REVUE DE STATISTIQUE APPLIQUÉE

J.-P. ASSELIN DE BEAUVILLE

Estimation non paramétrique de la densité et du mode exemple de la distribution Gamma

Revue de statistique appliquée, tome 26, nº 3 (1978), p. 47-70

<http://www.numdam.org/item?id=RSA_1978__26_3_47_0>

ESTIMATION NON PARAMÉTRIQUE DE LA DENSITÉ ET DU MODE EXEMPLE DE LA DISTRIBUTION GAMMA

J.-P. ASSELIN DE BEAUVILLE

Laboratoire d'Informatique Appliquée,
Université de Tours, Parc de Grandmont,
37 200 Tours

1. INTRODUCTION

Etant donné un échantillon x_i $(i = 1,\dots, n)$ issu d'une distribution inconnue, de densité de probabilité $f(x)$ bornée et continue sur R, comment obtenir une estimation de $f(x)$ à partir de la seule information contenue dans l'échantillon ?

Ce problème, que l'on désigne généralement par "estimation non paramétrique de la densité", a fait l'objet de nombreux travaux. (Pour une revue des différentes méthodes on pourra consulter [31], [32], [35] et [34]. Parmi les estimateurs proposés, la grande majorité d'entre eux se rangent dans une classe très importante constituée par les estimateurs bâtis à partir d'un noyau (ou estimateur à noyau général). Parmi les plus connus on peut citer :

a) L'estimateur de Parzen ([13]) et Rosenblatt ([15]) :

$$f_n(x) = (1/n\ \delta_n) \sum_{i=1}^{n} K\left(\frac{x - x_i}{\delta_n}\right)$$

où $K(y)$, le noyau, est une fonction mesurable sur R, non négative et bornée ; δ_n est le terme général d'une suite numérique positive.

Rosenblatt ([20]) a donné les conditions de convergence en moyenne quadratique de cet estimateur. P. Deheuvels ([21]) a, en outre, étudié la convergence ponctuelle presque sûre et uniforme presque sûre de $f_n(x)$.

b) L'estimateur séquentiel de H. Yamato ([22]) :

$$g_n(x) = (1/n) \sum_{i=1}^{n} (1/\delta_i)\, K\left(\frac{x - x_i}{\delta_i}\right)$$

Les conditions de convergence de $g_n(x)$ sont fournies dans ([22]) et ([21]).

Mots-Clés : **Estimation de la densité. Estimation à noyaux. Estimation du mode. Robustesse. Simulation.**

c) Les estimateurs séquentiels de P. Deheuvels ([23], [24], [25]) :

$$h_{1,n}(x) = (1/\sum_{i=1}^{n} \delta_i) \sum_{i=1}^{n} K\left(x, \frac{x - x_i}{\delta_i}\right)$$

$$h_{2,n}(x) = [\sum_{i=1}^{n} \delta_i H(\delta_i)]^{-1} \sum_{i=1}^{n} H(\delta_i) K\left(x, \frac{x - x_i}{\delta_i}\right)$$

où $K(x, y)$ est une fonction positive, bornée et mesurable sur R^2.

Ces deux estimateurs convergent en moyenne quadratique sous certaines hypothèses. On trouvera, en outre, dans ([21]) les conditions nécessaires et suffisantes de convergence p.s. et uniforme p.s.

d) L'estimateur séquentiel de G. Banon ([26]) :

$$1_n(x) = (1/b_n) \sum_{i=1}^{n} (\delta_i/b_i) \sum_{k=1}^{i} K\left(\frac{x - x_i}{\delta_i}\right)$$

où $b_n = \sum_{j=1}^{n} \delta_j$.

En plus des conditions de convergence en moyenne quadratique de $1_n(x)$, Banon défini un critère de comparaison des estimateurs de $f(x)$ basé sur le comportement asymptotique de la variance. Selon ce dernier critère, il montre que $1_n(x)$ est meilleur que $f_n(x)$ et $h_{1,n}(x)$.

Pour tous les estimateurs définis ci-dessus, il est nécessaire de définir un critère de choix permettant pour une densité $f(x)$ donnée, ou mieux pour une famille de densités, de définir la forme optimum de la fonction $K(.)$. On pourra, par exemple, chercher à minimiser une quantité proportionnelle à $E\int [f(x) - f_n(x)]^2 dx$.

Utilisant ce critère, Watson et Leadbetter ([19] ont déterminé les fonctions $K(.)$. optimales pour plusieurs densités $f(x)$ données. Ils ont, en particulier, mis en évidence deux inconvénients essentiels de cette première approche :

1) La forme de K dépend en grande partie de la densité $f(x)$ que l'on désire estimer. En d'autres termes, la méthode n'est pas robuste et si l'on possède suffisamment d'informations sur la forme de $f(x)$, on aura tout intérêt à utiliser une méthode d'estimation paramétrique (Méthode du maximum de vraisemblance par exemple).

2) L'expression de K peut, dans certains cas, être extrêmement compliquée.

On peut, toutefois, éviter ces écueils en estimant $f(x)$ sur un intervalle fini $[a, b]$ à partir d'un développement en série de fonctions orthogonales. L'estimateur sera donc défini par :

$$q_m(x) = \sum_{k=0}^{m} \hat{a}_k \varphi_k(x)$$

où les fonctions $\varphi_k(x)$, intégrables sur $[a, b]$ vérifient :

$$\int_a^b \varphi_k(x) \varphi_j(x) = \begin{cases} 0 \text{ si } k \neq j \\ 1 \text{ si } k = j. \end{cases}$$

Les coefficients $\hat{a}_k$ sont des estimateurs sans biais de :

$$a_k = \int_a^b f(x)\, \varphi_k(x)\, dx,$$

c'est-à-dire que :

$$\hat{a}_k = (1/n) \sum_{j=1}^{n} \varphi_k(x_j).$$

Sachant que, par hypothèse,

$$f(x) = \sum_{k=0}^{\infty} a_k\, \varphi_k(x)\,, \quad x \in [a, b]$$

il est clair que $q_m(x)$ est un estimateur biaisé de $f(x)$. Les résultats concernant le biais asymptotique et la convergence de $q_m(x)$ sont bien connus (on pourra consulter en particulier [10], [27], [28], [29], [30], [33].

Remarques :

1) On peut récrire $q_m(x)$ sous la forme suivante :

$$q_m(x) = (1/n) \sum_{j=1}^{n} \left[\sum_{k=0}^{m} \varphi_k(x_j)\, \varphi_k(x) \right].$$

Si on définit une fonction symétrique $K(x, y)$ telle que :

$$K(x, y) = \sum_{k=0}^{m} \varphi_k(x)\, \varphi_k(y) \text{ il vient :}$$

$$q_m(x) = (1/n) \sum_{j=1}^{n} K(x, x_j),$$

relation qui montre que $q_m(x)$ appartient en fait à la classe des estimateurs à noyau et qu'il peut donc être réexprimé sous la forme d'une somme pondérée de variables aléatoires indépendantes. En conséquence l'étude de $q_m(x)$ peut être faite à partir des théorèmes généraux régissant les estimateurs à noyau.

Il faut noter qu'ici le noyau $K(.,.)$ peut prendre des valeurs négatives.

2) La différence entre $q_m(x)$ et les estimateurs à noyau réside surtout dans leur utilisation pratique. En effet l'emploi de $q_m(x)$ ne nécessite, une fois choisie les fonctions $\varphi_k(x)$, que la détermination préalable du nombre m de termes et, comme on le verra plus loin, ce choix peut, dans certains cas au moins, être fait assez facilement, sans aucune information sur $f(x)$, à partir de la minimisation d'un critère adéquat. (Les théorèmes généraux justifiant l'existence d'une valeur optimum de m à partir du critère de minimisation de l'erreur quadratique intégrée (J) sont exposés dans [29]. On montre, en particulier, que $J(m)$ est une fonction monotone croissante pour m supérieur à sa valeur optimum). Cette remarque garantit la robustesse des estimateurs $q_m(x)$.

Par contre, l'emploi des estimateurs $f_n(x)$, $g_n(x)$, $h_{1,n}(x)$, $h_{2,n}(x)$ ou $1_n(x)$ ne peut se faire qu'après avoir choisi auparavant la fonction $K(.)$ optimale. Or, en l'absence de toute information sur $f(x)$, ce choix ne pourra être qu'arbitraire.

Ainsi, Kronmal, Tarter et Holcomb ([10], [16]) estiment f(x) à partir d'un développement en série de Fourier. Ils proposent trois estimateurs :

$$f_m(x) = c_o/2 + \sum_{k=1}^{m} \bar{c}_k \cos k\pi\,[(x - a)/(b - a)]$$

$$f_m^{(1)}(x) = \sum_{k=1}^{m} \bar{s}_k \sin k\pi\,[(x - a)/(b - a)]$$

$$f_m^{(2)}(x) = [f_m(x) + f_m^{(1)}(x)]/2.$$

Dans ces formules $\bar{c}_k$ et $\bar{s}_k$ représentent les moments trigonométriques de l'échantillon, c'est-à-dire :

$$\bar{c}_k = (c_o/n) \sum_{i=k}^{n} \cos k\pi\,[(x_i - a)/(b - a)]$$

$$\bar{s}_k = (c_o/n) \sum_{i=1}^{n} \sin k\pi\,[(x_i - a)/(b - a)]$$

$$c_o = 2/(b - a).$$

a et b sont deux constantes arbitraires telles, par exemple, que : $\forall\, i, a \leqslant x_i \leqslant b$.

Le nombre m de termes intervenant dans les expressions $f_m(x)$, $f_m^{(1)}(x)$ et $f_m^{(2)}(x)$ est déterminé par la minimisation de l'erreur quadratique moyenne intégrée.

L'inconvénient essentiel de ce type d'estimateur réside dans le fait que les fonctions $f_m(x)$, $f_m^{(1)}(x)$ et $f_m^{(2)}(x)$ peuvent prendre des valeurs négatives sur [a, b]. Cet inconvénient, signalé par Kronmal et Tarter, se produit très rarement en pratique dès que n est suffisamment grand mais sa probabilité d'apparition augmente, plus ou moins suivant f(x), lorsque n diminue ou lorsque l'étendue de l'échantillon est petite devant (b – a). Cette possible négativité de $f_m(x)$ limite pratiquement l'emploi de ces estimateurs à l'étude des très grands échantillons.

On peut contourner cette difficulté en remplaçant la série de Fourier par une série de Féjer. On sait, en effet, que la somme de Féjer correspondante est toujours positive ou nulle sur l'intervalle [a, b] considéré, quelle que soit la taille a et l'étendue de l'échantillon.

L'un des buts de cet article est, précisément, de montrer de quelle façon, en pratique, utiliser un tel estimateur.

Dans le paragraphe 2 on rappellera (en les démontrant dans le cas particulier de l'estimateur de Féjer) les résultats théoriques nécessaires. On vérifiera par exemple (formule (10) ci-dessous) que cet estimateur entre bien dans la catégorie des estimateurs à noyau avec $K(u) = \dfrac{1}{h(b - a)}\left[\dfrac{\sin hu/2}{\sin u/2}\right]^2$. On formulera, en outre, une règle d'obtention de la valeur optimum de m.

Dans le paragraphe 3 on étudiera le problème de l'estimation non paramétrique du mode x_o de f(x). On exposera deux façons d'aborder ce problème :

La première consiste à prendre pour mode la valeur de x qui rend maximum l'estimation de la densité f(x).

La seconde estime directement x_o en exploitant l'idée que, dans l'échantillon, on doit observer un groupement des valeurs dans le voisinage du mode.

Au paragraphe 4 on applique les résultats précédents à l'estimation du mode de la distribution gamma et on compare (par simulation) les résultats fournis par les deux méthodes précédentes avec ceux obtenus par l'application de la méthode d'estimation paramétrique des moments.

2. ESTIMATION DE LA DENSITE DE PROBABILITE

La méthode consiste à estimer, non pas la fonction de densité $f(x)$ elle-même, mais une fonction $g(x)$ telle que :

$$g(x) = \begin{cases} f(x), \forall\ x \in [a, b] \\ 0, \text{ partout ailleurs} \end{cases} \tag{1}$$

Le choix des constantes a et b permettra d'estimer la partie de la fonction $f(x)$ correspondant à : $a \leqslant x \leqslant b$. Il faut noter que $g(x)$ ne sera pas, en général, une fonction de densité de probabilité mais qu'elle sera continue sur $[a, b]$.

L'estimateur considéré est défini par :

$$g_m(x) = c_o/2 + \sum_{k=1}^{m} \left(1 - \frac{k}{m+1}\right) [\bar{c}_k \cos(k\, z(x)) + \bar{s}_k \sin(k\, z(x))] \tag{2}$$

où

$$z(x) = 2\pi(x - a)/(b - a) \tag{3}$$

$$\bar{c}_k = (c_o/n) \sum_{i=1}^{n} \cos(k\, z_i) \tag{4}$$

$$\bar{s}_k = (c_o/n) \sum_{i=1}^{n} \sin(k\, z_i) \tag{5}$$

$$c_o = 2/(b - a) \tag{6}$$

$$z_i = 2\pi(x_i - a)/(b - a). \tag{7}$$

a et b sont deux constantes arbitraires telles que :

$$\forall\ i = 1, 2, \ldots, n \text{ on ait } a \leqslant x_i \leqslant b$$

ou, d'une façon plus générale, telles que certaines valeurs x_i soient supérieures à b ou inférieures à a. Dans ce dernier cas, seules les x_i comprises entre a et b interviendront dans le calcul des moments trigonométriques $\bar{c}_k$ et $\bar{s}_k$. Dans tout ce qui suit on supposera, pour simplifier que, quel que soit i, $a \leqslant x_i \leqslant b$, et on écrira z pour $z(x)$. $g_m(x)$ est la somme de Féjer qui correspond au développement de Fourier suivant :

$$h_m(x) = c_o/2 + \sum_{k=1}^{m} [\bar{c}_k \cos(kz) + \bar{s}_k \sin(kz)]. \tag{8}$$

Désignons par c_k et s_k les coefficients de Fourier du développement de $g(x)$ soit :

$$c_k = c_o \int_a^b g(x) \cdot \cos(kz)\, dx$$
$$s_k = c_o \int_a^b g(x) \cdot \sin(kz) \cdot dx,$$

toutes les intégrales étant prises entre a et b, on remplacera dans tout le texte $\int_a^b$ par $\int$.

On a alors $E\,\bar{c}_k = c_k$ et $E\,\bar{s}_k = s_k$.

On sait, d'autre part, par le théorème de Féjer (cf. [17] p. 181) que :

$$\lim_{m \to \infty} E\, g_m(x) = g(x) \text{ (convergence uniforme).} \tag{9}$$

Il est alors facile de déduire de la relation (9) que $g_m(x)$ est, en général, un estimateur biaisé de $g(x)$. En effet :

$$g(x) - E\, g_m(x) = \sum_{k=m+1}^{\infty} \left(1 - \frac{k}{m+1}\right)[c_k \cos(kz) + s_k \sin(kz)].$$

THEOREME 1 : *$g_m(x)$ est toujours positif ou nul.*

En effet, en remplaçant $\bar{c}_k$ et $\bar{s}_k$ par leur définition dans (2) on obtient :

$$g_m(x) = c_o\,[1/2 + (1/n) \sum_{i=1}^{n} \sum_{k=1}^{m} \left(1 - \frac{k}{m+1}\right) \cos k\, v_i]$$

où

$$v_i = 2\pi(x_i - x)/(b-a).$$

Or il est facile de montrer, en multipliant par $2\sin(v_i/2)$ et en utilisant les formules classiques de trigonométrie que :

$$2(m+1)\sin(v_i/2) \sum_{k=1}^{m} \left(1 - \frac{k}{m+1}\right) \cos(k\, v_i) =$$
$$= -m \sin(v_i/2) + \sum_{k=1}^{m} \sin[(k+1/2)\, v_i].$$

Par le même procédé, appliqué à la somme de sinus, on obtient finalement :

$$2(m+1)\sin^2(v_i/2) \sum_{k=1}^{m} \left(1 - \frac{k}{m+1}\right) \cos(k\, v_i) =$$
$$= \sin^2[(m+1)\, v_i/2] - (m+1)\sin^2(v_i/2).$$

Il vient donc pour $g_m(x)$:

$$g_m(x) = \frac{c_o}{2n(m+1)} \sum_{i=1}^{n} \left(\frac{\sin[(m+1)\, v_i/2]}{\sin(v_i/2)}\right)^2 \tag{10}$$

ce qui démontre le théorème 1.

THEOREME 2 : *Si m est d'ordre inférieur à* $\sqrt{n}$ $(m = o(\sqrt{n}))$ *on a alors :*

1) $\lim\limits_{\substack{n \to \infty \\ m \to \infty}} \operatorname{Var} g_m(x) = 0$

2) $\lim\limits_{\substack{n \to \infty \\ m \to \infty}} J[g_m(x)] = 0$

avec $J[g_m(x)] = c_o E \int [g_m(x) - g(x)]^2 dx$.

La démonstration de ce théorème est analogue à celle donnée dans [10] : En partant de l'expression (10) on déduit que :

$$\operatorname{Var} g_m(x) = \left[\frac{c_o}{2n(m+1)}\right]^2 \sum_{i=1}^{n} \operatorname{Var}\left[\frac{[\sin(m+1)v_i/2]}{\sin(v_i/2)}\right]^2,$$

soit :

$$\operatorname{Var} g_m(x) = \frac{1}{n} \cdot \left[\frac{c_o}{2(m+1)}\right]^2 \operatorname{Var}\left[\frac{\sin[(m+1)v_i/2]}{\sin(v_i/2)}\right]^2.$$

Par définition de la variance on sait que :

$$\operatorname{Var}\left[\frac{\sin[(m+1)v_i/2]}{\sin(v_i/2)}\right]^2 \leqslant E\left[\frac{\sin[(m+1)v_i/2]}{\sin(v_i/2)}\right]^4$$

Donc :

$$\operatorname{Var}\left[\frac{\sin[(m+1)v_i/2]}{\sin(v_i/2)}\right]^2 \leqslant \int f(y)\left[\frac{\sin[\pi(m+1)(y-x)/(b-a)]}{\sin[\pi(y-x)/(b-a)]}\right]^4 dy$$

Puisque, d'autre part :

$$\frac{\sin[(m+1)t]}{\sin t} \leqslant m+1$$

il en résulte que :

$$\operatorname{Var} g_m(x) \leqslant (m+1)^2/[n(b-a)^2] \int f(y)\, dy$$

ce qui prouve la première partie du théorème 2.

La seconde partie de ce théorème peut se démontrer de la façon suivante :

$$E[g_m(x) - g(x)]^2 = E\left\{ g_m(x) - E\, g_m(x) - \sum_{k=m+1}^{\infty}\left(1 - \frac{k}{m+1}\right)[c_k \cos(kz) + s_k \sin(kz)]\right\}^2$$

$$\operatorname{Var} g_m(x) - 2E[g_m(x) - Eg_m(x)] \sum_{k=m+1}^{\infty} [c_k \cos(kz) + s_k \sin(kz)] + \left\{\sum_{k=m+1}^{\infty}\left(1 - \frac{k}{m+1}\right)[c_k \cos(kz) + s_k \sin(kz)]\right\}^2.$$

On vient de montrer que Var $g_m(x) \to 0$ lorsque m et n tendent vers l'infini.

Le second terme de la somme est évidemment nul puisque

$$E[g_m(x) - E\, gm(x)] = 0.$$

Le troisième terme tend lui aussi vers zéro puisque les coefficients de Fourier de g(x) tendent vers zéro lorsque $k \to \infty$ (cf. [17] p. 181-2).

Finalement $\lim\limits_{\substack{m\to\infty \\ n\to\infty}} E[g_m(x) - g(x)]^2 = 0$. En outre, la convergence de $E[g_m(x) - g(x)]^2$ étant uniforme :

$$\lim_{\substack{m\to\infty \\ n\to\infty}} J\,[g_m(x)] = \lim_{\substack{m\to\infty \\ n\to\infty}} c_o\, E \int [g_m(x) - g(x)]^2\, dx = 0\,,$$

ce qui termine la démonstration du théorème 2.

L'intérêt majeur de ce théorème est que la condition $m = o(\sqrt{n})$ implique que pratiquement le nombre de termes m qui intervient dans $g_m(x)$ doit toujours être relativement petit par rapport à la taille n de l'échantillon.

2.2. Calcul de $J[g_m(x)]$

On part de la définition de J :

$$J\,[g_m(x)] = c_o\, E \int [g(x) - g(x)]^2\, dx\,.$$

En remplaçant $g_m(x)$ par son expression (2) et en tenant compte des relations :

$$E\,\bar{c}_k = c_k \;\; ; \; E\,\bar{s}_k = s_k$$

$$\int \sin(kz)\,.\cos(jz)dx = 0$$

$$\int \cos(kz)\cos(jz)\,dx = \begin{cases} 0 & \text{si } k \neq j \\ 1/c_o & \text{si } k = j \end{cases}$$

$$\int \sin(kz)\sin(jz)\,dx = \begin{cases} 0 & \text{si } k \neq j \\ 1/c_o & \text{si } k = j \end{cases}$$

$$\int \cos(kz)\,dx = \int \sin(kz)\,dx = 0,$$

On obtient :

$$J\,[g_m(x)] = c_o \int g^2(x)\,dx - \frac{c_o^2}{2} + \sum_{k=1}^{m} \left(1 - \frac{k}{m+1}\right)^2 \left[\mathrm{Var}\,\bar{c}_k + \mathrm{Var}\,\bar{s}_k - \frac{m+1+k}{m+1-k}(c_k^2 + s_k^2)\right] \quad (11)$$

Remarque :

On sait par le théorème de Féjer que :

$$g(x) = c_o/2 + \sum_{k=1}^{\infty} \left(1 - \frac{k}{m+1}\right) [c_k \cos(kz) + s_k \sin(kz)].$$

En portant cette expression de g(x) dans $\int g^2(x)\,dx$ on obtient la formule de Parseval :

$$c_o \int g^2(x)\,dx = c_o^2/2 + \sum_{k=1}^{\infty} (c_k^2 + s_k^2).$$

THEOREME 3 : *L'erreur relative à l'emploi des séries de Féjer est supérieure à celle résultant de l'utilisation des séries de Fourier, c'est-à-dire que :*

$J[g_m(x)] > J[h_m(x)]$ où $h_m(x)$ est définie par la relation (8).

Démonstration :

L'emploi de l'estimateur de Fourier (8) conduit à une erreur quadratique :

$$J[h_m(x)] = c_o \int g^2(x)\,dx - c_o^2/2 + \sum_{k=1}^{m} [(\mathrm{Var}\,\bar{c}_k - c_k^2) + (\mathrm{Var}\,\bar{s}_k - s_k^2)].$$

On en déduit que :

$$J[g_m(x)] - J[h_m(x)] = \sum_{k=1}^{m} \{(\mathrm{Var}\,\bar{c}_k + \mathrm{Var}\,\bar{s}_k)\left[\left(1 - \frac{k}{m+1}\right)^2 - 1\right] + (c_k^2 + s_k^2)\left(\frac{k}{m+1}\right)^2\}.$$

En utilisant les formules (13) et (15) ci-dessous on montre que :

$\mathrm{Var}\,\bar{c}_k + \mathrm{Var}\,\bar{s}_k = (1/n)[c_o^2 - (c_k^2 + s_k^2)]$ et il vient donc :

$$J[g_m(x)] - J[h_m(x)] = \sum_{k=1}^{m} \left\{ \frac{c_o^2}{n}\left[\left(\frac{k}{m+1}\right)^2 - \left(\frac{2k}{m+1}\right)\right] + \frac{c_k^2 + s_k^2}{n(m+1)^2}[(n-1)k^2 + 2k(m+1)] \right\}$$

Pour démontrer le théorème il suffit de prouver que $J[g_m(x)] - J[h_m(x)] > 0$ ou encore que chaque terme de la somme est positif. Il revient au même d'étudier le signe de :

$$B_k = c_o^2[k^2 - 2k(m+1)] + (c_k^2 + s_k^2)[(n-1)k^2 + 2k(m+1)].$$

Le premier terme de B_k est négatif puisque :

$$k^2 - 2k(m+1) = [k-(m+1)]^2 - (m+1)^2 < 0.$$

Il est plus simple d'étudier le signe de $-B_k$:

$$-B_k = c_o^2[2k(m+1) - k^2] - (c_k^2 + s_k^2)[(n-1)k^2 + 2k(m+1)]$$

Les deux termes de $-B_k$ étant > 0, le signe de $-B_k$ sera le même que celui de l'expression obtenue en majorant le second terme.

Or l'inégalité $\mathrm{Var}\,\bar{c}_k + \mathrm{Var}\,\bar{s}_k \geqslant 0$ se traduit par $c_k^2 + s_k^2 \leqslant c_o^2$. En remplaçant $(c_k^2 + s_k^2)$ par sa borne supérieure dans $-B_k$ on obtient finalement : $-n\,k^2\,c_o^2$.

Cette quantité étant toujours négative on en déduit qu'à fortiori $-B_k < 0$ et donc que $B_k > 0$.

Finalement, tous les termes de la somme $J[g_m(x)] - J[h_m(x)]$ étant > 0, le théorème 3 est démontré.

Ainsi la généralisation de la méthode de Fourier s'accompagne d'une diminution d'efficacité en terme d'erreur moyenne quadratique intégrée.

2.3. Calcul des variances et covariances des $\bar{c}_k$ et $\bar{s}_k$

Il est possible dans (11) d'expliciter les expressions de Var $\bar{c}_k$ et Var $\bar{s}_k$ en fonction des coefficients de Fourier c_k et s_k.

On sait que $\mathrm{cov}(\bar{c}_k, \bar{c}_j) = E(\bar{c}_k \bar{c}_j) - c_k c_j$, or :

$$E(\bar{c}_k \bar{c}_j) = (c_o^2/n^2)\, E\left[\left(\sum_{i=1}^{n} \cos(k z_i)\right)\left(\sum_{i=1}^{n} \cos(j z_i)\right)\right].$$

En développant il vient :

$$E(\bar{c}_k \bar{c}_j) = (c_o^2/n^2)\, E\left[\sum_{i=1}^{n} \cos(k z_i) \cos(j z_i) + \sum_{i=1}^{n} \sum_{\substack{1=1 \\ 1 \neq i}}^{n} \cos(k z_i) \cos(j z_1)\right],$$

$$E(\bar{c}_k \bar{c}_j) = (c_o^2/2n)\, E[\cos(k+j)Z +$$

$$+ \cos(k-j)Z] + ((n-1)\, c_o^2/n)\, E[\cos(kZ)]\, E[\cos(jZ)].$$

Par définition de l'espérance mathématique :

$E[\cos(kZ)] = \int g(x) \cos[2k\pi(x-a)/(b-a)]\, dx = c_k/c_o$, puisqu'on ne tient compte que des valeurs échantillonnées x_i, de la variable aléatoire X, comprises entre a et b.

De même :

$$E[\cos(k+j)Z] = c_{k+j}/c_o$$

$$E[\cos(k-j)Z] = c_{k-j}/c_o.$$

Finalement :

$$E(\bar{c}_k \bar{c}_j) = (c_o/2n)(c_{k+j} + c_{k-j}) + (n-1)\, c_k c_j/n.$$

Il s'en suit que :

$$\mathrm{cov}(\bar{c}_k, c_j) = (1/n)\, [(c_o/2)(c_{k+j} + c_{k-j}) - c_k c_j] \tag{12}$$

$$\mathrm{Var}\, \bar{c}_k = (1/n)\, [c_o/2)(c_o + c_{2k}) - c_k^2]. \tag{13}$$

Par un raisonnement analogue on obtient :

$$\mathrm{cov}(\bar{s}_k, \bar{s}_j) = (1/n)\, [(c_o/2)(c_{j-k} - c_{j+k}) - s_k s_j] \tag{14}$$

$$\mathrm{Var}\, \bar{s}_k = (1/n)\, [(c_o/2)(c_o - c_{2k}) - s_k^2]. \tag{15}$$

Dans la suite on aura besoin de la covariance entre $\bar{c}_j$ et $\bar{s}_k$ soit :

$$\mathrm{cov}(\bar{c}_j, \bar{s}_k) = (1/n)\, [c_o/2)(s_{k+j} + s_{k-j}) - s_k c_j] \tag{16}$$

2.4. Détermination du nombre optimum de termes m

L'emploi de (2) pour estimer f(x) n'est possible qu'après avoir déterminé le nombre de termes m de la somme. Il est naturel de choisir m de sorte que l'erreur J soit minimum. D'autre part, compte-tenu du résultat trouvé au théorème 2, on aura toujours intérêt à donner à m une valeur assez faible. La règle adoptée pour déterminer la valeur optimum de m repose sur les deux remarques précédentes : A partir de $m = 1$ on augmente la valeur de m d'une unité jusqu'à ce que J augmente. On donne alors à m la valeur qui précède juste l'augmentation de J. Dans le cas où J(m) serait stationnaire ou croissante on prendrait $m = 1$. Pratiquement, on observe que si, pour une valeur donnée de n, J(m) possède un minimum, il est en général toujours voisin de $m = 1$. Ainsi il suffit de tester une quinzaine de valeurs de m pour être à peu près certain de détecter le minimum de J s'il existe. Donc on ajoutera à la somme (2) le $m^{\text{ième}}$ terme si et seulement si $\Delta J_m = J[g_m(x)] - J[g_{m-1}(x)]$ est négatif ou nul. Or on peut déduire de (11) que :

$$\Delta J_m = \left(1 - \frac{m}{m+1}\right)^2 [\text{Var}\,\bar{c}_m + \text{Var}\,\bar{s}_m - (2m+1)(c_m^2 + s_m^2)]. \quad (17)$$

Il suffit donc de tester la quantité :

$$\Delta_m = \text{Var}\,\bar{c}_m + \text{Var}\,\bar{s}_m - (2m+1)(c_m^2 + s_m^2),$$

soit, en utilisant (13) et (15) :

$$\Delta_m = (1/n)\,[c_o^2 - (c_m^2 + s_m^2)(2mn + n + 1)]. \quad (18)$$

En général on ignore la fonction f(x) et il n'est pas possible de calculer les coefficients de Fourier c_k et s_k. On remédie à cet inconvénient en substituant à l'expression (18) un estimateur sans biais de Δ_m.

Considérons les statistiques :

$$\hat{\text{Var}}\,\bar{c}_m = [1/n(n-1)] \sum_{i=1}^{n} [c_o \cos(m z_i) - \bar{c}_m]^2$$

$$\hat{\text{Var}}\,\bar{s}_m = [1/n(n-1)] \sum_{i=1}^{n} [c_o \sin(m z_i) - \bar{s}_m]^2$$

On sait que $E(\hat{\text{Var}}\,\bar{c}_m) = \text{Var}\,\bar{c}_m$ et $E(\text{Var}\,\bar{S}_m) = \text{Var}\,\bar{s}_m$.

Donc :

$$E(\bar{c}_m^2 + \bar{s}_m^2 - \hat{\text{Var}}\,\bar{c}_m - \hat{\text{Var}}\,\bar{s}_m) = E\,\bar{c}_m^2 + E\,\bar{s}_m^2 - E\,\bar{c}_m^2 + (E\,\bar{c}_m)^2 - E\,\bar{s}_m^2 + (E\,\bar{s}_m)^2 = c_m^2 + s_m^2.$$

D'autre part :

$$\hat{\text{Var}}\,\bar{c}_m + \hat{\text{Var}}\,\bar{s}_m = [1/(n-1)]\,[c_o^2 - (\bar{c}_m^2 + \bar{s}_m^2)].$$

Finalement l'expression :

$$\hat{\Delta}_m = (1/n)\,[c_o^2 - (2mn + n + 1)(n\,\bar{c}_m^2 + n\,\bar{s}_m^2 - c_o^2)/(n-1)]$$

est un estimateur sans biais de Δ_m.

En résumé, on ajoutera le $n^{ième}$ terme si et seulement si $\hat{\Delta}_m \leqslant 0$ soit, sous une forme équivalente, si : $(\bar{c}_m^2 + \bar{s}_m^2)/c_o^2 \geqslant (2m+1)/(2mn+n+1)$ (19) (on notera que le second membre de l'inégalité est voisin de 1/n).

3. ESTIMATION DU MODE

L'estimation non paramétrique du mode x_o de f(x) est la seule façon de procéder lorsqu'on ne dispose d'aucune information à priori sur la densité elle-même. Schématiquement on peut distinguer deux méthodes d'estimation non paramétrique :

La première (méthode indirecte) consiste à obtenir dans un premier temps une estimation f(x) de la densité et à prendre pour mode de valeur de x pour laquelle f(x) est maximum (cf. [13] par exemple).

La seconde (méthode directe) procède d'emblée à l'estimation de x_o en se basant sur le fait que, dans l'échantillon, on doit observer un groupement des valeurs dans le voisinage du mode (cf. [2], [4], [6], [7], [14] [18] par exemple).

3.1. Méthode indirecte

Soit X_{01} la variable aléatoire définie sur [a, b] par :

$$g_m(\hat{x}_{01}) = \sup_{a \leqslant x \leqslant b} g_m(x).$$

$\hat{x}_{01}$ représente le mode associé à l'échantillon par la fonction $g_m(x)$. Supposons que le mode x_o de f(x) est unique et qu'il appartient à [a, b] on a :

$$f(x_o) = \sup_x f(x) = \sup_{a \leqslant x \leqslant b} g(x) = g(x_o).$$

THEOREME 4 : *X_{01} est asymptotiquement convergent.*

C'est-à-dire que :

$$\lim_{\substack{m \to \infty \\ n \to \infty}} P[\,|x_o - \hat{x}_{01}| \geqslant \epsilon] = 0, \forall\, \epsilon > 0.$$

Démonstration :

f(x) étant continue par hypothèse, il en résulte que g(x) est uniformément continue sur [a, b]. Donc, ϵ étant un nombre positif donné, il existe un nombre $\eta > 0$, tel que si $|x_o - x| < \eta$, il en résulte toujours $|g(x_o) - g(x)| < \epsilon$.

Le mode x_o étant supposé unique, pour prouver que $\hat{x}_{01} \to x_o$ en probabilité il suffira donc de montrer que $g(\hat{x}_{01}) \to g(x_o)$ en probabilité.

Or :

$$|g(\hat{x}_{01}) - g(x_o)| \leqslant |g(\hat{x}_{01}) - g_m(\hat{x}_{01})| + |g_m(\hat{x}_{01}) - g(x_o)|.$$

Par définition du mode x_o :

$$|g(\hat{x}_{01}) - g_m(\hat{x}_{01})| \leqslant |g(x_o) - g_m(\hat{x}_{01})|.$$

D'autre part :

$$|g_m(\hat{x}_{01}) - g(x_o)| \leqslant \operatorname*{Sup}_x |g_m(x) - g(x)|,$$

Donc :

$$|g(\hat{x}_{01}) - g(x_o)| \leqslant 2 \operatorname*{Sup}_x |g_m(x) - g(x)|. \qquad (20)$$

Ainsi pour montrer que $g(\hat{x}_{01}) \to g(x_o)$ en probabilité il suffira de montrer que Sup $|g_m(x) - g(x)| \to 0$ en probabilité.

Or on sait que :

$$g_m(x) - E\, g_m(x) = \sum_{k=1}^{m} \left(1 - \frac{k}{m+1}\right)[(\bar{c}_k - c_k) \cos(kz) + (\bar{s}_k - s_k) \sin(kz)].$$

D'où on tire :

$$|g_m(x) - E\, g_m(x)| \leqslant \sum_{k=1}^{m} [\,|\left(1 - \frac{k}{m+1}\right) (\bar{c}_k - c_k) \cos(kz)| +$$

$$+ |\left(1 - \frac{k}{m+1}\right) . (\bar{s}_k - s_k) \sin(kz)|\,],$$

$$\text{et } \operatorname*{Sup}_x |g_m(x) - Eg_m(x)| \leqslant \sum_{k=1}^{m} [\operatorname*{Sup}_x |\left(1 - \frac{k}{m+1}\right) (\bar{c}_k - c_k) \cos(kz)| +$$

$$+ \operatorname*{Sup}_x |\left(1 - \frac{k}{m+1}\right) (\bar{s}_k - s_k) \sin(kz)|\,]$$

Mais on a aussi :

$$\operatorname*{Sup}_x |\left(1 - \frac{k}{m+1}\right) (\bar{c}_k - c_k) \cos(kz(x))| = |\left(1 - \frac{k}{m+1}\right) (\bar{c}_k - c_k)|$$

$$\operatorname*{Sup}_x |\left(1 - \frac{k}{m+1}\right) (\bar{s}_k - s_k) \sin(kz(x))| = |\left(1 - \frac{k}{m+1}\right) (\bar{s}_k - s_k)|.$$

Il vient donc :

$$\operatorname*{Sup}_x |g_m(x) - E\, g_m(x)| \leqslant \sum_{k=1}^{m} [\left(1 - \frac{k}{m+1}\right) (|\bar{c}_k - c_k| + |\bar{s}_k - s_k|)] \leqslant$$

$$\leqslant \sum_{k=1}^{m} (|\bar{c}_k - c_k| + |\bar{s}_k - s_k|).$$

D'autre part :

$$E\,[\operatorname*{Sup}_x |g_m(x) - E\, g_m(x)|)^2] \leqslant E\,[\sum_{k=1}^{m} (|\bar{c}_k - c_k|^2 +$$

$$+ |\bar{s}_k - s_k|^2 + 2\,|\bar{c}_k - c_k| . |\bar{s}_k - s_k|)] \leqslant$$

$$\leqslant \sum_{k=1}^{m} [\operatorname{Var} \bar{c}_k + \operatorname{Var} \bar{s}_k + 2\,|\operatorname{cov}(\bar{c}_k, \bar{s}_k)|].$$

Les relations (13), (15) et (16) permettent d'écrire :

$$\lim_{n\to\infty} \operatorname{Var} \bar{c}_k = \lim_{n\to\infty} \operatorname{Var} \bar{s}_k = \lim_{n\to\infty} |\operatorname{cov}(\bar{c}_k, \bar{s}_k)| = 0,$$

d'où :

$$\lim_{n\to\infty} \{E [(\sup_x |g_m(x) - E\, g_m(x)|)^2]\}^{1/2} = 0.$$

La relation (9) finalement permet de poser :

$$\lim_{\substack{m\to\infty \\ n\to\infty}} \{E [(\sup_x |g_m(x) - g(x)|)^2]\}^{1/2} = 0 \text{ ou de façon équivalente :}$$

$$\lim_{\substack{m\to\infty \\ n\to\infty}} P [\sup_x |g_m(x) - g(x)| < \epsilon] = 1, \forall\, \epsilon > 0.$$

Le théorème 4 est ainsi démontré.

La convergence asymptotique de $\hat{x}_{01}$ rend possible, le calcul approximatif, pour les grands échantillons, de Var X_{01}. Pour y parvenir on aura besoin du résultat suivant :

Soit $L(Y_1, Y_2, \ldots, Y_k)$ une fonction de k variables aléatoires telles que $E\, Y_i = \theta_i$ $(i = 1,2, \ldots, k)$ et que les variances et covariances soient d'ordre $1/n$. Le développement en série de Taylor de $L(.)$ autourde $(\theta = \theta_1, \theta_2, \ldots, \theta_k)$ donne :

$$L(Y_1, \ldots, Y_k) = L(\theta) + \sum_{i=1}^{k} L_i'(\theta)\,(y_i - \theta_i) + 0(1/n) \tag{21}$$

où $L_i'(\theta) = \partial L(.)/\partial Y_i$ évaluée au point θ et où $0(1/n)$ représente des termes d'ordre inférieur ou égal à $1/n$.

Si tous les $L_i'(\theta)$ ne sont pas nuls on a donc :

$$\operatorname{Var} L(.) = E[L(.) - E\, L(.)]^2 =$$
$$= \sum_{i=1}^{k} [L_i'(\theta)]^2 \operatorname{Var} Y_i + \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j\neq 1}}^{k} L_i'(\theta) \,.\, L_j'(\theta) \operatorname{cov}(Y_i, Y_j) + o(1/n).(22)$$

Revenons au cas de l'estimateur $\hat{X}_{01}$; par un développement de $g_m'(x)$ on obtient :

$$g_m'(\hat{x}_{01}) = 0 = g_m'(x_o) + (\hat{x}_{01} - x_o)\, g_m''(x_o) + \ldots$$

Si la taille n de l'échantillon est suffisamment grande on pourra poser :

$$\hat{x}_{01} - x_o = - g_m'(x_o)/g_m''(x_o), \tag{23}$$

et puisque le théorème 4 assure que pour n suffisamment grand $\hat{x}_{01}$ est voisin de x_o on peut supposer que $g_m''(x_o) < 0$. La relation (23) peut alors s'écrire :

$$\hat{x}_{01} - x_o = g_m'(x_o)/G_m''(x_o) \tag{24}$$

où $$G_m''(x_o) = - g_m''(x_o) > 0.$$

Finalement on tire de (24) :

$$\text{Var}\ \hat{X}_{01} = \text{Var}\ [g'_m(x_o)/G''_m(x_o)] \tag{25}$$

Dans cette dernière relation il est clair que $g'_m(x_o)$ et $G''_m(x_o)$ sont deux variables aléatoires, tandis que m est une constante définie par (18).

Supposons, pour le moment, que $g'_m(x_o)$ et $G''_m(x_o)$ ont des variances et covariances d'ordre 1/n, l'application de (22) à la relation (25) donne :

$$\text{Var}\ X_{01} = \frac{\text{Var}\ g'_m(x_o)}{[E\ G''_m(x_o)]^2} + \frac{[E\ g'_m(x_o)]^2\ \text{Var}\ G''_m(x_o)}{[E\ G''_m(x_o)]^4} - \frac{2\,[E\ g'_m(x_o)]\ \text{cov}\ [g'_m(x_o), G''_m(x_o)]}{[E\ G''_m(x_o)]^3}. \tag{26}$$

Il est possible d'exprimer les différents termes de (26) en fonction des coefficients de Fourier c_k et s_k.

En effet, partant de (2), après un calcul long mais ne présentant pas de difficultés on trouve :

$$E\ g'_m(x_o) = \pi\, c_o\ \{\sum_{k=1}^{m} k\left(1 - \frac{k}{m+1}\right)[s_k \cos(k\, z_o) - c_k \sin(k\, z_o)]\} \tag{27}$$

$$E\ G''_m(x_o) = (\pi\, c_o)^2\ \{\sum_{k=1}^{m} k^2\left(1 - \frac{k}{m+1}\right)[c_k \cos(k\, z_o) + s_k \sin(k\, z_o)]\} \tag{28}$$

$$\begin{aligned}\text{Var}\ g'_m(x_o) = (\pi\, c_o)^2\ \{&\sum_{k=1}^{m} k^2\left(1 - \frac{k}{m+1}\right)^2 [\text{Var}\ \bar{s}_k \cos^2(k\, z_o) + \\ &+ \text{Var}\ \bar{c}_k \sin^2(k\, z_o) - 2\text{cov}(\bar{c}_k, \bar{s}_k)\cos(k\, z_o)\sin(k\, z_o)] + \\ &+ \sum_{k=1}^{m}\sum_{\substack{j=1\\ j\neq k}}^{m} k\, j\left(1 - \frac{k}{m+1}\right)^2 [\text{cov}(\bar{s}_k, \bar{s}_j)\cos(k\, z_o)\cos(j\, z_o) + \\ &+ \text{cov}(\bar{c}_k, \bar{c}_j)\sin(k\, z_o)\sin(j\, z_o) - 2\text{cov}(\bar{s}_k, \bar{c}_j)\cos(k\, z_o)\sin(j\, z_o)]\}\end{aligned} \tag{29}$$

$$\begin{aligned}\text{Var}\ G''_m(x_o) = (\pi\, c_o)^4\ \{&\sum_{k=1}^{m} k^4\left(1 - \frac{k}{m+1}\right)^2 \text{Var}\ \bar{c}_k \cos^2(k\, z_o) + \\ &+ \text{Var}\ \bar{s}_k \sin^2(k\, z_o) + 2\,\text{cov}(\bar{c}_k, \bar{s}_k)\cos(k\, z_o)\sin(k\, z_o)] + \\ &+ \sum_{k=1}^{m}\sum_{\substack{j=1\\ j\neq k}}^{m} k^2 j^2\left(1 - \frac{k}{m+1}\right)^2 [\text{cov}(\bar{c}_k, \bar{c}_j)\cos(k\, z_o)\cos(j\, z_o) + \\ &+ \text{cov}(\bar{s}_k, \bar{s}_j)\sin(k\, z_o)\sin(j\, z_o) + 2\text{cov}(\bar{c}_k, \bar{s}_j)\cos(k\, z_o)\sin(j\, z_o)]\}\end{aligned} \tag{30}$$

$$\operatorname{cov}[g'_m(x_o), G''_m(x_o)] = (\pi c_o)^3 \{\sum_{k=1}^{m} k^3 \left(1 - \frac{k}{m+1}\right)^2 [\operatorname{cov}(\bar{c}_k, \bar{s}_k) \cos(k z_o) +$$

$$+ (\operatorname{Var} \bar{s}_k - \operatorname{Var} \bar{c}_k) \cos(k z_o) \sin(k z_o)] +$$

$$+ \sum_{k=1}^{m} \sum_{\substack{j=1 \\ j \neq k}}^{m} k j^2 \left(1 - \frac{k}{m+1}\right)^2 [\operatorname{cov}(\bar{c}_k, \bar{s}_j) \cos(k+j) z_o$$

$$+ \operatorname{cov}(\bar{s}_k, \bar{s}_j) \cos(k z_o) \sin(j z_o) -$$

$$- \operatorname{cov}(\bar{c}_k, \bar{c}_j) \sin(k z_o) \cos(j z_o)]\} \qquad (31)$$

Les formules (12) jusquà (16) expriment que les variances et covariances des $\bar{c}_k$ et $\bar{s}_k$ sont d'ordre 1/n, l'application de la relation (22) se trouve donc justifiée à postériori. Finalement les expressions (27) à (31) jointes à (26) permettent de calculer approximativement Var $\hat{X}_{01}$. On notera que l'on a :

$$\operatorname{Var} X_{01} = (1/n) \, . \, F(c_o, \ldots, c_1, \ldots, c_{2m}, s_1, \ldots, s_{2m}) + o(1/n)$$

où $F(c_o, \ldots, s_{2m})$ est une fonction non explicitée des coefficients de Fourier de $g(x)$.

3.2. Méthode directe

Afin d'illustrer cette façon de procéder on propose ci-dessous un algorithme d'estimation directe du mode x_o de $f(x)$. La justification intuitive de cet algorithme repose sur le fait, déjà cité, que dans l'échantillon on doit observer une aggrégation des points au voisinage de x_o.

Soit un échantillon de taille N' issu de f(x) et ordonné par valeurs non décroissante :

$$x_1 \leqslant x_2 \leqslant \ldots \leqslant x_{N'}.$$

Soient a et b deux points tels que :

$$a = x_{0,1} - 3\,\delta$$

$$b = x_{0,9} + 3\,\delta$$

où

$$\delta = \operatorname{Inf}[(x_{0,9} - x_{0,75}) \,; (x_{0,25} - x_{0,1})],$$

$x_{0,1}$ et $x_{0,9}$ = estimations des déciles inférieur et supérieur de f(x).

$x_{0,25}$ et $x_{0,75}$ = estimations des quartiles de f(x).

Ainsi, au moins pour les grands échantillons, et pour la grande majorité des distributions unimodales continues rencontrées en pratique, on aura :

$$P(a < x_o < b) = 1.$$

On réalise une partition de]a, b] en k intervalles égaux de longueur s = (b – a)/k. On prend comme première valeur de k : $k = k_{max}$ où

$$k_{max} = \begin{cases} \text{“}15 \log N'\text{”} & \text{pour } N' \geqslant 30 \\ \text{“}10 \log N'\text{”} & \text{pour } N' < 30 \end{cases}$$

(les guillemets signifiant que l'on prend la partie entière de l'expression). Les intervalles sont donc, pour une valeur donnée de k :

$$]a, a + s],]a + s, a + 2s], \dots]a + (k - 1) s, b].$$

On désigne par n_i ($i = 1, 2, \dots, \kappa$) l'effectif de chaque intervalle. On pose :

$$\sum_{i=1}^{k} n_i = N \text{ (où } N \leqslant N').$$

(N représente l'effectif de l'échantillon situé entre les points a et b).

Soit n l'effectif théorique d'un intervalle correspondant à une densité de probabilité uniforme sur]a, b], on a :

$$n = N/k.$$

On défini :

– une chaîne comme c points x_i appartenant à h ($h = 1, 2, \dots, k_{max}$) intervalles adjacents tels que pour chacun d'eux on ait $n_i \geqslant n$;

– un trou est constitué par h intervalles adjacents tels que pour chacun d'eux on ait $n_i < n$.

En général la partition de]a, b] donnera plusieurs chaînes séparées par des trous. L'algorithme proposé consiste, dans un premier temps, à rechercher la valeur de k (en commençant par k_{max} et en diminuant k d'une unité chaque fois) qui fournit une chaîne unique : la chaîne modale (elle peut être constituée par l'échantillon entier). Ce problème n'admet pas toujours de solution. En particulier on peut parvenir à la partition minimum (k = 3) et se trouver en présence de deux chaînes ; il est alors nécessaire de définir un critère de choix :

En présence de deux chaînes pour k = 3, la chaîne modale sera celle pour laquelle la quantité :

$$(1/c) \left[\prod_{i=1}^{1+c-1} (x_i - x_{i-1}) \right]$$

(où 1 représente l'indice de la plus petite valeur de la chaîne) sera la plus faible. Cette expression caractérise, en effet, le groupement des points de la chaîne.

Il peut se produire qu'au cours du traitement on rencontre une chaîne de un point (c = 1). Dans ce cas on adopte, suivant la valeur de k, l'une des attitudes suivantes :

– si k > 3, on diminue k d'une unité et on recommence le processus de recherche de la chaîne modale ;

– si k = 3, on assimile alors cette chaîne de un point à un trou. Dans un deuxième temps, on applique le processus définit ci-dessus aux points de la chaîne modale et ainsi de suite.

L'algorithme procède donc par itérations et il est nécessaire de déterminer un critère d'arrêt :

La $(j + 1)^{ième}$ itération ne sera pas réalisée si l'une ou l'autre des conditions suivantes est remplie :

– l'effectif $c^{(j)}$ ($j = 1, 2, \dots$) de la $j^{ième}$ chaîne modale est égal à $c^{(j-1)}$. (Dans ce cas une nouvelle itération conduirait à la même chaîne modale) ;

— l'effectif $c^{(j)}$ de la $j^{\text{ième}}$ chaîne modale est inférieur ou égal à une quantité $c_1^{(j-1)}$ définie par :

$$c_1^{(j-1)} = \begin{cases} c^{(j-1)}\,(1 - d^{(j-1)}) & \text{pour } N' \geqslant 30 \text{ (avec } c^{(o)} = N') \\ N'/3 & \text{pour } N' < 30 \end{cases}$$

où

$$d^{(j-1)} = \left[\frac{(x_{0,75}^{(j-1)} - x_{0,5}^{(j-1)}) + (x_{0,25}^{(j-1)} - x_{0,5}^{(j-1)})}{x_{0,75}^{(j-1)} - x_{0,25}^{(j-1)}} \right].$$

$d^{(j-1)}$ représente la valeur absolue du coefficient de symétrie de Yule relatif à la $(j-1)^{\text{ième}}$ chaîne modale ($0 \leqslant d^{(j-1)} < 1$). Ainsi, lorsque l'échantillon est important ($N' \geqslant 30$), la quantité $c_1^{(j-1)}$ sera d'autant plus grande que la chaîne traitée est symétrique. L'objectif visé ici étant d'augmenter le nombre de points participant à l'estimation de x_o pour un échantillon symétrique. Si l'échantillon de départ est petit ($N' < 30$), l'erreur standard de d devenant trop importante (cf. [5]) on n'en tient pas compte pour le calcul de $c_1^{(j-1)}$.

Finalement, on estime le mode x_o de $f(x)$ par la médiane $\hat{x}_{02}$ de la dernière chaîne modale.

(Le sous-programme Fortran correspondant à cet algorithme peut être obtenu en s'adressant à l'auteur).

4. APPLICATION A L'ESTIMATION DU MODE DE LA DISTRIBUTION GAMMA

On connaît de nombreux exemples d'utilisation de la distribution gamma en statistique appliquée (cf. par exemple, [8] et [9]). On s'intéressera particulièrement ici à l'estimation du mode de cette distribution. C'est, en effet, la meilleure caractéristique de position pour une distribution biaisée (puisque la plus probable).

Sous une forme très générale, on peut définir la distribution gamma par la fonction de densité de probabilité :

$$f(x) = [1/\sigma.\Gamma(p)]\,[(x-\alpha)/\sigma]^{p-1} \exp[-(x-\alpha)/\sigma]$$

où :

$$\sigma > 0\,;\, p > 1\,;\, \alpha \leqslant x \leqslant \infty\,;\, \Gamma(p) = \int_0^\infty \exp(-t)\, t^{p-1}\, dt.$$

α, σ et p sont respectivement les paramètres de position, de dispersion et de forme de la distribution. On supposera, pour simplifier, que p est entier [$\Gamma(p) = (p-1)!$].

Le mode x_o de $f(x)$ est défini par :

$$x_o = (p-1)\,\sigma + \alpha. \tag{32}$$

Les moments non centrés peuvent s'obtenir par :

$$\mu'_r = \int_\alpha^\infty x^r f(x)\,dx \text{ soit,}$$

$$\mu'_r = 1/(p-1)! \sum_{j=0}^{r} C_r^j\, \sigma^j \alpha^{r-j}\,(p+j-1)!$$

On en déduit les moments centrés par :

$$\mu_r = \int_{\alpha}^{\infty} (x - \mu'_1)^r f(x)\, dx \text{ soit,}$$

$$\mu_r = \sum_{j=0}^{r} C_r^j \mu'_{r-j} (-\mu'_1)^j .$$

Ainsi on trouve :

$$\mu'_1 = EX = p\sigma + \alpha \quad (33) ; \mu_4 = 3p(p+2)\sigma^4 \quad (36)$$

$$\mu_2 = \text{Var } X = p\,\sigma^2 \quad (34) ; \mu_5 = 4p(5p+6)\sigma^5 \quad (37)$$

$$\mu_3 = 2\,p\,\sigma^3 \quad (35) ; \mu_6 = 5p(3p^2 + 26p + 24)\sigma^6 . \quad (38)$$

4.1. Estimation par la méthode des moments

Soient m'_1, m_2 et m_3 les trois premiers moments empiriques. Pour un échantillon de taille n :

$$m'_1 \doteq \bar{x} = (1/n) \sum_{i=1}^{n} x_i$$

$$m_r = (1/n) \sum_{i=1}^{n} (x_i - \bar{x})^r .$$

En égalant ces trois moments empiriques aux moments théoriques correspondants on obtient les équations suivantes :

$$p\sigma + \alpha = m'_1$$

$$p\sigma^2 = m_2$$

$$2p\,\sigma^3 = m_3 .$$

La résolution de ce système fournit les estimations des paramètres α, σ et p d'où on tire l'estimateur de x_o :

$$X_{03} = m'_1 - m_3/2m_2 \quad (39)$$

D'autre part en poussant le développement de (21) jusqu'aux termes de degré deux en $(y_i - \theta_i)$ on obtient :

$$E\,[L(Y_1, \ldots, Y_k)] = L(\theta) + (1/2) \sum_{i=1}^{k} L''_i(\theta) \text{ Var } Y_i + o(1/n) \quad (40)$$

Il est alors possible d'utiliser les relations suivantes (cf. [3]) :

$$E\, m'_1 = \mu'_1 \quad (41) ; E\, m_3 = (n-3)\mu_3/n + o(1/n) \quad (43)$$

$$E\, m_2 = (n-1)\mu_2/n \quad (42) ; \text{Var } m_2 = (\mu_4 - \mu_2^2)/n + o(1/n) \quad (44)$$

afin d'obtenir, pour les grands échantillons, une expression approchée de l'espérance mathématique de X_{03}. L'application de (40) à la relation (39) de définition de $\hat{X}_{03}$, compte-tenu des expressions ci-dessus, conduit en effet à :

$$E\,\hat{X}_{03} = x_o + (1/n)\,[\mu_3/\mu_2 - \mu_3(\mu_4 - \mu_2^2)/2\mu_2^3] + o(1/n).$$

Soit avec (34), (35) et (36) :

$$E\,\hat{X}_{03} = x_o - 6\,\sigma/np + o(1/n). \tag{45}$$

De la même manière, l'emploi de (22) et des résultats suivants (cf. [3]) :

$$\mathrm{Var}\,m'_1 = \mu_2/n$$

$$\mathrm{Var}\,m_3 = (\mu_6 - \mu_3^2 + 9\mu_2^3 - 6\mu_2\mu_4)/n + o(1/n)$$

$$\mathrm{cov}(m'_1, m_2) = \mu_3/n + o(1/n)$$

$$\mathrm{cov}(m'_1, m_3) = (\mu_4 - 3\mu_2^2)/n + o(1/n)$$

$$\mathrm{cov}(m_2, m_3) = (\mu_5 - 4\mu_2\,\mu_3)/n + o(1/n),$$

permettent de trouver une expression approximative de la variance de X_{03}. Compte tenu des relations (33) à (38), on trouve tous calculs faits :

$$\mathrm{Var}\,\hat{X}_{03} = (\sigma^2/2\,np)\,(5p^2 + 16p + 24) + o(1/n). \tag{46}$$

Les formules (45) et (46) confirment que l'estimateur de x_o obtenu par la méthode des moments est asymptotiquement convergent. Cette convergence est d'autant plus rapide que p est grand, ou que l'on se rapproche de la loi normale puisque, pour de grandes valeurs de p, la distribution gamma est voisine de la loi normale de variance $p\sigma^2$. En conséquence, la précision des formules (45) et (46) sera d'autant plus grande que le produit np sera important.

Il faut enfin souligner que, contrairement aux deux méthodes non paramétriques précédentes, la méthode d'estimation des moments est paramétrique puisqu'elle suppose connue la forme de la densité f(x) et qu'elle procède par l'intermédiaire de l'estimation des paramètres α, σ et p.

4.2. Résultats expérimentaux

Puisqu'on ne possède pas de formule pour la variance asymptotique de l'estimateur X_{02} décrit au paragraphe trois, il est intéressant de comparer les efficacités des trois méthodes d'estimation du mode exposées plus haut à l'aide d'une simulation. Pour cela on échantillonne la distribution gamma pour les deux groupes de valeurs suivants des paramètres (on pourra consulter [1] pour connaître les algorithmes de simulation utilisés) :

a) $\alpha = 0 ; p = 3 ; \sigma = 2$

b) $\alpha = 0 ; p = 10 ; \sigma = 1/\sqrt{10}$

La première densité est identique à la loi du khi-deux à six degrés de liberté. C'est une loi très biaisée de mode $x_o = 4$, d'espérance mathématique EX = 6 et de variance Var X = 12,25. En outre, elle présente la particularité d'avoir un mode très aplati [(f(x_o) = 0,14] ce qui contribue à accentuer la difficulté d'estimation de x_o.

La seconde densité correspond à une loi proche de la loi normale de variance unité. C'est donc, à l'inverse du cas précédent, une loi plutôt symétrique de mode $x_o = 2,85$, d'espérance mathématique EX = 3,16 et de variance Var X = 1. Son mode est plus aigu [(f(x_o) = 0,41].

Dans chacun des deux cas (a) et (b), M = 100 échantillons de taille n = 50 ont été simulés et on a appliqué successivement à chacun de ces échantillons les trois estimateurs X_{01}, X_{02} et X_{03} décrits plus haut. Il est donc possible d'obtenir ainsi une estimation des distributions d'échantillonnage de ces trois statistiques. On calcule ensuite le biais et la dispersion de ces distributions. Pour chaque estimateur X_{0i} ($i = 1, 2, 3$) on calcule donc les quantités :

$$b_i = \bar{x}_{0i} - x_o$$

avec $\bar{x}_{0i} = (1/M) \sum_{j=1}^{M} \hat{x}_{0i,j}$.

b_i représente le biais de $\hat{X}_{0i}$.

$$s_i^2 = \sum_{j=1}^{M} (\hat{x}_{0i,j} - \bar{x}_{0i})^2/(M-1).$$

s_i^2 représente l'estimation de la variance de $\hat{X}_{0i}$.

$$q_i = (b_i^2 + s_i^2)^{1/2}.$$

q_i est l'erreur quadratique moyenne de $\hat{X}_{0i}$.

Pour déterminer la valeur de $\hat{X}_{01}$ à partir de connaissance de la fonction $g_m(x)$ il est nécessaire d'employer un algorithme de recherche du maximum (on pourrait également utiliser une méthode numérique pour résoudre l'équation $\partial g_m(x)/\partial x = 0$ qui ne peut être résolue explicitement). Il est équivalent de rechercher la valeur de x qui rend minimum $-g_m(x)$. On a fait appel pour cela à l'algorithme de minimisation de Nelder et Mead ([11]) dont il existe une version programmée en Fortran ([12]). Cet algorithme présente la particularité intéressante de ne nécessiter aucune hypothèse sur la fonction à minimiser ni sur ses dérivées. Afin de s'assurer que le minimum atteint n'est pas un minimum local on a légèrement modifié la programmation de cet algorithme. Pour cela, on calcule $-g_m(x)$ pour dix valeurs de x régulièrement espacées de part et d'autre du minimum $\hat{x}_{01}$: si pour une seule de ces valeurs x_h on a : $-g_m(x_h) < -g_m(\hat{x}_{01})$ on recommence le processus en l'affinant.

D'autre part il convient, toujours dans le cas de $\hat{X}_{01}$, de définir les constantes a et b qui limitent l'intervalle sur lequel on va estimer f(x). On a adopté les valeurs suivantes : $a = 0$ et $b = 27{,}7$ pour la loi du khi-deux à six degrés de liberté, $a = 0$, et $b = 8$ pour l'autre loi.

Ces valeurs ont été choisies d'une part, pour que $a < x_o < b$ et d'autre part de sorte que pour chacune des deux distributions on ait $(b - a)/(\mathrm{Var}\ X)^{1/2} = 8$.

On a regroupé dans les tableaux 1 et 2 les résultats obtenus :

TABLEAU 1
Résultats des simulations

f(x) \ X_{0i}	X_{01} (Méthode indirecte)	X_{02} (Méthode directe)	X_{03} (Méthode des moments
$\alpha = 0$ $p = 3$ $\sigma = 2$	$b = 0{,}69$ $s^2 = 0{,}66$ $q = 1{,}07$	$b = 0{,}63$ $s^2 = 0{,}82$ $q = 1{,}10$	$b = 0{,}18$ $s^2 = 0{,}85$ $q = 0{,}94$
$\alpha = 0$ $p = 10$ $\sigma = 1/\sqrt{10}$	$b = 0{,}13$ $s^2 = 0{,}06$ $q = 0{,}28$	$b = 0{,}09$ $s^2 = 0{,}07$ $q = 0{,}28$	$b = 0{,}03$ $s^2 = 0{,}06$ $q = 0{,}25$

TABLEAU 2

Distribution du nombre de termes optimum (m) **dans** g_m (x)

f(x) \ m	1	2	3	4	5	6	7	8	9	10	11
$\alpha = 0 ; p = 3$ $\sigma = 2$	0	13	33	19	19	8	4	2	1	0	1
$\alpha = 0 ; p = 10$ $\sigma = 1/\sqrt{10}$	0	44	33	10	6	4	1	0	1	1	0

L'examen du tableau 1 permet de constater, comme on pouvait s'y attendre, que la méthode des moments donne les meilleurs résultats en terme d'erreur quadratique moyenne q. Cet avantage étant plus marqué dans le cas où la densité échantillonnée est très biaisée. Il est toutefois remarquable que l'absence d'information à priori concernant f(x) ne pénalise pas outre mesure les estimateurs non paramétriques $\hat{X}_{01}$ et $\hat{X}_{02}$ puisque l'efficacité relative minimum $q_{\hat{X}_{03}}/q_{\hat{X}_{02}}$ est tout de même de 85 %.

Les estimateurs non paramétriques $\hat{X}_{01}$ et $\hat{X}_{02}$ sont pratiquement équivalents et l'on pourra choisir l'une ou l'autre des deux méthodes suivant que l'on désire ou non une estimation de f(x). Il faut noter à ce sujet que, si l'on prend comme unité le temps de calcul nécessaire à l'estimateur $\hat{X}_{03}$, il faut respectivement 2,5 et 10,2 fois plus de temps pour calculer $\hat{X}_{02}$ et $\hat{X}_{01}$. Le temps nécessaire pour $\hat{X}_{01}$ pourrait certainement être abaissé par l'utilisation d'un algorithme de minimisation plus performant, mais il est très probable qu'il ne sera jamais inférieur à celui qui est nécessaire au calcul de $\hat{X}_{02}$.

Les chiffres du tableau 2 montrent que l'emploi des séries de Féjer au lieu des séries de Fourier ne modifie pas l'observation faite par Kronmal et Tarter ([10]) suivant laquelle le nombre optimum m de termes à prendre pour estimer f(x) augmente en moyenne avec le biais de la distribution. On observe également, qu'en pratique on pourra se contenter de calculer les moments trigonométriques $\bar{c}_k$ et $\bar{s}_k$ jusqu'à k = 10.

Enfin, si on compare les valeurs founies par les formules (45) et (46) à celles obtenues expérimentalement pour $\hat{X}_{03}$ on trouve :

Pour p = 3 et $\sigma = 2$: $E\hat{X}_{03} - x_o = 0{,}08$ au lieu de b = 0,18

$\text{Var } X_{03} = 1{,}60$ au lieu de $s^2 = 0{,}85$

Pour p = 10 et $\sigma = 1/\sqrt{10}$: $EX_{03} - x_o = 0{,}00$ au lieu de b = 0,03

$\text{Var } X_{03} = 0{,}07$ au lieu de $s^2 = 0{,}06$.

Ces résultats confirment l'observation faite au paragraphe quatre, suivant laquelle la précision des formules (45) et (46) est meilleure si le produit np est grand. On observe, en effet, que l'accord avec l'expérience est moins bon dans le premier cas (np = 150) que dans le second (np = 500).

BIBLIOGRAPHIE

[1] ASSELIN de BEAUVILLE J.P. – Les sous-programmes usuels de simulation statique, *Rev. Stat. Appl.*, 22 (1974), 57-87.

[2] CHERNOFF H. – Estimation of the mode, *Ann. Inst. Stat. Math.*, 16 (1964), 31-41.

[3] CRAMER H. – *Mathematical Methods of Statistics,* Princeton, Princeton University Press, 1971.

[4] DALENIUS T. – The mode-A neglected statistical parameter, *J. Roy. Stat. Soc. Ser. A,* 128 (1965), 110-117.

[5] DAVID F.N. et JOHNSON N.L. – A test for skewness with ordered variables, *Ann. Eugen.*, 18 (1954), 351-353.

[6] EKBLOM H. – A Monte-Carlo investigation of mode estimators in small samples, *Appl. Stat.*, 21 (1972), 177-184.

[7] GRENANDER U. – Some direct estimates of the mode, *Ann. Math. Stat.*, 36 (1965), 131-138.

[8] GUPTA S.S. – Order statistics from the gamma distribution, *Technometrics,* 2 (1960), 243-262.

[9] GUPTA S.S. et GROLL P.A. – Gamma distribution in acceptance sampling based on life tests, *J. Amer. Stat. Ass.*, 56 (1961), 942-970.

[10] KRONMAL R. et TARTER M. – The estimation of probability densities and cumulatives by Fourier series methods, *J. Amer. Stat. Ass.*, Septembre (1968), 925-952.

[11] NELDER J.A. et MEAD R. – A simplex method for function minimisation, *Computer J.*, 7 (1965), 308-313.

[12] O'NEILL R. – Function minimisation using a simplex procedure, *Appl. Stat.*, 20 (1971), 338-345.

[13] PARZEN E. – On estimation of a probability density function and mode, *Ann. Math. Stat.*, 33 (1962), 1065-1076.

[14] ROBERTSON T. et CRYER J.D. – An iterative procedure for estimating the mode, *J. Amer. Stat. Ass.*, 69 (1974), 1012-1016.

[15] ROSENBLATT M. – Remarks on some non-parametric estimates of a density function, *Ann. Math. Stat.*, 27 (1956), 832-837.

[16] TARTER M. et HOLCOMB R. – After the histogram what ? A description of new computer methods for estimating the population density, *Proceeding A.C.M. National Meeting,* (1967), 511-519.

[17] VALIRON G. – *Théorie des fonctions,* Paris, Masson et Cie, 1955.

[18] VENTER J.H. – On estimation of the mode, *Ann. Math. Stat.*, 38 (1967), 1446-1455.

[19] WATSON G.S. et LEADBETTER M.R. – On the estimation of probability density, *Ann. Math. Stat.*, 34 (1963), 480-491.

[20] ROSENBLATT M. – Curve estimates, *Ann. Math. Stat.*, 42 (1971), 1815-1842.

[21] DEHEUVELS P. – Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité, *C.R. Ac. Sc.*, t. 278, Série A (29 avril 1974), 1217-1220.

[22] YAMATO H. – Sequential estimation of a continuous probability density function and mode, *Bull. Math. Statis. Jap.*, 14 (1972), 1-12.

[23] DEHEUVELS P. – Sur une famille d'estimateurs de la densité d'une variable aléatoire, *C.R. Ac. Sc.*, t. 276, série A (2 avril 1973), 1013-1015.

[24] DEHEUVELS P. – Sur l'estimation séquentielle de la densité, *C.R. Ac. Sc.*, t. 276, série A (16 avril 1973), 1119-1121.

[25] DEHEUVELS P. – Thèse présentée à l'Université de Paris VI, (1974).

[26] BANON G. – Sur un estimateur non paramétrique de la densité de probabilité, *Rev. Stat. Appl.*, 24 (1976), n° 4, 61-73.

[27] SCHWARTZ S.C. – Estimation of probability density by an orthogonal series, *Ann. Math. Statist.*, 38 (1967), 1261-1265.

[28] WATSON G.S. – Density estimation by orthogonal series, *Ann. Math. Statist.*, 40 (1969), 1496-1498.

[29] CRAIN B.R. – A note on density estimation using orthogonal expansions, *J.A.S.A.*, 68 (1973), 964-965.

[30] CRAIN B.R. – More on estimation of distributions using orthogonal espansions, *J.A.S.A.*, 71 (1976), 741-745.

[31] FUKUNAGA K. – Introduction to statistical pattern recognition, New-York et Londres, Academic Press, 1972, 165-195.

[32] WEGMAN E.J. – Non parametric probability density estimation I : A survey of available methods, *Technometrics*, 14 (1972), 533-546.

[33] BOSQ D. – Contribution à la théorie de l'estimation fonctionnelle, Publ. de l'Institut de Statistique de l'Université de Paris, XIX, (1970).

[34] REVESZ P. – On empirical density function, *Period. Math. Hungar.*, 2 (1972), 85-110.

[35] TARTER M.E. et KRONMAL R.A. – An introduction to the implementation and theory of nonparametric density estimation, *Amer. Statistician*, 30 (1976), n° 3, 105-112.