

REVUE DE STATISTIQUE APPLIQUÉE

PAUL DEHEUVELS

PIERRE HOMINAL

Estimation automatique de la densité

Revue de statistique appliquée, tome 28, n° 1 (1980), p. 25-55

[<http://www.numdam.org/item?id=RSA_1980__28_1_25_0>](http://www.numdam.org/item?id=RSA_1980__28_1_25_0)

© Société française de statistique, 1980, 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 AUTOMATIQUE DE LA DENSITE

Paul DEHEUVELS & Pierre HOMINAL

Ecole Pratique des Hautes Etudes & Université Paris VI

RESUME. — L'estimation de la densité à partir d'un échantillon $X_1, \dots, X_n$ nécessite le choix d'un ou plusieurs paramètres, extérieurement aux données. Dans le cas de l'estimateur du noyau de Parzen-Rosenblatt, ou de l'histogramme, ce choix se limite essentiellement à la détermination de la fenêtre δ_n . Nous développons ici plusieurs méthodologies permettant d'obtenir un choix de ce nombre en fonction des valeurs observées. Nous montrons la consistance des estimations obtenues, et introduisons un procédé permettant de déterminer la fenêtre entièrement à partir des données. Ce procédé fournit des estimations asymptotiquement optimales pour une famille de densités suffisamment régulières. Un travail de simulation sur des échantillons de taille $n = 10, 20, 50, 100$, ainsi que sur des échantillons expérimentaux montre que le procédé est efficace même pour des échantillons de taille aussi petite que $n = 10$. Deux programmes Fortran sont donnés permettant de réaliser concrètement la méthode. Des exemples sont exposés.

ENGLISH SUMMARY. — This paper deals with automatic nonparametric density estimation, that is to give procedures which enable to construct efficient density estimates for some general family of densities, without any exterior choice of parameters. Following results obtained in [8], we first develop general consistency results for simple automatic estimates, and give a general theorem on strong uniform consistency of these estimates. We then introduce a new automatic procedure for which we give Fortran programs, and show its efficiency for samples with size as low as $n = 10$.

Plan : I — Estimation automatique de la densité — Théorie — par Paul Deheuvels.

- 1) Généralités
- 2) Critères d'efficacité et premier choix de la fenêtre
- 3) Convergence uniforme presque sûre des estimations
- 4) Deuxième choix de la fenêtre dans le cas d'une estimation paramétrique
- 5) Troisième choix de la fenêtre
- 6) Méthode proposée pour le choix de la fenêtre

II — Estimation automatique de la densité — Résultats — par Pierre Hominal

- 1) Généralités
- 2) Résultats de simulation
- 3) Conclusion de la simulation
- 4) Programmes
- 5) Exemples de résultats

III — Bibliographie

Classification : Principale 62G05, Secondaires 62G20, 62G25, 62F11, 62XX04, 65CXX.

Mots Clés : Estimation, Méthodes non paramétriques, Estimation de la densité, Méthode du noyau.

Key Words : Estimation, nonparametric methods, density estimation, kernel methods.

I – ESTIMATION AUTOMATIQUE DE LA DENSITE – THEORIE

par Paul DEHEUELS

1. Généralités

On considère un échantillon $X_1, \dots, X_n$ de v.a. réelles indépendantes de même loi. Le problème statistique fondamental à résoudre est celui de l'estimation de la distribution de probabilité qui leur est associée, en fonction des valeurs numériques observées.

Dans le cas où on ne fait, extérieurement à l'échantillon, aucune hypothèse restreignant le choix de cette distribution à une famille particulière, l'estimation la plus naturelle de cette loi commune est donnée par la *loi de probabilité empirique* de l'échantillon, associant à chaque valeur observée parmi les $X_1, \dots, X_n$, une probabilité $1/n$. La fonction de répartition associée, appelée *fonction de répartition empirique* de l'échantillon est alors :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}},$$

où $\mathbf{1}_A$ désigne la fonction indicatrice de A égale à 1 sur A et 0 ailleurs.

L'usage de F_n est justifié par divers résultats, comme le théorème de Glivenko-Cantelli, prouvant que F_n converge vers la fonction de répartition F de la loi cherchée en tout point de continuité avec probabilité 1, lorsque $n \rightarrow \infty$. Par ailleurs, si on suppose que F est continue, des statistiques telles celles de Kolmogorov-Smirnov, $D_n = \sup_x |F_n(x) - F(x)|$, ont une loi indépendante de F , et telles que (voir par exemple [16], p. 378-463).

$$\lim_{n \rightarrow \infty} P(\sqrt{n} D_n > u) = 2 \sum_{k=1}^{+\infty} (-1)^{k+1} e^{-2k^2 u^2}$$

$$\lim_{n \rightarrow \infty} \sup \sqrt{\frac{2n}{\log \log n}} D_n = 1 \quad \text{p.s. (Chung, [4], Smirnov [17])}$$

Malheureusement, F_n n'est pas une statistique très commode pour décrire visuellement les caractères de l'échantillon, et même un œil exercé aura du mal à distinguer des distributions aussi différentes qu'une loi Gamma simple, et un mélange fortement bimodal de deux lois Gamma, au vu de leurs fonctions de répartition respectives (voir par exemple [13], p. 61-67, graphe 3).

Par contre, lorsque la loi admet une densité f , celle-ci donne un élément descriptif très utile et compréhensible visuellement. Le problème se pose alors d'estimer cette densité.

La principale difficulté que l'on rencontre alors est due à la nature de la liaison de f et F , à savoir que, si f est continue, $f = dF/dx$. Comme F_n est discontinue et de limite F , la façon la plus naturelle d'estimer f consiste à définir :

$$f_n(x) = \frac{1}{\delta_n} \left(F_n\left(x + \frac{1}{2} \delta_n\right) - F_n\left(x - \frac{1}{2} \delta_n\right) \right),$$

où $\{\delta_n, n \geq 1\}$ est une suite positive décroissant vers 0, fonction qui, si δ_n décroît vers 0 assez lentement, convergera vers f (voir [5]) avec probabilité 1.

On peut généraliser ce type d'estimation qui présente l'inconvénient de ne pas donner une densité continue, par :

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

où $K(t)$ est une densité sur $\mathbf{R}$, ce qui donne le cas précédent si $K = \mathbf{1}_{\left[-\frac{1}{2}, +\frac{1}{2}\right]}$.

On obtient ainsi l'estimateur à noyau de Parzen-Rosenblatt. Cet estimateur est de loin le plus utilisé, et est souvent appelé *fonction de densité empirique* de l'échantillon (voir [20], [12] pour des bibliographies et exposés de synthèse).

Cet estimateur pose alors de nouveaux problèmes, dûs au fait qu'il n'est pas défini de manière unique. En effet, il dépend d'une part de K (le noyau), et d'autre part de δ_n (la fenêtre). Il est cependant possible de constater ([7], [13], [18]) d'une part que le choix de K convenablement normalisé influe peu sur l'efficacité de l'estimation, d'autre part qu'en pratique on se limite le plus souvent à des noyaux comme les suivants :

- Noyau unité : $K = \mathbf{1}_{\left[-\frac{1}{2}, \frac{1}{2}\right]}$,
- Noyau triangulaire : $K(x) = 1 - |x|$, si $|x| \leq 1$,
- Noyau parabolique (Epanechnikov) : $K(x) = \frac{3}{2} (1 - x^2)$, si $|x| \leq 1$,
- Noyau normal : $K(x) = (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} x^2\right)$.

Les raisons de ce choix sont dues surtout à la facilité de construire numériquement les estimations (y compris pour le noyau normal, par transformation de Fourier rapide FFT, selon une méthode mise au point par Silverman, [19]).

Il reste alors à résoudre le problème principal, c'est-à-dire celui du choix de la fenêtre δ_n .

Ce choix, pour être acceptable, doit posséder les caractéristiques suivantes :

- (i) δ_n est une fonction de $X_1, \dots, X_n$ et $n : \delta_n = h(n; X_1, \dots, X_n)$;
- (ii) $\delta_n = h(n; X_1, \dots, X_n)$ est tel que :

$$\forall \lambda, \mu \in \mathbf{R}, \lambda \neq 0, h(n; \lambda X_1 + \mu, \dots, \lambda X_n + \mu) = |\lambda| h(n; X_1, \dots, X_n) ; \quad (1.3)$$

Cette dernière condition est essentielle, puisqu'elle exprime qu'un changement affine d'échelle ne modifie pas, compte tenu de la nouvelle unité, le choix de la fenêtre.

Nous développons dans le paragraphe suivant de tels choix, basés sur une méthode proposée dans [7].

2. Critères d'efficacité et premier choix de la fenêtre

Le critère d'efficacité que nous retiendrons ici est voisin du critère du M.I.S.E. (mean integrated square error), donné par :

$$M_n^2(\delta_n) = E \left\{ \int_{-\infty}^{+\infty} (f_n(x) - f(x))^2 dx \right\},$$

où f_n est donné par (1.1).

Si on sépare dans cette expression les termes faisant apparaître le biais et les termes faisant apparaître la variance, on obtient (voir [7]) que :

$$\begin{aligned} M_n^2(\delta_n) &= B_1^2(\delta_n) + B_2^2(\delta_n), \text{ où :} \\ B_1^2(\delta_n) &\sim B_{1*}^2(\delta_n) = \frac{1}{4} \delta_n^4 [[y^2 K]]^2 \int_{-\infty}^{+\infty} f''(x)^2 dx, \\ B_2^2(\delta_n) &\sim B_{2*}^2(\delta_n) = \frac{[[K^2]]}{n \delta_n}, \quad \text{si } \delta_n \rightarrow 0, \quad n \delta_n \rightarrow \infty, \end{aligned} \quad (2.1)$$

$$\text{avec} \quad [[y^2 K]] = \int_{-\infty}^{+\infty} y^2 K(y) dy, \quad [[K^2]] = \int_{-\infty}^{+\infty} K^2(y) dy.$$

Les conditions de validité de (2.1) ont été détaillées dans [7] (et [8] pour le cas multivarié). Retenons que ces formules sont vraies pour les noyaux (1.2), lorsque f admet sur $\mathbf{R}$ une dérivée seconde f'' continue bornée et de carré intégrable sur $\mathbf{R}$.

Le calcul exact de ces expressions pour les distributions usuelles (voir [7]) montre que, même pour les petites valeurs de n , $M_n^2(\delta_n)$ est très voisin de $M_{n*}^2(\delta_n) = B_{1*}^2(\delta_n) + B_{2*}^2(\delta_n)$. Pour ces raisons, nous introduisons le critère d'efficacité suivant (A.M.I.S.E. pour asymptotic mean integrated square error) :

$$M_{n*}^2(\delta_n) = \frac{1}{4} \delta_n^4 [[y^2 K]]^2 \theta + \frac{[[K^2]]}{n \delta_n}, \quad \text{où } \theta = \int_{-\infty}^{+\infty} f''(x)^2 dx. \quad (2.2)$$

Le choix de δ_n minimisant le critère de l'A.M.I.S.E. est alors :

$$\tilde{\delta}_n = \left(\frac{[[K^2]]}{[[y^2 K]]^2} \right)^{1/5} \frac{1}{\theta^{1/5} n^{1/5}}. \quad (2.3)$$

On sait alors (voir [7], (2.56)) que si $\theta = \int_{-\infty}^{+\infty} f''(x)^2 dx$,

$$M_{n*}^2(\tilde{\delta}_n) = \frac{5}{4} [[K^2]]^{5/4} [[y^2 K]]^{2/5} \theta^{1/5} n^{-4/5} \sim \inf_{\delta} M_n^2(\delta),$$

lorsque $n \rightarrow \infty$.

La conséquence de ces résultats est qu'on est assuré d'obtenir asymptotiquement des résultats de convergence raisonnable si on se contente de choisir une constante C indépendamment de l'échantillon, et si on pose :

$$\bar{\delta}_n(C) = \left(\frac{[[K^2]]}{[[y^2 K]]^2} \right)^{1/5} \frac{s}{C^{1/5} n^{1/5}},$$

où $s = s(X_1, \dots, X_n)$ désigne une statistique de l'échantillon estimant un paramètre de dispersion de la loi des X_i . Nous avons introduit ([7]) en prenant pour C la valeur $C = \int_{-\infty}^{+\infty} \varphi''(x)^2 dx = \frac{3}{8\sqrt{\pi}}$, où φ est la densité d'une loi normale $N(0,1)$, le choix suivant de δ_n :

$$\bar{\delta}_n = \left(\frac{[[K^2]]}{[[y^2 K]]^2} \right)^{1/5} \frac{s}{(3/(8\sqrt{\pi}))^{1/5} n^{1/5}}, \quad (2.4)$$

où $s = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

Ce choix donne, en particulier les valeurs suivantes pour δ_n :

- Noyau unité : $\bar{\delta}_n(C) = \left[\frac{144}{nC} \right]^{1/5} s \# \frac{(3.686)s}{n^{1/5}}$
- Noyau triangulaire : $\bar{\delta}_n(C) = \left[\frac{24}{nC} \right]^{1/5} s \# \frac{(2.576)s}{n^{1/5}}$
- Noyau parabolique : $\bar{\delta}_n(C) = \left[\frac{15}{nC} \right]^{1/5} s \# \frac{(2.345)s}{n^{1/5}}$
- Noyau normal : $\bar{\delta}_n(C) = \left[\frac{1}{2nC\sqrt{\pi}} \right]^{1/5} s \# \frac{(1.059)s}{n^{1/5}}$.

La justification de ces choix peut être déduite, dans le cas où X possède une variance σ^2 , de la convergence de s^2 vers σ^2 , et du fait que la valeur de $\bar{\delta}_n(C)$ est alors équivalente pour $n \rightarrow \infty$ à la valeur optimale $\tilde{\delta}_n$ si $C = \int_{-\infty}^{+\infty} f''(x)^2 dx$.

On remarquera que les valeurs données par (2.4) et (2.5) donnent le choix optimal (A.M.I.S.E.) de δ_n pour une loi normale $N(m, \sigma^2)$, lorsque $s = \sigma$.

On peut objecter que les choix (2.4) et (2.5) privilégient la loi normale ; il faut cependant constater d'une part que c'est le cas le plus courant, d'autre part que le coefficient C étant élevé à la puissance $1/5$, des choix de C très différents de $3/(8\sqrt{\pi})$ changent peu les valeurs obtenues.

Dans le tableau ci-dessous, nous donnons les valeurs qui pourraient être obtenues pour $\bar{\delta}_n(C)$, en remplaçant C par $\int_{-\infty}^{+\infty} f_0''(x)^2 dx$, où f_0 désigne la densité de la distribution réduite à une variance unité (cas du noyau unité) :

Loi exponentielle $C = \int_0^{+\infty} f''(x)^2 dx$	$\frac{(3.104)s}{n^{1/5}}$	Loi Beta $\sqrt{28} \beta(3, 3)$ $(C = \int_0^{\sqrt{28}} f''(x)^2 dx)$	$\frac{(3.835)s}{n^{1/5}}$
Loi gamma $\Gamma(r, r^{1/2})$ $r = 3$	$\frac{(2.180)s}{n^{1/5}}$	Loi normale $N(0, 1)$	$\frac{(3.686)s}{n^{1/5}}$

Dans [7], nous avons développé des estimations voisines de (2.4), obtenues en remplaçant s par ks où k est un paramètre de robustesse. Nous avons en particulier recommandé un choix robuste de δ_n , fondé sur une argumentation bayésienne.

Ce choix (voir [7], (2.71)) recommande une modification de (2.4) par :

$$\bar{\delta}_n = \left(\frac{[K^2]}{[y^2 K]} \right)^{1/5} \frac{ks}{(3/(8\sqrt{\pi}))^{1/5} n^{1/5}}, \quad \text{où } k = \left[\frac{10 \log R}{R^{10} - 1} \right]^{1/5}$$

$R > 1$ représente ici un paramètre d'incertitude, supposé tel que la valeur exacte $\theta = \int_{-\infty}^{+\infty} f_0''(x)^2 dx$ (où f_0 est la distribution réduite à une variance unité) est a priori uniformément distribuée sur $(3/(8\sqrt{\pi})) [1/R, R]$.

Nous avons en particulier justifié un choix robuste de δ_n , lorsque les hypothèses précédentes sont vérifiées, fourni par (0.5) $\bar{\delta}_n$ ([7], (2.71)).

Il reste à montrer que si $f_{n,\delta_n}(x) = \frac{1}{n\delta_n} \sum_{i=1}^n K\left(\frac{X_i - x}{\delta_n}\right)$, $f_{n,\bar{\delta}_n}$ converge vers f lorsque $n \rightarrow \infty$. Ceci sera développé dans [10], et dans le paragraphe suivant.

3. Convergence uniforme presque sûre des estimations

On suppose ici que f est uniformément continue sur $\mathbf{R}$.

On connaît (voir [5]) des conditions nécessaires et suffisantes de convergence de f_n lorsque δ_n est non aléatoire. Le résultat suivant les généralise :

Théorème 1. — Si $\{u_n, n \geq 1\}$ et $\{v_n, n \geq 1\}$ sont deux suites telles que $0 \leq u_n \leq v_n$, $\lim_{n \rightarrow \infty} \frac{nu_n}{\log n} = +\infty$, $\lim_{n \rightarrow \infty} v_n = 0$, alors si

$$f_{n,\delta_n}(x) = \frac{1}{n\delta_n} \sum_{i=1}^n K\left(\frac{X_i - x}{\delta_n}\right),$$

$$\lim_{n \rightarrow \infty} \left\{ \sup_x \left(\sup_{u_n \leq \delta_n \leq v_n} |f_{n,\delta_n}(x) - f(x)| \right) \right\} = 0 \quad \text{p.s.} \quad (3.1)$$

Preuve : Nous donnons ici une démonstration abrégée de ce résultat, lorsque K est un noyau unité. Tout d'abord, si $\delta_n = u_n$, $f_n(x) = f_{n,u_n}(x)$ convergera uniformément p.s. vers f (voir [5], [11]). Maintenant, on peut exprimer $f_{n,2u_n}(x)$ en fonction de

$$f_{n,u_n} \quad \text{par} \quad f_{n,2u_n}(x) = \frac{1}{2} \left(f_{n,u_n}\left(x - \frac{1}{2}\delta_n\right) + f_{n,u_n}\left(x + \frac{1}{2}\delta_n\right) \right),$$

et de même

$$f_{n,ku_n}(x) = \frac{1}{k} \sum_{i=0}^{k-1} f_{n,u_n}\left(x - \frac{1}{2}(k-1)u_n + iu_n\right).$$

On en déduit alors que :

$$\sup_x |f_{n,ku_n}(x) - f(x)| \leq \sup_x |f_{n,u_n}(x) - f(x)| + \sup_x \left(\sup_{|h| \leq \frac{1}{2}ku_n} |f(x+h) - f(x)| \right)$$

De ceci, on déduit que

$$\lim_{n \rightarrow \infty} \left\{ \sup_x \left(\sup_{1 \leq k \leq v_n/u_n} |f_{n,ku_n}(x) - f(x)| \right) \right\} = 0 \text{ p.s. ;}$$

Maintenant, on aurait pu aussi bien faire ce raisonnement avec $\delta_n = u_n/m$, ce qui prouve que

$$\lim_{n \rightarrow \infty} \left\{ \sup_x \left(\sup_{1 \leq k \leq m v_n/u_n} |f_{n,ku_n/m}(x) - f(x)| \right) \right\} = 0 \text{ p.s. ;}$$

Or (voir [6]) ceci implique que

$$\lim \left\{ \sup_{u_n \leq \delta_n \leq v_n} |f_{n,\delta_n}(x) - f(x)| \right\} = 0 \text{ p.s.}$$

ce qui prouve le résultat. Il est possible d'en déduire que (3.1) est vrai pour un noyau K intégrable au sens de Riemann quelconque (voir [6]). —

La démonstration précédente permet d'obtenir, en utilisant les résultats de [7] :

Théorème 2 : si $\{u_n, n \geq 1\}$ et $\{v_n, n \geq 1\}$ sont deux suites telles que

$$0 < u_n \leq v_n, \lim_{n \rightarrow \infty} \frac{n u_n}{\text{Log Log } n} = +\infty, \lim_{n \rightarrow \infty} v_n = 0,$$

alors, pour tout x ,

$$\lim_{n \rightarrow \infty} \left\{ \sup_{u_n \leq \delta_n \leq v_n} |f_{n,\delta_n}(x) - f(x)| \right\} = 0 \text{ p.s. —} \quad (3.2)$$

Corollaire : Si $\delta_n = h_n(n; X_1, \dots, X_n)$ dépend de l'échantillon, pour que $f_{n,\delta_n}(x)$ converge uniformément p.s. vers $f(x)$, il suffit que :

$$\lim_{n \rightarrow \infty} \frac{n \delta_n}{\text{Log } n} = +\infty \text{ p.s., et } \lim_{n \rightarrow \infty} \delta_n = 0 \text{ p.s. —} \quad (3.3)$$

Preuve : C'est une conséquence directe de (3.1) ; ce résultat généralise les théorèmes classiques (voir [7]) au cas où δ_n est aléatoire, même lorsque δ_n dépend de l'échantillon. —

Corollaire : Si X a une variance finie, $\lim_{n \rightarrow \infty} \sup_x |f_{n,\bar{\delta}_n}(x) - f(x)| = 0 \text{ p.s. —}$

Preuve : Si $\delta_n = \bar{\delta}_n$ est donné par (2.4), comme $\lim s^2 = \sigma^2 \text{ p.s.,}$ on en déduit que $\delta_n \sim A/n^{1/5} \text{ p.s.,}$ d'où le résultat par (3.1).

Nous pouvons maintenant donner trois nouveaux critères que devrait vérifier $\delta_n = h(n; X_1, \dots, X_n)$:

$$f_{n,\delta_n}(x) = \frac{1}{n \delta_n} \sum_{i=1}^n K\left(\frac{X_i - x}{\delta_n}\right) \quad \text{lorsque} \quad \delta_n = h(n; X_1, \dots, X_n), \quad (\text{iii})$$

converge uniformément vers $f(x)$ p.s. (resp. (iii') en probabilité).

$$\delta_n = h(n; X_1, \dots, X_n) \quad \text{est tel que, lorsque } n \rightarrow \infty, \quad (\text{iv})$$

$$\delta_n \sim \delta_n = \left(\frac{[[K^2]]}{[[y^2 K]]^2} \right)^{1/5} \frac{1}{\left(\int_{-\infty}^{+\infty} f''(\bar{x})^2 dx \right)^{1/5} n^{1/5}} \quad \text{p.s.} \quad (3.4)$$

(resp. (iv') en probabilité).

On remarquera que, par le corollaire, (iv) $\Rightarrow$ (iii). En général, sous-réserve que X admette des moment d'ordre 2 le choix de $\delta_n = \bar{\delta}_n$ vérifie la condition plus faible :

$$\delta_n = h(n; X_1, \dots, X_n) \quad \text{est tel que, lorsque } n \rightarrow \infty, \quad (\text{v})$$

$$\delta_n \sim C/n^{1/5}, \quad \text{où } 0 < C < +\infty \quad \text{p.s. (resp. (v') en probabilité).}$$

Ici (iv) $\Rightarrow$ (v) $\Rightarrow$ (iii) .

La recherche d'un choix de $\delta_n = h(n; x_1, \dots, X_n)$ vérifiant les conditions (i-ii-iii-iv-v) mène directement au problème de l'estimation du paramètre :

$$\theta = \int_{-\infty}^{+\infty} f''(x)^2 dx .$$

Nous développons ce point dans les paragraphes suivants.

4. Deuxième choix de la fenêtre dans le cas d'une estimation paramétrique

D'un point de vue pratique, *il faut recommander de procéder à une estimation non paramétrique de la densité simultanément à toute estimation paramétrique.*

Ceci signifie que, si on se dispose à représenter la distribution des v.a. d'un échantillon par un modèle paramétrique, soit une densité $f(x; \alpha)$, dépendant d'un paramètre $\alpha \in \mathbf{R}^p$ (*). Dans ce cas, si on dispose d'une estimation α_n de α , il est utile de comparer l'estimation $f(x; \alpha_n)$ de $f(x)$ à une estimation non paramétrique $f_{n,\delta_n}(x)$ de la forme (1.1). Pour construire une telle estimation, on pourra estimer $\theta = \int_{-\infty}^{+\infty} f''(\bar{x})^2 dx$ par $\theta_n = \int_{-\infty}^{+\infty} f''(\bar{x}; \alpha_n)^2 dx$, ce qui donne :

(1) On pourrait objecter que les fonctions de la forme (1.1) peuvent être représentées par $f(x; \omega)$ où ω appartient à un espace assez grand. Il s'agit là cependant d'une estimation non paramétrique. La distinction entre non paramétrique et paramétrique provient de la dimension de l'espace où ω prend ses valeurs ; si cette dimension est finie, il s'agit d'un problème paramétrique ; si cette dimension est infinie, le problème est non paramétrique. On ne peut ramener un cas à l'autre, car il n'existe pas de bijection bicontinue entre deux ouverts d'espaces $\mathbf{R}^p$ et $\mathbf{R}^q$ lorsque $p \neq q$ (bien que leurs cardinaux soient égaux).

$$\hat{\delta}_n = \left(\frac{[[K^2]]}{[[y^2 K]]^2} \right)^{1/5} \frac{1}{\left(\int_{-\infty}^{+\infty} f''(x; \alpha_n)^2 dx \right)^{1/5} n^{1/5}} \quad (4.1)$$

On obtient alors le théorème suivant :

Théorème 3. — Si $\alpha_0 \in \mathbb{R}^p$ et si, dans un voisinage V_0 de α_0 , $f(x; \alpha)$ admet une dérivée seconde $f''(x; \alpha)$ sur $\mathbb{R}$, telle que $\sup_{\alpha \in V_0} f''(x; \alpha)^2 \leq \psi_0(x)$, avec

$\int_{-\infty}^{+\infty} \psi_0(x) dx < \infty$, si enfin $\hat{\alpha}_n \rightarrow \alpha_0$ p.s. lorsque $n \rightarrow \infty$, alors $\hat{\theta}_n \rightarrow \theta$ p.s.,

et $f_{n, \hat{\delta}_n}(x) = \frac{1}{n \hat{\delta}_n} \sum_{i=1}^n K\left(\frac{X_i - x}{\hat{\delta}_n}\right)$ vérifie les conditions (i-iii-v) de (1.3) et (3.4). De plus, si $f(x) = f(x; \alpha_0)$, alors les conditions (i-iii-iv-v) sont vérifiées.

Preuve. — Il s'agit d'une application du théorème de Lebesgue et du théorème 1 (§ 3). (On suppose que $f''(x; \alpha)$ est continue en α).

On remarquera que la condition (ii) sera vérifiée ou non suivant la forme de l'estimation $\hat{\alpha}_n$ de α . Dans le cas de la loi normale $N(m, \sigma^2)$, on constatera que la valeur de $\hat{\delta}_n$ donnée par (4.1) coïncide avec celle de $\bar{\delta}_n$, lorsque on estime

$$\sigma^2 \text{ par } s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

La comparaison des estimations $f(x; \hat{\alpha}_n)$ et $f_{n, \hat{\delta}_n}(x)$ permet de rendre évidentes des erreurs de modélisation dues au choix de la famille paramétrée $f(x; \alpha)$.

Nous avons donné dans [13] des exemples manifestes de modélisations erronées qui n'étaient pas décelables à l'aide des tests usuels. Le tracé simultané des courbes $f(x; \hat{\alpha}_n)$ et $f_{n, \hat{\delta}_n}(x)$ permet de se rendre compte de l'erreur. Il est, par ailleurs possible de comparer $f_{n, \hat{\delta}_n}(x)$ à une estimation semblable construite à partir d'un échantillon simulé à partir de $f(x; \hat{\alpha}_n)$.

Lorsque f n'appartient pas à la famille paramétrée, sous réserve que l'estimateur $\hat{\alpha}_n$ converge néanmoins vers une limite admissible, le théorème 3 montre que l'estimation reste consistante, vérifiant (i-iii-v).

Comme $\bar{\delta}_n$, $\hat{\delta}_n$ ne permet pas, en cas de modélisation erronée, d'obtenir un choix de $\delta_n = h(n; X_1, \dots, X_n)$ vérifiant (i-ii-iii-iv-v). Nous essayons dans le paragraphe suivant de surmonter cette difficulté par des méthodes non paramétriques.

Nous avons introduit (4.1) dans [7], et montré dans [13], des exemples de ses applications. Le cas $\bar{\delta}_n$ peut être simplement considéré comme un cas particulier de $\hat{\delta}_n$.

5. Troisième choix de la fenêtre

On considère un noyau $H(u)$, vérifiant les conditions suivantes :

(i) $H(u) = H(-u)$, (ii) $H(u) \geq 0$, (iii) $H'' = 0$, (iv) H est borné et à support compact, (v) $[[H]] = 1$. (5.1)

Le noyau suivant vérifie (5.1) :

$$H(x) = \begin{cases} \frac{1}{4} (2 - x^2) & , \quad \text{si } 0 \leq |x| \leq 1 , \\ \frac{1}{4} (|x| - 2)^2 & , \quad \text{si } 1 \leq |x| \leq 2 , \\ 0 & , \quad \text{si } |x| \geq 2 . \end{cases}$$

On obtient :

$$H''(x) = \begin{cases} -\frac{1}{2} & , \quad \text{si } 0 < |x| < 1 , \\ +\frac{1}{2} & , \quad \text{si } 1 < |x| < 2 , \\ 0 & , \quad \text{si } |x| > 2 ; \end{cases}$$

$$[[H'']] = [[y^2 H'']] = 2$$

Il serait tout aussi possible de prendre le noyau de Silverman [18]. On définit alors :

$$h_n''(x) = \frac{1}{n \lambda_n^3} \sum_{i=1}^n H''\left(\frac{X_i - x}{\lambda_n}\right), \quad (5.2)$$

où $\{\lambda_n, n \geq 1\}$ est une suite positive.

Le résultat suivant peut alors être obtenu (voir [10]) :

Théorème 4. — Si $\lim_{n \rightarrow \infty} \lambda_n = 0$, $\lim_{n \rightarrow \infty} n \lambda_n^5 = +\infty$, alors

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{+\infty} h_n''(x)^2 dx = \frac{1}{4} [[y^2 H'']]^2 \int_{-\infty}^{+\infty} f''(x)^2 dx \text{ P.-}$$

Corollaire. — Si $f_{n, \check{\delta}_n}(x) = \frac{1}{n \check{\delta}_n} \sum_{i=1}^n K\left(\frac{X_i - x}{\check{\delta}_n}\right)$, où, si h_n'' est définie par (5.2),

$$\check{\delta}_n = \left(\frac{4 [[K^2]]}{[[y^2 H'']]^2 [[y^2 K]]^2} \right)^{1/5} \frac{1}{\left(\int_{-\infty}^{+\infty} h_n''(x)^2 dx \right)^{1/5} n^{1/5}}, \quad (5.3)$$

avec $\lim_{n \rightarrow \infty} \lambda_n = 0$, $\lim_{n \rightarrow \infty} n \lambda_n^5 = +\infty$,

alors $f_{n, \check{\delta}_n}$ satisfait les conditions (i-ii-iii'-iv'-v') de (1.3) et (3.4).

Malheureusement, l'estimation donnée par $\check{\delta}_n$ introduit un nouveau paramétrage par l'intermédiaire de la suite $\{\lambda_n, n \geq 1\}$; le choix de λ_n a en fait une influence considérable sur $\int_{-\infty}^{+\infty} h_n''(x)^2 dx$, et, à moins d'avoir des renseignements extérieurs sur f , il est impossible d'utiliser le résultat du corollaire pour construire une fenêtre pour un échantillon donné (n fixé).

On peut, cependant en déduire des applications à l'estimation séquentielle de la densité pour la construction des estimateurs optimaux (voir [9]).

Pour ces raisons, nous avons été amené à introduire de nouveaux estimateurs de la fenêtre pouvant donner des résultats efficaces pour les petits échantillons.

La définition des estimateurs est exposée dans le paragraphe suivant.

6. Méthode proposée pour le choix de la fenêtre

Si on considère, l'échantillon étant fixé, les variations de

$$\theta_n(\lambda) = \int_{-\infty}^{+\infty} h_n''(x) dx, \quad \text{où} \quad h_n''(x) = \frac{1}{n\lambda^3} \sum_{i=1}^n H''\left(\frac{X_i - x}{\lambda}\right),$$

par (5.1), on constate que :

$$- \text{ Lorsque } \lambda \rightarrow 0, \text{ pour une valeur assez petite, } \theta_n(\lambda) = \frac{[[H''^2]]}{n\lambda^3};$$

$$- \text{ Lorsque } \lambda \rightarrow +\infty, \theta_n(\lambda) \sim \frac{[[H''^2]]}{\lambda^5};$$

On est alors amené à considérer la fonction modifiée :

$$\hat{\theta}_n(\lambda) = \frac{4}{[[y^2 H'']]^2} \int_{-\infty}^{+\infty} h_n''(x)^2 dx - \frac{[[H''^2]]}{n\lambda^5}. \quad (6.1)$$

Le critère proposé consiste alors à poser :

$$\theta_n^D(\lambda) = \sup_{\lambda \geq 1} \hat{\theta}_n(\lambda), \quad \text{où} \quad I = \inf \{ \xi > 0; \forall \eta \geq \xi \hat{\theta}_n(\eta) > 0 \},$$

Nous avons défini ce critère, (d'où l'initiale D), ce qui a amené M. Hominal à proposer un critère voisin :

$$\theta_n^H(\lambda) = \sup_{\lambda \geq J} \hat{\theta}_n(\lambda), \quad \text{où} \quad J = \inf \{ \xi > 0; \forall \eta \geq \xi \hat{\theta}_n(\eta) \leq \hat{\theta}_n(\xi) \}.$$

$$\text{On définit alors } f_{n,\delta_n}(x) = \frac{1}{n\delta_n} \sum_{i=1}^n K\left(\frac{X_i - x}{\delta_n}\right), \quad \text{où}$$

$$\delta_n = \left(\frac{[[K^2]]}{[[y^2 K]]^2} \right)^{1/5} \frac{1}{\tilde{\theta}^{1/5} n^{1/5}}, \quad \text{où} \quad \tilde{\theta} = \theta_n^D \quad \text{ou} \quad \theta_n^H.$$

Les résultats théoriques de convergence de cet estimateur seront développés dans [10]. La partie qui suit est consacrée à l'aspect pratique de la méthode, et à l'étude du comportement des estimations pour les petits échantillons.

II – ESTIMATION AUTOMATIQUE DE LA DENSITE – RESULTATS PAR Pierre HOMINAL

1. Généralités

Nous avons développé dans ce qui suit les aspects pratiques de la méthode de M. Deheuvels. On utilise toujours le noyau unité pour les estimations.

Deux noyaux ont été utilisés pour le calcul de θ_n^D et θ_n^H :

$$\text{— Le noyau de Silverman} \quad H(x) = \begin{cases} \frac{1}{4} x^4 - \frac{1}{2} |x|^3 + \frac{1}{2} & \text{si } 0 \leq |x| \leq 1 \\ \frac{1}{4} |x| (2 - |x|^2) & \text{si } 1 \leq |x| \leq 2 \\ 0 & \text{si } 2 \leq |x| \end{cases}$$

(voir [18]) ;

$$\text{— Le noyau de Deheuvels} \quad H(x) = \begin{cases} \frac{1}{4} (2 - x^2) & \text{si } 0 \leq |x| \leq 1 \\ \frac{1}{4} (|x| - 2)^2 & \text{si } 1 \leq |x| \leq 2 \\ 0 & \text{si } 2 \leq |x|. \end{cases}$$

Ces noyaux vérifient $[[y^2 H'']] = 2$, et $[[H''^2]] = 1$.

Il s'agit de calculer pour chacun des noyaux :

$$\hat{\theta}_n(\lambda) = \int_{-\infty}^{+\infty} h_n''(x)^2 dx - \frac{1}{n\lambda^5} = \frac{2}{n^2\lambda^6} \sum_{1 \leq i < j \leq n} Q_{ij},$$

où
$$Q_{ij} = \int_{-\infty}^{+\infty} H''\left(\frac{x - X_i}{\lambda}\right) H''\left(\frac{x - X_j}{\lambda}\right) dx ;$$

$\theta_n(\lambda)$ est donc déterminé par les $\frac{1}{2} n(n-1)$ valeurs Q_{ij} . Posons $H_{ij} = X_j - X_i$, ($i < j$) ; on obtient :

— Noyau de Deheuvels :

H_{ij}	0	λ	2λ	3λ	4λ	∞
$4Q_{ij}$	$4\lambda - 5H_{ij}$	$-H_{ij}$	$3H_{ij} - 8\lambda$	$4\lambda - H_{ij}$	0	

— Noyau de Silverman : (On pose $F(a; b) = a^3(a^2 + 5b\lambda)$)

H_{ij}	0	λ	2λ	3λ
$\frac{10}{3} \lambda^4 Q_{ij}$	$4F(\lambda - H_{ij}; H_{ij})$ $-F(H_{ij}; \lambda - H_{ij})$	$-2F(H_{ij} - \lambda; 2\lambda - H_{ij})$ $-F(2\lambda - H_{ij}; H_{ij} - \lambda)$	$F(H_{ij} - 2\lambda; 3\lambda - H_{ij})$ $-2F(3\lambda - H_{ij}; H_{ij} - 2\lambda)$	
H_{ij}	3λ	4λ	∞	
$\frac{10}{3} \lambda^4 Q_{ij}$	$F(4\lambda - H_{ij}; H_{ij} - 3\lambda)$	0		

Remarque 1 : Si L est l'étendue de l'échantillon ($L = \sup_i X_i - \inf_i X_i$), le maximum de $\theta_n(\lambda)$ pour θ_n^H ou θ_n^D est atteint pour une valeur de λ inférieure à L . Ainsi, pour le noyau de Deheuvels, $\theta_n(\lambda)$ est décroissant pour $\lambda > \max(L, \lambda')$,

où $\lambda' = \frac{3}{n(n-1)} \sum_{i=1}^{n-1} i(n-i)H_{i,i+1}$, et on peut vérifier que pour $n > 2$, $\lambda' > L$, d'où le résultat.

Remarque 2 : Pour les motifs de rapidité de calcul, il nous a semblé intéressant de déterminer θ_n^H au lieu de θ_n^D . Dans la plupart des cas étudiés, les deux valeurs coïncident. Nous avons comparé en général les valeurs obtenues lorsqu'elles étaient distinctes.

Compte tenu des tableaux précédents, le calcul de $\theta_n(\lambda)$ puis de θ_n^H et θ_n^D est extrêmement rapide.

2. Résultats de simulation

La méthode a été essayée sur 50 échantillons :

- **Lois normales simulées :** 31 échantillons de taille 100 (numérotés de 1 à 31) ont été obtenus par le programme NORMAL de la bibliothèque STATLIB du CII - Iris 80 de l'Université Paris VII. Ce programme utilise la méthode des grands nombres.

2 échantillons de taille 100 (A et B) ont été tirés selon la procédure de Box & Muller [2].

2 échantillons de taille 10, 4 échantillons de taille 20, 4 échantillons de taille 50 (désignés par 10-1, 10-2, 20-1 . . . 20-4, 50-1 . . . 50-4) ont été tirés par la procédure de Box & Muller.

- **Lois gamma simulées :** 4 échantillons de taille 100 (GAM (2) . . . GAM (5)) ont été tirés par la méthode d'Ahrens & Dieter, [1]. Les paramètres sont $\Gamma(2,1)$, $\Gamma(2.5,1)$, $\Gamma(4,1)$, $\Gamma(5,1)$.

- **Echantillons expérimentaux :** Nous avons utilisé les trois échantillons de relevés de hauteurs de vagues en mer du nord (tailles 79, 72 et 94) numérotés VAG (1) à VAG (3), dont on trouvera les données dans [13].

Nous avons contrôlé les caractéristiques des échantillons, par les statistiques suivantes :

$$\text{MOYENNE} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \text{VARIANCE} = s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

$$\text{COEFF A} = a = \frac{\sum_{i=1}^n |X_i - \bar{X}|}{n \sum_{i=1}^n (X_i - \bar{X})^2}, \quad \text{COEFF B1} = \frac{M_3^2}{M_2^3} = \beta_1,$$

$$\text{COEFF B2} = \frac{M_4}{M_2^2} = \beta_2 \quad , \quad \text{où} \quad M_j = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^j$$

Le tableau 1 donne ces résultats. On y remarquera que pour les lois normales, MOYENNE # 0 , et VARIANCE # 1 ; ceci provient du fait que, pour faciliter la comparaison des estimations, tous ces échantillons ont été normalisés par leur moyenne et écart-type empiriques. Ceci ne change pas grand-chose aux fenêtres, mais fait que les tracés des intervalles de confiance de Kolmogorov-Smirnov pour les fonctions de répartition ne correspondent plus au seuil initial (voir § 4, 5).

TABLEAU I
(Voir page suivante)

Determination des valeurs caractéristiques
Echantillons loi normale
Echantillons loi gamma
Echantillons expérimentaux
Détermination
Echantillons loi normale

On peut vérifier la conformité des échantillons simulés aux hypothèses de normalité, en comparant les valeurs obtenues de a , β_1 , β_2 dans le tableau 1, aux niveaux critiques. Par exemple, la valeur critique au seuil de 5 % de β_1 est (Pearson & Hartley, [15]) 0.151 pour $n = 100$; les échantillons 15, 17, 25 ont des coefficients 0.180, 0.222, 0.153, ce qui reste normal pour 33 échantillons.

Une analyse statistique portant sur ces échantillons montre d'une manière générale que les valeurs de simulation sont acceptables.

Présentation des résultats :

Pour chaque échantillon, nous avons réalisé la construction des éléments suivants :

1) La fonction de répartition empirique, avec, pour les lois simulées, les bornes critiques du test de Kolmogorov-Smirnov à 10 % (ici < 10 % pour les lois normales) ;

2) Une tabulation et un graphe de $\hat{\theta}_n(\lambda)$ en trois parties, pour le noyau de Deheuvels (D) et pour le noyau de Silverman (S) :

- Partie 1 : Intervalle $[0, 0.015]$ (Graphe 1) avec un pas de 0.0001
- Partie 2 : Intervalle $[0.015, \lambda_0]$ (Graphe 2) avec un pas de $\lambda_0/500$;
Ici, λ_0 désigne la plus grande racine de $\hat{\theta}(\lambda) = 0$.
- Partie 3 : Intervalle $[\lambda_0, L]$ (Graphe 3) avec un pas de $L/500$;
Ici L désigne l'étendue empirique de l'échantillon.

3) Une détermination de $\lambda_0 =$ plus grande racine de $\hat{\theta}_n(\lambda) = 0$,

Echantillons	Racine λ .		Maximum $\tilde{\lambda}$ Deheuvels		Maximum local $\tilde{\lambda}$ Hominal		$\delta_n(\tilde{\lambda})$ Deheuvels		$\delta_n(\tilde{\lambda}_0)$ Hominal	
	D	S	D	S	D	S	D	S	D	S
Lois normales							Optimum (1.510) (1.510)			
A	.592	.626	.862	.892			1.887	1.903		
B	.412	.474	.646	.684	.664		1.672	1.669	1.674	
1	.405	.423	.511	.569	.579		1.588	1.626	1.590	
2	.572	.598	.817	.848			1.861	1.876		
3	.495	.529	.770	.781	.775		1.750	1.769		
4	.464	.465	.707	.775	.772		1.803	1.811	1.813	
5	.567	.595	.775	.868	.812		1.904	1.930		
6	.472	.495	.735	.793	.785	.795	1.751	1.757	1.752	
7	.521	.549	.735	.747	.737		1.713	1.713		
8	.532	.567	.803	.845			1.830	1.849		
9	.552	.581	.788	.824			1.933	1.960		
10	.542	.572	.748	.757	.761		1.745	1.734	1.746	
11	.195	.202	.229	.227	.741	.626	.860	.816	1.791	1.736
12	.475	.481	.593	.591		.593	1.621	1.558		
13	.545	.528	.785	.812			1.908	1.916		
14	.497	.537	.713	.768	.723		1.745	1.764	1.746	
15	.483	.484	.784	.608	.791	.803	1.793	1.774		1.798
16	.438	.482	.624	.634	.641		1.577	1.563		
17	.376	.408	.466	.762	.699		1.609	1.675	1.674	
18	.554	.575	.747	.730	.750	.732	1.826	1.814		
19	.342	.343	.424	.591	.521		1.446	1.511	1.468	
20	.308	.311	.358	.340	.613	.623	1.467	1.488	1.710	1.700
21	.468	.490	.652	.853	.766	.855	1.826	1.851	1.832	
22	.487	.512	.785	.837	.800		1.802	1.805		
23	.552	.581	.788	.824			1.933	1.960		
24	.474	.475	.607	.576	.617		1.715	1.666	1.716	
25	.491	.562	.811	.858	.843		1.798	1.809	1.800	
26	.583	.611	.835	.903	.869	.904	1.932	1.949	1.934	
27	.549	.581	.819	.866	.824		1.858	1.864		
28	.453	.437	.788	.813	.804		1.750	1.751		
29	.551	.588	.757	.795	.776		1.918	1.942	1.919	
30	.552	.590	.783	.846			1.859	1.882		
31	.433	.473	.588	.678	.637	.681	1.597	1.620	1.603	
Lois normales							Optimum	2.475		
10-1	1.10	1.145	1.464	1.512			3.983	4.103		
10-2	1.005	1.054	1.242	1.279			3.676	3.668		
							Optimum	2.126		
20-1	.731	.795	1.034	1.071			2.740	2.786		
20-2	.781	.804	1.077	.974			2.983	3.001		
20-3	.750	.815	0.992	1.065			2.683	2.742		
20-4	.840	.891	1.100	1.155			2.907	2.993		
							Optimum	1.747		
50-1	.648	.682	.877	.921	.878	.923	2.164	2.194	2.139	2.133
50-2	.224	.655	.241	.872	.871		1.190	2.079		
50-3	.501	.538	.613	.632	.625	.956	1.983	2.093		
50-4	.699	.687	.979	.940	.989		2.351	2.349		
Lois gamma							Optimum	1.029		
G(2)	.541	.566	.663	.748	.708		1.893	1.917	1.904	
							Optimum	?		
G(2.50)	.676	.732	.935	1.008	.992	1.011	2.319	2.334	2.328	
							Optimum	2.151		
G(4)	.607	.658	.848	.939	.890	.942	2.313	2.362	2.329	
							Optimum	2.618		
G(5)	.189	.487	.229	.612	.595		.830	1.812	1.933	
Lois expérimentales										
V1	.093	.094	.113	.117	.423	.404	.375	.364	1.014	.987
V2	.115	.121	.153	.152			.441	.450		
V3	.236	.223	.347	.369	.361		.890	.905	.894	

$$\tilde{\lambda} = \text{Point} \quad \text{où} \quad \hat{\theta}_n(\lambda) = \theta_n^D ; \quad \tilde{\theta} = \theta_n^D ;$$

$$\bar{\lambda} = \text{Point} \quad \text{où} \quad \hat{\theta}_n(\bar{\lambda}) = \theta_n^H ; \quad \bar{\theta} = \theta_n^H ;$$

$\delta_n(\lambda)$ et $\theta_n(\lambda)$, choix respectifs des fenêtres correspondant à θ_n^D et θ_n^H ; pour le noyau de Deheuvels (D), et celui de Silverman (S).

Ces résultats sont tabulés dans le tableau suivant. On a évité d'y transcrire les valeurs fournies par $\theta_n^H, \bar{\lambda}_n, \delta_n(\bar{\lambda})$ (méthode Hominal) lorsque ces valeurs étaient différentes de $\theta_n^D, \lambda_n, \delta_n(\lambda)$ (méthode Deheuvels). Dans certains échantillons, on obtient pour un des deux noyaux (Deheuvels ou Silverman) des valeurs identiques, et pour l'autre des valeurs distinctes.

3. Conclusions de la simulation

— Le choix du noyau de Deheuvels s'impose du fait de sa grande simplicité, de sa rapidité de calcul, et de ce qu'il donne des résultats pratiquement identiques, même pour de très petits échantillons, à ceux du noyau de Silverman.

— Le choix de θ_n^H au lieu de θ_n^D est numériquement préférable, du fait qu'il nécessite un temps de calcul plus réduit (de l'ordre de 50 %). De plus, les valeurs obtenues sont le plus souvent très proches.

— *La méthode de Deheuvels donne d'excellents résultats même pour des échantillons de taille très faible (n = 10).*

— Dans le cas de la loi normale $N(0, 1)$, on peut calculer la valeur δ_n exacte optimale pour le critère du M.I.S.E. Ceci est obtenu à partir des formules de Deheuvels [7], et donne en particulier :

$$\text{Pour } n = 10 \quad \delta_n^{\text{Opt}} = 2.475$$

$$n = 20 \quad \delta_n^{\text{Opt}} = 2.126$$

$$n = 50 \quad \delta_n^{\text{Opt}} = 1.747$$

$$n = 100 \quad \delta_n^{\text{Opt}} = 1.510$$

Compte tenu de la normalisation par l'écart-type, les courbes obtenues avec un DELTA OPTIMAL sont en fait obtenues pour les échantillons normaux avec un δ_n de la forme $\delta_n = \delta_n^{\text{OPT}}$ s, variante du choix (2.4). Les courbes continues tracées correspondent à l'estimation paramétrique de la loi normale. *La méthode de Deheuvels fournit des valeurs toujours très voisines des valeurs optimales*, avec une légère tendance à la surestimation (donnant un lissage un peu plus accentué que l'optimum), qui s'atténue au fur et à mesure que la taille de l'échantillon augmente. Ce défaut est accentué par le choix θ_n^H au lieu de θ_n^D . Cependant, dans deux cas (loi normale $n = 100$, $N^\circ 11$, et loi Gamma $G(5) \Gamma(5,1)$) le choix θ_n^D donne une fenêtre de l'ordre de la moitié de la fenêtre optimale, alors que le choix de θ_n^H reste acceptable. Il peut être éventuellement recommandé de comparer les résultats de la méthode de Deheuvels avec ses variantes θ_n^D et θ_n^H , afin d'obtenir deux résultats d'estimation si ces valeurs sont très différentes.

4. Programmes

Les deux programmes suivants (FORTRAN) fournissent $\hat{\theta}_n(\tilde{\lambda})$ et $\hat{\theta}_n(\bar{\lambda})$ et construisent les estimations de la densité correspondantes.

5. Exemples de résultats

L'ensemble des résultats de simulation ne peut être reproduit dans son intégralité ici (il peut être consulté à l'ISUP, t.45-55, 4 Place Jussieu, Université Paris VI, 75230 PARIS CEDEX 05). Nous en avons extrait les cas suivants :

— (10-2), (20-3), (50-3, 3), 11-16-17-7-4 (ces derniers échantillons étant choisis en prenant dans l'ensemble des 33 échantillons, ceux dont le choix de la fenêtre était le plus voisin supérieurement du nombre de la forme $k(0.1)$ immédiatement inférieur ; par exemple l'échantillon 16 à un $\delta_n(\lambda) = 1.577$, et il n'y a pas d'autre échantillon dont le $\delta_n(\lambda)$ soit compris entre 1.5 et 1.577), G(2), et les échantillons expérimentaux VAG(1), VAG(2), VAG(3).

Tous les échantillons sont reproduits à côté de leur fonction de répartition empirique et des estimations comparées (Optimale, Deheuvels, Hominal le cas échéant). Nous avons tenu à fournir dans ce lot les courbes les plus "mauvaises" afin de donner un aperçu de l'efficacité de la méthode (voir en particulier (50-2) et 11).

III. BIBLIOGRAPHIE

Les méthodes d'estimation concrète de la densité pour des échantillons de taille finie ont été abordées dans [3], [14], [18], [21].

- [1] AHRENS & DIETER, 1974. — Computer methods for sampling from gamma, beta, Poisson and binomial distributions, *Computing*, Vol. 12, p. 223-246.
- [2] BOX & MULLER, 1958. — A note on the generation of random normal deviates, *Ann. Math. Statist.*, Vol. 29, p. 610-611.
- [3] BRETAGNOLLE & HUBER, 1979. — Estimation des densités : risque minimax, *Z. Wahr. und Verw. Geb.*, 47, p.119-137.
- [4] CHUNG, 1948. — An estimate concerning the Kolmogorov limit distribution, *Trans. Amer. Math. Soc.*, 64, p. 205-233.
- [5] DEHEUVELS, 1974. — Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité, *C.R. Acad. Sci. Paris*, t. 278, Ser. A, p. 1217-1220.
- [6] — 1974. — Thèse de doctorat d'Etat (Université Paris VI)
- [7] — 1977. — Estimation non paramétrique de la densité par histogrammes généralisés, *Revue de Statistique Appliquée*, Vol. 25, p. 5-42.
- [8] — 1977. — Estimation non paramétrique de la densité par histogrammes généralisés II, *Publ. Inst. Statist. Univ. Paris*, Vol. 22, p. 1-24.
- [9] — 1979. — Estimation séquentielle de la densité, *Contrib. en Prob. y Est. Mat. Ens. de la Mat. y Analisis*, Univ. Granada, p. 156-169.
- [10] — 1980. — Objective density estimation (A paraître)
- [11] DEVROYE & WAGNER, 1979. — The strong uniform consistency of kernel density estimates, *Multivariate Analysis V*, Krishnaiah edit., North Holland.

- [12] HOMINAL, 1979. — Thèse de III^{ème} Cycle, Université Paris VI
- [13] HOMINAL & DEHEUVELS, 1979. — Estimation non paramétrique de la densité compte tenu d'informations sur le support, *Revue de Statistique Appliquée*, Vol. 27, p. 47-68.
- [14] NADAYARA, 1974. — On the integral mean square error of some non parametric estimates for the density function, *Theor. Prob. Appl.*, Vol. 19, p. 133-141.
- [15] PEARSON & HARTLEY, 1976. — Biometrika tables for statisticians, Vol. I-II, Cambridge Univ. Press.
- [16] RENYI, 1966. — Calcul des probabilités, Dunod.
- [17] SMIRNOV, 1944. — Approximate laws of distribution of random variables from empirical data, *Uspehi Mat. Nauk*, 10, p. 179-206.
- [18] SILVERMAN, 1978. — Choosing the window width when estimating a density, *Biometrika* 65, 1, p. 1-11.
- [19] — 1979. — Communication privée
- [20] WERTZ & SCHNEIDER, 1979. — Statistical density estimation : a bibliography, *International Statistical Review*, 47, p. 155-175.
- [21] WOODROOFE, 1970. — On choosing a delta sequence, *Ann. Math. Statist.*, Vol. 41, p. 1665-1671.

DETERMINATION DES VALEURS CARACTERISTIQUES

* Echantillons de loi normale :

+	+	MOYENNE	VARIANCE	COEFF A	COEFF B1	COEFF B2	+
+	+	+	+	+	+	+	+
*!	10 1	!	.0000	!	.9993E+00	!	.8166F+00
*!	10 2	!	.0001	!	.9994E+00	!	.7950F+00
*!	10 3	!	.0002	!	.1000E+01	!	.7884F+00
*!	10 4	!	.0003	!	.1000E+01	!	.8135F+00
*!	10 5	!	.0004	!	.1000E+01	!	.7950F+00
*!	10 6	!	.0005	!	.1000E+01	!	.8141F+00
*!	10 7	!	.0006	!	.1000E+01	!	.7946F+00
*!	10 8	!	.0007	!	.1000E+01	!	.7941F+00
*!	10 9	!	.0008	!	.1000E+01	!	.8082F+00
*!	10 10	!	.0009	!	.1000E+01	!	.8418F+00
*!	10 11	!	.0010	!	.1000E+01	!	.9557E+03
*!	10 12	!	.0011	!	.1000E+01	!	.7984F+00
*!	10 13	!	.0012	!	.1000E+01	!	.6860E+02
*!	10 14	!	.0013	!	.1000E+01	!	.8078F+00
*!	10 15	!	.0014	!	.1000E+01	!	.9272E+02
*!	10 16	!	.0015	!	.1000E+01	!	.1153E+00
*!	10 17	!	.0016	!	.1000E+01	!	.7752F+00
*!	10 18	!	.0017	!	.1000E+01	!	.8324F+00
*!	10 19	!	.0018	!	.1000E+01	!	.5937E+03
*!	10 20	!	.0019	!	.1000E+01	!	.8009F+00
*!	10 21	!	.0020	!	.1000E+01	!	.1237E+00
*!	10 22	!	.0021	!	.1000E+01	!	.7769F+00
*!	10 23	!	.0022	!	.1000E+01	!	.1042F+00
*!	10 24	!	.0023	!	.1000E+01	!	.8066F+00
*!	10 25	!	.0024	!	.1000E+01	!	.2198E+02
*!	10 26	!	.0025	!	.1000E+01	!	.8127F+00
*!	10 27	!	.0026	!	.1000E+01	!	.1706E+01
*!	10 28	!	.0027	!	.1000E+01	!	.7903F+00
*!	10 29	!	.0028	!	.1000E+01	!	.2594F+01
*!	10 30	!	.0029	!	.1000E+01	!	.8418F+00
*!	10 31	!	.0030	!	.1000E+01	!	.9557E+03
*!	10 32	!	.0031	!	.1000E+01	!	.8312E+02
*!	10 33	!	.0032	!	.1000E+01	!	.1532E+00
*!	10 34	!	.0033	!	.1000E+01	!	.7955F+00
*!	10 35	!	.0034	!	.1000E+01	!	.8241F+00
*!	10 36	!	.0035	!	.1000E+01	!	.3472E+03
*!	10 37	!	.0036	!	.1000E+01	!	.4485E+03
*!	10 38	!	.0037	!	.1000E+01	!	.7931F+00
*!	10 39	!	.0038	!	.1000E+01	!	.4347E+01
*!	10 40	!	.0039	!	.1000E+01	!	.8485F+00
*!	10 41	!	.0040	!	.1000E+01	!	.6702E+02
*!	10 42	!	.0041	!	.1000E+01	!	.1036E+01
*!	10 43	!	.0042	!	.1000E+01	!	.2629F+01
*!	10 44	!	.0043	!	.1000E+01	!	.7848F+00
*!	10 45	!	.0044	!	.1000E+01	!	.3988F+01

* Echantillons de loi gamma :

+	+	MOYENNE	VARIANCE	COEFF A	COEFF B1	COEFF B2	+
+	+	+	+	+	+	+	+
*!	G(2)	!	2.190	!	.2165E+01	!	.7765F+00
*!	G(2,50)	!	2.701	!	.2818F+01	!	.1257E+01
*!	G(4)	!	3.794	!	.4257E+01	!	.1743E+01
*!	G(5)	!	4.881	!	.4592E+01	!	.5381F+01
*!	G(6)	!	5.968	!	.4927E+01	!	.1037F+02
*!	G(7)	!	7.055	!	.5262E+01	!	.1374F+02

* Echantillons experimentaux :

+	+	MOYENNE	VARIANCE	COEFF A	COEFF B1	COEFF B2	+
+	+	+	+	+	+	+	+
*!	V1	!	2.510	!	.5632E+00	!	.7412F+00
*!	V2	!	1.706	!	.2574E+00	!	.1770E+01
*!	V3	!	2.129	!	.9185E+00	!	.4951F+01
*!	V4	!	1.706	!	.2574E+00	!	.8905E+01
*!	V5	!	2.129	!	.9185E+00	!	.1803F+01
*!	V6	!	2.510	!	.5632E+00	!	.4649F+01

* Echantillons de loi normale :

+	+	MOYENNE	VARIANCE	COEFF A	COEFF B1	COEFF B2	+
+	+	+	+	+	+	+	+
*!	10-1	!	.0000	!	.1000E+01	!	.8475F+00
*!	10-2	!	.0001	!	.1000E+01	!	.7792F+00
*!	20-1	!	.0002	!	.1000E+01	!	.7770F+00
*!	20-2	!	.0003	!	.1000E+01	!	.8042F+00
*!	20-3	!	.0004	!	.1000E+01	!	.8011F+00
*!	20-4	!	.0005	!	.1000E+01	!	.8167F+00
*!	50-1	!	.0006	!	.1000E+01	!	.7974F+00
*!	50-2	!	.0007	!	.1000E+01	!	.8154F+00
*!	50-3	!	.0008	!	.1000E+01	!	.8067F+00
*!	50-4	!	.0009	!	.1000E+01	!	.8374F+00
*!	50-5	!	.0010	!	.1000E+01	!	.7970F+00
*!	50-6	!	.0011	!	.1000E+01	!	.8233F+00
*!	50-7	!	.0012	!	.1000E+01	!	.7799F+00
*!	50-8	!	.0013	!	.1000E+01	!	.8313F+00
*!	50-9	!	.0014	!	.1000E+01	!	.7970F+00
*!	50-10	!	.0015	!	.1000E+01	!	.8242F+00
*!	50-11	!	.0016	!	.1000E+01	!	.7950F+00
*!	50-12	!	.0017	!	.1000E+01	!	.8141F+00
*!	50-13	!	.0018	!	.1000E+01	!	.7946F+00
*!	50-14	!	.0019	!	.1000E+01	!	.7941F+00
*!	50-15	!	.0020	!	.1000E+01	!	.8082F+00
*!	50-16	!	.0021	!	.1000E+01	!	.8418F+00
*!	50-17	!	.0022	!	.1000E+01	!	.9557E+03
*!	50-18	!	.0023	!	.1000E+01	!	.7984F+00
*!	50-19	!	.0024	!	.1000E+01	!	.6860E+02
*!	50-20	!	.0025	!	.1000E+01	!	.8078F+00
*!	50-21	!	.0026	!	.1000E+01	!	.9272E+02
*!	50-22	!	.0027	!	.1000E+01	!	.1153E+00
*!	50-23	!	.0028	!	.1000E+01	!	.7752F+00
*!	50-24	!	.0029	!	.1000E+01	!	.8324F+00
*!	50-25	!	.0030	!	.1000E+01	!	.5937E+03
*!	50-26	!	.0031	!	.1000E+01	!	.8009F+00
*!	50-27	!	.0032	!	.1000E+01	!	.1237E+00
*!	50-28	!	.0033	!	.1000E+01	!	.7769F+00
*!	50-29	!	.0034	!	.1000E+01	!	.1042F+00
*!	50-30	!	.0035	!	.1000E+01	!	.8066F+00
*!	50-31	!	.0036	!	.1000E+01	!	.2198E+02
*!	50-32	!	.0037	!	.1000E+01	!	.8127F+00
*!	50-33	!	.0038	!	.1000E+01	!	.1706E+01
*!	50-34	!	.0039	!	.1000E+01	!	.7903F+00
*!	50-35	!	.0040	!	.1000E+01	!	.2594F+01
*!	50-36	!	.0041	!	.1000E+01	!	.8418F+00
*!	50-37	!	.0042	!	.1000E+01	!	.9557E+03
*!	50-38	!	.0043	!	.1000E+01	!	.8312E+02
*!	50-39	!	.0044	!	.1000E+01	!	.1532E+00
*!	50-40	!	.0045	!	.1000E+01	!	.7955F+00
*!	50-41	!	.0046	!	.1000E+01	!	.8241F+00
*!	50-42	!	.0047	!	.1000E+01	!	.3472E+03
*!	50-43	!	.0048	!	.1000E+01	!	.4485E+03
*!	50-44	!	.0049	!	.1000E+01	!	.7931F+00
*!	50-45	!	.0050	!	.1000E+01	!	.4347E+01
*!	50-46	!	.0051	!	.1000E+01	!	.8485F+00
*!	50-47	!	.0052	!	.1000E+01	!	.6702E+02
*!	50-48	!	.0053	!	.1000E+01	!	.1036E+01
*!	50-49	!	.0054	!	.1000E+01	!	.2629F+01
*!	50-50	!	.0055	!	.1000E+01	!	.7848F+00
*!	50-51	!	.0056	!	.1000E+01	!	.3988F+01


```

5 S=S-5*H+D4
  GOTO 2
6 CONTINUE
  IF(S.LE.0.) GOTO 9
  S=(S/D)**.2/D
  IF(S.GT.SP) GOTO 7
  IF(K.EQ.1) GOTO 8
  K=1
  D=D+DPAS+DPAS
  DPAS=DPAS/100
  GOTO 1
7 SMAX=AMAX1(SMAX,S)
8 SP=S
  WRITE (6,6000)
6000 FORMAT(/11X,7(' '), 'ATTENTION : AUCUNE VALEUR NEGATIVE RENCONTR',
  * 'EE DANS LE BALAYAGE',7(' '))
9 D=(288.*N)**.2/SMAX
  D2=D/2

```

C
C
C
C

```

-----
  DETERMINATION DES ABSCISSES ET DES SAUTS DE L'HISTOGRAMME.
-----

  J=0
  DO 10 I=1,N
    K=J+1
    J=K+1
    VI=V(I)
    ABS(K)=VI-D2
    SAUT(K)=1
    ABS(J)=VI+D2
10  SAUT(J)=-1.
    DO 14 I=2,N2
      II=I-1
      VI=ABS(I)
      DO 11 J=1,II
        IF(VI.LT.ABS(J)) GOTO 12
11  CONTINUE
      GOTO 14
12  M=I-J
      H=SAUT(I)
      DO 13 K=1,M
        L=I-K
        ABS(L+1)=ABS(L)
13  SAUT(L+1)=SAUT(L)
        ABS(J)=VI
        SAUT(J)=H
14  CONTINUE
      H=N*D
      S=0.
      DO 15 I=1,N2
        S=S+SAUT(I)
15  SAUT(I)=S/H
      RETURN
  END

```

```

C      -----
C      |
C      | CONSTRUCTION DE L'ESTIMATION NON PARAMETRIQUE DE LA DENSITE PAR LA
C      | FENETRE MOBILE DETERMINEE A PARTIR DE L'ECHANTILLON PAR LA METHODE
C      | DE MR DEHEUVELS.
C      |
C      | REALISATION P. HOMINAL - JUIN 1979 - UNIVERSITE PARIS VI.
C      |
C      |-----
C      |
C      | ENTREES : N TAILLE DE L'ECHANTILLON TRAITE.
C      |             N2=2*N.
C      |             V TABLEAU DES N VALEURS ORDONNEES DE L'ECHANTILLON.
C      |
C      | SORTIES : ABS TABLEAU DES N2 ABSCISSES DES SAUTS DE L'HISTOGRAMME.
C      |             SAUT TABLEAU DES N2 VALEURS DE L'HISTOGRAMME EN ABS.
C      |             D LONGUEUR ESTIMEE DE LA FENETRE.
C      |
C      |-----
C      | DIMENSIONNER LES TABLEAUX V,ABS ET SAUT DANS LE PROGRAMME D'APPEL.
C      |-----
C      | SUBROUTINE HISTO (N, N2, V, ABS, SAUT, D)
C      | DIMENSION V (N), ABS (N2), SAUT (N2)
C      |-----
C      | DETERMINATION DE LA FENETRE (METHODE HOMINAL).
C      |-----
C      | II=N-1
C      | K=0
C      | D=V(N)-V(1)
C      | DPAS=AMIN1(.05,D/100)
C      | 1 M=INT(D/DPAS)
C      |   SP=0.
C      |   D=D+DPAS
C      |   DO 8 L=1,M
C      |     D=D-DPAS
C      |     D2=D+D
C      |     D3=D2+D
C      |     D4=D3+D
C      |     S=0.
C      |     DO 6 I=1,II
C      |       VI=V(I)
C      |       J=I
C      |     2 J=J+1
C      |       IF(J.GT.N) GOTO 6
C      |       H=V(J)-VI
C      |       IF(H.GE.D4) GOTO 6
C      |       IF(H.LT.D3) GOTO 3
C      |       S=S+D4-H
C      |       GOTO 2
C      |     3 IF(H.LT.D2) GOTO 4
C      |       S=S+3*H-2*D4
C      |       GOTO 2
C      |     4 IF(H.LT.D) GOTO 5
C      |       S=S-H
C      |       GOTO 2
C      |     5 S=S-5*H+D4
C      |       GOTO 2
C      |     6 CONTINUE
C      |       IF(S.LE.0.) GOTO 9
C      |       S=(S/D)**.2/D
C      |       IF(S.GT.SP) GOTO 8
C      |       IF(K.EQ.1) GOTO 9
C      |       K=1

```

```

      D=D+DPAS+DPAS
      DPAS=DPAS/100
      GOTO 1
8 SP=S
  WRITE(6,6000)
6000 FORMAT(/11X,7(' '), 'ATTENTION :AUCUNE VALEUR NEGATIVE RENCONTR',
  * 'EE DANS LE BALAYAGE',7(' '))
9 D=(288.*N)**.2/S
  D2=D/2

```

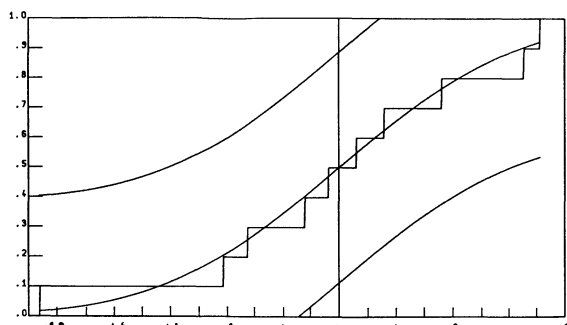
C
C
C
C
C

DETERMINATION DES ABSCISSES ET DES SAUTS DE L'HISTOGRAMME.

```

      J=0
      DO 10 I=1,N
      K=J+1
      J=K+1
      VI=V(I)
      ABS(K)=VI-D2
      SAUT(K)=1.
      ABS(J)=VI+D2
10 SAUT(J)=-1.
      DO 14 I=2,N2
      II=I-1
      VI=ABS(I)
      DO 11 J=1,II
      IF(VI.LT.ABS(J)) GOTO 12
11 CONTINUE
      GOTO 14
12 M=I-J
      H=SAUT(I)
      DO 13 K=1,M
      L=I-K
      ABS(L+1)=ABS(L)
13 SAUT(L+1)=SAUT(L)
      ABS(J)=VI
      SAUT(J)=H
14 CONTINUE
      H=N*D
      S=0.
      DO 15 I=1,N2
      S=S+SAUT(I)
15 SAUT(I)=S/H
      RETURN
      END

```

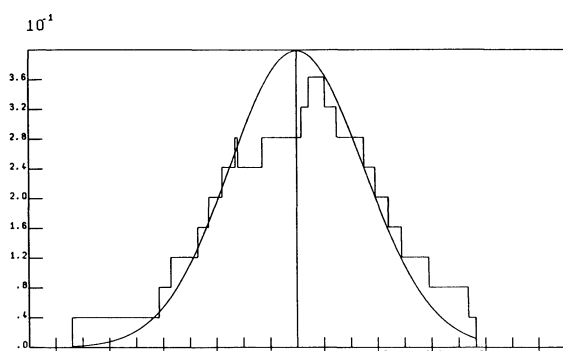



TEST DE KOLMOGOROFF SUR L ECHANTILLON 10-2

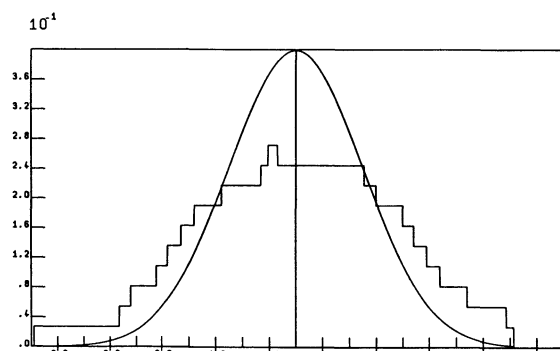
```

*****
* FTIUD DES MAXIMA DE THETA DANS LE DOMAINE ADMISSIBLE (D0,...
* POUR LE MOYAU DE DEHEUELS (D)
*
* ECHANTILLON 10-2 EFFECTIF : 10
*
*****
* VALEURS DE L ECHANTILLON
*
* -2.118 -.820 -.646 -.740 -.072 .125 .322 .724 1.307 1.418
*
* ETUDE FAITE AVEC UN PAS ENTRE VALEURS DE DELTA VALANT : .00035
*
* MAXIMUM RELATIF LE PLUS A DROITE : 1.2418 THETA : .46379
*
* MAXIMUM ABSOLU : 1.2418 THETA : .46379
*
* D0 EST ENTRE D= .1005E+01 (T= .1471E+00) ET D= .1004E+01 (T= -.2475E-01)
*
*****
* LONGUEUR ESTIMEE DE LA FENETRE ( METHODE HOMINAL ) : 3.676
* LONGUEUR ESTIMEE DE LA FENETRE (METHODE DEHEUELS) : 3.676
*
*****

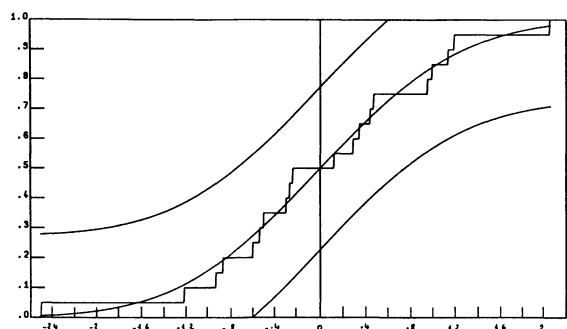
```



DENSITE DE L ECHANTILLON 10-2 (DELTA OPTIMAL)



DENSITE DE L ECHANTILLON 10-2 (METHODE DEHEUELS)

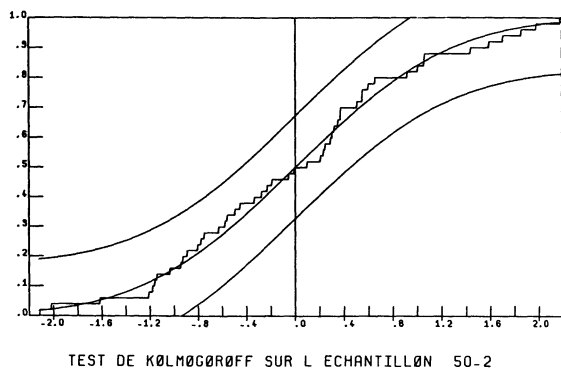
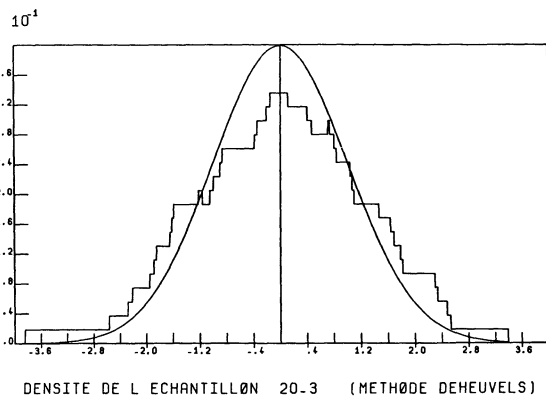
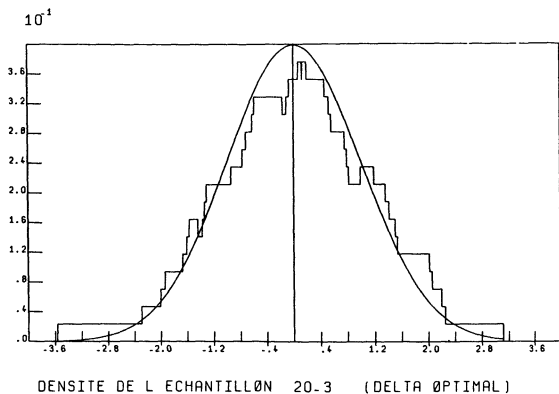


TEST DE KOLMOGOROFF SUR L ECHANTILLON 20-3

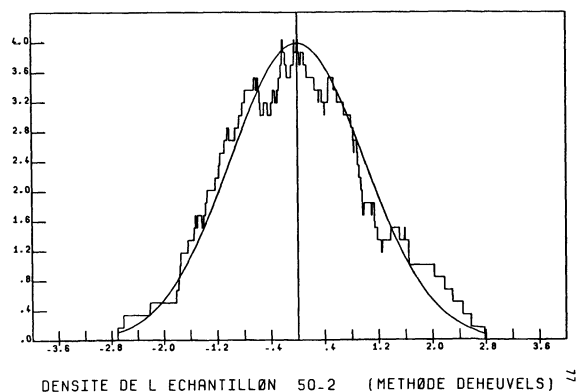
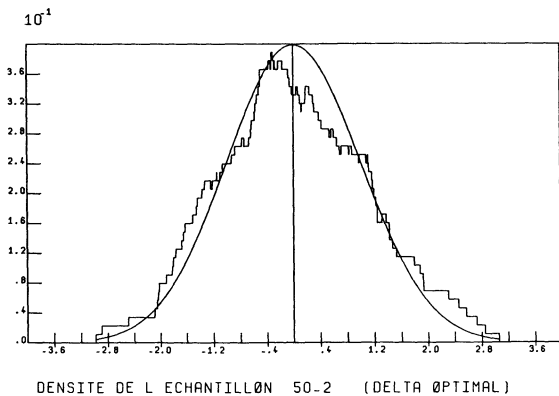
```

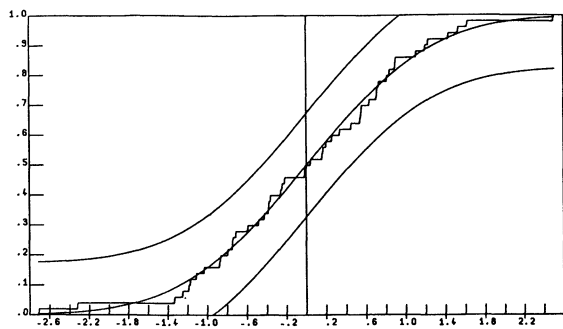
*****
* FTIUD DES MAXIMA DE THETA DANS LE DOMAINE ADMISSIBLE (D0,...
* POUR LE MOYAU DE DEHEUELS (D)
*
* ECHANTILLON 20-3 EFFECTIF : 20
*
*****
* VALEURS DE L ECHANTILLON
*
* -2.500 -1.225 -.938 -.871 -.004 -.541 -.506 -.306 -.274 -.244
* .123 .293 .348 .443 .476 .955 .995 1.135 1.191 2.053
*
* ETUDE FAITE AVEC UN PAS ENTRE VALEURS DE DELTA VALANT : .00046
*
* MAXIMUM RELATIF LE PLUS A DROITE : .9920 THETA : .55322
*
* MAXIMUM ABSOLU : .9920 THETA : .55322
*
* D0 EST ENTRE D= .7507E+00 (T= .1432E+00) ET D= .7494E+00 (T= -.1454E+00)
*
*****
* LONGUEUR ESTIMEE DE LA FENETRE ( METHODE HOMINAL ) : 2.683
* LONGUEUR ESTIMEE DE LA FENETRE (METHODE DEHEUELS) : 2.683
*
*****

```



```
*****
*
*   ETUDE DES MAXIMA DE THETA DANS LE DOMAINE ADMISSTRIE (NO,...
*   POUR LE MOYAU DE DEHEUVELS (D)
*
*   ECHANTILLON 50-2                      EFFECTIF : 50
*
*****
*
*   VALEURS DE L ECHANTILLON
*
*   -2.111 -2.020 -1.016 -1.213 -1.177 -1.159 -1.145 -1.036 -.946 -.927
*   -.893 -.803 -.775 -.750 -.672 -.580 -.562 -.495 -.454 -.437
*   -.270 -.220 -.189 -.058 -.005 .097 .205 .233 .244 .289
*   .301 .320 .350 .368 .371 .507 .545 .548 .592 .650
*   .919 1.002 1.053 1.059 1.435 1.584 1.702 1.853 1.928 2.186
*
*****
*
*   ETUDE FAITE AVEC UN PAS ENTRE VALEURS DE DELTA VALANT : .00043
*
*   MAXIMUM RELATIF LE PLUS A DROITE : .8705      THETA : .57763
*
*   MAXIMUM ABSOLU : .2405      THETA : -1.03824
*
*   ON EST ENTRE DE : .2744E+00 (T = .5235E+00) ET DE : .2747E+00 (T = .5015E+00)
*
*****
*
*   LONGUEUR ESTIMEE DE LA FENETRE (METHODE NOMINAL) : 2.139
*   LONGUEUR ESTIMEE DE LA FENETRE (METHODE DEHEUVELS) : 1.190
*
*****
```



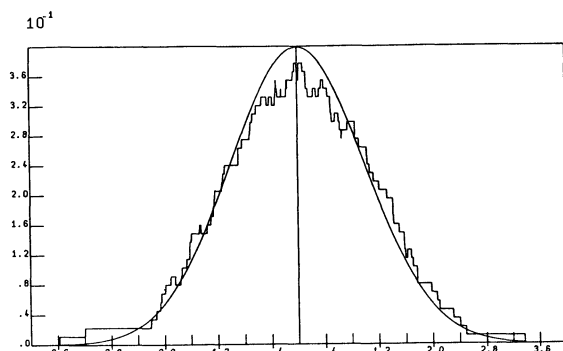


TEST DE KOLMOGOROFF SUR L ECHANTILLON 50-3

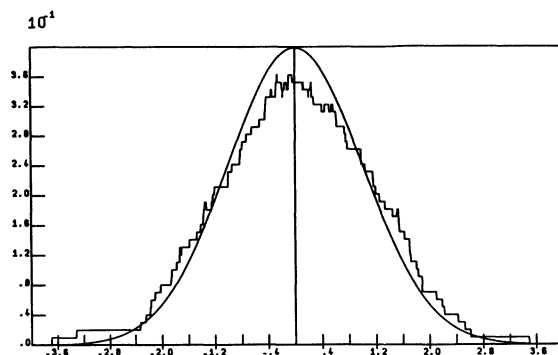
```

*****
*   ETUDE DES MAXIMA DE THETA DANS LE DOMAINE ADMISSIBLE (DO,...
*   POUR LE NOYAU DE DEHEUVELS (D)
*
*   ECHANTILLON 50-3   EFFECTIF : 50
*****
*
*   VALEURS DE L ECHANTILLON
*
*   -2.703 -2.320 -1.331 -1.208 -1.185 -1.170 -1.109 -1.027 -.874 -.862
*   -.780 -.743 -.702 -.713 -.593 -.481 -.422 -.381 -.375 -.362
*   -.257 -.242 -.214 -.012 -.011 .048 .162 .171 .201 .258
*   .339 .456 .542 .550 .558 .637 .710 .715 .729 .818
*   .839 .898 .901 1.113 1.149 1.232 1.435 1.536 1.822 2.502
*
*****
*   ETUDE FAITE AVEC UN PAS ENTRE VALEURS DE DELTA VALANT : .00050
*   MAXIMUM RELATIF LE PLUS A DROITE : .6245   THETA : .62296
*   MAXIMUM ABSOLU : .6130   THETA : .62318
*
*   DO EST ENTRE D = .5000E+00 (T = .2641E+00) ET D = .5005E+00 (T = .1885E+00)
*****
*
*   LONGUEUR ESTIMEE DE LA FENETRE ( METHODE NOMINAL ) : 1.983
*   LONGUEUR ESTIMEE DE LA FENETRE ( METHODE DEHEUVELS ) : 1.983
*****

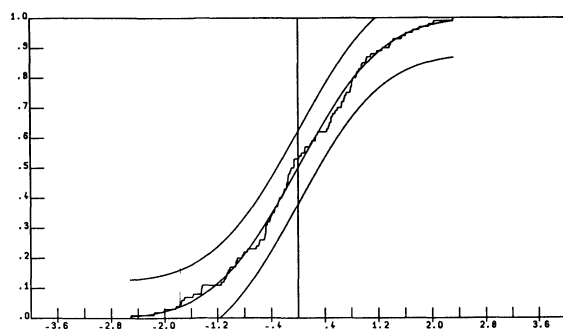
```



DENSITE DE L ECHANTILLON 50-3 (DELTA OPTIMAL)



DENSITE DE L ECHANTILLON 50-3 (METHODE DEHEUVELS)

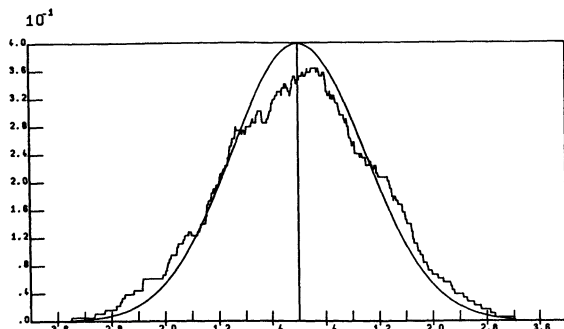


TEST DE KOLMOGOROFF SUR L ECHANTILLON N° 11

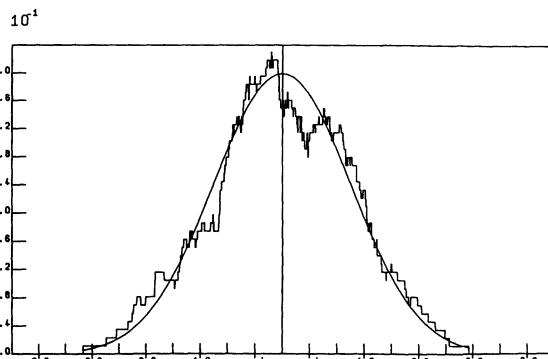
```

*****
*   ETUDE DES MAXIMA DE THETA DANS LE DOMAINE ADMISSIBLE (DO,...
*   POUR LE NOYAU DE DEHEUVELS (D)
*
*   ECHANTILLON N° 11   EFFECTIF : 100
*****
*
*   VALEURS DE L ECHANTILLON
*
*   -2.503 -2.156 -2.006 -1.832 -1.772 -1.747 -1.704 -1.570 -1.438 -1.435
*   -1.429 -1.138 -1.095 -1.073 -1.062 -1.040 -1.015 -.931 -.923 -.905
*   -.831 -.810 -.751 -.619 -.584 -.572 -.498 -.492 -.480 -.479
*   -.476 -.454 -.416 -.409 -.386 -.350 -.340 -.300 -.292 -.275
*   -.237 -.215 -.204 -.169 -.168 -.147 -.145 -.137 -.111 -.099
*   -.057 -.057 -.053 .022 .055 .104 .110 .184 .206 .266
*   .267 .291 .433 .458 .471 .497 .498 .522 .560 .591
*   .657 .660 .695 .704 .725 .789 .795 .810 .811 .831
*   .881 .886 .918 .959 .964 1.021 1.033 1.094 1.181 1.263
*   1.368 1.389 1.424 1.555 1.614 1.713 1.816 1.957 2.027 2.315
*
*****
*   ETUDE FAITE AVEC UN PAS ENTRE VALEURS DE DELTA VALANT : .00048
*   MAXIMUM RELATIF LE PLUS A DROITE : .7413   THETA : .60063
*   MAXIMUM ABSOLU : .2287   THETA : 1.25147
*
*   DO EST ENTRE D = .1950E+00 (T = .4182E+00) ET D = .1945E+00 (T = .6923E+00)
*****
*
*   LONGUEUR ESTIMEE DE LA FENETRE ( METHODE NOMINAL ) : 1.791
*   LONGUEUR ESTIMEE DE LA FENETRE ( METHODE DEHEUVELS ) : .860
*****

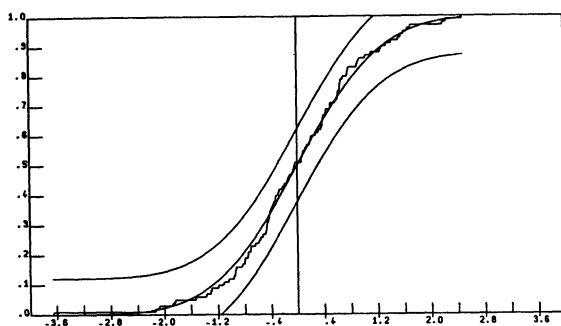
```



DENSITE DE L ECHANTILLON N° 11 (METHODE H0MINAL)



DENSITE DE L ECHANTILLON N° 11 (METHODE DEHEUVELS)

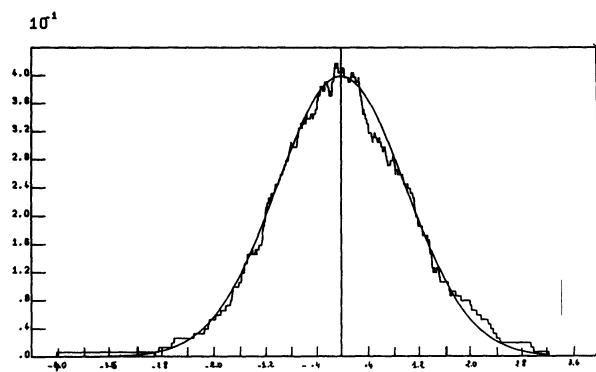


TEST DE KOLMOGOROFF SUR L ECHANTILLON N° 16

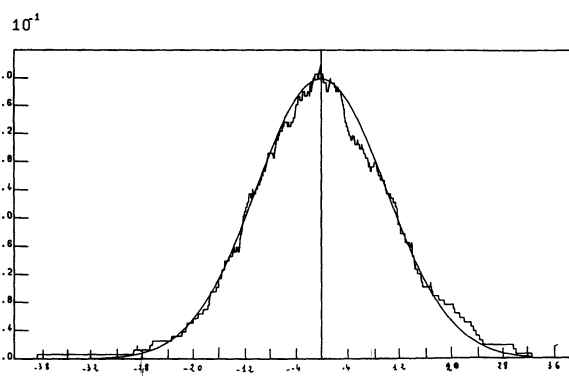
```

.....
*
*   ETUDE DES MAXIMA DE THETA DANS LE DOMAINE ADMISSIBLE (D0,...
*   POUR LE NIVAU DE DEHEUVELS (D)
*
*   ECHANTILLON N° 16          EFFECTIF : 100
*
*   VALEURS DE L'ECHANTILLON
*
*   -3.651 -2.131 -2.076 -1.875 -1.837 -1.831 -1.367 -1.308 -1.290 -1.194
*   -1.112 -1.037 -0.963 -0.945 -0.938 -0.926 -0.835 -0.798 -0.753
*   -0.750 -0.704 -0.704 -0.638 -0.580 -0.559 -0.528 -0.478 -0.469 -0.468
*   -0.457 -0.441 -0.427 -0.421 -0.417 -0.397 -0.375 -0.354 -0.332 -0.328
*   -0.280 -0.268 -0.220 -0.192 -0.154 -0.133 -0.108 -0.068 -0.058 -0.045
*   -0.029 -0.065 -0.073 -0.086 -0.090 -0.124 -0.147 -0.191 -0.208 -0.216
*   -0.284 -0.289 -0.299 -0.343 -0.384 -0.394 -0.401 -0.430 -0.434 -0.503
*   -0.504 -0.583 -0.615 -0.615 -0.619 -0.645 -0.650 -0.655 -0.665 -0.696
*   -0.752 -0.761 -0.776 -0.904 -0.919 -0.932 -1.025 -1.109 -1.250 -1.870
*   1.348 1.417 1.523 1.561 1.604 1.655 1.707 2.188 2.232 2.471
*
*   ETUDE FAITE AVEC UN PAS ENTRE VALEURS DE DELTA VALANT : 0.00050
*   MAXIMUM RELATIF LE PLUS A DROITE : 0.6409   THETA : 0.68219
*   MAXIMUM ABSOLU : 0.6239   THETA : 0.68223
*   D0 EST ENTRE D = 0.4384E+00 (T = 0.1574E+00) ET D = 0.4379E+00 (T = 0.2657E+00)
*
*   LONGUEUR ESTIMEE DE LA FENETRE (METHODE H0MINAL) : 1.577
*   LONGUEUR ESTIMEE DE LA FENETRE (METHODE DEHEUVELS) : 1.577
*
.....

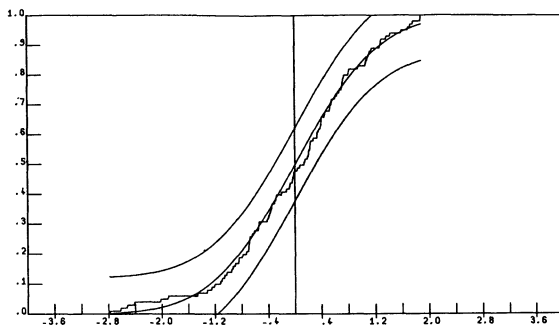
```



DENSITE DE L ECHANTILLON N° 16 (DELTA OPTIMAL)



DENSITE DE L ECHANTILLON N° 16 (METHODE DEHEUVELS)



TEST DE KOLMOGOROFF SUR L ECHANTILLON N° 17

ETUDE DES MAXIMA DE THETA DANS LE DOMAINE ADMISSIBLE (D0,...
POUR LE ROYAL DE SILVERMAN (S)

ECHANTILLON N° 17 EFFECTIF : 100

VALEURS DE L ECHANTILLON

-2.775	-2.045	-2.517	-2.449	-2.026	-1.914	-1.467	-1.300	-1.251	-1.207
-1.117	-1.003	-1.054	-1.052	-0.946	-0.955	-0.893	-0.841	-0.834	-0.784
-0.731	-0.605	-0.581	-0.677	-0.771	-0.675	-0.637	-0.615	-0.590	-0.542
-0.537	-0.408	-0.390	-0.376	-0.364	-0.350	-0.345	-0.308	-0.284	-0.274
-0.154	-0.124	-0.176	-0.070	-0.129	-0.024	-0.014	-0.008	-0.040	-0.076
0.144	0.157	0.151	0.210	0.212	0.233	0.236	0.243	0.282	0.341
0.341	0.365	0.368	0.400	0.402	0.404	0.445	0.460	0.528	0.535
0.552	0.558	0.614	0.644	0.676	0.709	0.710	0.717	0.725	0.739
0.811	0.813	0.851	0.903	0.970	1.053	1.108	1.123	1.151	1.273
1.282	1.287	1.369	1.429	1.597	1.695	1.726	1.771	1.883	1.884

ETUDE FAITE AVEC UN PAS ENTRE VALEURS DE DELTA VALANT : 0.0093

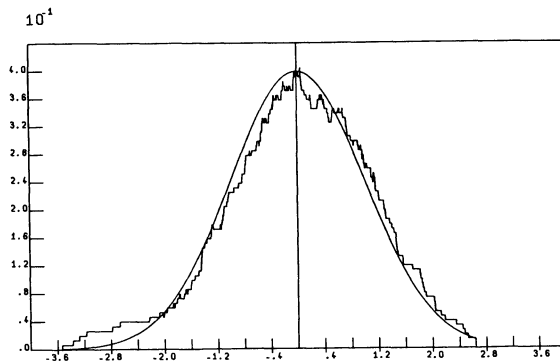
MAXIMUM RELATIF LE PLUS A DROITE : 0.7617 THETA : 0.64215

MAXIMUM ABSOLU : 0.7617 THETA : 0.64215

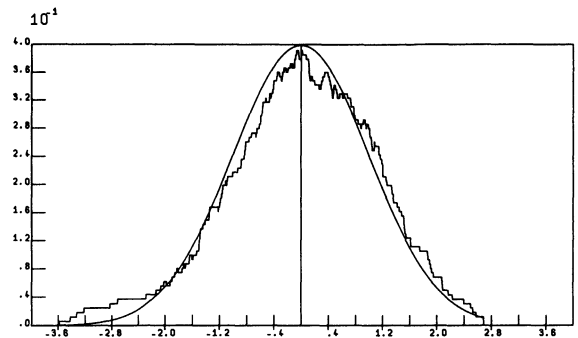
D0 EST ENTRE D = 0.83E+00 (T = 0.2324E+00) ET D = 0.074E+00 (T = 0.2901E+00)

LONGUEUR ESTIMEE DE LA FENETRE (METHODE NOMINAL) : 1.675

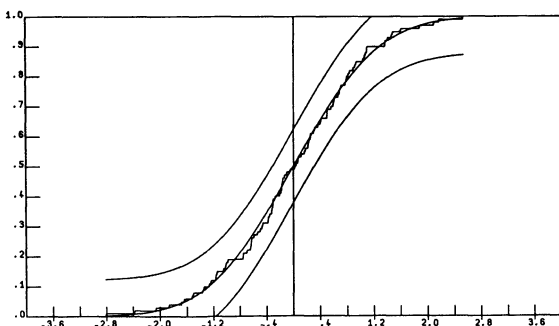
LONGUEUR ESTIMEE DE LA FENETRE (METHODE DEHEUVELS) : 1.675



DENSITE DE L ECHANTILLON N° 17 (DELTA OPTIMAL)



DENSITE DE L ECHANTILLON N° 17 (METHODE DEHEUVELS)



TEST DE KOLMOGOROFF SUR L ECHANTILLON N° 7

ETUDE DES MAXIMA DE THETA DANS LE DOMAINE ADMISSIBLE (D0,...
POUR LE ROYAL DE DEHEUVELS (D)

ECHANTILLON N° 7 EFFECTIF : 100

VALEURS DE L ECHANTILLON

-2.804	-2.372	-2.058	-1.865	-1.683	-1.640	-1.525	-1.503	-1.363	-1.350
-1.267	-1.240	-1.168	-1.160	-1.147	-1.049	-1.019	-1.007	-0.981	-0.751
-0.747	-0.688	-0.643	-0.631	-0.626	-0.621	-0.588	-0.536	-0.490	-0.404
-0.453	-0.375	-0.374	-0.353	-0.324	-0.320	-0.312	-0.311	-0.307	-0.276
-0.202	-0.185	-0.173	-0.172	-0.160	-0.157	-0.128	-0.098	-0.041	0.025
0.040	0.068	0.076	0.117	0.167	0.181	0.219	0.221	0.235	0.255
0.258	0.303	0.318	0.369	0.397	0.412	0.459	0.502	0.503	0.559
0.591	0.603	0.612	0.665	0.689	0.696	0.709	0.772	0.779	0.788
0.809	0.835	0.885	0.915	0.934	1.036	1.057	1.076	1.093	1.096
1.368	1.368	1.403	1.440	1.479	1.601	1.881	2.091	2.164	2.524

ETUDE FAITE AVEC UN PAS ENTRE VALEURS DE DELTA VALANT : 0.0050

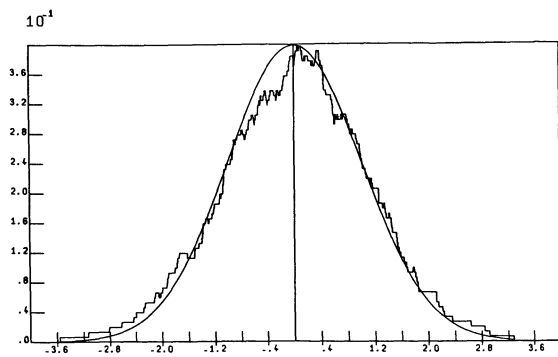
MAXIMUM RELATIF LE PLUS A DROITE : 0.7366 THETA : 0.62802

MAXIMUM ABSOLU : 0.7346 THETA : 0.62802

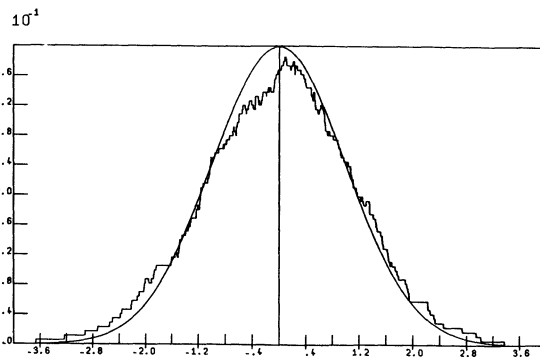
D0 EST ENTRE D = 0.5211E+00 (T = 0.1327E+00) ET D = 0.5206E+00 (T = 0.2205E+00)

LONGUEUR ESTIMEE DE LA FENETRE (METHODE NOMINAL) : 1.713

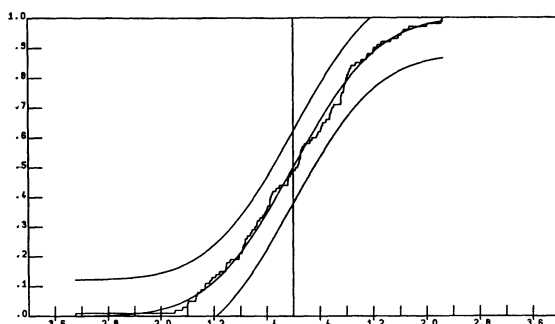
LONGUEUR ESTIMEE DE LA FENETRE (METHODE DEHEUVELS) : 1.713



DENSITE DE L ECHANTILLON N° 7 (DELTA OPTIMAL)



DENSITE DE L ECHANTILLON N° 7 (METHODE DEHEUVELS)

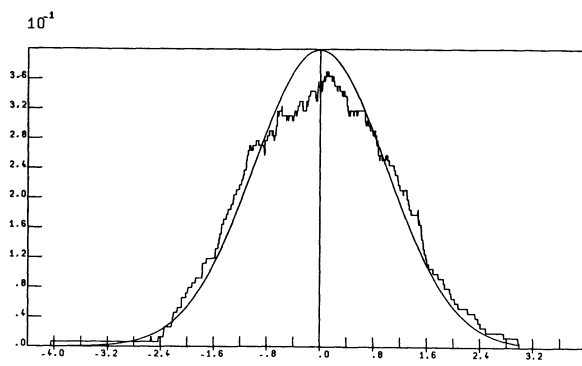


TEST DE KOLMOGOROFF SUR L ECHANTILLON N° 4

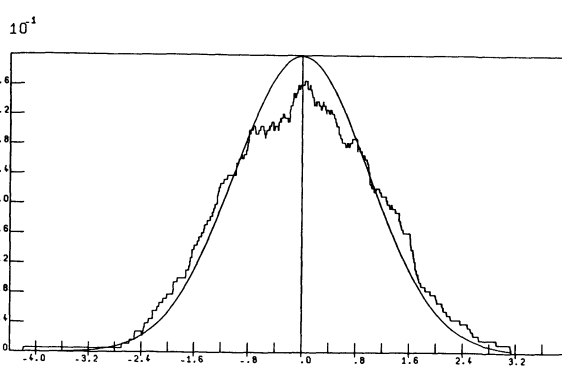
```

.....
*
*   ETUDE DES MAXIMA DE THETA DANS LE DOMAINE ADMISSIBLE (D),...
*   POUR LE NOYAU DE DEHEUVELS (D)
*
*   ECHANTILLON N° 4                      EFFECTIF : 100
*
*   VALEURS DE L ECHANTILLON
*
*   -3.289 -1.791 -1.675 -1.598 -1.591 -1.475 -1.460 -1.446 -1.389 -1.327
*   -1.327 -1.258 -1.230 -1.188 -1.114 -1.019 -1.011 -1.009 -.951 -.820
*   -.818 -.776 -.761 -.751 -.727 -.727 -.693 -.649 -.624 -.582
*   -.551 -.540 -.501 -.456 -.424 -.402 -.385 -.351 -.351 -.340
*   -.325 -.305 -.253 -.205 -.076 -.074 -.071 -.036 -.006 .038
*   .074 .090 .095 .111 .115 .115 .139 .176 .256 .284
*   .359 .373 .411 .421 .445 .494 .552 .552 .572 .586
*   .613 .715 .719 .721 .732 .761 .767 .771 .791 .793
*   .812 .833 .848 .879 .941 1.000 1.101 1.118 1.197 1.212
*   1.266 1.321 1.464 1.575 1.619 1.653 1.738 2.010 2.224 2.239
*
*   ETUDE FAITE AVEC UN PAS ENTRE VALEURS DE DELTA VALANT : .00050
*
*   MAXIMUM RELATIF LE PLUS A DROITE : .7717      THETA : .59337
*
*   MAXIMUM ABSOLU                      : .7067      THETA : .59650
*
*   ON EST ENTRE D = .4637E+00 (T = .2158E+00) ET D = .4632E+00 (T = .2347E+00)
*
*   LONGUEUR ESTIMEE DE LA FENETRE (METHODE NOMINAL) : 1.813
*   LONGUEUR ESTIMEE DE LA FENETRE (METHODE DEHEUVELS) : 1.803
*
.....

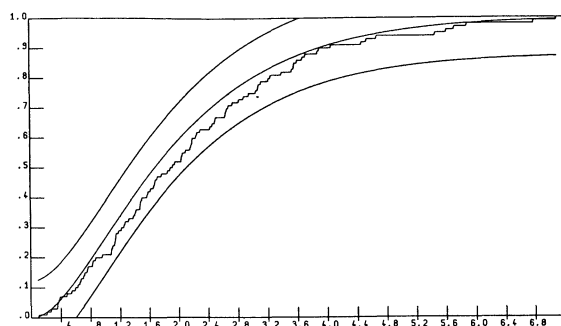
```



DENSITE DE L ECHANTILLON N° 4 (DELTA OPTIMAL)



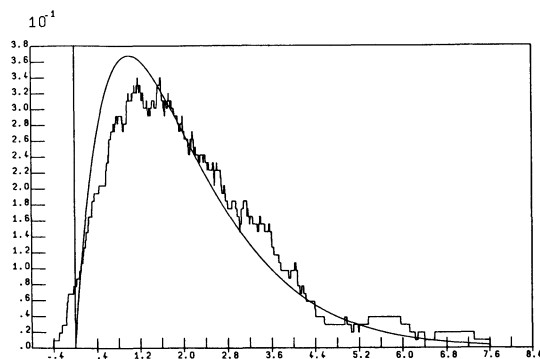
DENSITE DE L ECHANTILLON N° 4 (METHODE DEHEUVELS)



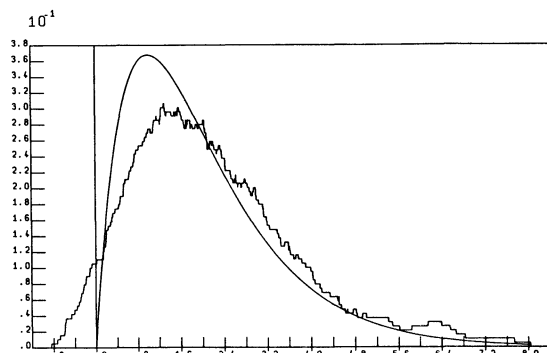
TEST DE KOLMOGOROFF SUR L ECHANTILLON GAM(2)

ETUDE DES MAXIMA DE THETA DANS LE DOMAINE ADMISSIBLE (D0... POUR LE NOYAU DE DEHEUELS (D)									
ECHANTILLON GAM(2)					EFFECTIF : 100				
VALEURS DE L ECHANTILLON									
.104	.216	.272	.355	.363	.371	.386	.490	.548	.596
.634	.664	.678	.710	.724	.760	.767	.833	.838	.875
.967	1.088	1.091	1.102	1.132	1.141	1.146	1.148	1.184	1.285
1.255	1.275	1.318	1.385	1.409	1.417	1.474	1.478	1.482	1.506
1.574	1.585	1.620	1.661	1.679	1.685	1.718	1.749	1.852	1.889
1.924	1.952	2.042	2.043	2.054	2.101	2.165	2.174	2.187	2.190
2.241	2.252	2.293	2.439	2.465	2.499	2.504	2.643	2.645	2.664
2.680	2.740	2.823	2.887	2.958	3.051	3.084	3.090	3.125	3.219
3.251	3.399	3.538	3.564	3.566	3.608	3.679	3.721	3.880	3.898
4.059	4.466	4.541	4.670	5.455	5.611	5.714	5.884	6.793	7.096

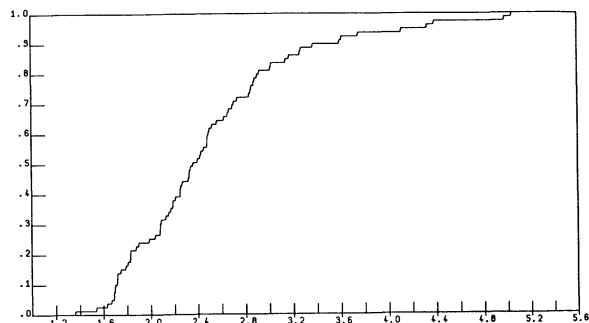
ETUDE FAITE AVEC UN PAS ENTRE VALEURS DE DELTA VALANT : .00050	
MAXIMUM RELATIF LE PLUS A DROITE : .7083	THETA : .56480
MAXIMUM ABSOLU : .6633	THETA : .56823
D0 EST ENTRE D = .5418E+00 (T = .1793E+00) ET D = .5413E+00 (T = -.2068E+00)	
LONGUEUR ESTIMEE DE LA FENETRE (METHODE MINIMAL) : 1.904	
LONGUEUR ESTIMEE DE LA FENETRE (METHODE DEHEUELS) : 1.893	



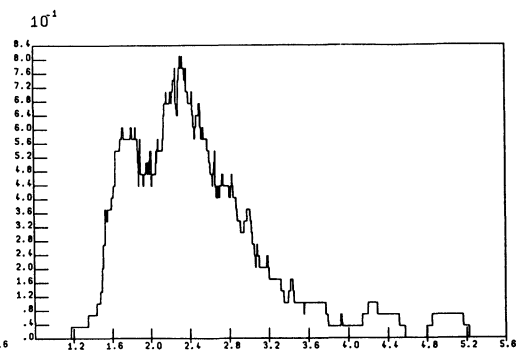
DENSITE DE L ECHANTILLON GAM(2) (DELTA OPTIMAL)



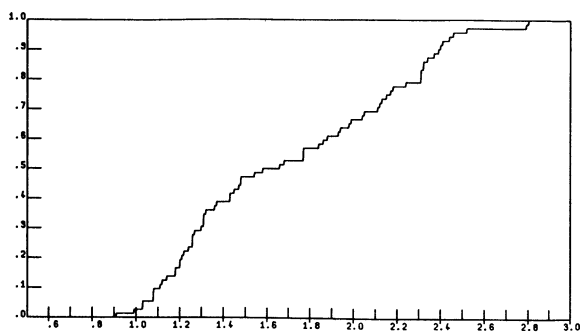
DENSITE DE L ECHANTILLON GAM(2) (METHODE DEHEUELS)



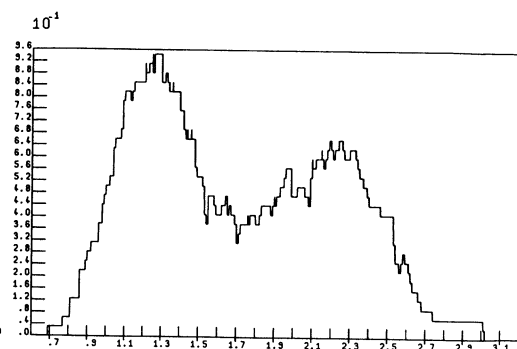
REPARTITION EMPIRIQUE DE L ECHANTILLON VAG(1)



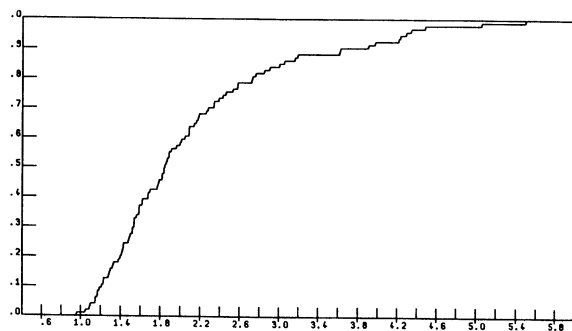
DENSITE DE L ECHANTILLON VAG(1) (METHODE DEHEUELS)



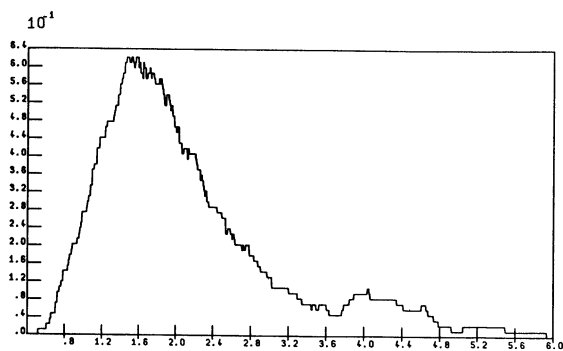
REPARTITIØN EMPIRIQUE DE L ECHANTILLØN VAG(2)



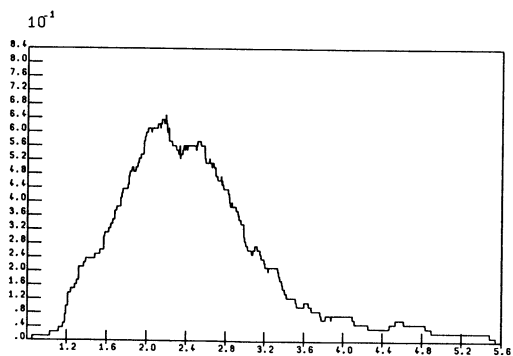
DENSITE DE L ECHANTILLØN VAG(2) (METHØDE DEHEUVELS)



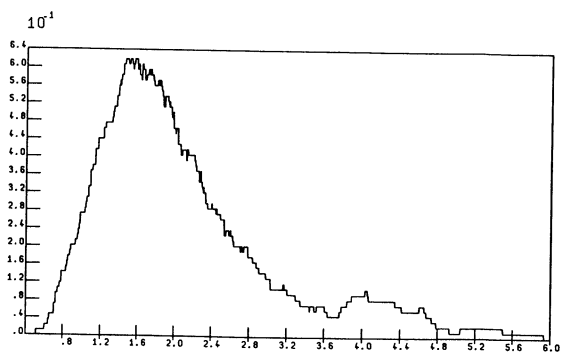
REPARTITIØN EMPIRIQUE DE L ECHANTILLØN VAG(3)



DENSITE DE L ECHANTILLØN VAG(3) (METHØDE DEHEUVELS)



DENSITE DE L ECHANTILLØN VAG(1) (METHØDE HØMNAL)



DENSITE DE L ECHANTILLØN VAG(3) (METHØDE HØMNAL)