons la fenêtre bootstrapée h_{boot} par la minimisation de $BMISE(h, h_0)$ sous h qui s'écrit sous la forme :

$$BMISE(h, h_0) = \frac{1}{B}\sum_{j=1}^{B}\int\left(f_h^j(x) - f_{h_0}(x)\right)^2 dx. \tag{10}$$

3 Choix du noyau

En générale on utilise le noyau gaussien en raison de sa simplicité et ses propriétés asymptotiques. Cependant l'inconvénient principale de ce noyau est qu'il attribue des poids positif pour des coordonnées qui sont à l'extérieur du support lorsque on cherche l'estimateur de la densité de probabilité d'une variable aléatoire bornée ou semi bornée. Cela cause le problème du biais aux bornes et donne un estimateur non consistant. La solution consiste à utiliser des noyaux asymétriques et adaptés qui n'assignent aucun poids à l'extérieur du support. Chen ((1999)[3] et (2000)[4]) propose respectivement le noyau $Beta$ pour les densités à support compact et le noyau gamma pour les densités à variables à support positif (c'est-à-dire sur $[0, +\infty[$).

3.1 Noyau Gamma

L'estimateur de la densité à noyau gamma est défini par :

$$f_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_{(\rho_h(x),h)}(X_i).$$

La forme du noyau gamma est définie comme suit [1] :

$$K_{(\rho_h(x),h)}(t) = \frac{t^{\rho_h(x)-1}e^{-t/h}}{h^{\rho_h(x)}\Gamma(\rho_h(x))}; \quad \rho_h(x) = \begin{cases} x/h, & \text{Si } x \geq 2h\,; \\ \frac{1}{4}(x/h)^2 + 1, & \text{Si } x \in [0, 2h[. \end{cases} \tag{11}$$

4 Simulations et résultats

Afin d'étudier l'influence de la technique bootstrap sur la qualité (performance) de l'estimateur non paramétrique d'une densité par la méthode du noyau, au sens de la minimisation du $AMISE$, on a effectué quelques simulations. Ces simulations sont organisées selon les étapes suivantes :
Générer un n-échantillon i.i.d d'une densité de probabilité cible ; Construire B échantillons de Bootstrap ; Estimer h^* et h_{boot} par une méthode de sélection ; Estimer $f_{h^*}(x)$ et $f_{h_{boot}}(x)$; Calculer l'$AMISE(f_{h^*}, f)$ notée $AMISE^*$ et $AMISE(f_{h_{boot}}, f)$ notée $AMISE_{boot}$.
Pour le calcul de h_{boot} on a utilisé l'algorithme suivant :
Générer un n-échantillon i.i.d d'une densité de probabilité cible ; Construire B échantillons de Bootstrap ; Estimer h_b, $(b = 1, ..., B)$ pour chaque échantillon de bootstrap par les trois méthodes ; Calculer h_{boot} à partir de $h_{boot} = \frac{1}{B} \sum_{b=1}^{B} h_b$.

4.1 Estimation du paramètre de lissage

Nous allons présenter les résultats (Tableau 1 et Tableau 2) des simulations obtenus pour l'estimation du paramètre de lissage d'une densité dans le cas : d'une loi normale et d'une loi exponentielle par les méthodes de sélection : UCV, BCV et SJ en appliquant la technique de bootstrap ainsi que leur AMISE associé.

Discussion

Loi normale :

Les résultats obtenus (tableau 1), nous permettent de constater que :
La méthode BCV (biased cross validation) est meilleure au sens du AMISE. L'augmentation de n entraîne la décroissance de h^*, h_{boot}, $AMISE^*$ et de $AMISE_{boot}$. La bootstrap réduit l'$AMISE$ pour les méthodes UCV et SJ, mais pas pour la BCV. Ceci est dû au biais important généré par cette méthode. On constate aussi que l'amélioration apportée par la bootstrap est indépendant de nombre de réplications B.

Loi exponentielle :

Les résultats obtenus (tableau 2) nous permettent de conclure que :
La méthode UCV est la meilleure au sens du critère du AMISE. On constate que les méthodes BCV, UCV et SJ nous fournissent des AMISE plus petits que ceux obtenus par la technique de bootstrap. Contrairement à loi normale, l'estimation de la densité de la loi exponentielle dépend du nombre de réplications B de bootstrap. En effet, on constate que pour $B \in \{10, 50\}$ l'AMISE diminue (sauf pour la BCV). pour d'autre valeurs de B la technique de bootstrap échoue. Ceci peut être expliqué par la présence d'un biais important au voisinage de zéro.

Paramètre	n	B	h^*	h_{boot}	$AMISE^*$	$AMISE_{boot}$
h_{UCV}	1000	10	0.2637520	0.2626835	0.0131875	0.0127912
		50	0.2713318	0.2643178	0.0133676	0.0133859
		100	0.2690290	0.2624879	0.0138132	0.0134523
		250	0.2741768	0.2644120	0.0135987	0.0130460
		500	0.2782387	0.2638493	0.0134459	0.0129456
	5000	10	0.2029686	0.1963668	0.0040273	0.0038783
		50	0.1987253	0.1959832	0.0038926	0.0038732
		100	0.1996528	0.1979249	0.0039821	0.0038867
		250	0.2017549	0.1963248	0.0041217	0.0038927
		500	0.2004525	0.1942246	0.0039784	0.0038623
	10000	10	0.1665237	0.1663452	0.0021732	0.0020348
		50	0.1663478	0.1660485	0.0023756	0.0020235
		100	0.1669836	0.1663073	0.0020124	0.0020099
		250	0.1669893	0.1665347	0.0022384	0.0020222
		500	0.1666928	0.1663834	0.0021386	0.0020237
h_{BCV}	1000	10	0.2803133	0.2906413	0.0128093	0.0130671
		50	0.2816399	0.2923112	0.0129138	0.0134826
		100	0.2796954	0.2902684	0.0134912	0.0137561
		250	0.2808413	0.2920106	0.0125973	0.0133967
		500	0.2811230	0.2924904	0.0133174	0.0133478
	5000	10	0.2038746	0.2090397	0.0039286	0.0039874
		50	0.1972895	0.2032374	0.0036334	0.0039947
		100	0.1996865	0.2043401	0.0035789	0.0039176
		250	0.2023900	0.2058912	0.0037923	0.0039009
		500	0.1989920	0.2013265	0.0038469	0.0039384
	10000	10	0.1619373	0.1650896	0.0023391	0.0023739
		50	0.1618341	0.1682934	0.0023598	0.0023912
		100	0.1611349	0.1690093	0.0023237	0.0024374
		250	0.1620075	0.1644465	0.0023883	0.0023982
		500	0.1617254	0.1624787	0.0023113	0.0024120
h_{SJ}	1000	10	0.2603793	0.2509944	0.0129250	0.0125867
		50	0.2609745	0.2432584	0.0131968	0.0129331
		100	0.2618081	0.2434982	0.0136932	0.0126986
		250	0.2605905	0.2428563	0.0130537	0.0122063
		500	0.2635642	0.2452238	0.0131986	0.0127583
	5000	10	0.1887019	0.1887826	0.0042946	0.0039931
		50	0.1886753	0.1889731	0.0037947	0.0032656
		100	0.1883298	0.1886132	0.0037365	0.0035413
		250	0.1882765	0.1883589	0.0039764	0.0036914
		500	0.1885491	0.1884050	0.0038349	0.0037190
	10000	10	0.1589912	0.1589214	0.0020907	0.0020589
		50	0.1589735	0.1589373	0.0022064	0.0021975
		100	0.1589273	0.1589910	0.0021834	0.0021758
		250	0.1589974	0.1589327	0.0021681	0.0021497
		500	0.1589017	0.1589130	0.0020899	0.0020512

Table 1: *Résultats des simulations effectuées sur la loi Normale(0,1) pour déterminer les fenêtres optimales h^* et h_{boot}.*

Paramétre	n	B	h^*	h_{boot}	$AMISE^*$	$AMISE_{boot}$
h_{UCV}	1000	10	0.0602429	0.0593424	0.2800012	0.2859504
		50	0.0585438	0.0597847	0.2791145	0.2865065
		100	0.0595894	0.0601354	0.2793140	0.2803535
		250	0.0606347	0.0596781	0.2829039	0.2859024
		500	0.0614829	0.6251291	0.2783237	0.2812265
	5000	10	0.0217138	0.0214390	0.0575054	0.0604128
		50	0.0220979	0.0218324	0.0587671	0.0603145
		100	0.0215915	0.0216328	0.0531080	0.0563274
		250	0.0209312	0.0218050	0.0543056	0.0582154
		500	0.0221009	0.0218694	0.0578763	0.0601623
	10000	10	0.0189512	0.0187808	0.0172925	0.0175139
		50	0.0187356	0.0184546	0.0159463	0.0175871
		100	0.0190909	0.0185561	0.0155283	0.0163543
		250	0.0188876	0.0188732	0.0164934	0.0179058
		500	0.0189219	0.0188053	0.0165632	0.0175967
h_{BCV}	1000	10	0.0957601	0.0923327	0.3177505	0.3197029
		50	0.0937565	0.0965328	0.3209693	0.3211784
		100	0.0967867	0.0954992	0.3166561	0.3104783
		250	0.0973491	0.0973780	0.3323125	0.3324214
		500	0.0948340	0.0934896	0.3259604	0.3274038
	5000	10	0.0574679	0.0567974	0.0931346	0.0957324
		50	0.0569907	0.0565325	0.0887760	0.0965220
		100	0.0573248	0.0559323	0.0917338	0.0928959
		250	0.0570746	0.0569642	0.0925603	0.0931216
		500	0.0567344	0.0563477	0.0909375	0.0925467
	10000	10	0.0199352	0.0197987	0.0161917	0.0163978
		50	0.0209462	0.0197764	0.0167019	0.0170754
		100	0.0198665	0.0199819	0.0167005	0.0172789
		250	0.0197796	0.0195088	0.0161098	0.0167899
		500	0.0196036	0.0197593	0.0164572	0.0168256
h_{SJ}	1000	10	0.0992531	0.0981050	0.3217277	0.338063
		50	0.0997447	0.0989786	0.3260613	0.337089
		100	0.0997769	0.0989617	0.2747429	0.308961
		250	0.0996543	0.0987863	0.2823170	0.316849
		500	0.0994548	0.0981204	0.3205451	0.375097
	5000	10	0.0413247	0.0405765	0.1682984	0.1701790
		50	0.0410946	0.0409486	0.1844357	0.1910789
		100	0.0394319	0.0396464	0.1770978	0.1916037
		250	0.0414704	0.0390642	0.1682692	0.1808675
		500	0.0408061	0.0404292	0.1787701	0.1847453
	10000	10	0.0342617	0.0348648	0.0886295	0.0900169
		50	0.0346901	0.0346344	0.0783287	0.0887651
		100	0.0341375	0.0346091	0.0796114	0.0878765
		250	0.0345903	0.0340978	0.0823089	0.0840856
		500	0.0348746	0.0347605	0.0882560	0.0906524

TABLE 2: *Résultats des simulations effectuées sur la loi exponentielle ($\lambda = 1$) pour déterminer les largeurs de fenêtres optimales h^* et h_{boot}.*

4.1.1 Comparaison des différents noyaux

On veut calculer l'erreur quadratique moyenne intégrée commise lors de l'estimation d'une densité de probabilité f par la méthode du noyau, en s'intéressant à l'influence de la technique de bootstrap sur le noyau utilisé pour l'estimation de f. Les noyaux utilisés sont : Noyau gaussien, noyau Epanechnikov (qui minimise le AMISE) il est de la forme : $\frac{3}{4\sqrt{5}}(1 - \frac{u^2}{5}), |u| \leq \sqrt{5}$ le noyau biweight qui est de la forme : $\frac{15}{16}(1 - u^2)^2, |u| \leq 1$ et le noyau gamma (11). Les résultats obtenus dans le cas sont rangés dans les tableaux 3 et 4.

n	h	Gauss.	Epane.	Biwei.
1000	h_{UCV}	0.0133	0.0129	0.0385
	h_{BCV}	0.0130	0.0127	0.0343
	h_{SJ}	0.0133	0.0129	0.0373
5000	h_{UCV}	0.0037	0.0035	0.0093
	h_{BCV}	0.0037	0.0035	0.0087
	h_{SJ}	0.0037	0.0035	0.0091
10000	h_{UCV}	0.0022	0.0021	0.0054
	h_{BCV}	0.0023	0.0023	0.0063
	h_{SJ}	0.0021	0.0020	0.0051

TABLE 3: *Résultats des simulations effectués sur la loi Normale(0,1) pour déterminer les AMISE pour différents noyaux.*

Discussion
Loi normale
Les résultats obtenus (tableau 3), nous permettent de conclure que :
Le choix du noyau n'est pas important au sens de l'$AMISE$, car les résultats sont proches. La convergence de l'$AMISE$ vers zéro est indépendante du noyau mais dépend essentiellement de la taille n de l'échantillon.
Enfin, les résultats confirment l'optimalité du noyau Epanechnikov.
Loi exponentielle
Les résultats obtenus (tableau 4), nous permettent de conclure que :
Le noyau *gamma* améliore la qualité de l'estimateur de la densité au sens de l'$AMISE$. En effet, la valeur de l'AMISE diminue de plus de 90% par rapport aux noyaux biweight et Epanechnikov.
la diminution de l'$AMISE$ engendrée par la technique bootstrap dépend de la taille n de l'échantillon et du nombre de réplications B.

h	n	B	Epanich.		Biwei.		Gamma	
			$amise$	$amise_b$	$amise$	$amise_b$	$amise$	$amise_b$
h_{UCV}	1000	10	0.8158900	0.8172653	0.8773973	0.8785201	0.0161366	0.0161424
		50	0.9144138	0.9119522	0.8007408	0.7992636	0.0181240	0.0183341
		100	0.8194956	0.8185549	0.8556567	0.8549958	0.0245195	0.0245780
		250	0.8900072	0.8917226	0.8411298	0.8424278	0.0160835	0.0165922
		500	0.8159415	0.8158520	0.8421741	0.8411209	0.0165029	0.0166452
	2500	10	0.0784133	0.0784048	0.0821921	0.0852894	0.0077749	0.0077787
		50	0.0804945	0.0803803	0.0864820	0.0836296	0.0095261	0.0094218
		100	0.0893738	0.0893033	0.0849527	0.0848954	0.0075513	0.0075001
		250	0.0816903	0.0816685	0.0721977	0.0721732	0.0094949	0.0094628
		500	0.0933548	0.0933838	0.0851609	0.0851780	0.0041001	0.0040925
	5000	10	0.0083725	0.0083767	0.0091097	0.0091140	0.0003582	0.0003638
		50	0.0083762	0.0083800	0.0087144	0.0087180	0.0005486	0.0005489
		100	0.0073624	0.0073624	0.0084080	0.0084075	0.0003509	0.0003508
		250	0.0082011	0.0081918	0.0087243	0.0087127	0.0006140	0.0006094
		500	0.0079490	0.0079508	0.0073386	0.0073403	0.0003270	0.0003275
h_{SJ}	1000	10	0.9274914	0.9287839	0.7721525	0.7730992	0.0966429	0.0942967
		50	0.8490935	0.8476073	0.7999490	0.7986728	0.0751576	0.0750330
		100	0.9726110	0.9715667	0.8303785	0.8296706	0.1237604	0.1234712
		250	0.9325880	0.9337777	0.8133108	0.8144709	0.1048228	0.1038240
		500	0.8515873	0.8493262	0.8121030	0.8108708	0.0770548	0.0769996
	2500	10	0.0846891	0.0846696	0.0856098	0.0855945	0.0089344	0.0089224
		50	0.0898376	0.0896810	0.0928338	0.0926633	0.0058991	0.0058833
		100	0.0921760	0.0921180	0.0820543	0.0819981	0.0048394	0.0048152
		250	0.0765944	0.0765527	0.0833211	0.0832884	0.0070466	0.0070308
		500	0.0832602	0.0832713	0.0834229	0.0834372	0.0077377	0.0076767
	5000	10	0.0085693	0.0085716	0.0085685	0.0085717	0.0004133	0.0004130
		50	0.0078618	0.0078635	0.0088440	0.0088477	0.0005012	0.0005009
		100	0.0086987	0.0086979	0.0079939	0.0079935	0.0003543	0.0003533
		250	0.0087981	0.0087871	0.0087314	0.0087200	0.0003701	0.0003792
		500	0.0077783	0.0077824	0.0088969	0.0088999	0.0003740	0.0003743

TABLE 4: *Résultats des simulations effectués sur la loi Exponentielle de paramètre $\lambda = 1$ pour déterminer les AMISE pour différents noyaux.*

5 Conclusion

Ce travail est une contribution au problème du choix du paramètre de lissage par la technique de bootstrap lors de l'estimation de la densité de probabilité par la méthode du noyau. Nous avons utilisé la technique de bootstrap pour estimer le paramètre de lissage et nous avons étudié l'efficacité et la robustesse de cette technique qui ont été mesurées numériquement par une étude de simulation sur des densités tests présentant différents aspects (loi normale, loi exponentielle). Les résultats obtenus montrent que la technique de bootstrap classique améliore les performances de l'estimateur à noyau de la loi normale au sens de l'AMISE pour les méthodes de sélection du paramètre de lissage UCV et SJ et elle échoue pour la BCV, à cause de la présence du biais dans cette méthode. A cause de la présence d'un biais important au voisinage de zéro, la technique de bootstrap échoue au sens du AMISE dans l'estimation des densités à queues lourdes comme la loi exponentielle. L'étude montre aussi que le choix du noyau est important pour les densités cibles à support compact comme la loi exponentielle.

Références

[1] T. Bouezmarni and O. Scaillet (2003). *Consistency of Asymmetric Kernel Density Estimators and Smoothed Histograms with Application to Income Data*, Econometric Theory, 21, 390-412.

[2] A. W. Bowman (1984). emphAn alternative method of crooss-validation for the smoothing density estimates, Biometrika, 71, 553-560.

[3] S. Chen (1999). *A Beta Kernel Estimation for Density Functions*, Computational Statistics and Data Analysis, 31, 131-145.

[4] S. Chen (2000). *Probability Density Functions Estimation Using Gamma Kernels*, Annals of the Institute of Statistical Mathematics, 52, 471-480.

[5] B. Efron (1979). *Bootstrap Methods : Another Look at the Jackknife*, Annals of Statistics, 7, 1-26.

[6] J.J. Faraway and M. Jhun (1990). *Bootstrap Choice of Bandwidth for Density Estimation*, J. Amer. Statist. Assoc., 85, 1119-1122.

[7] E. Parzen (1962). *On estimation of a probability density function and mode*, Ann. Math. Statist, 33, 1065-1076.

[8] M. Rudemo (1982). *Empirical choice of histogram and kernel density estimators*, Scandinavian Journal of Statistics, 9, 65-78.

[9] D. W. Scott and G.R. Terrell (1987). *Biased and Unbiased Cross-Validation in Density Estimation* Journal of the American Statistical Association, 82, 1131-1146.

[10] S.J Sheather and M.C. Jones (1991). *A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation*, J. Roy. Statist. Soc., B53, 683-690.

[11] C. Stone (1984). *An Asymptotically Optimal Window Selection Rule for Kernel Density Estimates*, The Annals of Statistics, 12, 1285-1297.