

REVUE DE STATISTIQUE APPLIQUÉE

PIERRE HOMINAL

PAUL DEHEUVELS

Estimation non paramétrique de la densité compte- tenu d'informations sur le support

Revue de statistique appliquée, tome 27, n° 3 (1979), p. 47-68

http://www.numdam.org/item?id=RSA_1979__27_3_47_0

© Société française de statistique, 1979, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

ESTIMATION NON PARAMÉTRIQUE DE LA DENSITÉ COMPTE-TENU D'INFORMATIONS SUR LE SUPPORT

Pierre HOMINAL et Paul DEHEUVELS
Ecole Pratique des Hautes Etudes et Université Paris VI

RESUME

L'estimation de la densité de probabilité par la méthode du noyau de Parzen-Rosenblatt présente des inconvénients dans le cas où le support de la distribution est borné. Le support de la distribution associée à l'estimation ne coïncide généralement pas avec le support réel, et ceci est gênant en pratique, si on dispose d'informations a priori à son sujet. Un cas important est obtenu pour l'estimation de la densité d'une variable aléatoire positive à partir d'un petit échantillon. On observe en pratique dans ce cas des "bavures" dans les nombres négatifs, l'estimation accordant une probabilité non nulle en général à \mathbb{R}^- , en contradiction avec l'hypothèse de positivité.

Le but de ce qui suit est d'introduire et d'étudier des estimations efficaces de la densité dans ces circonstances. De nombreux exemples d'applications pratiques sont joints, montrant l'intérêt concret de ces méthodes.

Cet article se réfère essentiellement à [1] et [4] pour les détails des calculs.

ENGLISH SUMMARY

This paper deals mainly with the problem of giving in practice an efficient modification of the Parzen-Rosenblatt kernel density estimate, when the support of the distribution is wholly or partially known. It is shown that with a simple modification of the estimates near the boundary points of the support of the distribution, it is possible to obtain efficient estimates. A general methodology of density estimation is discussed.

I – INTRODUCTION

Le statisticien appliqué est souvent confronté au problème de l'estimation d'une distribution de probabilité à partir d'un échantillon X_1, \dots, X_N . Le plus souvent, il est amené à faire des hypothèses compte tenu d'informations extérieures sur la nature de cette distribution, permettant de limiter ses recherches à une estimation paramétrique.

Mots-clés : Estimation de la densité, Estimation paramétrique, Estimation non paramétrique, Méthode du noyau, Critère du MISE, Estimation fonctionnelle.

Key Words : Density estimation, parametrical and non-parametrical estimation, functional estimation, MISE Efficiency criterion.

Ainsi, s'il a des raisons a priori de croire que le phénomène physique générant les v.a. de l'échantillon est obtenu par la superposition d'un nombre important d'aléas indépendants, il supposera que la loi est normale ; s'il s'agit d'une superposition multiplicative, il supposera que la loi est lognormale. On pourrait ainsi multiplier les exemples. Le point important de la méthodologie classique est que le statisticien n'est, en général, que très rarement certain de l'hypothèse qu'il avance. Il cherchera en conséquence à vérifier la validité de cette hypothèse, en faisant des tests sur l'échantillon. Il fera, ainsi, dans le cas d'une hypothèse de normalité, un ou plusieurs tests comme le test de Geary ou Shapiro-Wilks, etc. Il arrive alors très fréquemment que ces tests donnent des résultats incertains, ne permettant pas une conclusion définitive. L'attitude adoptée est alors variable suivant le degré de rigueur du statisticien qui cherchera le plus souvent à adopter par pragmatisme un optimisme généreux au sujet des résultats des tests, acceptant avec beaucoup de facilité des hypothèses trop simples sans trop de vérifications.

Il va sans dire que ceci mène généralement à la construction de colosses aux pieds d'argile, les conclusions d'une étude pouvant dépendre de manière très importantes des hypothèses de modélisation de base.

Ces problèmes peuvent cependant être partiellement résolus par l'usage systématique d'estimations non paramétriques de la densité de la densité simultanément à toute autre estimation paramétrique. Nous montrons plus loin par des exemples tirés de données réelles, à quel point ces méthodes peuvent se révéler utiles pour une modélisation efficace d'un phénomène observé.

Dans [1], [2], nous avons comparé les méthodes classiques d'estimation non paramétrique de la densité. Une bibliographie de ce type d'estimation peut être obtenue dans WERTZ, 1978, [7], WEGMAN, [5], [6].

La seule méthode commode simple et robuste permettant d'obtenir une estimation de la densité pour une v.a. réelle, ne nécessitant pas un choix multiple de paramètres, est la méthode du noyau de Parzen-Rosenblatt. Si X_1, \dots, X_n est un échantillon de v.a. indépendantes de même loi de densité f sur \mathbf{R} , on estime $f(x)$ par :

$$f_n(x) = \frac{1}{n\delta_n} \sum_{i=1}^n K\left(\frac{X_i - x}{\delta_n}\right), \quad \text{où :} \quad (1.1)$$

– $K(u)$ est généralement supposé tel que :

$$(i) \quad K(u) \geq 0 \quad ; \quad (ii) \quad \int K(u) \, du = 1 \quad ; \quad (iii) \quad |K(u)| \leq C \quad ;$$

– δ_n est un paramètre réel positif, appelé généralement fenêtre, ou largeur de la fenêtre.

Quelques points pratiques importants méritent d'être soulignés, concernant f_n :

- Le choix de K ne présente, en général pas une très grande importance, pourvu qu'il soit convenablement normalisé par δ_n (Voir [1] p. 31) ;
- Le choix de δ_n présente une importance considérable pour l'efficacité de l'estimation. Un choix de δ_n trop petit implique la présence de fluctuations aléatoires trop importantes, un choix de δ_n trop grand élimine les aléas, mais introduit des biais de lissage trop importants.

On trouvera dans [1], [2], une discussion sur les choix optimaux de ces paramètres. De nombreux auteurs se sont penchés sur le problème du choix pratique de δ_n . Citons, parmi les articles récents, SILVERMAN, 1978, [8]. Nous utiliserons ici la méthodologie introduite dans [1].

– $f_n(x)$ est une densité. On appelle souvent $f_n(x)$ définie par (1.1), densité empirique de l'échantillon. Ceci peut être justifié comme suit :

La fonction de répartition empirique de l'échantillon X_1, \dots, X_n est généralement définie par :

$$F_n(x) = \frac{\text{Nb} \{X_i < x ; 1 \leq i \leq n\}}{n} ; \quad (1.2)$$

La dérivée par différentiation finie de F_n est définie par :

$$f_n(x) = \frac{1}{\delta_n} \left(F_n \left(x + \frac{1}{2} \delta_n \right) - F_n \left(x - \frac{1}{2} \delta_n \right) \right) = \frac{1}{n\delta_n} \sum_{i=1}^n K \left(\frac{X_i - x}{\delta_n} \right) \quad \text{p.p.}, \quad (1.3)$$

où le noyau $K(u)$ est le noyau uniforme :

$$K(u) = \begin{cases} 1 & \text{si } |u| \leq \frac{1}{2} \\ 0 & \text{si } |u| > \frac{1}{2} \end{cases}.$$

Nous appellerons estimateur de la fenêtre mobile l'estimateur défini par (1.3).

Nous présentons maintenant dans les figures 1 et 2, les résultats obtenus à partir de deux échantillons de tailles respectives 94 et 72, de hauteurs de vagues en mer du nord. On utilise l'estimateur de la fenêtre mobile, et une estimation paramétrique à l'aide d'une loi Gamma. A titre d'information, on a comparé la distribution paramétrique estimée Gamma à l'échantillon par les tests de Kolmogorov-Smirnov et du χ^2 , en utilisant l'approximation consistant à supposer que les tests sont voisins de ceux qu'on obtiendrait pour tester une distribution d'hypothèse.

Dans le premier cas (fig. 1) les tests du χ^2 et de Kolmogorov-Smirnov acceptent l'hypothèse à 90 % ; dans le second, le test du χ^2 rejette l'hypothèse, tandis que Kolmogorov-Smirnov l'accepte à 90 %. L'intérêt de la méthode est évident pour indiquer qu'un modèle bimodal devrait être préféré à un modèle Gamma dans le cas de la figure 2, alors que le modèle Gamma est acceptable dans le cas de la figure 1.

Les deux estimations non paramétriques obtenues dans les figures 1 et 2 ne sont, cependant, pas complètement satisfaisantes, en ce qu'elles possèdent des "bavures" importantes en dessous de 0. On obtient, en effet que $f_n(x)$ est positive pour l'estimateur de la fenêtre mobile, dès que x atteint la valeur :

$$x = X_{1,n} - \frac{1}{2} \delta_n, \quad \text{où } X_{1,n} = \inf_{1 \leq i \leq n} X_i.$$

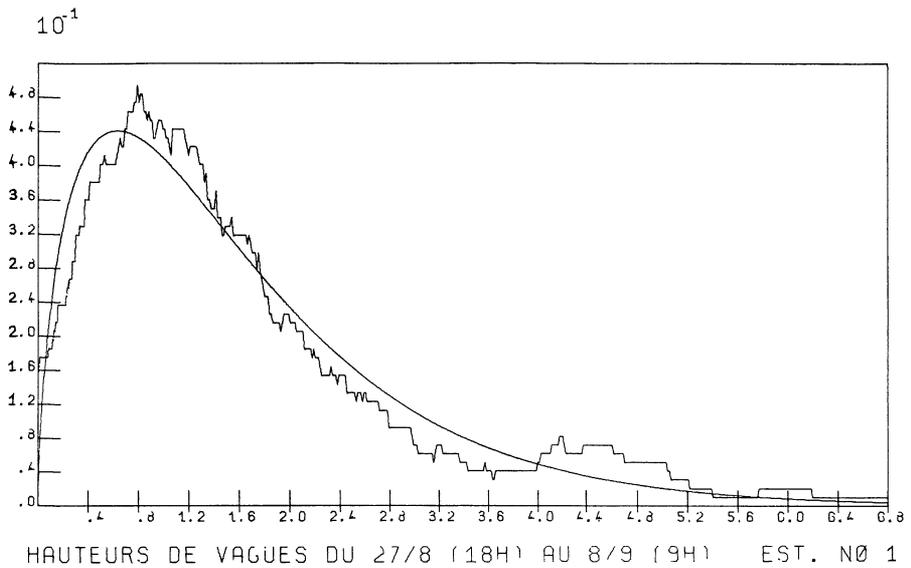


Figure 1

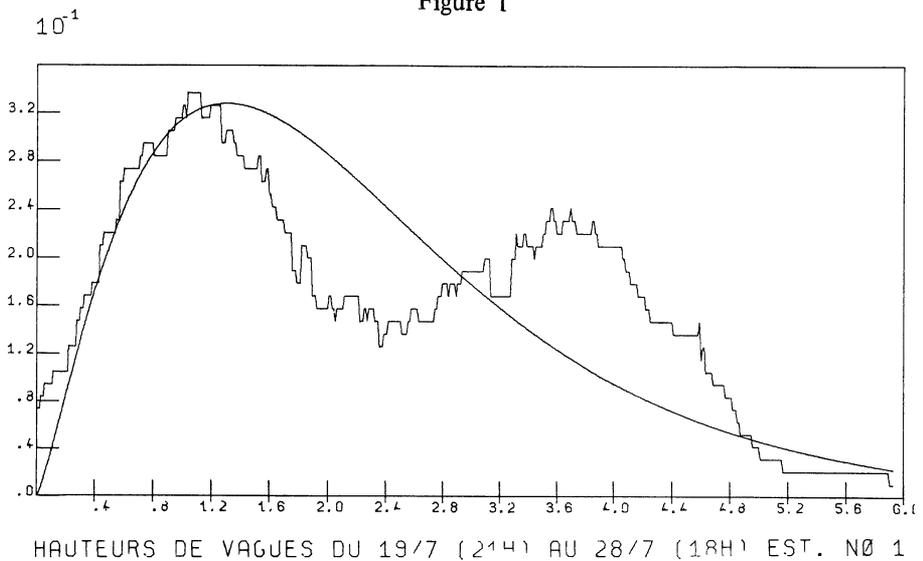


Figure 2

Il est, naturellement, absurde d'imaginer des hauteurs de vagues négatives, et il est nécessaire d'aménager l'estimation au voisinage de 0, pour tenir compte de la positivité de la distribution. Les estimations obtenues dans les figures 1 et 2 ont été obtenues en tronquant la partie négative, c'est-à-dire, par :

$$h_n(x) = \begin{cases} f_n(x) & , \text{ si } x \geq 0, \\ 0 & , \text{ si } x < 0. \end{cases} \quad (1.4)$$

L'inconvénient du choix introduit par (1.4) est que h_n présente généralement une forte discontinuité à l'origine, et que, en général, $\int_{\mathbf{R}} h_n(x) dx < 1$.

Pour celà, compte tenu de (1.3), et du fait que la positivité de la distribution n'introduit pas de distortions si $x > \frac{1}{2} \delta_n$, nous avons été amenés à introduire des estimations de la forme suivante :

$$f_n^*(x) = \begin{cases} f_n(x) = \frac{1}{\delta_n} \left(F_n \left(x + \frac{1}{2} \delta_n \right) - F_n \left(x - \frac{1}{2} \delta_n \right) \right), & \text{si } x > \frac{1}{2} \delta_n \\ g_n(x) = \frac{1}{\delta_n} \left(a(x) F_n \left(x + \frac{1}{2} \delta_n \right) + b(x) F_n(y(x)) \right), & \text{si } 0 \leq x \leq \frac{1}{2} \delta_n \end{cases} \quad (1.5)$$

L'étude de l'estimateur (1.5) est développée dans [4]. Nous présentons dans le § 3 suivant les valeurs de a, b, y optimales suivant le critère du MISE (Minimum Integrated Square Error), avec l'hypothèse que la densité à estimer $f(x)$ est identique à Lx^r , r connu, pour $0 \leq x \leq \delta_n \leq A(r)$.

Pour réaliser une estimation de ce type automatiquement, une programme a été écrit. Nous présentons dans le § 4 les résultats obtenus à partir de données réelles ou simulées.

II – ESTIMATEUR DE PARZEN-ROSENBLATT POUR DES DISTRIBUTIONS SUR R :

Nous utilisons les résultats suivants, dont la justification détaillée est faite dans [1].

Si $f_n(x) = \frac{1}{n\delta_n} \sum_{i=1}^n K \left(\frac{X_i - x}{\delta_n} \right)$ est l'estimation utilisée, où

$K(u)$ vérifie les propriétés (i) (ii) (iii) (1.1), et, de plus :

(iv) $K(u) = K(-u)$; (v) K est nul en dehors d'un compact ;

Si, de plus, la densité $f(x)$ à estimer est telle que :

$f^{(p)} \in L^1(V(I)) \cap L^2(V(I))$, $0 \leq p \leq 3$, où I est un intervalle de R , et $V(I)$ un ouvert contenant I , alors, si on pose :

$$M_I^2 = \int_I E ((f_n(x) - f(x))^2) dx = B_{1,I}^2 + B_{2,I}^2, \quad \text{où :}$$

$$B_{1,I} = \int_I (E(f_n(x)) - f(x))^2 dx, \quad B_{2,I} = \int_I \text{Var}(f_n(x)) dx,$$

alors, lorsque $\delta_n \rightarrow 0$, et $n\delta_n \rightarrow +\infty$,

$$B_{1,I}^2 = \frac{1}{4} \delta_n^4 \left(\int_I y^2 K(y) dy \right)^2 \left(\int_I f''(x)^2 dx \right) + o(\delta_n^4), \quad (2.1)$$

$$B_{2,I}^2 = \frac{1}{n\delta_n} \left(\int_I K^2(y) dy \right) + o\left(\frac{1}{n\delta_n}\right).$$

Le résultat important en pratique est (voir [1]) que les résultats obtenus en choisissant δ_n tel que M_I^2 soit minimum, sont, même pour de petits échantillons, très voisins de ceux qu'on obtient en minimisant l'expression obtenue en se limitant aux premiers termes de (2.1), à savoir :

$$M_I^2(*) = \frac{1}{2} \delta_n^4 \left(\int_I y^2 K(y) dy \right)^2 \left(\int_I f''(x)^2 dx \right) + \frac{1}{n \delta_n} \left(\int_I K^2(y) dy \right). \quad (2.2)$$

Le choix optimal asymptotique de K est fourni par le noyau d'Epanechnikov :

$$K(y) = \frac{3}{4\sqrt{5}} (1 - y^2/5), \text{ si } |y| \leq \sqrt{5}, \text{ et ([1], p. 31), l'efficacité asymptotique}$$

relative du noyau unité utilisé dans (1.3) est 94,24 % ; ceci nous permet d'utiliser l'estimateur à fenêtre mobile (1.3), avec un choix de δ_n donné par la minimisation de $M^2(*)$, fournissant, par [1], (2.56), un choix optimal :

$$\delta_n = \left(\frac{144}{n \int_I f''(x)^2 dx} \right)^{1/5} \quad (2.3)$$

En fait, il reste un paramètre inconnu pour obtenir δ_n , à savoir :

$$\theta = \int_I f''(x)^2 dx. \text{ Pour obtenir cette expression, nous utilisons la méthodologie suivante :}$$

– on effectue une hypothèse paramétrique sur la densité $f(x) = f(x; \alpha)$;

– on estime α par $\hat{\alpha}$ par une estimation paramétrique ;

– on construit l'estimateur à fenêtre mobile (1.3), déterminé par :

$$\delta_n = \left(\frac{144}{n \int_I f''(x; \hat{\alpha})^2 dx} \right)^{1/5}. \quad (2.4)$$

Cette méthodologie permet d'obtenir une estimation efficace comme dans les figures 1 et 2, en utilisant $I = \mathbf{R}$. Nous développons dans le paragraphe suivant, le cas d'un support borné.

III – ESTIMATION MODIFIEE DANS LE CAS DE SUPPORTS BORNES :

1) Choix de δ_n

On supposera ici que la densité à estimer a un support $I = [0, +\infty[$, et que l'on décompose I en la réunion de $I_1 = \left[0, \frac{1}{2} \delta_n \right]$ et $I_2 = \left[\frac{1}{2} \delta_n, +\infty \right[$.

L'estimateur (1.3) de la fenêtre mobile s'écrit alors :

$$f_n(x) = \begin{cases} \frac{1}{\delta_n} \left(F_n \left(x + \frac{1}{2} \delta_n \right) - F_n \left(x - \frac{1}{2} \delta_n \right) \right), & \text{si } x \in I_2. \\ \frac{1}{\delta_n} F_n \left(x + \frac{1}{2} \delta_n \right), & \text{si } x \in I_1. \end{cases} \quad (3.1)$$

La décomposition de M_I^2 en $M_{I_1}^2 + M_{I_2}^2$ montre que, si f est assez régulière, le résultat (2.1) peut s'appliquer pour I_2 :

$$M_{I_2}^2 = \frac{\delta_n^4}{576} \int_{\mathbb{R}} f''(x)^2 dx + \frac{1}{n\delta_n} + o(\delta_n^4) + o\left(\frac{1}{n\delta_n}\right). \quad (3.2)$$

Sur I_1 , le développement de $M_{I_1}^2$ dépend de la régularité de f au voisinage de 0. Si on suppose que $f(x) = Lx^r$ sur $[0, \delta_n[$, on obtient, en effectuant le changement de variable $X = x/\delta_n$:

$$M_{I_1}^2 = L^2 \delta_n^{2r+1} \int_0^{\frac{1}{2}} \left\{ \left(X + \frac{1}{2} \right)^{r+1} - (r+1) X^r \right\} dX + \frac{\delta_n^r}{n} \frac{L}{r+2} \left(1 - \frac{1}{2^{r+2}} \right) + o(\delta_n^{2r+1}) + o\left(\frac{\delta_n^r}{n}\right). \quad (3.3)$$

En additionnant (3.2) et (3.3), on obtient, si $r = 0, 1, 2$, ou si $r \geq 2.5$:

$$M_I^2 \sim K_1 \delta_n^{2r+1} + \frac{1}{n\delta_n} \quad (3.4)$$

si $r = 0$ ou 1 (cas des lois uniforme, exponentielle, triangulaire).

$$M_I^2 \sim K_2 \delta_n^4 + \frac{1}{n\delta_n}, \quad (3.5)$$

si $r = 2$ ou si $r \geq 2.5$;

L'utilisation de ces résultats pour le choix d'une fenêtre δ_n optimale conduit à un critère d'erreur M_I^2 décroissant dans le premier cas comme $n^{-\frac{2r+1}{2(r+1)}}$, tandis que dans le deuxième cas, on obtient $n^{-\frac{4}{5}}$; on constate ainsi, que dans le premier cas, le biais introduit par le comportement de f au voisinage de l'extrémité du support suffit à détériorer l'efficacité de l'estimation.

La conclusion logique de ce qui précède est, qu'à moins que l'ordre de décroissance de la densité au voisinage de l'origine soit suffisamment élevé ($f(0) = f'(0) = f''(0) = 0$), il est nécessaire, pour conserver l'efficacité de l'estimation, de modifier celle-ci dans I_1 . Nous développons ceci dans l'alinéa suivant.

2) Modifications de l'estimateur à fenêtre mobile

Nous comparons quatre modifications de l'estimateur (1.3) $f_n(x)$ sur l'intervalle $I_1 = \left[0, \frac{1}{2} \delta_n\right]$:

$$a) \quad \bar{f}_n(x) = \frac{1}{x + \frac{1}{2} \delta_n} F_n \left(x + \frac{1}{2} \delta_n\right);$$

$$b) \quad f_n^*(x) = \frac{1}{\delta_n} \left(F_n \left(x + \frac{1}{2} \delta_n\right) - F_n \left(\frac{1}{2} \delta_n - x\right) \right);$$

(estimateur de la ficelle I)

$$c) \quad f_n^{**}(x) = \frac{1}{\delta_n} \left(F_n \left(x + \frac{1}{2} \delta_n\right) + F_n \left(\frac{1}{2} \delta_n - x\right) \right);$$

(estimateur de la ficelle II)

$$d) \quad \hat{f}_n(x) = \hat{f}_n(x; y, a, b) = \frac{1}{\delta_n} \left(a F_n \left(x + \frac{1}{2} \delta_n\right) + b F_n(y) \right); \quad (3.6)$$

La justification de ces estimateurs peut être établie comme suit :

a) f_n est une normalisation différente de $f_n(x) = \frac{1}{\delta_n} \left(F_n \left(x + \frac{1}{2} \delta_n\right) \right)$, sans biais pour $r = 0$;

b) f_n^* et f_n^{**} sont obtenus à partir des valeurs de F_n aux points $x + \frac{1}{2} \delta_n$ et $\frac{1}{2} \delta_n - x$, tels que la somme des longueurs des intervalles $\left[0, x + \frac{1}{2} \delta_n\right]$, et $\left[0, \frac{1}{2} \delta_n - x\right]$ soit égale à δ_n . Le nom donné à ces estimateurs correspond à l'idée d'une ficelle de longueur δ_n , pivotant autour du point 0, et d'extrémités $x + \frac{1}{2} \delta_n$, $\frac{1}{2} \delta_n - x$. Lorsque $r = 0$, f_n^{**} est sans biais.

c) $\hat{f}_n(x; y, a, b)$, où a, b, y sont des fonctions de x et r , est une formulation générale, englobant les cas précédents.

Pour des raisons de simplicité des estimations, nous avons été amenés à la détermination des choix optimaux de a, b, y , avec les hypothèses suivantes :

- a et b sont supposés indépendants de x ;
- $y = y(x, a, b)$ est une fonction de x, a, b ;
- $f(x) = Lx^r$, sur un intervalle $[0, A(r)]$, tel que : $0 < \delta_n \leq A(r)$.

Le problème se ramène, compte tenu de (3.6), et avec les notations introduites dans (2.1) :

- à déterminer, pour chaque couple (a, b) , la fonction $y(x, a, b)$ qui minimise l'erreur quadratique moyenne en x : $E \left((\hat{f}_n(x, y, a, b) - f(x))^2 \right)$;

– à déterminer le couple (a, b) tel que les termes de plus haut degré des développements asymptotiques de :

$$B_{1,I_1}^2 = \int_{I_1} (E(\hat{f}_n(x; y(x, a, b), a, b)) - f(x))^2 dx, \text{ et}$$

$$B_{2,I_1}^2 = \int_{I_1} \text{Var}(\hat{f}_n(x; y(x, a, b), a, b)) dx,$$

aient une somme minimale, δ_n étant fixé.

La détermination de $a, b, y(x, a, b)$ est établie dans [4]. Les résultats obtenus sont les suivants :

- on n'obtient pas d'estimateur optimal si $r < 0$, ce qui correspond à une densité infinie à l'origine ;
- pour $r > 4$, l'estimateur optimal est nul ($a = b = 0$), ce qui s'explique par le fait que la part de la variance est prépondérante sur celle du biais ;
- pour $0 < r < 4$, on obtient :

a) *Détermination de a* : $a = 2 \left(\frac{r}{r+1} \right)^r$, si $0 < r < 1$,

$$a = \frac{r+1}{2^r}, \quad \text{si } 1 \leq r < 4.$$

b) *Détermination de b* : $b = \frac{(r+1)v^r}{\left(v + \frac{1}{2}\right)^{r+1}} - a$, où v est solution de l'équation :

$$\left(u + \frac{1}{2}\right)^{r+2} - \left(v + \frac{1}{2}\right)^{r+2} = \frac{1}{2} \left(1 - \left(\frac{1}{2}\right)^{r+2} + \frac{r+2}{2^{r+1}a}\right), \text{ et :}$$

$$u = \frac{1}{2}, \text{ si } 0.32 \leq r < 4,$$

$$u \text{ est solution de } u^r \left(v + \frac{1}{2}\right)^{r+1} = v^r \left(u + \frac{1}{2}\right)^{r+1}, \text{ si } 0 < r \leq 0.31.$$

c) *Détermination de A(r)* :

$$A(r) = \left\{ \frac{a}{-b} \left(\frac{1}{2}\delta_n\right)^{r+1} - \frac{r+1}{2L(n-1)} \right\}^{1/(r+1)}, \text{ si } 0 < r \leq 0.44,$$

$$A(r) = \delta_n \text{ si } 0.45 \leq r < 4;$$

on remarquera que, dans le premier cas, on a $A(r) < 2\delta_n$.

d) *Détermination de y* : Posons $C_1 = \frac{(r+1)b}{2L(n-1)}$, $C_2 = \frac{(r+1)(2a+b)}{2L(n-1)}$,

et
$$G_r(x, t) = t \left(x + \frac{1}{2} \delta_n \right)^{n+1} - \frac{n}{n-1} (r+1) \delta_n x^r ;$$
 alors :

$$y = x + \frac{1}{2} d_n, \text{ si } -C_2 \leq G_r(x, a+b) \leq -C_1,$$

$$y = \left\{ -\frac{1}{b} (G_r(x, a) + C_2) \right\}^{1/(r+1)}, \text{ si } G_r(x, a+b) \leq -C_2,$$

$$y = \left\{ -\frac{1}{b} (G_r(x, a) + C_1) \right\}^{1/(r+1)}, \text{ si } -C_1 \leq G_r(x, a+b).$$

Malgré la complexité apparente de ces résultats, leur programmation ne pose pas de problème particulier. La mise en œuvre concrète de l'estimation consistera à comparer les points de l'échantillon aux fonctions $x + \frac{1}{2} \delta_n$ et $y(x, a, b)$, et d'associer des accroissements de \hat{f}_n d'ordre a/δ_n ou b/δ_n .

3) Comparaison des estimations $\bar{f}_n, f^{**}, f^*, \hat{f}_n, f_n$:

On dira que l'estimateur g_n est meilleur que h_n , si les valeurs asymptotiques des erreurs quadratiques intégrées sur I_1 de h_n sont supérieures à celles de g_n : on notera : $g_n \succ h_n$. On obtient que :

- pour $0.25 \leq r \leq 0.74$, $f_n \succ f^{**} \succ \bar{f}_n$
- pour $0.75 \leq r \leq 1$, $f_n^* \succ f_n \succ f_n^{**} \succ \bar{f}_n$
- pour $1 < r < 4$, $\hat{f}_n \succ f_n^* \succ f_n \succ f_n^{**} \succ f_n$.

On constate donc que \hat{f}_n s'impose pour $1 < r < 4$; le biais intégré dans I_1 associé à \hat{f}_n décroît comme $n^{-9/5}$ au lieu de $n^{-(2r+1)/5}$ pour les quatre autres estimations, tandis que sa variance est du même ordre, $n^{-1-r/5}$.

La mise en œuvre de \hat{f}_n est, cependant, rendue délicate par le fait que r est inconnu. Pour obtenir un estimateur de ce type, il est nécessaire :

- soit de faire une hypothèse sur r ;
- soit d'estimer r , en faisant l'hypothèse que, dans l'intervalle $[0, \delta_n[$, l'approximation $f(x) \# Lx^r$ est valable.

En pratique, on constate que :

a) Si la valeur de r obtenue est inférieure à 0.4, l'estimation obtenue pour r étant d'autant plus imprécise que r est voisin de 0, il est préférable d'estimer r dans un intervalle le plus grand possible (par exemple $[0, 2 \delta_n[$) ; il faut cependant éviter d'inclure dans cet intervalle le mode de la distribution.

Dans ce cas, aucun choix concret ne s'impose vraiment, sauf pour des échantillons très importants, où \hat{f}_n semble donner de meilleurs résultats. On a, en définitive le choix entre des estimations fortement biaisées (f_n^*, f_n), ou fortement fluctuantes ($\hat{f}_n, \bar{f}_n, f_n^{**}$).

b) Si la valeur de r obtenue est supérieure à 0,4, \hat{f}_n donne d'excellents résultats, mêmes pour de petits échantillons ($n \leq 100$). Lorsque r est voisin de 1 ($0,80 \leq r \leq 1,20$), l'utilisation de f_n^* au lieu de \hat{f}_n peut être faite sans perte notable d'efficacité ; f_n^* est sans biais pour $r = 1$, et correspond, dans ce cas, à la valeur asymptotique de f_n au coefficient b près.

c) Le cas $r = 0$ correspond à un cas isolé par rapport au modèle étudié. On pourra, dans ce cas, utiliser valablement f_n^{**} .

IV – APPLICATION AU MODELE PARAMETRIQUE DE LA LOI GAMMA

La loi Gamma généralisée, de densité :

$$f(x, r, \lambda, d) = \frac{\lambda^r}{\Gamma(r)} (x - d)^{r-1} \exp(-\lambda(x - d)), \text{ si } x > d,$$

représente un modèle général pratique pour modéliser des distributions unimodales. Le programme présenté ici compare les estimations paramétrique et non paramétrique de la densité à partir de ce modèle.

Il réalise les trois opérations suivantes :

1) Une estimation des paramètres r, λ, d , par le maximum de vraisemblance (JOHNSON & KOTZ, [3]), On peut procéder, au choix, en supposant que les trois paramètres sont inconnus, ou connus en partie.

La distribution obtenue est comparée à la distribution empirique par les tests du χ^2 et de Kolmogorov-Smirnov.

2) Une estimation non paramétrique par l'estimateur à fenêtre mobile (1.3). La détermination de $\delta_n = \delta_n(r)$ par (2.3) n'est ici possible que dans les cas $r = 1, 2, 3$, ou $r > 3.5$; on obtient, avec $\lambda = 1$:

– pour $r = 1, \delta_n(1) = \left(\frac{288}{n}\right)^{1/5},$

– pour $r = 2, \delta_n(2) = \left(\frac{576}{5n}\right)^{1/5},$

– pour $r = 3, \delta_n(3) = \left(\frac{768}{n}\right)^{1/5},$

– pour $r > 3.5$, en notant $[r]$ la partie entière de r , et f_r sa partie fractionnaire,

$$\delta_n(r) = \left\{ \frac{384}{n} 2^{2f_r} (f_r + 1)(f_r + 2) Q(r) \frac{\Gamma(f_r + 1)^2}{\Gamma(2f_r + 1)} \right\}^{1/5}, \quad (4.1)$$

avec $Q(r) = 1$, si $r = 3$, et

$$Q(r) = \prod_{j=1}^{[r]-3} \frac{2(r-j)}{2(r-j)-5}, \text{ si } [r] > 3.$$

Dans le cas où r est non entier et compris entre 1 et 3.5, un travail par simulation a montré qu'il était possible d'utiliser en pratique (4.1) pour $r > 2.70$, en remplaçant f_r par $f_r - 1$, si $r < 3$; si $1 < r \leq 2.70$, il suffit, soit de procéder à des interpolations linéaires entre $\delta_n(1)$, $\delta(2)$, $\delta(3)$, soit de choisir $\delta_n(1)$ si r appartient à $[1, 1.50] \cup [2.25, 2.70]$, et $\delta_n(2)$, autrement.

L'estimation non paramétrique a été établie sur l'échantillon sous forme réduite, par la transformation $\lambda(X_1 - d)$.

3) Une comparaison des estimations f_n , f_n^* , et \hat{f}_n , ce dernier étant déterminé soit par le maximum de vraisemblance, soit par les moindres carrés à partir de la fonction de répartition empirique.

Présentation des résultats

Nous avons traité dans ce qui suit trois échantillons :

a) *Echantillon 1* : Obtenu par simulation d'une loi Gamma de paramètre $r = 4$, avec une taille $n = 100$. Le tableau 1 résume les résultats numériques d'estimation, le tableau 2, compare diverses estimations :

Estimation 1 : f_n , Estimation 2 : f_n^* , Estimation 3 : \hat{f}_n ;

On compare également les estimations entre 0 et $\frac{1}{2} \delta_n$.

Le graphe 3 montre une visualisation du test de Kolmogorov-Smirnov. Les graphes 4, 5, 6 montrent les estimations paramétrique et non-paramétrique dans chacun des cas.

Les graphes 7 et 8 comparent les résultats d'estimation dans l'intervalle $0, \frac{1}{2} \delta_n$, où r est estimé par le maximum de vraisemblance (7), et par les moindres carrés (8).

Dans le cas présent, la courbe d'estimation paramétrique a été remplacée par la densité réelle, afin de permettre une comparaison avec les estimations.

b) *Echantillon 2* : Obtenu avec des données réelles de hauteurs de vagues en mer du Nord. Taille de l'échantillon 72.

c) *Echantillon 3* : Obtenu avec des données réelles de hauteurs de vagues en mer du Nord. Taille de l'échantillon 94.

Les échantillons 2 et 3 sont présentés dans le même ordre que l'échantillon 1.

Les programmes ont été exécutés sur un ordinateur CII Iris 80 et un traceur Benson.

V – CONCLUSION

Il est utile d'insister sur le fait que la procédure décrite précédemment peut s'exécuter sans choix extérieur, les paramètres d'estimation étant tous déterminés directement à partir des données.

Les trois exemples traités in extenso correspondent aux situations suivantes :

Echantillon 1 : On vérifie que les estimations non paramétriques de la densité fournissent dans un cas où la distribution est déjà connue, une très bonne estimation, même pour un petit échantillon.

Echantillon 2 : Cet échantillon a une distribution qui peut être valablement modélisé par une loi Gamma. On vérifie la concordance des estimations paramétriques et non paramétrique.

Echantillon 3 : Cet échantillon est clairement bimodal, néanmoins cet état n'est pas clairement décelé par les tests. L'estimation non paramétrique met de façon très directe ceci en évidence. On se rend compte de l'erreur de modélisation qui aurait pu être faite avec une loi Gamma.

Pour les trois échantillons, on constate l'efficacité remarquable des modifications introduits au voisinage des bornes du support.

Ces méthodes peuvent être utilisées sans modification de principe, pour des hypothèses paramétriques plus générales ou différentes de celles que nous avons utilisées ici. Le fait qu'on utilise une estimation paramétrique d'une loi Gamma n'intervient pas dans l'étude des modifications aux extrémités du support.

REFERENCES

- [1] DEHEUVELS P. (1977). — Estimation non paramétrique de la densité par histogrammes généralisés, *Rev. Stat. appl.* 25, p. 5-42.
- [2] DEHEUVELS P. (1977). — Estimation non paramétrique de la densité par histogrammes généralisés II, *Publ. Inst. Statist. Univ. Paris*, XXII, f. 1-2, P. 1-24.
- [3] JOHNSON N.L., KOTZ S. (1970). — *Continuous univariate distributions*, T.I, Houghton Mifflin
- [4] HOMINAL P. (1979). — Thèse à l'Université Paris VI (Preprint).
- [5] WEGMAN E. (1972). — Non parametric density estimation I, A summary of available methods, *Technometrics* 14, p. 533-546.
- [6] WEGMAN E. (1972). — Non parametric density estimation II, A comparison of density estimation methods, *J. Statist. Comp. And Simulation* 1, p. 225-246.
- [7] WERTZ W. (1978). — *Statistical density estimation*, A survey, Vandenhoeck & Ruprecht.
- [8] B.W. SILVERMAN, (1978). — Choosing a window when estimating a density, *Biometrika*, 65, p. 1-12.

ECHANTILLON N° 1 : SIMULATION DE VARIABLE GAMMA $\alpha = 4$

***** SIMULATION DE VARIABLE GAMMA (A=7, B=1, G=0) *****
 ***** COMPARAISON DES ESTIMATEURS ENTRE G ET DELTAN/2 *****
 ESTIMATEUR 3 : R = 1.85 (VRAISEMBLANCE) A OPT. = 30 ET B OPT. = 53

I	ESTIMATION			DENSITE	POINT ESTIMATION			DENSITE
	1	2	3		1	2	3	
1	0.001	0.019	0.000	0.04	0.001	0.000	0.000	0.000
2	0.028	0.019	0.035	0.004	0.001	0.000	0.000	0.000
3	0.055	0.019	0.005	0.004	0.001	0.000	0.000	0.000
4	0.083	0.019	0.005	0.004	0.001	0.000	0.000	0.000
5	0.110	0.023	0.009	0.009	0.003	0.000	0.000	0.000
6	0.138	0.023	0.009	0.009	0.004	0.000	0.000	0.000
7	0.165	0.023	0.009	0.009	0.005	0.001	0.001	0.001
8	0.193	0.023	0.009	0.009	0.007	0.001	0.001	0.001
9	0.221	0.023	0.009	0.009	0.008	0.001	0.001	0.001
10	0.248	0.023	0.009	0.009	0.011	0.002	0.002	0.002
11	0.276	0.023	0.009	0.009	0.013	0.003	0.003	0.003
12	0.304	0.023	0.009	0.009	0.015	0.004	0.004	0.004
13	0.331	0.023	0.009	0.009	0.015	0.004	0.004	0.004
14	0.359	0.037	0.023	0.013	0.019	0.007	0.007	0.007
15	0.386	0.037	0.023	0.013	0.019	0.007	0.007	0.007
16	0.414	0.042	0.033	0.023	0.021	0.008	0.008	0.008
17	0.441	0.042	0.033	0.023	0.021	0.008	0.008	0.008
18	0.469	0.042	0.033	0.023	0.025	0.011	0.011	0.011
19	0.496	0.042	0.033	0.023	0.027	0.012	0.012	0.012
20	0.524	0.042	0.033	0.023	0.029	0.014	0.014	0.014

***** COMPARAISON DES ESTIMATEURS ENTRE G ET DELTAN/2 *****
 ESTIMATEUR 3 : R = 1.76 (INDRES CARRES A OPT. = 82 ET B OPT. = 450)

I	ESTIMATION			DENSITE	POINT ESTIMATION			DENSITE
	1	2	3		1	2	3	
1	0.001	0.019	0.000	0.04	0.001	0.000	0.000	0.000
2	0.028	0.019	0.035	0.004	0.001	0.000	0.000	0.000
3	0.055	0.019	0.005	0.004	0.001	0.000	0.000	0.000
4	0.083	0.019	0.005	0.004	0.001	0.000	0.000	0.000
5	0.110	0.023	0.009	0.009	0.003	0.000	0.000	0.000
6	0.138	0.023	0.009	0.009	0.004	0.000	0.000	0.000
7	0.165	0.023	0.009	0.009	0.005	0.001	0.001	0.001
8	0.193	0.023	0.009	0.009	0.007	0.001	0.001	0.001
9	0.221	0.023	0.009	0.009	0.008	0.001	0.001	0.001
10	0.248	0.023	0.009	0.009	0.011	0.002	0.002	0.002
11	0.276	0.023	0.009	0.009	0.013	0.003	0.003	0.003
12	0.304	0.023	0.009	0.009	0.015	0.004	0.004	0.004
13	0.331	0.023	0.009	0.009	0.015	0.004	0.004	0.004
14	0.359	0.037	0.023	0.013	0.019	0.007	0.007	0.007
15	0.386	0.037	0.023	0.013	0.019	0.007	0.007	0.007
16	0.414	0.042	0.033	0.023	0.021	0.008	0.008	0.008
17	0.441	0.042	0.033	0.023	0.021	0.008	0.008	0.008
18	0.469	0.042	0.033	0.023	0.025	0.011	0.011	0.011
19	0.496	0.042	0.033	0.023	0.027	0.012	0.012	0.012
20	0.524	0.042	0.033	0.023	0.029	0.014	0.014	0.014

Tableau 2

***** SIMULATION DE VARIABLE GAMMA (A=6, B=1, G=0) *****
 ***** COMPARAISON DES ESTIMATEURS ENTRE G ET DELTAN/2 *****
 ESTIMATEUR 3 : R = 1.85 (VRAISEMBLANCE) A OPT. = 30 ET B OPT. = 53

VALEURS DE L'ECHANTILLON

3.79 1.74 2.98 4.54 4.18 4.16 3.76 3.59 3.24 2.71 2.58 3.76 1.82 5.45 4.69
 2.63 6.56 4.51 4.77 3.27 3.53 4.77 4.46 3.99 5.06 2.81 7.73 1.39 4.43 3.24
 3.51 4.2 2.03 1.93 4.55 2.96 3.49 4.80 2.88 4.98 5.68 1.64 1.47 3.24 6.60
 3.93 2.52 4.24 3.77 7.13 3.7 2.45 3.05 3.16 3.54 5.11 3.13 6.74 5.29
 1.13 3.38 3.97 2.01 2.50 6.87 6.4 2.84 3.56 5.04 3.15 2.67 6.08 3.25 8.29
 1.71 3.11 4.35 1.07 4.52 2.56 4.70 2.89 3.42 6.54 2.74 3.38 6.26 3.49 4.93
 3.57 5.85 1.35 2.94 6.35 2.89 1.85 2.10 1.93 8.60

***** HYPOTHESE NULLE : ECHANTILLON ISSU D'UNE LOI GAMMA *****
 PARAMETRES ESTIMES : A = 3.667 B = 1.004 G = 0.68
 PARAMETRES CROISES : A = 4.000 B = 1.000 G = 0.68
 TEST DE KOLMOGOROFF-SMIRNOFF : 6.21 (ACCEPTATION : 8.11)
 TEST D'ADEQUATION DU CHI-2 : 6.21 (ACCEPTATION : 8.11)

***** ESTIMATION PAR LA FENETRE MOBILE SUR L'ECHANTILLON MOBILE PAR (G) B *****
 L'ANGULEUR DE LA FENETRE, DELTAN = 2.15 (DELTAN > 2C OBSERVATIONS)

***** SIMULATION DE VARIABLE GAMMA (A=7, B=1, G=0) *****
 ***** COMPARAISON DES ESTIMATEURS ENTRE G ET DELTAN/2 *****
 ESTIMATEUR 3 : R = 1.85 (VRAISEMBLANCE) A OPT. = 30 ET B OPT. = 53

VALEURS DE L'ECHANTILLON

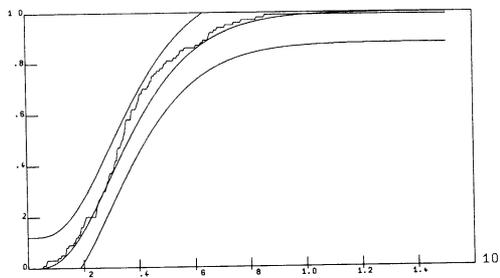
3.79 1.74 2.98 4.54 4.18 4.16 3.76 3.59 3.24 2.71 2.58 3.76 1.82 5.45 4.69
 2.63 6.56 4.51 4.77 3.27 3.53 4.77 4.46 3.99 5.06 2.81 7.73 1.39 4.43 3.24
 3.51 4.2 2.03 1.93 4.55 2.96 3.49 4.80 2.88 4.98 5.68 1.64 1.47 3.24 6.60
 3.93 2.52 4.24 3.77 7.13 3.7 2.45 3.05 3.16 3.54 5.11 3.13 6.74 5.29
 1.13 3.38 3.97 2.01 2.50 6.87 6.4 2.84 3.56 5.04 3.15 2.67 6.08 3.25 8.29
 1.71 3.11 4.35 1.07 4.52 2.56 4.70 2.89 3.42 6.54 2.74 3.38 6.26 3.49 4.93
 3.57 5.85 1.35 2.94 6.35 2.89 1.85 2.10 1.93 8.60

***** HYPOTHESE NULLE : ECHANTILLON ISSU D'UNE LOI GAMMA *****
 PARAMETRES ESTIMES : A = 3.667 B = 1.004 G = 0.68
 PARAMETRES CROISES : A = 4.000 B = 1.000 G = 0.68
 TEST DE KOLMOGOROFF-SMIRNOFF : 6.21 (ACCEPTATION : 8.11)
 TEST D'ADEQUATION DU CHI-2 : 6.21 (ACCEPTATION : 8.11)

***** ESTIMATION PAR LA FENETRE MOBILE SUR L'ECHANTILLON MOBILE PAR (G) B *****
 L'ANGULEUR DE LA FENETRE, DELTAN = 2.15 (DELTAN > 2C OBSERVATIONS)

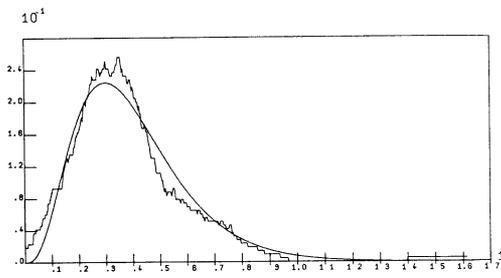
Tableau 1

ECHANTILLON N° 1 : SIMULATION DE VARIABLE GAMMA $a = 4$



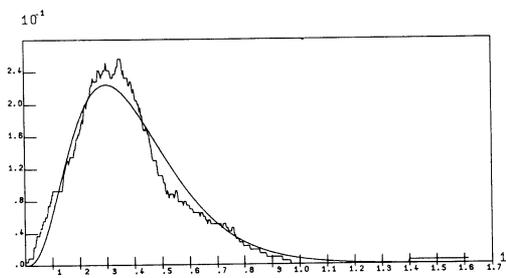
SIMULATION DE VARIABLE GAMMA (A=4, B=1, G=0) TEST K-S

Graphe 3



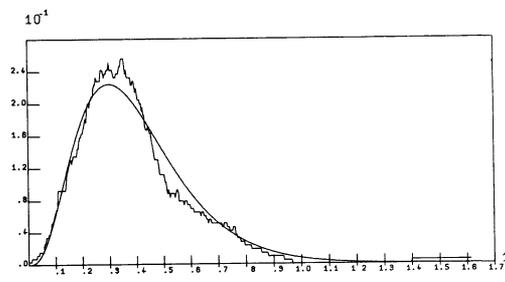
SIMULATION DE VARIABLE GAMMA (A=4, B=1, G=0) EST. N° 1

Graphe 4



SIMULATION DE VARIABLE GAMMA (A=4, B=1, G=0) EST. N° 2

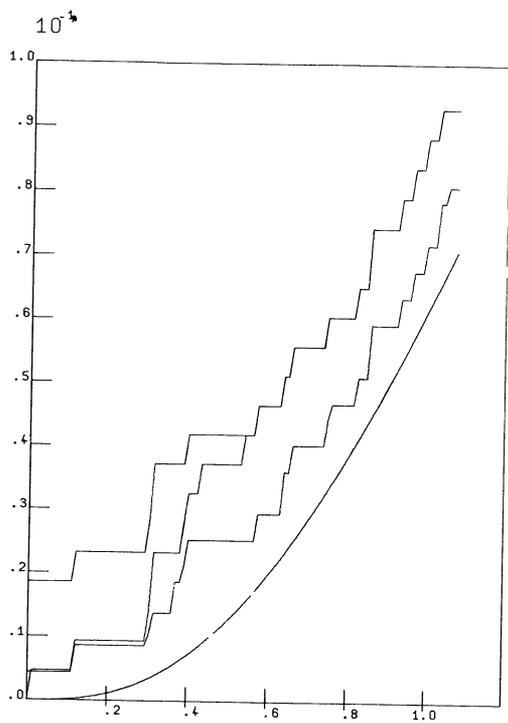
Graphe 5



SIMULATION DE VARIABLE GAMMA (A=4, B=1, G=0) EST. N° 3

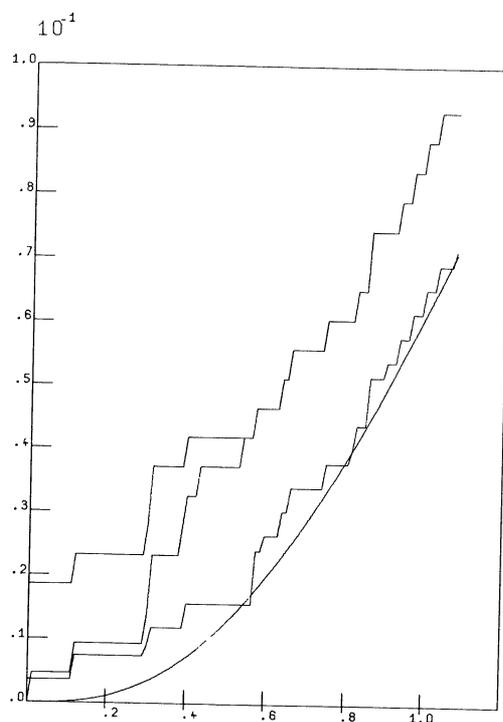
Graphe 6b

ECHANTILLON N° 1 : SIMULATION DE VARIABLE GAMMA $a = 4$



VARIABLE GAMMA (A=4, B=1, G=0)

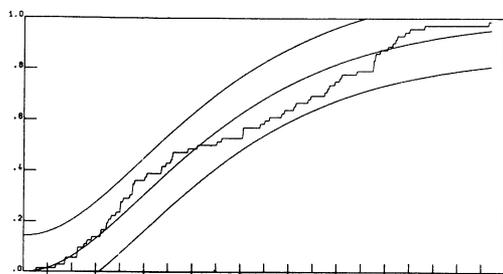
Grappe 7



VARIABLE GAMMA (A=4, B=1, G=0)

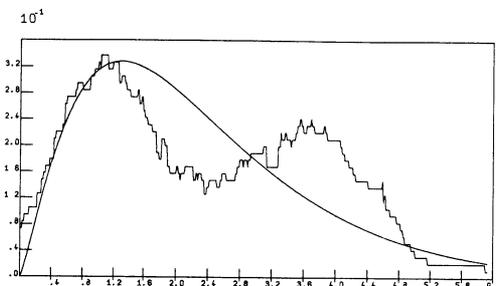
Grappe 8

ECHANTILLON N° 2 : HAUTEURS DE VAGUES EN MER DU NORD DU 19/7 au 28/7



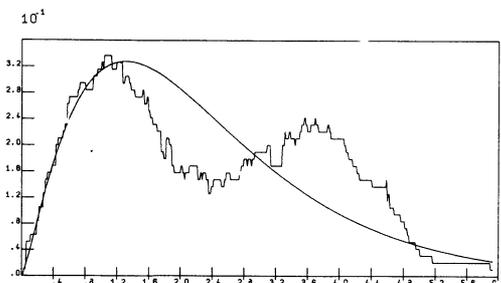
HAUTEURS DE VAGUES DU 19/7 (21H) AU 28/7 (18H) TEST K-S

Graphe 3



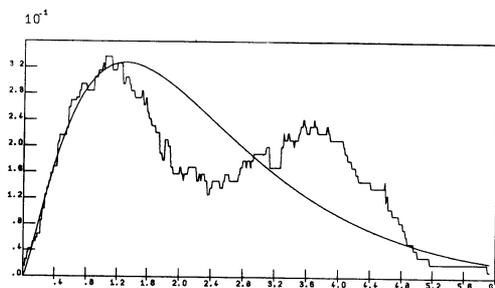
HAUTEURS DE VAGUES DU 19/7 (21H) AU 28/7 (18H) EST. N° 1

Graphe 4



HAUTEURS DE VAGUES DU 19/7 (21H) AU 28/7 (18H) EST. N° 2

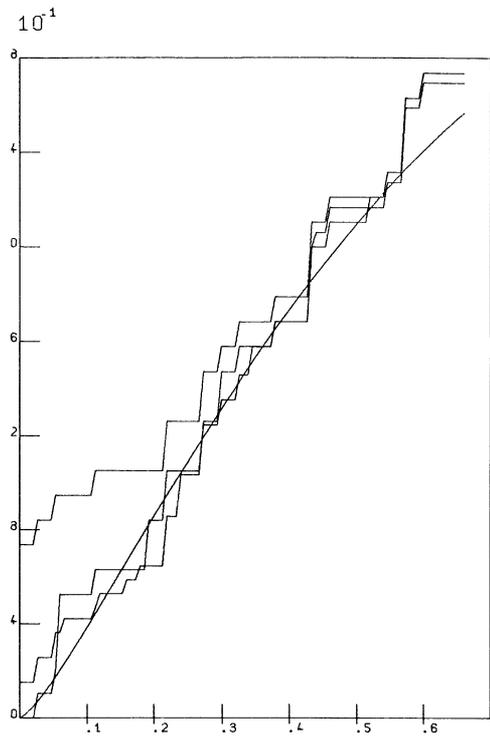
Graphe 5



HAUTEURS DE VAGUES DU 19/7 (21H) AU 28/7 (18H) EST. N° 3

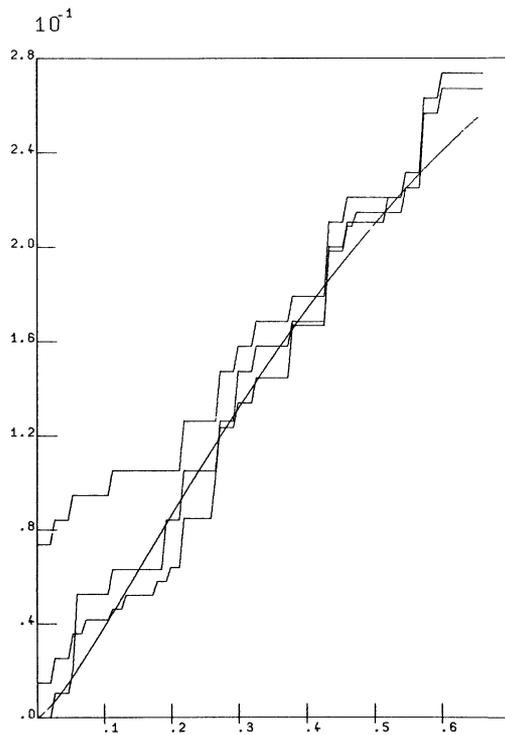
Graphe 6a

ECHANTILLON N° 2 : HAUTEURS DE VAGUES EN MER DU NORD DU 19/7 au 28/7



VAGUES DU 19/7 (21H) AU 28/7 (18H)

Graphe 7



VAGUES DU 19/7 (21H) AU 28/7 (18H)

Graphe 8

ECHANTILLON N° 3 : HAUTEURS DE VAGUES EN MER DU NORD DU 27/8 au 8/9

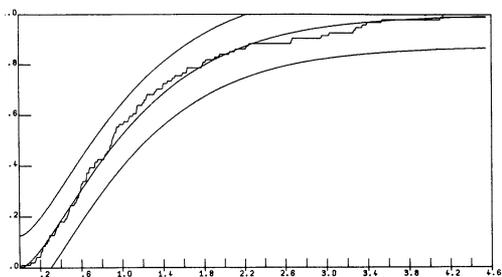
HAUTEURS DE VAGUES DU 27/8 (18h) AU 8/5 (9h)

COMPARAISON DES ESTIMATEURS ENTRE C ET DELTAN/2

ESTIMATEUR 3 : R = .41 (VRAISBLANCE) A BPT. = 1.0C ET B BPT. = .38

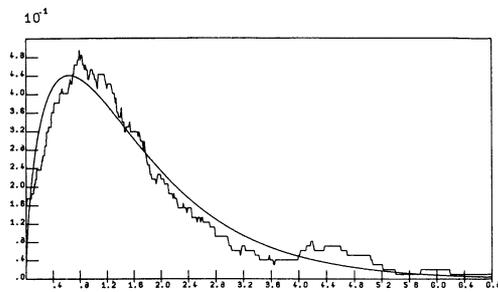
POINT	ESTIMATION			DENSITE			ILX*RI*GAMMA
	1	2	3	1	2	3	
1001	1.55	1.60	1.13	.90	.90	.90	.321
1002	1.58	1.62	1.14	.91	.91	.91	.321
1003	1.61	1.65	1.15	.92	.92	.92	.321
1004	1.64	1.68	1.16	.93	.93	.93	.321
1005	1.67	1.71	1.17	.94	.94	.94	.321
1006	1.70	1.74	1.18	.95	.95	.95	.321
1007	1.73	1.77	1.19	.96	.96	.96	.321
1008	1.76	1.80	1.20	.97	.97	.97	.321
1009	1.79	1.83	1.21	.98	.98	.98	.321
1010	1.82	1.86	1.22	.99	.99	.99	.321
1011	1.85	1.89	1.23	1.00	1.00	1.00	.321
1012	1.88	1.92	1.24	1.01	1.01	1.01	.321
1013	1.91	1.95	1.25	1.02	1.02	1.02	.321
1014	1.94	1.98	1.26	1.03	1.03	1.03	.321
1015	1.97	2.01	1.27	1.04	1.04	1.04	.321
1016	2.00	2.04	1.28	1.05	1.05	1.05	.321
1017	2.03	2.07	1.29	1.06	1.06	1.06	.321
1018	2.06	2.10	1.30	1.07	1.07	1.07	.321
1019	2.09	2.13	1.31	1.08	1.08	1.08	.321
1020	2.12	2.16	1.32	1.09	1.09	1.09	.321
1021	2.15	2.19	1.33	1.10	1.10	1.10	.321
1022	2.18	2.22	1.34	1.11	1.11	1.11	.321
1023	2.21	2.25	1.35	1.12	1.12	1.12	.321
1024	2.24	2.28	1.36	1.13	1.13	1.13	.321
1025	2.27	2.31	1.37	1.14	1.14	1.14	.321
1026	2.30	2.34	1.38	1.15	1.15	1.15	.321
1027	2.33	2.37	1.39	1.16	1.16	1.16	.321
1028	2.36	2.40	1.40	1.17	1.17	1.17	.321
1029	2.39	2.43	1.41	1.18	1.18	1.18	.321
1030	2.42	2.46	1.42	1.19	1.19	1.19	.321
1031	2.45	2.49	1.43	1.20	1.20	1.20	.321
1032	2.48	2.52	1.44	1.21	1.21	1.21	.321
1033	2.51	2.55	1.45	1.22	1.22	1.22	.321
1034	2.54	2.58	1.46	1.23	1.23	1.23	.321
1035	2.57	2.61	1.47	1.24	1.24	1.24	.321
1036	2.60	2.64	1.48	1.25	1.25	1.25	.321
1037	2.63	2.67	1.49	1.26	1.26	1.26	.321
1038	2.66	2.70	1.50	1.27	1.27	1.27	.321
1039	2.69	2.73	1.51	1.28	1.28	1.28	.321
1040	2.72	2.76	1.52	1.29	1.29	1.29	.321
1041	2.75	2.79	1.53	1.30	1.30	1.30	.321
1042	2.78	2.82	1.54	1.31	1.31	1.31	.321
1043	2.81	2.85	1.55	1.32	1.32	1.32	.321
1044	2.84	2.88	1.56	1.33	1.33	1.33	.321
1045	2.87	2.91	1.57	1.34	1.34	1.34	.321
1046	2.90	2.94	1.58	1.35	1.35	1.35	.321
1047	2.93	2.97	1.59	1.36	1.36	1.36	.321
1048	2.96	3.00	1.60	1.37	1.37	1.37	.321
1049	2.99	3.03	1.61	1.38	1.38	1.38	.321
1050	3.02	3.06	1.62	1.39	1.39	1.39	.321
1051	3.05	3.09	1.63	1.40	1.40	1.40	.321
1052	3.08	3.12	1.64	1.41	1.41	1.41	.321
1053	3.11	3.15	1.65	1.42	1.42	1.42	.321
1054	3.14	3.18	1.66	1.43	1.43	1.43	.321
1055	3.17	3.21	1.67	1.44	1.44	1.44	.321
1056	3.20	3.24	1.68	1.45	1.45	1.45	.321
1057	3.23	3.27	1.69	1.46	1.46	1.46	.321
1058	3.26	3.30	1.70	1.47	1.47	1.47	.321
1059	3.29	3.33	1.71	1.48	1.48	1.48	.321
1060	3.32	3.36	1.72	1.49	1.49	1.49	.321
1061	3.35	3.39	1.73	1.50	1.50	1.50	.321
1062	3.38	3.42	1.74	1.51	1.51	1.51	.321
1063	3.41	3.45	1.75	1.52	1.52	1.52	.321
1064	3.44	3.48	1.76	1.53	1.53	1.53	.321
1065	3.47	3.51	1.77	1.54	1.54	1.54	.321
1066	3.50	3.54	1.78	1.55	1.55	1.55	.321
1067	3.53	3.57	1.79	1.56	1.56	1.56	.321
1068	3.56	3.60	1.80	1.57	1.57	1.57	.321
1069	3.59	3.63	1.81	1.58	1.58	1.58	.321
1070	3.62	3.66	1.82	1.59	1.59	1.59	.321
1071	3.65	3.69	1.83	1.60	1.60	1.60	.321
1072	3.68	3.72	1.84	1.61	1.61	1.61	.321
1073	3.71	3.75	1.85	1.62	1.62	1.62	.321
1074	3.74	3.78	1.86	1.63	1.63	1.63	.321
1075	3.77	3.81	1.87	1.64	1.64	1.64	.321
1076	3.80	3.84	1.88	1.65	1.65	1.65	.321
1077	3.83	3.87	1.89	1.66	1.66	1.66	.321
1078	3.86	3.90	1.90	1.67	1.67	1.67	.321
1079	3.89	3.93	1.91	1.68	1.68	1.68	.321
1080	3.92	3.96	1.92	1.69	1.69	1.69	.321
1081	3.95	3.99	1.93	1.70	1.70	1.70	.321
1082	3.98	4.02	1.94	1.71	1.71	1.71	.321
1083	4.01	4.05	1.95	1.72	1.72	1.72	.321
1084	4.04	4.08	1.96	1.73	1.73	1.73	.321
1085	4.07	4.11	1.97	1.74	1.74	1.74	.321
1086	4.10	4.14	1.98	1.75	1.75	1.75	.321
1087	4.13	4.17	1.99	1.76	1.76	1.76	.321
1088	4.16	4.20	2.00	1.77	1.77	1.77	.321
1089	4.19	4.23	2.01	1.78	1.78	1.78	.321
1090	4.22	4.26	2.02	1.79	1.79	1.79	.321
1091	4.25	4.29	2.03	1.80	1.80	1.80	.321
1092	4.28	4.32	2.04	1.81	1.81	1.81	.321
1093	4.31	4.35	2.05	1.82	1.82	1.82	.321
1094	4.34	4.38	2.06	1.83	1.83	1.83	.321
1095	4.37	4.41	2.07	1.84	1.84	1.84	.321
1096	4.40	4.44	2.08	1.85	1.85	1.85	.321
1097	4.43	4.47	2.09	1.86	1.86	1.86	.321
1098	4.46	4.50	2.10	1.87	1.87	1.87	.321
1099	4.49	4.53	2.11	1.88	1.88	1.88	.321
1100	4.52	4.56	2.12	1.89	1.89	1.89	.321
1101	4.55	4.59	2.13	1.90	1.90	1.90	.321
1102	4.58	4.62	2.14	1.91	1.91	1.91	.321
1103	4.61	4.65	2.15	1.92	1.92	1.92	.321
1104	4.64	4.68	2.16	1.93	1.93	1.93	.321
1105	4.67	4.71	2.17	1.94	1.94	1.94	.321
1106	4.70	4.74	2.18	1.95	1.95	1.95	.321
1107	4.73	4.77	2.19	1.96	1.96	1.96	.321
1108	4.76	4.80	2.20	1.97	1.97	1.97	.321
1109	4.79	4.83	2.21	1.98	1.98	1.98	.321
1110	4.82	4.86	2.22	1.99	1.99	1.99	.321
1111	4.85	4.89	2.23	2.00	2.00	2.00	.321
1112	4.88	4.92	2.24	2.01	2.01	2.01	.321
1113	4.91	4.95	2.25	2.02	2.02	2.02	.321
1114	4.94	4.98	2.26	2.03	2.03	2.03	.321
1115	4.97	5.01	2.27	2.04	2.04	2.04	.321
1116	5.00	5.04	2.28	2.05	2.05	2.05	.321
1117	5.03	5.07	2.29	2.06	2.06	2.06	.321
1118	5.06	5.10	2.30	2.07	2.07	2.07	.321
1119	5.09	5.13	2.31	2.08	2.08	2.08	.321
1120	5.12	5.16	2.32	2.09	2.09	2.09	.321
1121	5.15	5.19	2.33	2.10	2.10	2.10	.321
1122	5.18	5.22	2.34	2.11	2.11	2.11	.321
1123	5.21	5.25	2.35	2.12	2.12	2.12	.321
1124	5.24	5.28	2.36	2.13	2.13	2.13	.321
1125	5.27	5.31	2.37	2.14	2.14	2.14	.321
1126	5.30	5.34	2.38	2.15	2.15	2.15	.321
1127	5.33	5.37	2.39	2.16	2.16	2.16	.321
1128	5.36	5.40	2.40	2.17	2.17	2.17	.321
1129	5.39	5.43	2.41	2.18	2.18	2.18	.321
1130	5.42	5.46	2.42	2.19	2.19	2.19	.321
1131	5.45	5.49	2.43	2.20	2.20	2.20	.321
1132	5.48	5.52	2.44	2.21	2.21	2.21	.321
1133	5.51	5.55	2.45	2.22	2.22	2.22	.321
1134	5.54	5.58	2.46	2.23	2.23	2.23	.321
1135	5.57	5.61	2.47	2.24	2.24	2.24	.321
1136	5.60	5.64	2.48	2.25	2.25	2.25	.321
1137	5.63	5.67	2.49	2.26	2.26	2.26	.321
1138	5.66	5.70	2.50	2.27	2.27	2.27	.321
1139	5.69	5.73	2.51	2.28	2.28	2.28	.321
1140	5.72	5.76	2.52	2.29	2.29	2.29	.321
1141	5.75	5.79	2.53	2.30	2.30	2.30	.321
1142	5.78	5.82	2.54	2.31	2.31	2.31	.321
1143	5.81	5.85	2.55	2.32	2.32	2.32	.321
1144	5.84	5.88	2.56	2.33	2.33	2.33	.321
1145	5.87	5.91	2.57	2.34	2.34	2.34	.321
1146	5.90	5.94	2.58	2.35	2.35	2.35	.321
1147	5.93	5.97	2.59	2.36	2.36	2.36	.321
1148	5.96	6.00	2.60	2.37	2.37	2.37	.321
1149	5.99	6.03	2.61	2.38	2.38	2.38	.321
1150	6.02	6.06	2.62	2.39	2.39	2.39	.321
1151	6.05	6.09	2.63	2.40	2.40	2.40	.321
1152	6.08	6.12	2.64	2.41	2.41	2.41	.321
1153	6.11	6.15	2.65	2.42	2.42	2.42	.321
1154	6.14	6.18	2.66	2.43	2.43	2.43	.321
1155	6.17	6.21	2.67	2.44	2.44	2.44	.321
1156	6.20	6.24	2.68	2.45	2.45	2.45	.321
1157	6.23	6.27	2.69	2.46	2.46	2.46	.321
1158	6.26	6.30	2.70	2.47	2.47	2.47	.321
1159	6.29	6.33	2.71	2.48	2.48	2.48	.321
1160	6.32	6.36	2.72	2			

ECHANTILLON N° 3 : HAUTEURS DE VAGUES EN MER DU NORD DU 27/8 au 8/9



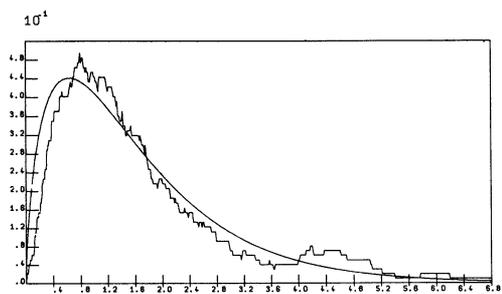
HAUTEURS DE VAGUES DU 27/8 (18H) AU 8/9 (9H) TEST K-S

Graphe 3



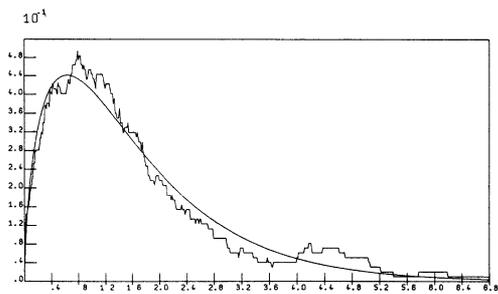
HAUTEURS DE VAGUES DU 27/8 (18H) AU 8/9 (9H) EST. N° 1

Graphe 4



HAUTEURS DE VAGUES DU 27/8 (18H) AU 8/9 (9H) EST. N° 2

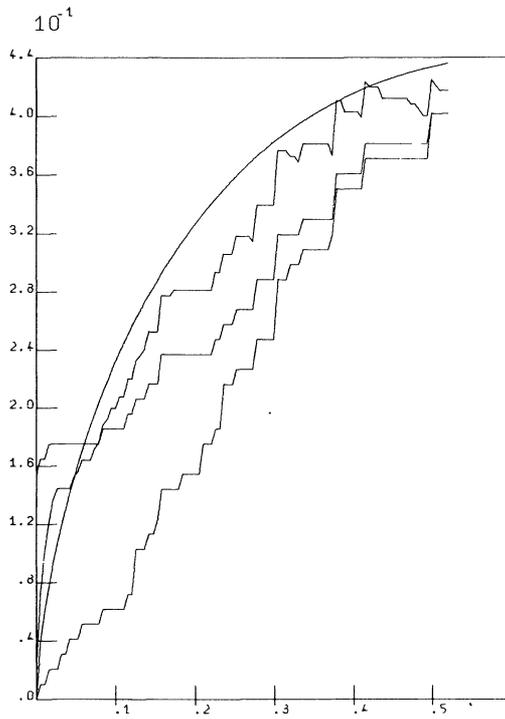
Graphe 5



HAUTEURS DE VAGUES DU 27/8 (18H) AU 8/9 (9H) EST. N° 3

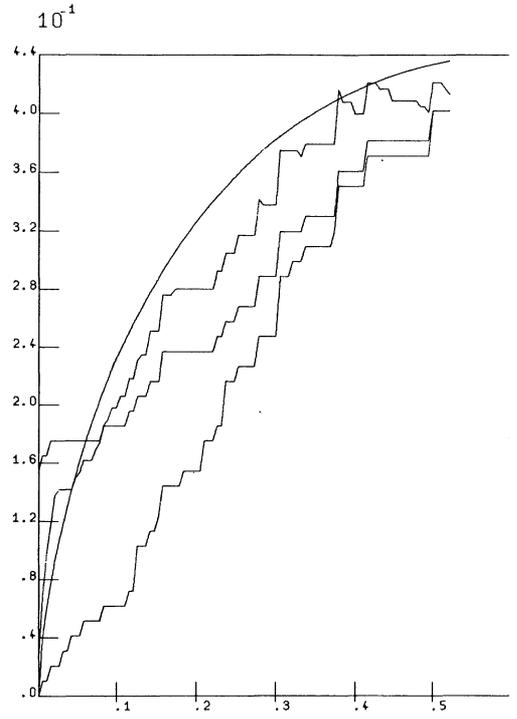
Graphe 6a

ECHANTILLON N° 3 : HAUTEURS DE VAGUES EN MER DU NORD DU 27/8 au 8/9



VAGUES DU 27/8 (18H) AU 8/9 (9H)

Graphe 7



VAGUES DU 27/8 (18H) AU 8/9 (9H)

Graphe 8