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, nº 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 Parsen-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 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 wholely 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 $X_1, \ldots, X_N$. 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 nécessistant pas un choix multiple de paramètres, est la méthode du noyau de Parzen-Rosenblatt. Si $X_1, \ldots, X_n$ est un échantillon de v.a. indépendantes de même loi de densité f sur $\mathbf{R}$, on estime f(x) par :

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

– K(u) est généralement supposé tel que :

$$\text{(i)} \quad K(u) \geqslant 0 \quad ; \quad \text{(ii)} \int K(u)\, du = 1 \quad ; \quad \text{(iii)} \ |K(u)| \leqslant C \ ;$$

– 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 $X_1, \ldots, X_n$ est généralement définie par :

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

La dérivée par différentiation finie de F_n est définie par :

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

où le noyau K(u) est le noyau uniforme : $K(u) = \begin{cases} 1 & \text{si} \quad |u| \leqslant \frac{1}{2} \\ 0 & \text{si} \quad |u| > \frac{1}{2} \end{cases}$.

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 :

$$x = X_{1,n} - \frac{1}{2}\delta_n, \quad \text{où} \quad X_{1,n} = \operatorname*{Inf}_{1 \leqslant i \leqslant n} X_i.$$

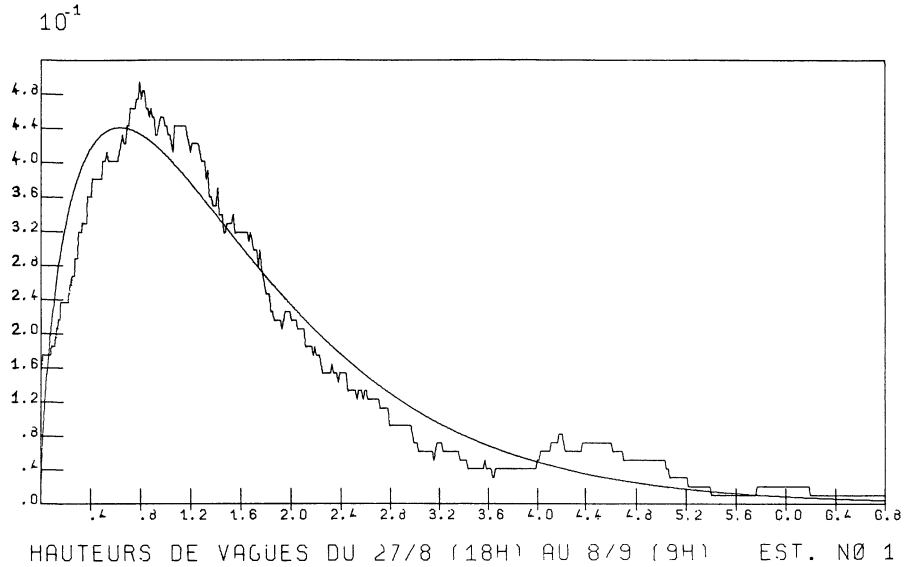


Figure 1

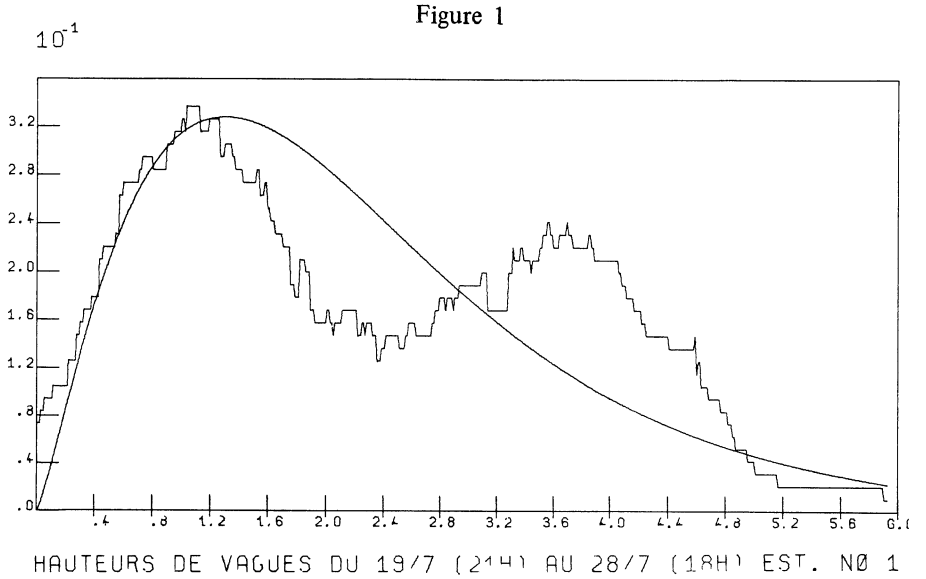


Figure 2

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 :

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

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 :

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

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 $0 \leqslant x \leqslant \delta_n \leqslant A(r)$.

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 :

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

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

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

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

$$\begin{aligned} B_{1,I}^2 &= \frac{1}{4}\delta_n^4 \left(\int_I y^2 K(y)\, dy\right)^2 \left(\int_I f''(x)^2\, dx\right) + o(\delta_n^4), \\ B_{2,I}^2 &= \frac{1}{n\delta_n}\left(\int_I K^2(y)\, dy\right) + o\left(\frac{1}{n\delta_n}\right). \end{aligned} \tag{2.1}$$

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 :

$K(y) = \frac{3}{4\sqrt{5}}(1 - y^2/5)$, si $|y| \leqslant \sqrt{5}$, 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 :

$\theta = \int_I f''(x)^2\,dx$. 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 :

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

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

$$M_{I_2}^2 = \frac{\delta_n^4}{576}\int_R f''(x)^2\,dx + \frac{1}{n\delta_n} + o(\delta_n^4) + o\left(\frac{1}{n\delta_n}\right). \tag{3.2}$$

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 $X = x/\delta_n$:

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

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

$$M_I^2 \sim K_1\,\delta_n^{2r+1} + \frac{1}{n\delta_n} \tag{3.4}$$

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

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

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

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'alinea 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) $$\bar{f}_n(x) = \frac{1}{x + \frac{1}{2}\delta_n} F_n\left(x + \frac{1}{2}\delta_n\right):$$

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

(estimateur de la ficelle I)

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

(estimateur de la ficelle II)

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

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 : $0 < \delta_n \leqslant A(r)$.

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 : $E((\hat{f}_n(x\,,y\,,a\,,b) - f(x))^2)$;

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

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

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

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

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

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

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

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

$$u \text{ est solution de } u^r\left(v + \frac{1}{2}\right)^{r+1} = v^r\left(u + \frac{1}{2}\right)^{r+1}, \text{ si } 0 < r \leqslant 0{,}31.$$

c) *Détermination de* A(r) :

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

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

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

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

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

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

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

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

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 $\overline{f}_n$, 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 $0.25 \leqslant r \leqslant 0.74$, $\quad f_n \succ f^{**} \succ \overline{f}_n$

– pour $0.75 \leqslant r \leqslant 1$, $\quad f_n^* \succ f_n \succ f_n^{**} \succ \overline{f}_n$

– pour $1 < r < 4$, $\quad \hat{f}_n \succ f_n^* \succ f_n \succ f_n^{**} \succ f_n$.

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 ($\hat{f}_n$, $\overline{f}_n$, 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 ($n \leqslant 100$). Lorsque r est voisin de 1 ($0.80 \leqslant r \leqslant 1{,}20$), 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 :

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

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,

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

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

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

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 $1 < r \leqslant 2.70$, 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 $\lambda(X_i - d)$.

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, lé 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 histrogrammes 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.

SIMULATION DE VARIABLE GAMMA (A=4, B=1, G=0) NOMBRE DE VALEURS : 100

VALEURS DE L ECHANTILLON

3.79	1.74	2.98	.54	4.18	4.06	3.76	3.59	3.24	2.71	2.58	3.76	1.82	5.50	.69					
2.63	6.56	4.51	4.07	3.27	3.53	7.47	4.46	3.99	15.06	2.81	7.73	1.39	4.43	3.24					
3.51	4.04	2.03	1.93	4.59	2.96	3.49	4.80	2.88	4.98	5.68	1.64	1.47	3.24	6.60					
3.93	2.52	4.14	2.47	3.77	7.13	1.37	2.45	3.05	3.16	3.54	5.41	3.13	6.74	5.29					
1.19	3.38	3.97	2.01	2.50	6.87	.64	2.84	3.56	5.04	3.19	2.07	6.08	3.25	8.29					
1.71	3.11	4.35	1.07	4.52	2.56	4.70	2.49	3.42	6.54	2.74	3.38	6.26	3.49	4.93					
3.57	5.55	1.39	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 = 3.867 B = 1.004 G = -.058
PARAMETRES CHOISIS : A = 4.000 B = 1.000 G = .000

TEST DE KOLMOGOROFF-SMIRNOFF : .101 (ACCEPTATION : 8LI)
TEST D ADEQUATION DU CHI-2 : 6.21 (ACCEPTATION : 8LI)

ESTIMATION PAR LA FENETRE MOBILE SUR L ECHANTILLON MODIFIE PAR (.-G)*B
LONGUEUR DE LA FENETRE, DELTAN = 2.161 (DELTAN > 20 OBSERVATIONS)

POINT	ESTM	DENS	POINT	ESTM	DENS	POINT	ESTM	DENS	POINT	ESTM	DENS
.00	.019	.000	3.13	.242	.223	7.47	.042	.039	11.82	.000	.002
.08	.019	.000	3.31	.237	.221	7.65	.037	.035	11.99	.000	.002
.16	.023	.001	3.48	.256	.216	7.82	.028	.032	12.17	.000	.002
.25	.023	.002	3.65	.233	.211	8.00	.023	.029	12.34	.000	.001
.33	.037	.004	3.83	.223	.203	8.17	.023	.026	12.51	.000	.001
.41	.042	.008	4.00	.219	.195	8.34	.019	.023	12.69	.000	.001
.45	.042	.012	4.17	.200	.187	8.52	.019	.021	12.86	.000	.001
.57	.047	.018	4.35	.167	.177	8.69	.014	.018	13.03	.000	.001
.65	.051	.024	4.52	.163	.168	8.86	.009	.016	13.21	.000	.001
.74	.056	.032	4.69	.130	.158	9.04	.009	.015	13.38	.000	.001
.82	.065	.040	4.87	.112	.148	9.21	.009	.013	13.56	.000	.001
.90	.074	.049	5.04	.112	.138	9.39	.005	.012	13.73	.000	.000
.98	.084	.059	5.22	.088	.128	9.56	.005	.010	13.90	.000	.000
1.06	.093	.069	5.39	.088	.119	9.73	.000	.009	14.08	.005	.000
1.22	.093	.089	5.56	.088	.110	9.91	.000	.008	14.25	.005	.000
1.39	.098	.112	5.74	.079	.101	10.08	.000	.007	14.42	.005	.000
1.57	.130	.134	5.91	.074	.093	10.25	.000	.006	14.60	.005	.000
1.74	.135	.154	6.08	.070	.086	10.43	.000	.006	14.77	.005	.000
1.92	.158	.172	6.26	.065	.078	10.60	.000	.005	14.94	.005	.000
2.09	.177	.188	6.43	.065	.071	10.78	.000	.004	15.12	.005	.000
2.26	.195	.201	6.61	.056	.065	10.95	.000	.004	15.29	.005	.000
2.44	.223	.211	6.78	.051	.059	11.12	.000	.003	15.47	.005	.000
2.61	.228	.218	6.95	.051	.054	11.30	.000	.003	15.64	.005	.000
2.78	.242	.222	7.13	.051	.048	11.47	.000	.003	15.81	.005	.000
2.96	.242	.224	7.30	.051	.044	11.64	.000	.002	15.99	.005	.000

Tableau 1

COMPARAISON DES ESTIMATEURS ENTRE 0. ET DELTAN/2

ESTIMATEUR 3 : R = 1.45 (VRAISEMBLANCE) A 8PT. = .90 ET B 8PT. = -.53

POINT	ESTIMATION 1	2	3	DENSITE LX**R	GAMMA	POINT	ESTIMATION 1	2	3	DENSITE LX**R	GAMMA
.00	.019	.000	.004	.000	.000	.552	.042	.042	.025	.032	.016
.028	.019	.005	.004	.000	.000	.579	.047	.047	.029	.034	.018
.055	.019	.005	.004	.001	.000	.607	.047	.047	.029	.036	.020
.083	.019	.005	.004	.002	.000	.634	.047	.047	.032	.039	.023
.110	.023	.009	.009	.003	.000	.662	.056	.056	.040	.041	.025
.138	.023	.009	.009	.004	.000	.690	.056	.056	.040	.044	.027
.165	.023	.009	.009	.005	.001	.717	.056	.056	.040	.046	.030
.193	.023	.009	.009	.007	.001	.745	.060	.060	.044	.049	.033
.221	.023	.009	.009	.008	.001	.772	.060	.060	.047	.051	.035
.248	.023	.009	.009	.010	.002	.800	.060	.060	.047	.054	.038
.276	.023	.009	.009	.011	.003	.827	.065	.065	.051	.057	.041
.303	.028	.014	.010	.013	.003	.855	.070	.070	.055	.060	.044
.331	.037	.023	.014	.015	.004	.883	.074	.074	.059	.062	.047
.359	.037	.023	.014	.017	.005	.910	.074	.074	.059	.065	.051
.386	.037	.028	.019	.019	.007	.938	.079	.079	.063	.068	.054
.414	.042	.033	.025	.021	.008	.965	.084	.084	.068	.071	.057
.441	.042	.037	.025	.023	.009	.993	.084	.084	.068	.074	.060
.469	.042	.037	.025	.025	.011	1.020	.088	.088	.072	.077	.064
.496	.042	.037	.025	.027	.012	1.048	.093	.093	.078	.080	.067
.524	.042	.037	.025	.029	.014	1.076	.093	.093	.081	.083	.071

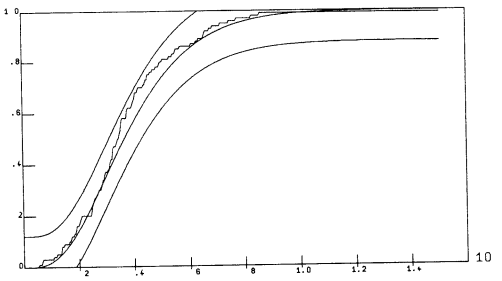
COMPARAISON DES ESTIMATEURS ENTRE 0. ET DELTAN/2

ESTIMATEUR 3 : R = 1.76 (MOINDRES CARRES) A 8PT. = .82 ET B 8PT. = -.50

POINT	ESTIMATION 1	2	3	DENSITE LX**R	GAMMA	POINT	ESTIMATION 1	2	3	DENSITE LX**R	GAMMA
.00	.019	.000	.004	.000	.000	.552	.042	.042	.016	.024	.016
.028	.019	.005	.004	.000	.000	.579	.047	.047	.024	.026	.018
.055	.019	.005	.004	.000	.000	.607	.047	.047	.026	.028	.020
.083	.019	.005	.004	.001	.000	.634	.047	.047	.028	.030	.023
.110	.023	.009	.007	.001	.000	.662	.056	.056	.034	.032	.025
.138	.023	.009	.007	.002	.000	.690	.056	.056	.034	.035	.027
.165	.023	.009	.007	.003	.001	.717	.056	.056	.034	.037	.030
.193	.023	.009	.007	.004	.001	.745	.060	.060	.038	.040	.033
.221	.023	.009	.007	.005	.001	.772	.060	.060	.038	.042	.035
.248	.023	.009	.007	.006	.002	.800	.060	.060	.038	.045	.038
.276	.023	.009	.007	.007	.003	.827	.065	.065	.044	.048	.041
.303	.028	.014	.009	.008	.003	.855	.070	.070	.048	.051	.044
.331	.037	.023	.012	.010	.004	.883	.074	.074	.051	.054	.047
.359	.037	.023	.012	.011	.005	.910	.074	.074	.054	.057	.051
.386	.037	.028	.012	.013	.007	.938	.079	.079	.058	.060	.054
.414	.042	.033	.016	.014	.008	.965	.084	.084	.061	.063	.057
.441	.042	.037	.016	.016	.009	.993	.084	.084	.061	.066	.060
.469	.042	.037	.016	.018	.011	1.020	.088	.088	.065	.069	.064
.496	.042	.037	.016	.020	.012	1.048	.093	.093	.069	.072	.067
.524	.042	.037	.016	.021	.014	1.076	.093	.093	.071	.076	.071

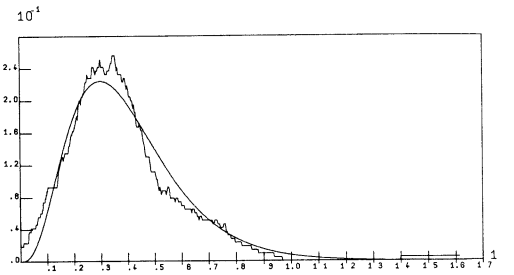
Tableau 2

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



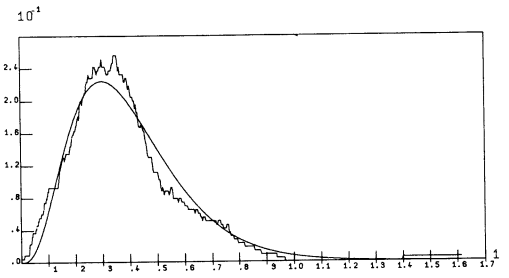
SIMULATIØN DE VARIABLE GAMMA (A=4, B=1, G=0) TEST K-S

Graphe 3



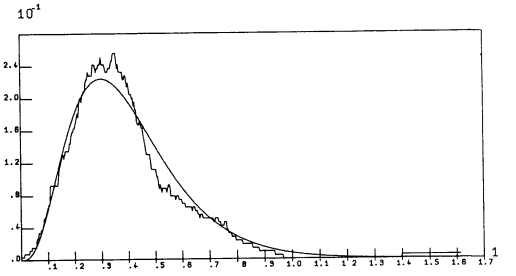
SIMULATIØN DE VARIABLE GAMMA (A=4, B=1, G=0) EST. NØ 1

Graphe 4



SIMULATIØN DE VARIABLE GAMMA (A=4, B=1, G=0) EST. NØ 2

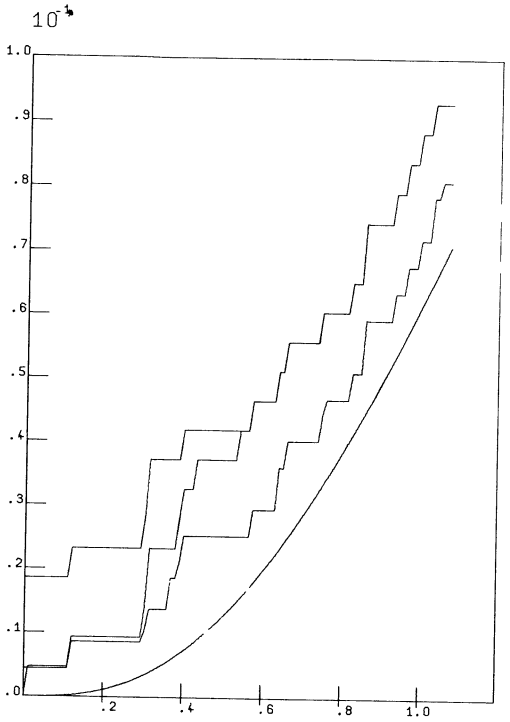
Graphe 5



SIMULATIØN DE VARIABLE GAMMA (A=4, B=1, G=0) EST. NØ 3

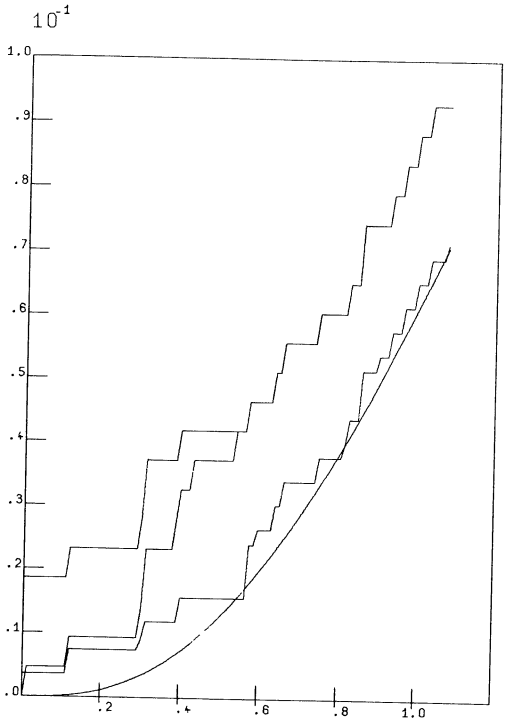
Graphe 6b

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



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

Graphe 7



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

Graphe 8

```
*****************************************************************************
*  HAUTEURS DE VAGUES DU 19/7 (21H) AU 28/7 (18H)   NOMBRE DE VALEURS :  72 *
*****************************************************************************

*****************************************************************************
*                          VALEURS DE L ECHANTILLON                         *
*---------------------------------------------------------------------------*
*1.11  .99 1.24 1.43 1.94 1.98 1.77 1.58 2.39 2.37 2.44 2.40 2.34 2.79 2.80*
*2.31 2.46 2.18 2.31 1.77 2.17 1.84 1.77 1.48 1.26 1.88 1.43 1.31 1.20 1.21*
*1.14  .91 1.08 1.27 1.22 1.03 1.26 1.30 1.08 1.31 1.26 1.45 1.66 2.04 2.24*
*2.32 2.31 2.52 2.12 2.41 2.32 2.13 1.99 1.93 2.15 2.11 2.05 1.68 1.86 1.54*
*1.36 1.31 1.47 1.18 1.12 1.03 1.18 1.20 1.08 1.48 1.32 1.37                *
*****************************************************************************

*****************************************************************************
*          HYPOTHESE NULLE : ECHANTILLON ISSU D UNE LOI GAMMA               *
*---------------------------------------------------------------------------*
*        PARAMETRES ESTIMES  :   A = 2.305  B = 2.711  G =   .857           *
*        PARAMETRES CHOISIS  :   A = 2.305  B = 2.711  G =   .857           *
*---------------------------------------------------------------------------*
*        TEST DE KOLMOGOROFF-SMIRNOFF : .109  (ACCEPTATION : OUI)           *
*        TEST D ADEQUATION DU CHI-2 :  10.08  (ACCEPTATION : NON)           *
*****************************************************************************

*****************************************************************************
*  ESTIMATION PAR LA FENETRE MOBILE SUR L ECHANTILLON MODIFIE PAR (.+G)*B   *
*   LONGUEUR DE LA FENETRE, DELTAN = 1.320  ( DELTAN > 26 OBSERVATIONS )    *
*****************************************************************************
* +------+----+----+ +------+----+----+ +------+----+----+ +------+----+----+ *
* |POINT|ESTM|DENS| |POINT|ESTM|DENS| |POINT|ESTM|DENS| |POINT|ESTM|DENS| *
* +------+----+----+ +------+----+----+ +------+----+----+ +------+----+----+ *
* |  .00|.074|.000| | 1.38|.284|.327| | 2.90|.168|.189| | 4.42|.137|.072| *
* |  .05|.084|.016| | 1.44|.274|.326| | 2.96|.189|.183| | 4.48|.137|.069| *
* |  .10|.095|.038| | 1.50|.274|.324| | 3.02|.189|.176| | 4.54|.137|.066| *
* |  .15|.105|.062| | 1.56|.263|.321| | 3.08|.189|.170| | 4.60|.116|.063| *
* |  .20|.105|.086| | 1.62|.242|.317| | 3.14|.168|.165| | 4.66|.105|.060| *
* |  .25|.126|.109| | 1.68|.232|.313| | 3.20|.168|.159| | 4.72|.095|.058| *
* |  .30|.158|.132| | 1.74|.221|.309| | 3.26|.168|.153| | 4.78|.084|.055| *
* |  .35|.168|.153| | 1.80|.179|.304| | 3.32|.221|.148| | 4.84|.063|.053| *
* |  .40|.179|.174| | 1.87|.200|.299| | 3.38|.211|.142| | 4.90|.053|.050| *
* |  .45|.211|.192| | 1.93|.158|.293| | 3.45|.211|.137| | 4.96|.042|.048| *
* |  .50|.221|.210| | 1.99|.158|.287| | 3.51|.221|.132| | 5.03|.032|.046| *
* |  .55|.232|.226| | 2.05|.158|.281| | 3.57|.242|.127| | 5.09|.032|.044| *
* |  .60|.274|.241| | 2.11|.158|.275| | 3.63|.221|.122| | 5.15|.032|.042| *
* |  .65|.274|.255| | 2.17|.168|.268| | 3.69|.242|.117| | 5.21|.021|.040| *
* |  .71|.284|.269| | 2.23|.147|.262| | 3.75|.221|.113| | 5.27|.021|.039| *
* |  .77|.295|.281| | 2.29|.158|.255| | 3.81|.221|.108| | 5.33|.021|.037| *
* |  .83|.284|.292| | 2.35|.126|.248| | 3.87|.221|.104| | 5.39|.021|.035| *
* |  .89|.284|.302| | 2.41|.147|.242| | 3.93|.211|.100| | 5.45|.021|.034| *
* |  .95|.316|.309| | 2.47|.147|.235| | 3.99|.211|.096| | 5.51|.021|.032| *
* | 1.01|.326|.316| | 2.53|.137|.228| | 4.05|.211|.092| | 5.57|.021|.031| *
* | 1.08|.337|.320| | 2.59|.158|.221| | 4.11|.179|.088| | 5.63|.021|.029| *
* | 1.14|.316|.324| | 2.66|.147|.215| | 4.17|.168|.085| | 5.69|.021|.028| *
* | 1.20|.326|.326| | 2.72|.147|.208| | 4.24|.158|.081| | 5.75|.021|.027| *
* | 1.26|.326|.328| | 2.78|.168|.202| | 4.30|.147|.078| | 5.82|.021|.025| *
* | 1.32|.305|.328| | 2.84|.168|.195| | 4.36|.147|.075| | 5.88|.021|.024| *
* +------+----+----+ +------+----+----+ +------+----+----+ +------+----+----+ *
*****************************************************************************
```

Tableau 1

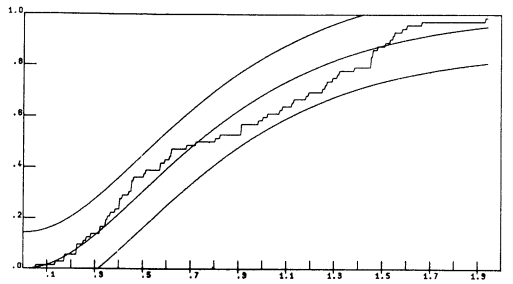
```
*****************************************************************************
*                HAUTEURS DE VAGUES DU 19/7 (21H) AU 28/7 (18H)             *
*****************************************************************************

*****************************************************************************
*              COMPARAISON DES ESTIMATEURS ENTRE G. ET DELTAN/2             *
*---------------------------------------------------------------------------*
* ESTIMATEUR 2 : R =  .97 ( VRAISEMBLANCE ) A OPT. = 1.00 ET B OPT. =  .56 *
*---------------------------------------------------------------------------*
* +------+-----------+-----+-----+-----+ +------+-----------+-----+-----+-----+ *
* |POINT| ESTIMATION |  DENSITE  | |POINT| ESTIMATION |  DENSITE  | *
* |     |  1 |  2 |  3 |LX**R|GAMMA| |     |  1 |  2 |  3 |LX**R|GAMMA| *
* |-----|----|----|----|-----|-----| |-----|----|----|----|-----|-----| *
* |  .00|.074|.000|.015|.000|.000| |  .38|.168|.152|.145|.149|.148| *
* |  .17|.074|.000|.015|.008|.004| |  .35|.168|.152|.150|.151|.155| *
* |  .34|.084|.011|.026|.016|.010| |  .32|.168|.168|.158|.158|.162| *
* |  .51|.084|.011|.026|.023|.017| |  .38|.179|.168|.168|.165|.169| *
* |  .68|.095|.053|.042|.031|.024| |  .40|.179|.168|.168|.172|.176| *
* |  .85|.095|.053|.042|.038|.031| |  .42|.179|.168|.168|.179|.182| *
* |  .02|.095|.053|.042|.045|.039| |  .44|.211|.200|.200|.186|.188| *
* |  .18|.105|.063|.053|.052|.047| |  .45|.211|.200|.200|.193|.195| *
* |  .35|.105|.063|.053|.060|.055| |  .47|.221|.211|.211|.200|.201| *
* |  .52|.105|.063|.053|.067|.063| |  .49|.221|.211|.211|.207|.207| *
* |  .69|.105|.063|.059|.074|.071| |  .50|.221|.211|.211|.214|.212| *
* |  .86|.105|.063|.065|.081|.079| |  .52|.221|.221|.211|.221|.218| *
* |  .03|.105|.084|.065|.088|.087| |  .54|.232|.232|.221|.228|.223| *
* |  .20|.126|.105|.086|.095|.095| |  .55|.232|.232|.221|.234|.228| *
* |  .37|.126|.105|.104|.102|.103| |  .57|.263|.263|.253|.241|.234| *
* |  .54|.126|.105|.104|.109|.111| |  .59|.263|.263|.253|.248|.238| *
* |  .71|.147|.126|.125|.116|.118| |  .60|.274|.274|.270|.255|.243| *
* |  .88|.147|.126|.125|.123|.126| |  .62|.274|.274|.270|.262|.248| *
* |  .05|.158|.147|.135|.131|.133| |  .64|.274|.274|.270|.269|.252| *
* |  .21|.158|.147|.135|.138|.141| |  .66|.274|.274|.270|.275|.257| *
* +------+-----+----+----+-----+-----+ +------+-----+----+----+-----+-----+ *
*****************************************************************************

*****************************************************************************
*              COMPARAISON DES ESTIMATEURS ENTRE G. ET DELTAN/2             *
*---------------------------------------------------------------------------*
* ESTIMATEUR 3 : R = 1.01 (MOINDRES CARRES) A OPT. = 1.00 ET B OPT. =  .56 *
*---------------------------------------------------------------------------*
* +------+-----------+-----+-----+-----+ +------+-----------+-----+-----+-----+ *
* |POINT| ESTIMATION |  DENSITE  | |POINT| ESTIMATION |  DENSITE  | *
* |     |  1 |  2 |  3 |LX**R|GAMMA| |     |  1 |  2 |  3 |LX**R|GAMMA| *
* |-----|----|----|----|-----|-----| |-----|----|----|----|-----|-----| *
* |  .00|.074|.000|.015|.000|.000| |  .38|.168|.152|.145|.139|.148| *
* |  .17|.074|.000|.015|.007|.004| |  .35|.168|.152|.145|.146|.155| *
* |  .34|.084|.011|.025|.013|.010| |  .32|.168|.158|.145|.153|.162| *
* |  .51|.084|.011|.025|.020|.017| |  .38|.179|.168|.167|.160|.169| *
* |  .68|.095|.053|.036|.027|.024| |  .40|.179|.168|.167|.167|.176| *
* |  .85|.095|.053|.042|.034|.031| |  .42|.179|.168|.167|.174|.182| *
* |  .02|.095|.053|.042|.041|.039| |  .44|.211|.200|.198|.181|.188| *
* |  .18|.105|.063|.046|.048|.047| |  .45|.211|.200|.198|.188|.195| *
* |  .35|.105|.063|.052|.055|.055| |  .47|.221|.211|.215|.195|.201| *
* |  .52|.105|.063|.052|.062|.063| |  .49|.221|.211|.215|.202|.207| *
* |  .69|.105|.063|.052|.069|.071| |  .50|.221|.211|.215|.209|.212| *
* |  .86|.105|.063|.058|.076|.079| |  .52|.221|.221|.215|.216|.218| *
* |  .03|.105|.084|.064|.083|.087| |  .54|.232|.232|.225|.223|.223| *
* |  .20|.126|.105|.085|.090|.095| |  .55|.232|.232|.225|.230|.228| *
* |  .37|.126|.105|.085|.097|.103| |  .57|.263|.263|.257|.238|.234| *
* |  .54|.126|.105|.085|.104|.111| |  .59|.263|.263|.257|.245|.238| *
* |  .71|.147|.126|.124|.111|.118| |  .60|.274|.274|.267|.252|.243| *
* |  .88|.147|.126|.124|.118|.126| |  .62|.274|.274|.267|.259|.248| *
* |  .05|.158|.147|.134|.125|.133| |  .64|.274|.274|.267|.266|.252| *
* |  .21|.158|.147|.134|.132|.141| |  .66|.274|.274|.267|.273|.257| *
* +------+-----+----+----+-----+-----+ +------+-----+----+----+-----+-----+ *
*****************************************************************************
```

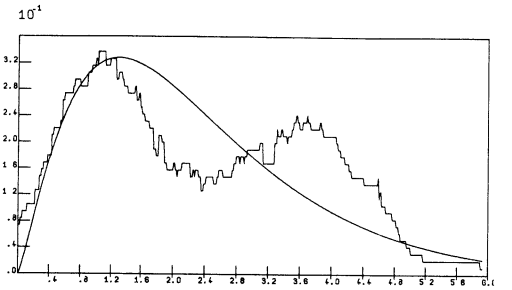
Tableau 2

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



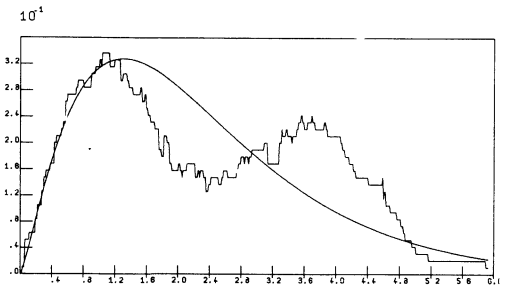
HAUTEURS DE VAGUES DU 19/7 (21H) AU 28/7 (18H) TEST K-S

Graphe 3



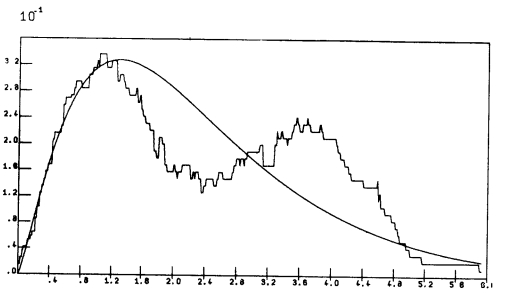
HAUTEURS DE VAGUES DU 19/7 (21H) AU 28/7 (18H) EST. NØ 1

Graphe 4



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

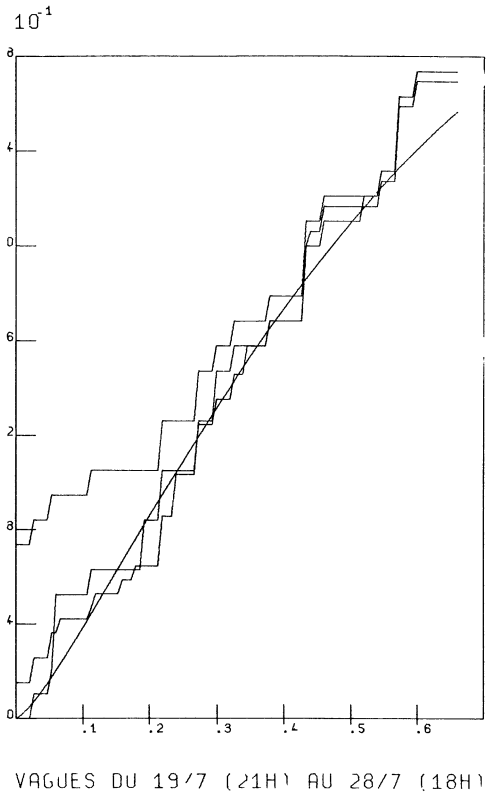
Graphe 5



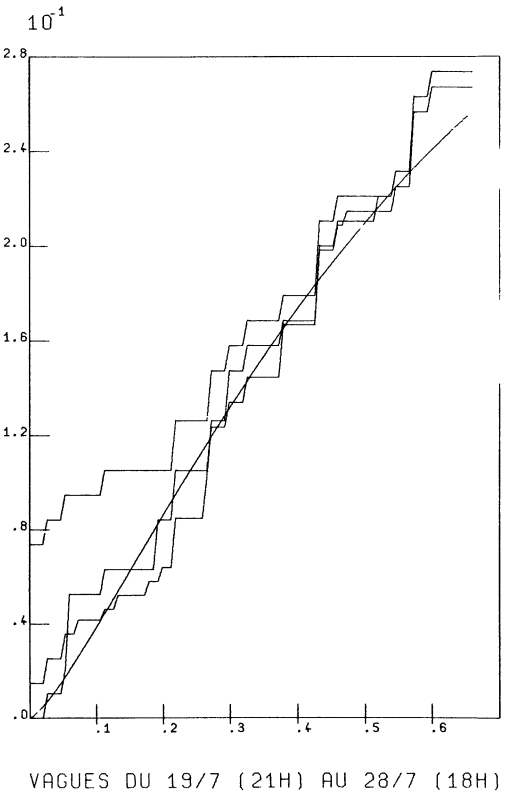
HAUTEURS DE VAGUES DU 19/7 (21H) AU 28/7 (18H) EST. NØ 3

Graphe 6a

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



Graphe 7



Graphe 8

HAUTEURS DE VAGUES DU 27/8 (18H) AU 8/9 (9H) NOMBRE DE VALEURS : 94

VALEURS DE L ECHANTILLON

1.59	1.42	1.29	1.43	1.30	1.50	1.38	1.52	2.00	1.56	1.52	1.41	1.21	1.32	1.78
1.87	1.70	1.68	1.54	1.43	1.89	1.40	1.62	2.01	2.09	2.19	2.14	2.53	2.39	2.46
2.43	2.09	1.77	2.09	2.18	1.84	1.79	1.96	2.72	2.73	3.19	2.85	3.00	3.61	4.21
5.05	5.49	4.48	4.34	4.29	3.97	4.23	3.62	3.16	2.58	2.90	2.34	2.26	2.05	1.86
1.59	1.59	1.48	1.10	1.05	1.19	.96	1.09	1.17	1.15	1.23	1.17	1.28	1.23	1.33
1.15	1.54	1.49	1.68	1.54	2.58	3.90	3.05	2.76	2.28	1.89	1.91	1.82	2.16	2.34
1.85	1.82	1.84	1.62											

HYPOTHESE NULLE : ECHANTILLON ISSU D UNE LOI GAMMA

PARAMETRES ESTIMES : A = 1.638 B = 1.381 G = .942
PARAMETRES CHOISIS : A = 1.638 B = 1.381 G = .942

TEST DE KOLMOGOROFF-SMIRNOFF : .058 (ACCEPTATION : OUI)
TEST D ADEQUATION DU CHI-2 : 3.18 (ACCEPTATION : OUI)

ESTIMATION PAR LA FENETRE MOBILE SUR L ECHANTILLON MODIFIE PAR (.-G)*B
LONGUEUR DE LA FENETRE, DELTAN = 1.042 (DELTAN > 39 OBSERVATIONS)

POINT	ESTM	DENS	POINT	ESTM	DENS	POINT	ESTM	DENS	POINT	ESTM	DENS
.00	.155	.000	1.38	.351	.344	3.19	.072	.096	5.00	.052	.021
.04	.175	.136	1.45	.340	.331	3.26	.062	.091	5.07	.031	.020
.08	.175	.204	1.52	.330	.317	3.34	.062	.085	5.15	.021	.018
.12	.196	.254	1.60	.320	.304	3.41	.052	.081	5.22	.021	.017
.16	.237	.293	1.67	.309	.291	3.48	.041	.076	5.29	.021	.016
.20	.237	.325	1.74	.278	.278	3.55	.041	.072	5.36	.021	.015
.24	.258	.351	1.81	.247	.265	3.63	.041	.067	5.44	.010	.014
.28	.268	.372	1.89	.216	.253	3.70	.041	.064	5.51	.010	.013
.32	.320	.389	1.96	.227	.241	3.77	.041	.060	5.58	.010	.013
.36	.330	.403	2.03	.216	.230	3.84	.041	.056	5.65	.010	.012
.40	.361	.415	2.10	.206	.218	3.92	.041	.053	5.73	.010	.011
.44	.381	.424	2.18	.175	.208	3.99	.041	.050	5.80	.021	.010
.47	.381	.430	2.25	.155	.197	4.06	.062	.047	5.87	.021	.010
.51	.402	.435	2.32	.155	.187	4.13	.072	.044	5.94	.021	.009
.58	.402	.440	2.39	.155	.177	4.20	.082	.042	6.02	.021	.009
.65	.433	.441	2.47	.134	.168	4.28	.062	.039	6.09	.021	.008
.73	.454	.439	2.54	.124	.159	4.35	.062	.037	6.16	.021	.007
.80	.495	.434	2.61	.134	.151	4.42	.072	.035	6.23	.010	.007
.87	.454	.427	2.68	.124	.143	4.49	.072	.032	6.31	.010	.007
.94	.443	.417	2.76	.113	.135	4.57	.072	.030	6.38	.010	.006
1.02	.433	.407	2.83	.093	.128	4.64	.062	.029	6.45	.010	.006
1.09	.443	.396	2.90	.093	.121	4.71	.052	.027	6.52	.010	.005
1.16	.443	.383	2.97	.093	.114	4.78	.052	.025	6.60	.010	.005
1.23	.423	.371	3.05	.062	.108	4.86	.052	.024	6.67	.010	.005
1.31	.402	.358	3.12	.062	.102	4.93	.052	.022	6.74	.010	.004

Tableau 1

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

COMPARAISON DES ESTIMATEURS ENTRE C* ET DELTAN/2

ESTIMATEUR 3 : R = .41 (VRAISEMBLANCE) A OPT. = 1.20 ET B OPT. = -.38

POINT	ESTIMATION 1	ESTIMATION 2	ESTIMATION 3	DENSITE LX**R	DENSITE GAMMA	POINT	ESTIMATION 1	ESTIMATION 2	ESTIMATION 3	DENSITE LX**R	DENSITE GAMMA
.000	.155	.000	.019	.000	.000	.267	.268	.227	.319	.321	.367
.013	.165	.010	.109	.093	.070	.280	.289	.247	.340	.327	.374
.027	.175	.021	.145	.124	.107	.294	.289	.247	.340	.334	.380
.040	.175	.031	.145	.147	.137	.307	.320	.289	.377	.340	.385
.053	.175	.041	.160	.165	.163	.320	.320	.299	.373	.346	.391
.067	.175	.052	.164	.181	.185	.334	.330	.309	.381	.352	.396
.080	.175	.052	.176	.195	.205	.347	.330	.309	.381	.358	.400
.093	.186	.062	.200	.208	.223	.361	.330	.309	.381	.363	.405
.107	.186	.062	.208	.220	.240	.374	.361	.351	.411	.369	.409
.120	.196	.072	.220	.231	.255	.387	.361	.351	.411	.374	.412
.134	.206	.103	.236	.241	.269	.401	.361	.351	.403	.379	.416
.147	.216	.113	.253	.251	.283	.414	.361	.351	.399	.385	.419
.160	.237	.144	.278	.260	.295	.427	.381	.371	.420	.390	.422
.174	.237	.144	.281	.269	.306	.441	.381	.371	.412	.395	.425
.187	.237	.155	.281	.277	.317	.454	.381	.371	.412	.399	.427
.200	.237	.155	.281	.285	.326	.467	.381	.371	.412	.404	.429
.214	.237	.175	.281	.293	.336	.481	.381	.371	.405	.409	.431
.227	.247	.186	.294	.300	.344	.494	.381	.371	.401	.414	.433
.240	.258	.216	.306	.307	.352	.507	.402	.402	.422	.418	.435
.254	.268	.227	.319	.314	.360	.521	.402	.402	.418	.423	.436

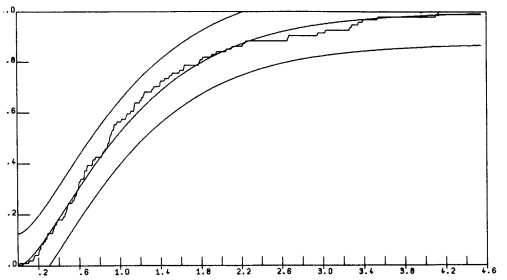
COMPARAISON DES ESTIMATEURS ENTRE C* ET DELTAN/2

ESTIMATEUR 3 : R = .42 (MOINDRES CARRES) A OPT. = 1.20 ET B OPT. = -.39

POINT	ESTIMATION 1	ESTIMATION 2	ESTIMATION 3	DENSITE LX**R	DENSITE GAMMA	POINT	ESTIMATION 1	ESTIMATION 2	ESTIMATION 3	DENSITE LX**R	DENSITE GAMMA
.000	.155	.000	.025	.000	.000	.267	.268	.227	.317	.319	.367
.013	.165	.010	.105	.090	.070	.280	.289	.247	.338	.326	.374
.027	.175	.021	.142	.121	.107	.294	.289	.247	.338	.332	.380
.040	.175	.031	.142	.143	.137	.307	.320	.289	.375	.338	.385
.053	.175	.041	.158	.162	.163	.320	.320	.299	.375	.344	.391
.067	.175	.052	.162	.178	.185	.334	.330	.309	.383	.350	.396
.080	.175	.052	.174	.192	.205	.347	.330	.309	.379	.356	.400
.093	.186	.062	.190	.205	.223	.361	.330	.309	.379	.362	.405
.107	.186	.062	.206	.217	.240	.374	.361	.351	.416	.368	.409
.120	.196	.072	.219	.228	.255	.387	.361	.351	.408	.373	.412
.134	.206	.103	.235	.238	.269	.401	.361	.351	.400	.379	.416
.147	.216	.113	.251	.248	.283	.414	.361	.351	.400	.384	.419
.160	.237	.144	.276	.257	.295	.427	.381	.371	.421	.389	.422
.174	.237	.144	.280	.266	.306	.441	.381	.371	.417	.394	.425
.187	.237	.155	.280	.274	.317	.454	.381	.371	.409	.399	.427
.200	.237	.155	.280	.283	.326	.467	.381	.371	.409	.404	.429
.214	.237	.175	.280	.290	.336	.481	.381	.371	.409	.409	.431
.227	.247	.186	.292	.298	.344	.494	.381	.371	.401	.414	.433
.240	.258	.216	.305	.305	.352	.507	.402	.402	.421	.418	.435
.254	.268	.227	.317	.312	.360	.521	.402	.402	.413	.423	.436

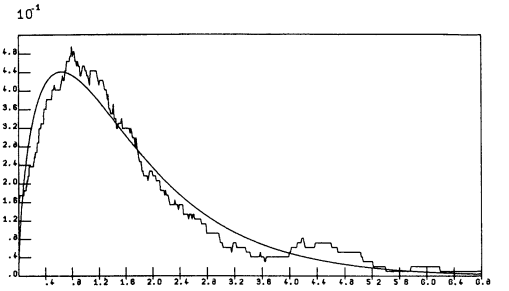
Tableau 2

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



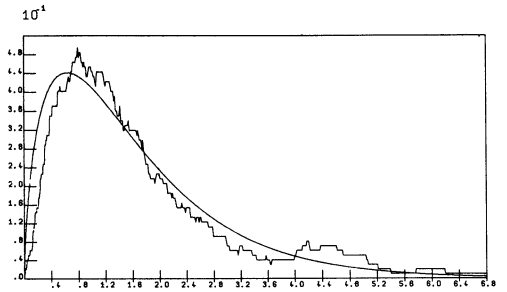
HAUTEURS DE VAGUES DU 27/8 (18H) AU 8/9 (9H) TEST K-S

Graphe 3



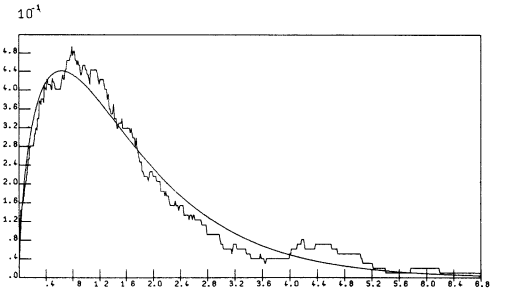
HAUTEURS DE VAGUES DU 27/8 (18H) AU 8/9 (9H) EST. NØ 1

Graphe 4



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

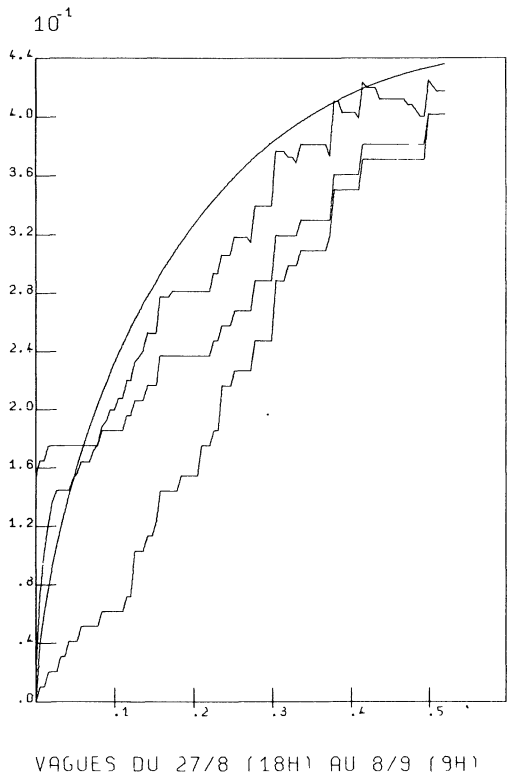
Graphe 5



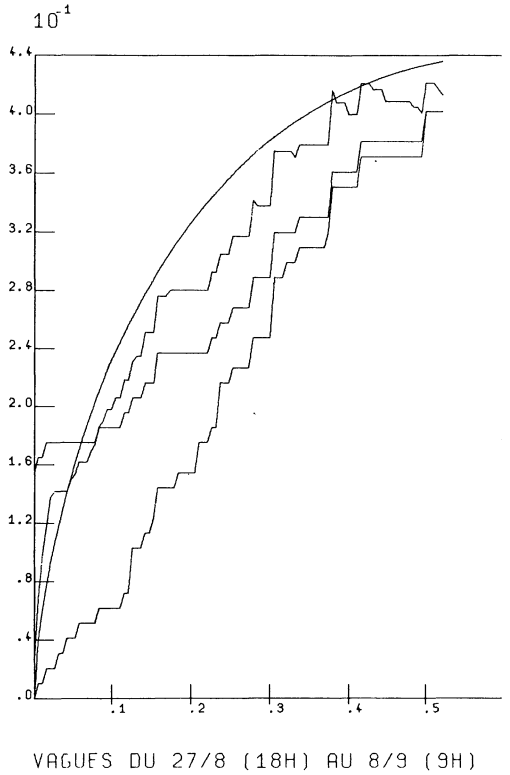
HAUTEURS DE VAGUES DU 27/8 (18H) AU 8/9 (9H) EST. NØ 3

Graphe 6a

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



Graphe 7



Graphe 8