Estimation de densités unimodales

Anne-Laure FOUGÈRES

Département de mathématiques et de statistique

Université Laval, Québec, Canada G1K 7P4

E-mail: fougeres@mat.ulaval.ca

Key words and phrases: Unimodal distributions; nonparametric density estimation; Grenander's estimator; monotone decreasing rearrangement; kernel estimation.

AMS 1991 classification: 62G05, 62G07.

Running title: Estimation de densités unimodales.

Résumé

Cet article propose un nouvel estimateur non paramétrique unimodal d'une densité de probabilité unimodale de mode connu. La solution la plus classique de ce problème est l'estimateur du maximum de vraisemblance sous contrainte de monotonie, considéré par Grenander (1956). L'approche que nous envisageons est fondée sur un réarrangement unimodal de l'estimateur à noyau de la densité. Nous étudions les propriétés asymptotiques de cet estimateur et examinons, à l'aide de simulations, son comportement pour de petites tailles d'échantillons.

Abstract

This paper proposes a new nonparametric unimodal estimator of a unimodal probability density function, in the case where the mode is known. The classical solution to this problem is the maximum likelihood estimator under monotonicity constraint, considered by Grenander (1956). Our approach is based on a unimodal rearrangement of the kernel estimator of the density. Asymptotic properties of this estimator are studied, and its small-sample behaviour is examined through simulations.

1. Introduction

L'estimation non paramétrique d'une densité de probabilité est un vaste problème, sur lequel ont porté de nombreux travaux, et pour lequel la variété de méthodes suggérées peut parfois sembler déroutante. Citons parmi elles la méthode du noyau, des méthodes dites locales (histogramme, régressogramme), ainsi que des méthodes d'estimation par projection (voir par exemple Bosq et Lecoutre (1987), ou Devroye (1987)). Toutes ces techniques font appel à divers paramètres, dont le choix s'avère souvent relativement difficile. Une façon de remédier partiellement à ce problème consiste à prendre en compte, lorsque cela est possible, toute information *a priori* sur la régularité de la fonction que l'on cherche à estimer, et plus spécifiquement sur la forme que la densité cherchée est susceptible d'avoir. L'hypothèse d'unimodalité apparaît dans ce cadre parmi les hypothèses peu restrictives les mieux pressenties. Pour exploiter une information de ce type sur la densité f étudiée, il semble légitime d'exiger de l'estimateur de f qu'il vérifie également cette propriété de régularité.

L'objectif que nous nous fixons ici est d'estimer une densité de probabilité unimodale de mode supposé connu par un estimateur *intrinsèque*, c'est-à-dire par un estimateur qui soit lui-même une densité unimodale. Après avoir rappelé brièvement en section 2 quelques propriétés des lois unimodales, nous présenterons le principal estimateur intrinsèque étudié à ce jour, appelé *estimateur de Grenander*, et mentionnerons divers résultats le concernant. Évoquant ensuite les limites d'ordre pratique de cet estimateur, nous proposerons en section 3 un nouvel estimateur de la densité, noté $f_{n\downarrow}$, fondé sur un réarrangement unimodal de l'estimateur à noyau. La section 4 sera consacrée à l'étude des propriétés théoriques de ce nouvel estimateur et établira un résultat de continuité, de convergence forte et uniforme, ainsi qu'un résultat concernant son comportement asymptotique. L'aspect pratique de $f_{n\downarrow}$ sera ensuite envisagé en section 5, et nous observerons dans quelle mesure il permet de pallier certains défauts à la fois de l'estimateur de Grenander, et de l'estimateur à noyau classique. Une brève discussion fera l'objet de la section 6, dans laquelle nous évoquerons de nouvelles perspectives de recherche. Les démonstrations sont données en annexe.

2. Estimateur de Grenander

Les distributions unimodales représentent un vaste ensemble de lois de probabilité, possédant d'intéressantes propriétés, et ayant fait l'objet de nombreuses études répertoriées entre autres par Dharmadhikari et Joag-Dev (1988), ainsi que Bertin, Cuculescu et Theodorescu (1996). Rappelons qu'une fonction de répartition F définie sur $\mathbb{R}$ est dite *unimodale* s'il existe un réel M, appelé *mode* de F, tel que F soit convexe sur $(-\infty, M)$ et concave sur (M, ∞). Une telle distribution peut posséder plusieurs modes, qui forment alors un intervalle. La proposition suivante (voir par exemple Dharmadhikari et Joag-Dev (1988) pour une démonstration) réunit quelques conséquences simples de la définition précédente.

Proposition 1

(i) *Une fonction de répartition F unimodale est absolument continue, sauf peut-être en son mode M. Ses dérivées à droite et à gauche existent, sauf peut-être en M.*

(ii) *Si F est absolument continue, alors elle est unimodale de mode M si et seulement si sa densité f est croissante sur $(-\infty, M)$ et décroissante sur (M, ∞).*

Nous nous intéressons dans cet article à des lois de probabilité absolument continues, admettant une densité f unimodale, de mode connu M. Remarquons que l'on peut, d'un point de vue théorique, se restreindre sans perte de généralité à l'étude des densités unimodales de mode en 0, et définies sur $[0, \infty)$. Le passage au cas d'une densité unimodale de mode M se fait alors en traitant séparément les deux parties monotones que sont les restrictions de f à $(-\infty, M)$ et à (M, ∞). Aussi considérerons-nous dans la suite de cette section, ainsi qu'en sections 3 et 4, des densités de probabilité définies sur $\mathbb{R}^+$ et décroissantes. Le problème est alors le suivant : comment estimer de telles densités de façon intrinsèque, c'est-à-dire par un estimateur qui soit lui-même une densité sur $\mathbb{R}^+$, décroissante ?

La contrainte de monotonie assure l'existence d'un estimateur du maximum de vraisemblance. Grenander (1956) a montré qu'il s'obtenait comme la pente du plus petit majorant concave de la fonction de répartition empirique ; il s'agit donc d'un histogramme à largeur de fenêtre variable, dépendant des données. Cet estimateur, appelé également *estimateur de Grenander*, et que nous noterons f_n^G, a suscité diverses études, et possède un grand nombre de propriétés asymptotiques. Des ouvrages tels que Barlow, Bartholomew, Bremner et Brunk (1972) ou Robertson, Wright et Dykstra (1988) passent en revue ces propriétés. Prakasa Rao (1969) a notamment obtenu le comportement asymptotique de f_n^G en un point donné, résultat que nous rappelons dans le théorème suivant. Précisons d'abord que si X et X_n, $n \in \mathbb{N}$, sont des variables aléatoires sur $\mathbb{R}$, la notation $X_n \xrightarrow{\mathcal{L}} X$ sera employée pour signifier qu'asymptotiquement, X_n a la même loi que X.

Théorème 1 (Prakasa Rao, 1969)

Soit f une densité de probabilité définie sur $\mathbb{R}^+$ et décroissante; soit $x \in \mathbb{R}^+$ et $f_n^G(x)$ l'estimateur du maximum de vraisemblance de $f(x)$, basé sur n observations indépendantes. On suppose que f est dérivable en x, avec $f'(x) \neq 0$. On a *alors*

$$n^{1/3}|f(x)f'(x)/2|^{-1/3}(f_n^G(x) - f(x)) \xrightarrow{\mathcal{L}} Z,$$

où Z est la pente du plus petit majorant concave du processus $W(t) - t^2$ sur $\mathbb{R}$, pris en $t = 0$, W dénotant un mouvement brownien standard avec $W(0) = 0$.

Prakasa Rao montre également que cette variable aléatoire Z est distribuée comme la localisation du maximum du processus $(W(t) - t^2, t \in \mathbb{R})$. Il est à noter que Groeneboom (1989) a obtenu de façon explicite la densité de probabilité de cette dernière variable aléatoire. L'étude du comportement asymptotique de la norme L^1 a été effectuée par Groeneboom (1985) sous contrainte de monotonie stricte, et généralisée par Wang (1995) au cas où la densité admet à la fois des plateaux et des parties de décroissance stricte. Devroye et Gyorfi (1985, p. 213) et Devroye (1987, p. 145) récapitulent les principales propriétés asymptotiques de la norme L^1 pour l'estimateur de Grenander. Le comportement asymptotique de la distance L^2 a été étudié pour des densités de probabilité uniformes par Groeneboom et Pyke (1983). Enfin, Birgé (1989) a montré que f_n^G possédait la propriété d'optimalité minimax pour le risque L^1 sur l'ensemble des densités de probabilité définies sur $[0, 1]$, décroissantes et bornées par une constante donnée.

L'estimateur de Grenander possède donc des propriétés théoriques indéniables. Il présente toutefois des inconvénients d'un point de vue pratique. D'une part, bien que cet histogramme évite le problème classique du choix des classes, celles-ci étant déterminées par les observations, il fournit une estimation discontinue. Dans le cadre de l'estimation d'une densité de probabilité continue, la discontinuité de f_n^G peut s'avérer un inconvénient. D'autre part, l'estimateur de Grenander se comporte mal au voisinage du mode. On peut notamment montrer, comme le suggère Datta (1995), que pour toute densité de probabilité définie sur $[0,1]$, bornée et décroissante, on a $\mathbb{E}\ f_n^G(0) = \infty$. Ceci exhibe, concernant les densités bornées, un autre point par rapport auquel l'estimateur de Grenander n'est pas intrinsèque. La section 5 nous fournira d'ailleurs des illustrations claires du comportement médiocre de f_n^G au voisinage du mode.

3. Un nouvel estimateur

La démarche que nous effectuons ici est tout à fait différente de celle de Grenander, et consiste à construire un estimateur unimodal à partir d'un estimateur de la densité, modifié pour le rendre décroissant. L'approche que nous envisageons est fondée sur l'estimateur à noyau de la densité, dont nous nous bornerons ici à rappeler la définition et un résultat de convergence ; pour une synthèse sur ce sujet, on pourra par exemple se référer à Devroye (1987), ou Berlinet et Devroye (1990). Un *noyau* désigne toute application K de $\mathbb{R}$ dans $\mathbb{R}$ bornée, intégrable par rapport à la mesure de Lebesgue, positive, et d'intégrale égale à 1. Soit K un tel noyau, et $(X_1, \ldots, X_n)$ un échantillon aléatoire issu d'une loi de probabilité de densité f ; l'*estimateur à noyau* de f est l'application $f_n : \mathbb{R} \mapsto \mathbb{R}_+$ définie par

$$x \mapsto f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) , \qquad (1)$$

où h_n est un nombre réel strictement positif, appelé “largeur de fenêtre”, ou “paramètre de lissage”. Nous nous placerons tout au long de cet article sous l'hypothèse usuelle que h_n tend vers 0 et $nh_n/\log n$ tend vers l'infini quand n tend vers l'infini ; cette hypothèse sera donc systématiquement sous-entendue dans les énoncés des résultats. Abou Jaoudé (1974) a montré qu'il n'existait pas d'estimateur uniformément convergent de la densité : si $\mathcal{C}$ désigne l'ensemble des densités de probabilité continues sur $\mathbb{R}$, alors il est impossible de trouver un estimateur f_n de la densité tel que :

$$\forall f \in \mathcal{C}, \quad \sup_{x \in \mathbb{R}} |f_n(x) - f(x)| \quad \xrightarrow[n\to\infty]{} \quad 0 \ \text{ en probabilité.}$$

Toutefois, si l'on restreint la classe de fonctions f, en considérant notamment la classe des fonctions f continues et à support compact, on peut obtenir des résultats de convergence de l'estimateur à noyau f_n vers f (voir par exemple Bosq et Lecoutre (1987, chapitre 4)). Mentionnons en particulier le résultat suivant :

Théorème 2 *Soit K un noyau à variation bornée et à support compact, et f une densité de probabilité continue. Alors presque-sûrement, l'estimateur à noyau f_n défini par* (1) *converge uniformément sur tout compact vers f :*

$$\forall a < b, \sup_{x \in [a,b]} |f_n(x) - f(x)| \quad \xrightarrow[n\to\infty]{} \quad 0 \ \ p.s.$$

L'estimateur à noyau d'une densité unimodale n'est pas un estimateur unimodal, et notre objectif consiste donc à le modifier pour le rendre intrinsèque. Nous utiliserons pour cela la notion de réarrangement monotone, étudiée par Hardy, Littlewood et Pólya (1952, chapitre 10). Considérons une densité f bornée sur $\mathbb{R}$ et désignons par B son maximum ; pour tout borélien A de $\mathbb{R}$, on notera $\lambda(A)$ la mesure de Lebesgue sur $\mathbb{R}$ de A. Soit r l'application de $\mathbb{R}_*^+$ dans $\mathbb{R}^+$, continue à droite, qui à u associe $r(u) = \lambda\{t : f(t) > u\}$. On appelle alors *réarrangement décroissant* de f, et l'on note $f_\downarrow$, l'inverse généralisé (continu à droite) de r, défini sur $\mathbb{R}$ à valeurs dans $\mathbb{R}^+$, tel que :

$$f_\downarrow(x) = \inf\{u \in \mathbb{R}_*^+ : r(u) \leq x\}.$$

Si $\{u \in \mathbb{R}_*^+ : r(u) \leq x\}$ est l'ensemble vide, on pose $f_\downarrow(x) = 0$. L'inverse généralisé de r étant noté $r^{\leftarrow}(x)$, on a par définition $f_\downarrow(x) = r^{\leftarrow}(x)$. La figure 1 illustre la définition précédente.

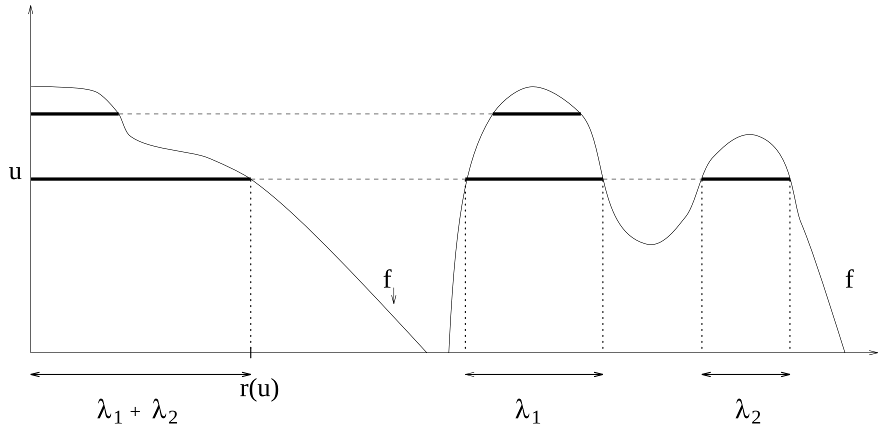

Figure 1: *Graphe d'une densité de probabilité f et du réarrangement décroissant associé $f_\downarrow$.*

Remarque 1 L'application $f_{\downarrow}$ est une densité de probabilité, continue à droite, décroissante sur $\mathbb{R}^+$, telle que $B = f_{\downarrow}(0)$; elle a pour support $[0, \Lambda]$, où Λ désigne la longueur du support de f.

Cette remarque nous conduit à proposer un nouvel estimateur unimodal de toute densité de probabilité f définie sur $\mathbb{R}^+$, unimodale de mode en 0, qui n'est autre que le réarrangement décroissant de l'estimateur à noyau f_n défini par (1). Il s'écrit par conséquent de la façon suivante, pour tout $x \in \mathbb{R}^+$:

$$f_{n\downarrow}(x) \;=\; \inf\{u \in \mathbb{R}^+_* \,:\, r_n(u) \leq x\}, \quad \text{où} \;\; r_n(u) = \lambda\{t \,:\, f_n(t) > u\}. \tag{2}$$

La section qui suit est consacrée à l'étude des propriétés théoriques de cet estimateur, et sa mise en pratique sera effectuée en section 5.

4. Propriétés du nouvel estimateur

Nous présenterons dans un premier temps un résultat de continuité de l'estimateur $f_{n\downarrow}$ précédemment défini, puis établirons la convergence uniforme presque-sûre de $f_{n\downarrow}$ vers f. Nous mentionnerons enfin un résultat lié au comportement asymptotique de la quantité $f_{n\downarrow} - f$.

Théorème 3 *Si K est un noyau continu et f une densité définie sur $\mathbb{R}$, alors l'estimateur $f_{n\downarrow}$ défini par (2) est continu sur $\mathbb{R}^+$.*

La démonstration de ce résultat fait l'objet de l'annexe 1. Remarquons que l'hypothèse de continuité du noyau n'est que très peu restrictive, et que l'utilisation des noyaux usuels conduit à un estimateur $f_{n\downarrow}$ à la fois unimodal et continu. Cet estimateur possède de plus la propriété de convergence forte et uniforme, comme l'établit le théorème suivant :

Théorème 4 *Soit K un noyau à variation bornée et à support compact, et f une densité de probabilité définie sur $\mathbb{R}^+$, continue, bornée, unimodale de mode en 0. Alors $f_{n\downarrow}$ est un estimateur presque-sûrement uniformément convergent sur $\mathbb{R}^+$:*

$$\sup_{x\in\mathbb{R}^+} |f_{n\downarrow}(x) - f(x)| \quad \xrightarrow[n\to\infty]{} \quad 0 \ \ p.s.$$

La démonstration de ce résultat est présentée en annexe 2. Elle découle du fait que pour toute densité f continue bornée, le réarrangement décroissant de l'estimateur à noyau de f converge fortement uniformément sur $\mathbb{R}^+$ vers celui de f, et de ce qu'une densité définie sur $\mathbb{R}^+$ et décroissante coïncide avec son réarrangement décroissant.

Corollaire 1 *Sous les hypothèses du théorème 4, on a*

$$\int_0^{+\infty} |f_{n\downarrow}(x) - f(x)|\, dx \quad \xrightarrow[n\to\infty]{} \quad 0 \quad p.s.$$

Ce résultat découle directement du théorème 4 et du théorème de Glick (1974) (voir par exemple Devroye et Gyorfi, 1985, p. 10).

Nous présentons à présent un résultat lié au comportement asymptotique local du réarrangement décroissant de l'estimateur à noyau $f_{n\downarrow}$. Considérant un réel positif x fixé, nous souhaiterions obtenir un résultat de convergence en loi de la quantité $(f_{n\downarrow} - f)(x) = (r_n{}^{\leftarrow} - r^{\leftarrow})(x)$ convenablement renormalisée. Le résultat mentionné ici, qui constitue un premier pas dans ce sens, décrit le comportement asymptotique en loi, à u fixé, de la quantité $(r_n - r)(u)$ renormalisée. Nous évoquons au préalable l'argument nous assurant de la proximité entre $f_{n\downarrow} - f$ et sa quantité inverse. La déduction rigoureuse du comportement asymptotique local de $f_{n\downarrow} - f$ à partir de celui de $r_n - r$ requiert l'utilisation de techniques liées à l'obtention d'une représentation de Bahadur-Kiefer (voir par exemple Shorack et Wellner, 1986), et fera l'objet d'un développement dans un futur proche. Nous pouvons toutefois formuler l'argument heuristique suivant, en nous plaçant sous des hypothèses de régularité convenables : la quantité $(r_n{}^{\leftarrow} - r^{\leftarrow})(x)$ tendant vers 0 (théorème 4), nous pouvons écrire un développement limité de r au voisinage de $r^{\leftarrow}(x)$, sous la forme $r \circ r_n{}^{\leftarrow}(x) = r(r^{\leftarrow}(x)) + (r_n{}^{\leftarrow} - r^{\leftarrow})(x)\, r'(r^{\leftarrow}(x)) + o((r_n{}^{\leftarrow} - r^{\leftarrow})(x))$. L'expression de $(r_n{}^{\leftarrow} - r^{\leftarrow})(x)$ ainsi obtenue, et la relation $r \circ r_n{}^{\leftarrow}(x) = x - (r_n - r)(r_n{}^{\leftarrow}(x))$ entraînent alors $(f_{n\downarrow} - f)(x) = \dfrac{-1}{r'(r^{\leftarrow}(x))}(r_n - r)(r_n{}^{\leftarrow}(x)) + o((r_n{}^{\leftarrow} - r^{\leftarrow})(x))$ p.s., ce qui nous assure une certaine proximité entre $f_{n\downarrow} - f$ et la quantité $r_n - r$.

Intéressons-nous dès lors au comportement de $(r_n - r)(u)$ pour un certain $u \in (0, f(0)]$ fixé ; dans le résultat qui suit, nous supposerons que la largeur de fenêtre h_n est du même ordre que $n^{-1/3}$. En l'absence de cette hypothèse, il apparaît dans le développement asymptotique un terme de biais que l'on ne peut négliger, empêchant l'obtention d'un résultat asymptotique du type $\alpha_n^{-1}(r_n - r)(u) \xrightarrow{\mathcal{L}} Z$, quelle que soit la renormalisation α_n. On pourrait alors chercher des résultats asymptotiques de la forme $\alpha_n^{-1}\,(r_n - r)(u) - B_n \xrightarrow{\mathcal{L}} Z'$; remarquons toutefois que sous cette hypothèse, la vitesse de convergence obtenue sera la même que celle de l'estimateur de Grenander f_n^G (voir théorème 1).

Nous exposons ici le résultat obtenu lorsque le noyau est uniforme sur $[-1,1]$, cas pour lequel l'écriture est simplifiée, et nous référons le lecteur à Fougères (1996) pour le cas à noyau quelconque, ainsi que pour les démonstrations de ces résultats. Ces dernières s'appuient essentiellement sur un théorème d'approximation de la fonction de répartition empirique, démontré par Komlós, Major et Tusnády en 1975 (voir par exemple Csörgő et Révész, 1981, p. 133).

Théorème 5 *Soit f une densité de probabilité sur $\mathbb{R}^+$ continue, bornée et décroissante; soit $u \in (0, f(0)]$ et $x = \sup\{t :\, f(t) \geq u\}$; on suppose qu' au voisinage de x, f est dérivable, et de dérivée bornée strictement négative. Soit f_n l'estimateur à noyau uniforme sur $[-1,1]$ de la densité, et $h_n = n^{-1/3}$ la largeur de fenêtre. Alors on a*

$$n^{1/3}(r_n - r)(u) \xrightarrow{\mathcal{L}} Z,$$

où Z est la variable aléatoire

$$\begin{aligned} & \lambda\left\{c > 1 :\, W(c+1) - W(c-1) > c\frac{-2\,f'(x)}{\sqrt{f(x)}}\right\} \\ + \quad & \lambda\left\{0 < c < 1 :\, W(c+1) + W(1-c) > c\frac{-2\,f'(x)}{\sqrt{f(x)}}\right\} \\ - \quad & \lambda\left\{c > 1 :\, W(c+1) - W(c-1) \leq c\frac{2\,f'(x)}{\sqrt{f(x)}}\right\} \end{aligned}$$

$$- \quad \lambda\left\{0<c<1 : W(1-c)+W(c+1) \leq c\frac{2\,f'(x)}{\sqrt{f(x)}}\right\},$$

et où W est un mouvement brownien.

5. Simulations

Cette section est consacrée à la mise en pratique de l'estimateur $f_{n\downarrow}$ introduit en section 3. Des simulations ayant été réalisées à partir de différentes lois de probabilité à densité unimodale, nous présentons à des fins illustratives les résultats obtenus pour l'estimateur que nous proposons, ainsi que pour l'estimateur de Grenander f_n^G et l'estimateur à noyau classique f_n. Nous confronterons dans un premier temps les résultats obtenus par les deux estimateurs unimodaux f_n^G et $f_{n\downarrow}$, qui asseoiront les avantages de $f_{n\downarrow}$ du point de vue pratique. Notons que le noyau et la largeur de fenêtre utilisés pour le calcul de $f_{n\downarrow}$ sont ici choisis de façon purement arbitraire ; rappelons par ailleurs que si le choix du noyau n'affecte généralement que dans une moindre mesure le comportement de l'estimateur à noyau, il n'en est pas de même pour le choix du paramètre de lissage. Cette remarque nous a conduit, sans entrer dans le domaine polémique de la recherche de la largeur de fenêtre optimale, à examiner à partir des simulations ce qu'apportait, de ce point de vue, l'utilisation de $f_{n\downarrow}$ par rapport à l'estimation à noyau classique. Nous nous attacherons ainsi dans un deuxième temps à examiner les résultats obtenus par f_n et $f_{n\downarrow}$ pour différents choix arbitraires de largeur de fenêtre ; ces résultats feront ressortir une certaine robustesse du réarrangement décroissant vis-à-vis du paramètre de lissage, mettant en exergue une amélioration pratique évidente.

Les distributions unimodales considérées sont la loi normale, la loi de Cauchy et la loi double exponentielle (de densité $x \mapsto 1/2 \exp\{-|x|\}$) ; pour chacune de ces lois, 500 échantillons aléatoires de taille n ont été simulés, et ce, pour trois tailles d'échantillon $n = 25, 50$ et 100 ; pour chacun de ces échantillons, les estimateurs f_n^G et f_n ont été évalués en chaque point d'une subdivision de [-5,5] et de pas $\delta = 0.01$, notée $(x_1, \ldots, x_p)$, où $p = 1000$. L'approximation de $f_{n\downarrow}$ qui a été utilisée est celle fournie par le vecteur $(f_n(x_1), \ldots, f_n(x_p))$ réordonné en ordre croissant à gauche du mode, puis décroissant à droite. (Pour de plus amples détails sur l'approximation numérique du réarrangement $f_{n\downarrow}$, voir Fougères, 1996). Si $\widehat{f}$ désigne indifféremment f_n^G, f_n ou $f_{n\downarrow}$, notons $\left(\widehat{f}_j(x_1), \ldots, \widehat{f}_j(x_p)\right)$ le vecteur ainsi obtenu pour le $j$ème échantillon, $j = 1, \ldots, 500$. La qualité de l'ajustement entre f et son estimation a été mesurée par les erreurs L^1, L^2 et L^∞, la distance L^1 ayant un caractère visuel appréciable, en tant qu'aire entre les deux fonctions, et la distance L^∞ devant fournir une information (globale) sur la distance maximale entre f et son estimateur, calculés au même point. Ces distances L^1, L^2 et L^∞ ont été approximées, pour chaque échantillon ($j = 1, \ldots, 500$), respectivement par les quantités $\sum_{i=1}^{p} \delta\, |\widehat{f}_j(x_i) - f(x_i)|$, $\sum_{i=1}^{p} \delta\, |\widehat{f}_j(x_i) - f(x_i)|^2$ et $\max_{i=1,\ldots,p} |\widehat{f}_j(x_i) - f(x_i)|$. L'examen des résultats nous a permis de constater que pour les trois estimateurs considérés, l'écart maximal entre f et son estimateur était pratiquement toujours réalisé au voisinage du mode. Par suite, la distance uniforme ne nous a pas semblé a posteriori fournir un critère adapté de mesure de la qualité globale de l'estimation. De plus, les calculs effectués pour l'erreur quadratique conduisant à des résultats tout à fait similaires à ceux de l'erreur L^1, nous ne présenterons que ces derniers.

La figure suivante présente, pour un échantillon aléatoire de taille 100 d'une loi de Cauchy, les estimations obtenues respectivement par f_n^G (en trait gras) et $f_{n\downarrow}$ (en pointillé) ; un noyau gaussien et une largeur de fenêtre égale à 0.21 sont utilisés pour le calcul de $f_{n\downarrow}$; la densité à estimer apparaît en trait plein.

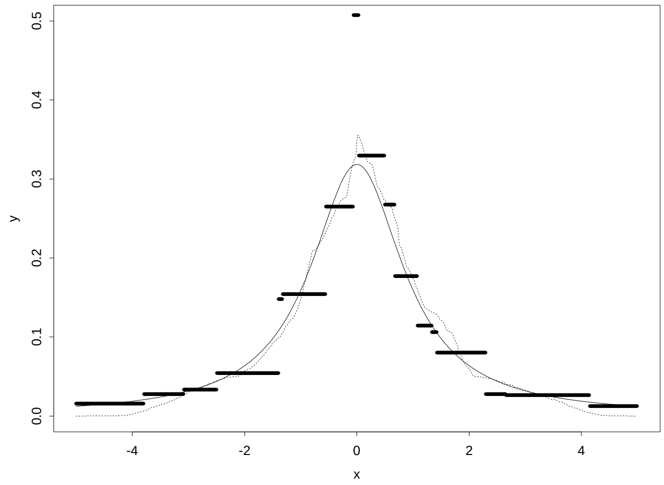

Figure 2: *Estimateur de Grenander et réarrangement unimodal de l'estimateur à noyau, obtenus à partir d'un échantillon aléatoire de taille 100 d'une loi de Cauchy.*

La figure 3 permet d'évaluer de façon plus précise la qualité de l'estimation faite par $f_{n\downarrow}$ par rapport à celle de f_n^G. Elle présente les diagrammes en boîte des 500 erreurs L^1 calculées pour chaque estimateur et chaque loi, pour une taille d'échantillon égale à 50 :

les trois premiers diagrammes correspondent aux erreurs L^1 de l'estimateur de Grenander (noté 'fgr' sur la figure), respectivement pour la loi normale, la loi de Cauchy et la loi double exponentielle, et les trois derniers aux erreurs L^1 de $f_{n\downarrow}$ (noté 'fnd' sur la figure) pour ces mêmes lois. Les résultats obtenus pour les tailles d'échantillon $n = 25$ et $n = 100$ étant similaires, ils ne sont pas présentés ici.

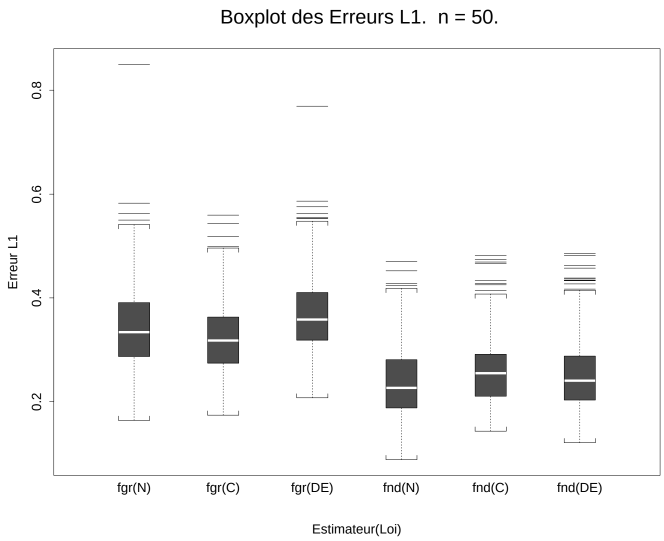

Figure 3: *Diagrammes en boîte représentant la distribution observée des erreurs* L^1 *de l'estimateur de Grenander et du réarrangement unimodal de l'estimateur à noyau de la densité d'une loi normale, d'une loi de Cauchy, et d'une loi double exponentielle, construits à partir de 500 échantillons aléatoires de taille 50 issus de ces lois.*

L'examen de ce diagramme permet aisément de constater un meilleur comportement du réarrangement unimodal de l'estimateur à noyau par rapport à l'estimateur de Grenander.

La distribution observée des erreurs L^1 présente en effet, pour les trois lois de probabilité considérées, une médiane plus petite pour $f_{n\downarrow}$ que pour f_n^G, ainsi qu'une moins grande dispersion. Les diagrammes témoignent enfin de l'existence de valeurs aberrantes beaucoup plus importantes pour l'estimateur de Grenander.

Intéressons-nous à présent à ce qu'apporte l'estimateur $f_{n\downarrow}$ par rapport à l'estimateur à noyau f_n. Rappelons pour cela qu'un problème important en estimation à noyau est de trouver un compromis entre un paramètre de lissage trop petit, privilégiant une approximation locale, mais entraînant de fortes oscillations de f_n, et un paramètre de lissage trop grand, qui conduira à un estimateur lisse, mais ayant tendance à sous-estimer de façon importante la densité au voisinage du mode. Remarquons de plus que $f_{n\downarrow}$ coïncide avec f_n dès lors que l'estimateur à noyau est unimodal, ce qui a lieu d'autant plus souvent que la largeur de fenêtre est grande, f_n étant alors plus lisse.

À partir de chaque échantillon simulé, issu des 3 distributions de probabilité normale, de Cauchy et double exponentielle, nous avons évalué les estimations fournies par f_n et $f_{n\downarrow}$, pour différents choix arbitraires de largeur de fenêtre. Nous avons ainsi considéré trois estimateurs à noyau, notés 'fn1', 'fn2' et 'fn3' sur les figures, correspondant aux paramètres de lissage 0.27, 0.7, et 1.2, et leurs réarrangements unimodaux respectifs, notés 'fnd1', 'fnd2' et 'fnd3'. Le noyau utilisé est un noyau gaussien. La figure 4 présente les diagrammes en boîte des erreurs L^1 calculées pour ces six estimateurs, à partir de 500 échantillons de taille

50 d'une loi de Cauchy. Les résultats étant véritablement similaires pour les deux autres lois, et quelle que soit la taille de l'échantillon ($n = 25, 50$ ou 100), nous ne présentons pas les autres diagrammes.

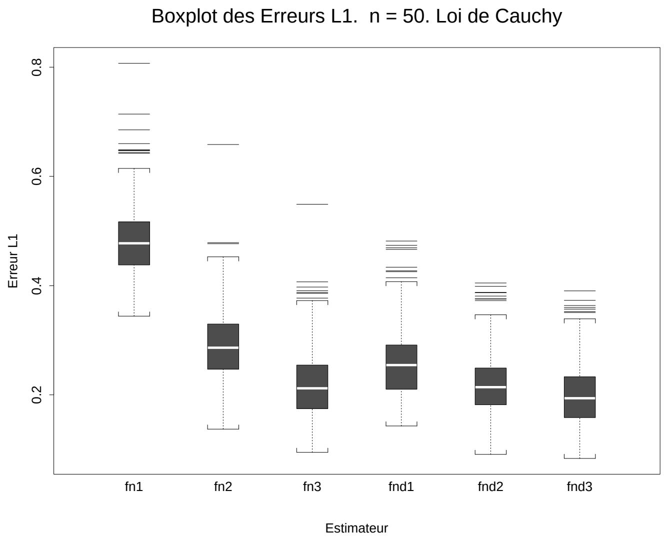

Figure 4: *Diagrammes en boîte représentant la distribution observée des erreurs* L^1 *de l'estimateur à noyau pour trois largeurs de fenêtre* $(0.27, 0.7,$ *et* $1.2)$ *et des réarrangements unimodaux correspondants, construits à partir de 500 échantillons aléatoires de taille 50 issus d'une loi de Cauchy.*

Ces diagrammes rendent nettement compte d'une variation moins importante de l'erreur L^1 entre 'fnd1', 'fnd2' et 'fnd3' qu'entre 'fn1', 'fn2' et 'fn3'. Le fait de réarranger l'estimateur à noyau semble donc rendre le choix (difficile) du paramètre de lissage moins crucial. Nous terminerons cette section par les deux figures suivantes, qui illustrent de façon éloquente ce phénomène. La figure 5 présente, pour un échantillon aléatoire de taille 50 d'une loi normale, les estimateurs à noyau 'fn1' et 'fn3', et leurs réarrangements unimodaux 'fnd1' et 'fnd3' sont présentés en figure 6. Ceci témoigne clairement de la plus grande robustesse de $f_{n\downarrow}$ face à la variation du paramètre de lissage. Le graphe en pointillé correspond à une largeur de fenêtre égale à 1.2, et les deux graphes en trait plein représentent la densité à estimer d'une part, et d'autre part l'estimateur correspondant à la largeur de fenêtre égale à 0.27, la confusion entre les deux graphes n'étant pas possible.

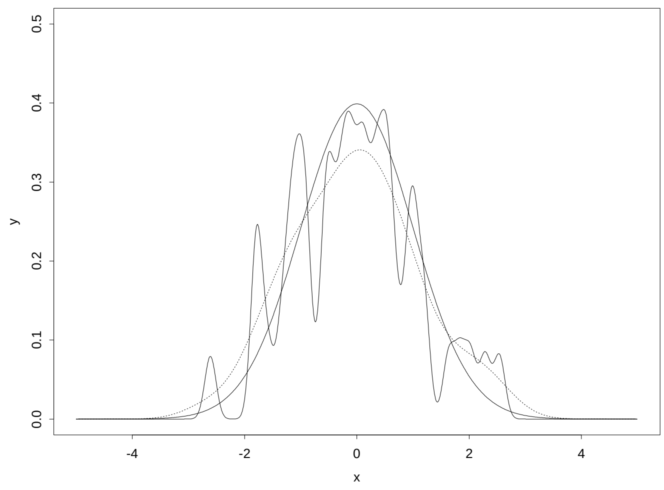

Figure 5: *Estimateurs à noyau* fn1 *et* fn3 *de paramètres de lissage respectifs* 0.27 *et* 1.2*, obtenus à partir d'un échantillon aléatoire de taille 50 d'une loi normale.*

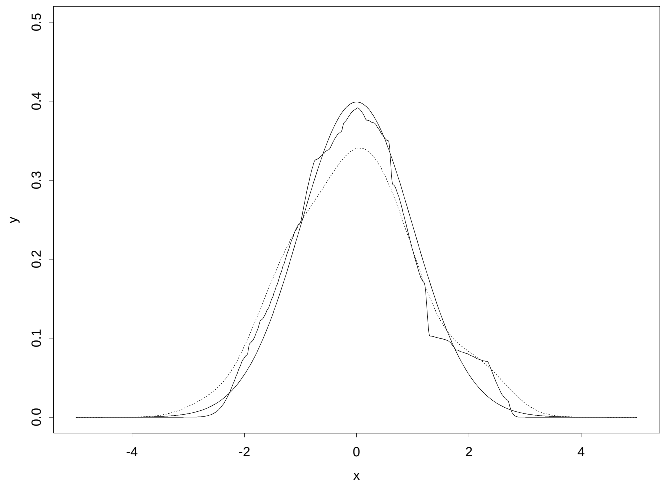

Figure 6: *Réarrangements unimodaux* fnd1 *et* fnd3 *de paramètres de lissage respectifs* 0.27 *et* 1.2*, obtenus à partir d'un échantillon aléatoire de taille 50 d'une loi normale, identique à celui de la figure 5.*

6. Discussion

Dans cet article, nous nous sommes intéressés à l'estimation de densité pour des distributions unimodales de mode connu, au moyen d'un estimateur unimodal. Nous avons proposé dans ce cadre un nouvel estimateur, fondé sur un réarrangement unimodal de l'estimateur à noyau classique. Outre le fait que cet estimateur possède des propriétés théoriques intéressantes, les différentes simulations effectuées ont permis de constater qu'il conduisait à des résultats nettement meilleurs que ceux de l'estimateur de Grenander, ce dernier étant le principal estimateur unimodal étudié à ce jour. Il apporte enfin une amélioration conséquente à l'estimation à noyau classique, dans la mesure où il présente une certaine robustesse vis-à-vis du paramètre de lissage.

Ces différents résultats positifs motivent une recherche plus approfondie à différents niveaux : d'un point de vue pratique tout d'abord, une étude expérimentale plus complète est à envisager, qui fera l'objet d'un travail dans un futur proche, mettant en compétition d'autres estimateurs unimodaux que celui de Grenander, comme par exemple celui récemment proposé par Datta (1995), et dans laquelle le problème du choix du paramètre de lissage serait étudié. Il conviendrait également dans cette plus large comparaison d'inclure la situation, dont l'intérêt pratique est clair, où le mode est inconnu. L'extension à ce cas pourrait notamment s'envisager via une méthode indirecte, utilisant au préalable un estimateur du mode. Un développement théorique parallèle à la mise en œuvre de cette procédure s'impose de façon évidente, et constituera le sujet d'une prochaine étude. Un travail auquel nous souhaitons finalement nous intéresser consiste d'une part à compléter les résultats obtenus concernant le comportement asymptotique local de l'estimateur que nous avons proposé, et d'autre part à obtenir un résultat assurant la supériorité de $f_{n\downarrow}$ par rapport à f_n en terme d'erreur L_1.

Annexe 1

Démonstration du Théorème 3

L'estimateur à noyau f_n conservant les propriétés de régularité de K (voir par exemple Bosq et Lecoutre (1987)), il suffit de montrer que le réarrangement décroissant préserve la continuité. Soit g une densité bornée continue sur $\mathbb{R}$; remarquons tout d'abord que $r : u \mapsto \lambda\{t : g(t) > u\}$ est strictement décroissante sur $(0, B]$, où B est le maximum de g. Dans le cas contraire, il existe en effet $u \in (0, B]$ et $\delta > 0$ tels que $r(u) = r(u + \delta)$, ce qui entraîne que $\lambda\{t : u < g(t) \leq u + \delta\} = 0$; la continuité de g rend cette égalité impossible. Supposons que $g_\downarrow$ ait une discontinuité en $s \in \mathbb{R}^+$; il existe $\varepsilon > 0$ tel que pour tout $t < s$, $r(v) \leq t \Rightarrow r(v - \varepsilon) \leq s$; soit encore, par passage à la limite quand t croît vers s :

$$\exists \varepsilon > 0 \, : \, r(v) \leq s \Rightarrow r(v - \varepsilon) \leq s. \tag{3}$$

Deux cas de figure se présentent : si s est atteint, c'est-à-dire s'il existe $\bar{v}$ tel que $r(\bar{v}) = s$, alors la relation (3) entraîne l'existence d'un $\varepsilon > 0$ tel que $r(\bar{v} - \varepsilon) \leq r(\bar{v})$, ce qui est en contradiction avec la décroissance stricte de r. Si s n'est pas atteint, la continuité à droite de r en tout point entraîne qu'il existe u tel que $r(u) \leq s$ et tel que $r(u - \delta) > s$ pour tout $\delta > 0$. Pour $\delta = \varepsilon$, on obtient une contradiction avec la relation (3).

Annexe 2

Démonstration du Théorème 4

Dans cette démonstration, nous montrons la convergence presque-sûre uniforme sur $\mathbb{R}^+$ de $f_{n\downarrow}$ vers $f_\downarrow$, pour toute densité f continue bornée. Lorsque cette densité est de plus définie sur $\mathbb{R}^+$ et décroissante, elle coïncide avec son réarrangement décroissant $f_\downarrow$, et l'on obtient de facto le théorème 4.

Les fonctions $f_{n\downarrow}$ et $f_\downarrow$ étant décroissantes, l'étude de la convergence p.s. uniforme sur $\mathbb{R}^+$ se réduit à celle de la convergence p.s. uniforme sur tout compact de $\mathbb{R}^+$. La monotonie de $f_{n\downarrow}$ et $f_\downarrow$, ainsi que la continuité de $f_\downarrow$ (conséquente au théorème 3) assure la convergence p.s. uniforme sur tout compact de $f_{n\downarrow}$ vers $f_\downarrow$, dès que l'on a convergence p.s. ponctuelle de $f_{n\downarrow}$ vers $f_\downarrow$ (lemme de Dini). De plus, les applications $f_{n\downarrow}$ et $f_\downarrow$ étant par définition les inverses respectifs des fonctions décroissantes r_n et r, il suffit de montrer la convergence de r_n vers r en tout point de continuité de r (voir par exemple Billingsley (1979, p. 287)).

Soit t un point de continuité de r et soit $\varepsilon > 0$; l'application f étant intégrable, il existe un compact $[a,b]$ tel que $|r(t) - \lambda\{x \in [a,b] : f(x) > t\}| \leq \varepsilon$. Notons $\tilde{r}(t) = \lambda\{x \in [a,b] : f(x) > t\}$, et $\tilde{r}_n(t) = \lambda\{x \in [a,b] : f_n(x) > t\}$. Alors $r_n(t) \geq \tilde{r}_n(t)$ p.s., et par ailleurs, $\tilde{r}$ est continue en t dès que r l'est. Montrons dans un premier temps la convergence p.s. de $\tilde{r}_n(t)$ vers $\tilde{r}(t)$; la convergence p.s. uniforme sur tout compact de f_n vers f, assurée par le théorème 2, entraîne, à partir d'un certain rang, l'inégalité $\tilde{r}_n(t) \leq \tilde{r}(t-\varepsilon)$ p.s., et l'on en déduit $\limsup_{n\to\infty} \tilde{r}_n(t) \leq \tilde{r}(t-\varepsilon)$ p.s. On obtient de façon analogue l'inégalité $\liminf_{n\to\infty} \tilde{r}_n(t) \geq \tilde{r}(t+\varepsilon)$ p.s., et la continuité de $\tilde{r}$ en t permet de conclure que $\lim_{n\to\infty} \tilde{r}_n(t) = \tilde{r}(t)$ p.s.

La définition de $\tilde{r}$ et l'inégalité $r_n(t) \geq \tilde{r}_n(t)$ p.s. entraînent alors immédiatement $\liminf_{n\to\infty} r_n(t) \geq r(t)$ p.s. On obtient finalement l'inégalité $\limsup_{n\to\infty} r_n(t) \leq r(t)$ p.s. en raisonnant par l'absurde et en utilisant le même type d'arguments que précédemment.

Bibliographie

Abou-Jaoudé, S. (1974). Sur un théorème de non-existence d'estimateurs convergeant en probabilité. *C. R. Acad. Sc. Paris Sér. A* **278**, 1445-1448.

Barlow, R.E., D.J. Bartholomew, J.M. Bremner, and H.D. Brunk (1972). *Statistical Inference under Order Restrictions.* Wiley, New York.

Berlinet, A. and L. Devroye (1990). Estimation d'une densité: Un point sur la méthode du noyau. *Statist. anal. données* **14**, 1-32.

Bertin, E., I. Cuculescu, and R. Theodorescu (1996). *Unimodality of Probability Measures.* Kluwer, New York (in print).

Billingsley, P. (1979). *Probability and Measure.* Wiley, New York.

Birgé, L. (1989). The Grenander estimator: A nonasymptotic approach. *Ann. Statist.* **17**, 1532-1549.

Bosq, D. and J.-P. Lecoutre (1987). *Théorie de l'estimation fonctionnelle.* Econometrica, Paris.

Csörgő, M. and P. Révész (1981). *Strong Approximations in Probability and Statistics.* Academic Press, New York.

Datta, S. (1995). A minimax optimal estimator for continuous monotone densities. *J. Statist. Plann. Inf.* **46**, 183-193

Devroye, L. (1987). *A Course in Density Estimation.* Birkhäuser, Boston.

Devroye, L. and L. Györfi (1985). *Nonparametric Density Estimation: The L_1 View.* Wiley, New York.

Dharmadhikari, S. and K. Joag-Dev (1988). *Unimodality, Convexity and Applications.* Academic Press, New York.

Fougères, A.-L. (1996). *Estimation non paramétrique de densités unimodales et de fonctions de dépendance.* Thèse de doctorat inédite, Université Paul Sabatier, Toulouse (France).

Glick, N. (1974). Consistency conditions for probability estimators and integrals of density estimators. *Utilitas Mathematica* **6**, 61-74.

Grenander, U. (1956). On the theory of mortality measurement. *Skand. Aktuarietidskr.* **39**, 125-153.

Groeneboom, P. (1985). *Estimating Monotone Density.* Proceedings of the Berkeley Conference in Honor of J. Neyman and J. Kiefer **2**. L. M. Le Cam, R. A. Olsen (Eds). Wadsworth. 539-555.

Groeneboom, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probab. Th. Rel. Fields* **81**, 79-109.

Groeneboom, P. and R. Pyke (1983). Asymptotic normality of statistics based on the convex minorant of empirical distribution functions. *Ann. Probab.* **11**, 328-345.

Hardy, G., J. Littlewood and G. Pólya (1952). *Inequalities.* University Press, Cambridge.

Komlós, J., P. Major and G. Tusnády (1975). An approximation of partial sums of independent R. V.'s and the sample DF. I. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **32**, 111-131.

Prakasa Rao, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31**, 23-36.

Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes.* Springer-Verlag, New York.

Robertson, T., F.T. Wright, and R.L. Dykstra (1988). *Order Restricted Statistical Inference.* Wiley, New York.

Shorack, G. R. and J. A. Wellner (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

Wang, Y. (1995). The L_1 theory of estimation of monotone and unimodal densities. *Nonparam. Statist.* **4**, 249-261.