

Estimation de la densité par les histogrammes et les estimateurs à noyau *

Raymond Moché †

10 juillet 2006

1 Position du problème

Le problème s'énonce très simplement :

on dispose d'un échantillon (x_1, \dots, x_n) de taille n d'une loi à densité f inconnue. Que peut-on dire de f à partir dudit échantillon ?

Précisons que dire que (x_1, \dots, x_n) est un échantillon d'une loi à densité inconnue f signifie que nous supposons qu'il s'agit des valeurs prises simultanément par n variables aléatoires (X_1, \dots, X_n) , indépendantes et suivant cette loi de densité f . Plus précisément encore, mais cette fois nous entrons dans le formalisme du Calcul des Probabilités, nous supposons qu'il existe un espace de probabilité (Ω, \mathcal{A}, P) , des variables aléatoires (X_1, \dots, X_n) définies sur cet espace de probabilité, suivant la loi de densité f et indépendantes et un point ω de Ω tels que ¹

$$(1) \quad \boxed{(x_1, \dots, x_n) = (X_1(\omega), \dots, X_n(\omega))}.$$

(x_1, \dots, x_n) s'appelle l'échantillon observé, (X_1, \dots, X_n) l'échantillon théorique.

Le problème de l'estimation de la densité inconnue f est très difficile ² et n'a reçu que

*Atelier tenu lors des Journées académiques de l'IREM de Lille sur le thème du Hasard les 15 et 16 avril 2004 au CRDP du Nord-Pas de Calais, 3 rue Jean Bart à Lille.

†Raymond.Moche@univ-lille1.fr

¹Par exemple, les fonctions rand des caulettes fournissent des suites de nombres qui peuvent être interprétées comme une suite $(X_1(\omega), \dots, X_n(\omega))$, X_1, \dots, X_n étant des variables aléatoires indépendantes uniformément réparties sur $[0, 1]$, c'est à dire admettant comme densité la fonction égale à 1 sur $[0, 1]$ et à 0 ailleurs.

²parce que nous allons nous placer dans le cadre général de la statistique *non-paramétrique*. Dans les livres de statistique élémentaire, c'est beaucoup plus simple parce que l'on se place dans le cadre de la *statistique paramétrique*. On suppose par exemple que l'échantillon considéré suit une loi à densité inconnue, mais qui est une loi normale. Comme les lois normales sont entièrement déterminées par

des solutions asymptotiques (lorsque l'on peut faire grandir indéfiniment la taille n de l'échantillon) sous des hypothèses restrictives. On se réservera donc de faire tendre vers $+\infty$ la taille n de l'échantillon. Voilà pourquoi on supposera disposer d'un échantillon infini $(x_n, n \geq 1)$. L'équivalent de l'égalité (1) s'écrira dans ce cadre

$$(2) \quad \boxed{(x_n, n \geq 1) = (X_n(\omega), n \geq 1).}$$

A taille d'échantillon fixée, on ne peut avoir qu'une idée vague de f , grâce à divers estimateurs dont l'histogramme, bien connu au lycée.

1.1 Distances entre des courbes

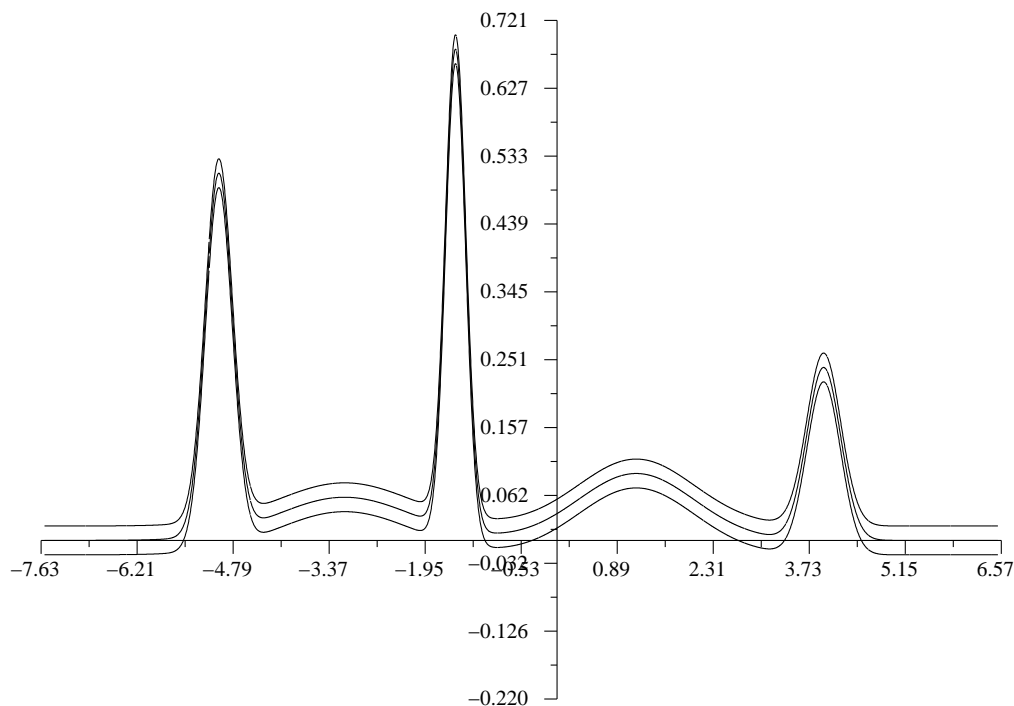
L'étude de la convergence vers f d'une suite $(\hat{f}_n, n \geq 1)$ d'estimateurs de f , par exemple une suite d'histogrammes, posera le problème de la définition d'une distance d entre f et \hat{f}_n ³. Nous n'utiliserons que la distance d_{sup} de la convergence uniforme, que l'on peut utiliser si la densité f et l'estimateur \hat{f}_n sont des fonctions bornées. Cette distance est définie par

$$(3) \quad d_{sup}(f, \hat{f}_n) = \sup_{x \in \mathbf{R}} |f(x) - \hat{f}_n(x)|.$$

A titre d'exemple, nous avons tracé ci-dessous le graphe d'une densité de probabilité notée f , ainsi que les graphes translatés parallèlement à l'axe des ordonnées de $+0.3$ puis de -0.3 . Une fonction g est à une distance de la convergence uniforme de f inférieure ou égale à 0.3 si et seulement si son graphe se trouve dans le tuyau dont le centre est le graphe de f .

deux paramètres, il suffit d'estimer ces paramètres. Ce cas est particulièrement simple parce que les paramètres en question sont l'espérance et la variance que l'on estime à l'aide de l'espérance empirique et de la variance empirique. Ceci dit, à taille d'échantillon fixée, on ne peut avoir d'estimation exacte. Les résultats exacts sont nécessairement des résultats asymptotiques.

³Il s'agira donc d'une distance entre les graphes de deux fonctions.



1.2 Convergences stochastiques

De plus, nous verrons que $d(f, \hat{f}_n)$ est une variable aléatoire ⁴. Il en résulte que la suite $(d(f, \hat{f}_n), n \geq 1)$ pourra converger ou non vers 0 suivant divers modes de convergence stochastiques. Comme nous voulons restreindre au maximum le sujet traité, nous n'utiliserons que la convergence presque sûre qui est la convergence ordinaire pour presque tous les échantillons. En fait, nous la confondrons avec cette dernière car les probabilistes considèrent qu'un événement de probabilité 0 ne se produit pas, dans la pratique.

Résultats espérés

On attend des résultats du style :

Si la densité f appartient à telle classe \mathcal{F} de densités de probabilité bornées, si $(\hat{f}_n, n \geq 1)$ désigne une suite d'estimateurs bornés de f , la suite $(d_{sup}(f, \hat{f}_n), n \geq 1)$ converge presque sûrement vers 0.

\mathcal{F} s'appelle en général l'ensemble des densités possibles. Par exemple, si on suppose que

⁴C'est évident, en fait, puisque \hat{f}_n est calculée à partir de l'échantillon qui dépend du hasard.

la loi inconnue de l'échantillon est une loi normale, \mathcal{F} est l'ensemble des densités de la forme $x \mapsto \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$. Plus loin, on citera un théorème dans lequel \mathcal{F} est l'ensemble des densités uniformément continues sur \mathbf{R} . On voit que le sujet, qui actuellement peut être considéré comme épuisé comme thème de recherche, est foisonnant : lorsque l'on veut étudier une suite d'estimateurs, on recherchera une classe \mathcal{F} de densités de probabilité possibles ⁵, une distance d ⁶ et un mode de convergence stochastique ⁷ qui satisfont un énoncé du type précédent, ce qui fait beaucoup de théorèmes possibles!

En simplifiant au maximum, nous allons essayer de montrer comment des idées qui paraissent maintenant simples se sont imposées peu à peu et ont permis d'obtenir des résultats assez satisfaisants .

Nous nous contenterons de faire des simulations et d'observer le comportement asymptotique de quelques estimateurs de la densité calculés à partir d'échantillons simulés dont nous connaissons la loi, donc la densité f , ce qui nous permettra de savoir de manière très empirique s'ils convergent ou non vers f et d'avoir une idée des vitesses de convergence.

Ces simulations seront faites à l'aide de scilab ⁸.

2 Rappels sur les densités de probabilité

Nous rappelons qu'une variable aléatoire X admet une fonction donnée f comme densité de probabilité si quels que soient les réels a et b tels que $a < b$, la probabilité pour que X prenne ses valeurs dans l'intervalle $[a, b]$ est égale à $\int_a^b f(x) dx$, ce que l'on note

$$(4) \quad P(a \leq X \leq b) = \int_a^b f(x) dx.$$

Il s'ensuit que la fonction f prend nécessairement ses valeurs dans l'intervalle $[0, +\infty[$, qu'elle est intégrable sur \mathbf{R} et que $\int_{-\infty}^{+\infty} f(x) dx = 1$ ⁹.

Les programmes actuels ne font référence qu'à deux sortes de lois à densité : la loi uniforme sur l'intervalle $[0, 1]$ et les lois exponentielles. Mais il faut bien comprendre que

⁵La classe \mathcal{F} devra être la plus grande possible.

⁶parce que d'autres distances que la distance d_{sup} sont intéressantes

⁷parce que la convergence presque sûre est une convergence forte et que l'on peut passer à côté de résultats moins bons mais intéressants quand même en utilisant des modes de convergence stochastique moins exigeants

⁸SCILAB, développé en France par l'INRIA et l'ENPC, est téléchargeable à l'adresse : <http://scilabsoft/inria.fr/>.

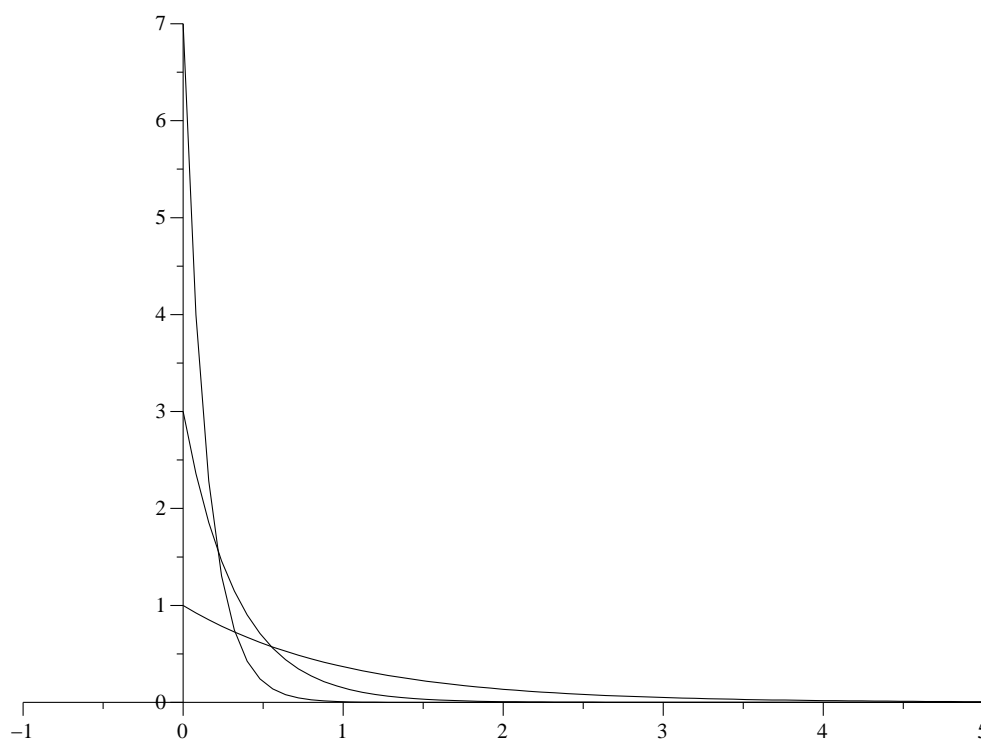
⁹Bien sûr, j'aurais dû dire que la fonction f était supposée mesurable, que les égalités (4) impliquent seulement que f est presque sûrement à valeurs dans $[0, +\infty[$ et que les intégrales utilisées sont des intégrales de Lebesgue. Mais il se trouve que ces intégrales sont aussi des intégrales de Riemann dans tous les cas usuels. Nous omettons ces précisions dans la suite. Tout ce qui est écrit dans ce papier est exact et se justifie dans le cadre de l'intégrale de Lebesgue.

l'ensemble des lois de probabilité qui admettent une densité est infini. Deux densités de probabilité, c'est à dire deux fonctions ≥ 0 dont l'intégrale sur \mathbf{R} est égale à 1 définissent deux lois de probabilité différentes dès qu'elles prennent des valeurs différentes sur un intervalle de longueur > 0 , même si cette longueur est extrêmement petite.

Dans ce qui suit, les densités qui seront considérées ou qui apparaîtront à la suite de calculs seront toujours simples, plus précisément continues, sauf peut-être en un nombre fini de points de discontinuité où elles auront une limite à droite et une limite à gauche.

Nous allons maintenant exhiber quelques-unes d'entre elles. D'autres, fort nombreuses, apparaîtront ensuite.

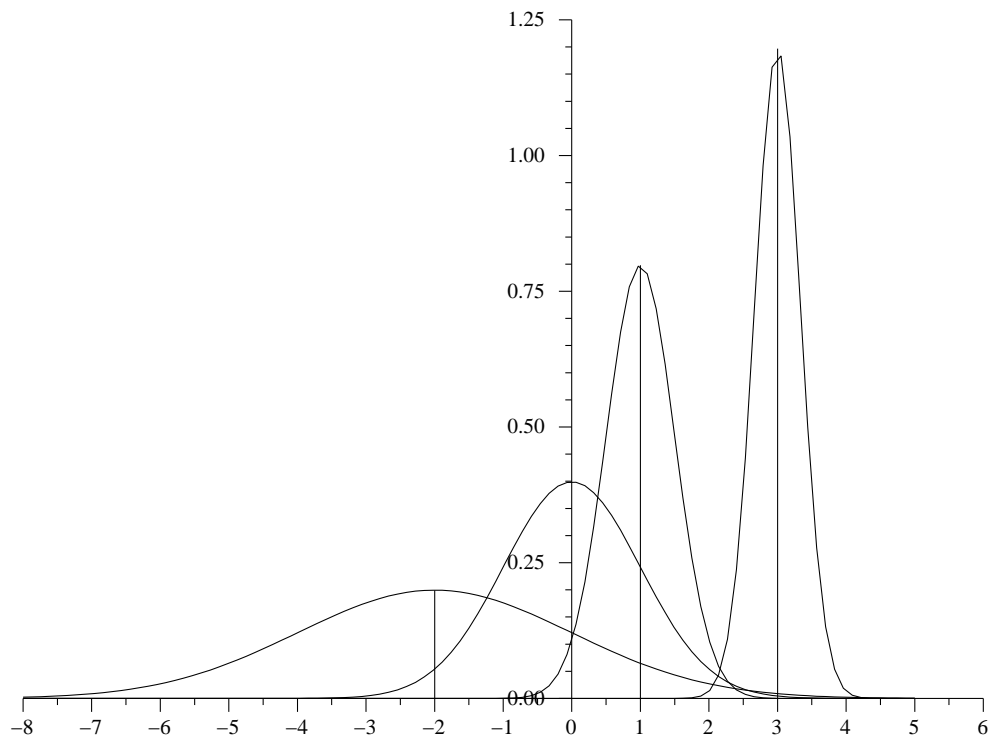
Lois exponentielles Voici maintenant le graphe des densités des lois $\mathcal{E}(1)$, $\mathcal{E}(3)$ et $\mathcal{E}(7)$ (c'est à dire les graphes des densités de variables aléatoires suivant ces lois). Il s'agit des fonctions nulles sur $] -\infty, 0[$ et égales respectivement à $\exp(-x)$, $3 \exp(-3x)$ et $7 \exp(-7x)$ sur $[0, +\infty[$. On les reconnaît facilement par leur valeur en 0.



Lois normales

Nous avons tracé successivement (de la gauche vers la droite) les graphes de la densité de

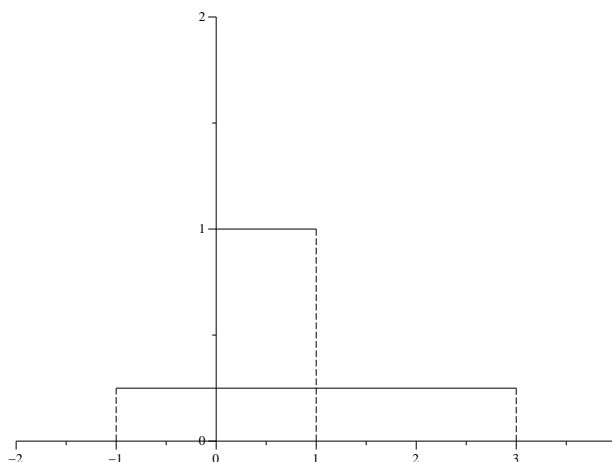
la loi normale $\mathcal{N}(-2, 4)$, $\mathcal{N}(0, 1)$, qui est la célèbre courbe en cloche, $\mathcal{N}(1, \frac{1}{4})$ et $\mathcal{N}(3, \frac{1}{4})$. Le second paramètre désigne la variance et non l'écart-type ¹⁰.



Lois uniformes

Sont tracés ci-dessous les graphes des densités de la loi uniforme sur $[0, 1]$ et sur $[-1, 3]$, notées respectivement $\mathcal{U}(0, 1)$ et $\mathcal{U}(-1, 3)$.

¹⁰Pour des raisons irréfutables concernant les lois normales multidimensionnelles, c'est une erreur de noter $\mathcal{N}(m, \sigma)$ la loi normale de moyenne m et d'écart-type σ : il faut la noter $\mathcal{N}(m, \sigma^2)$.

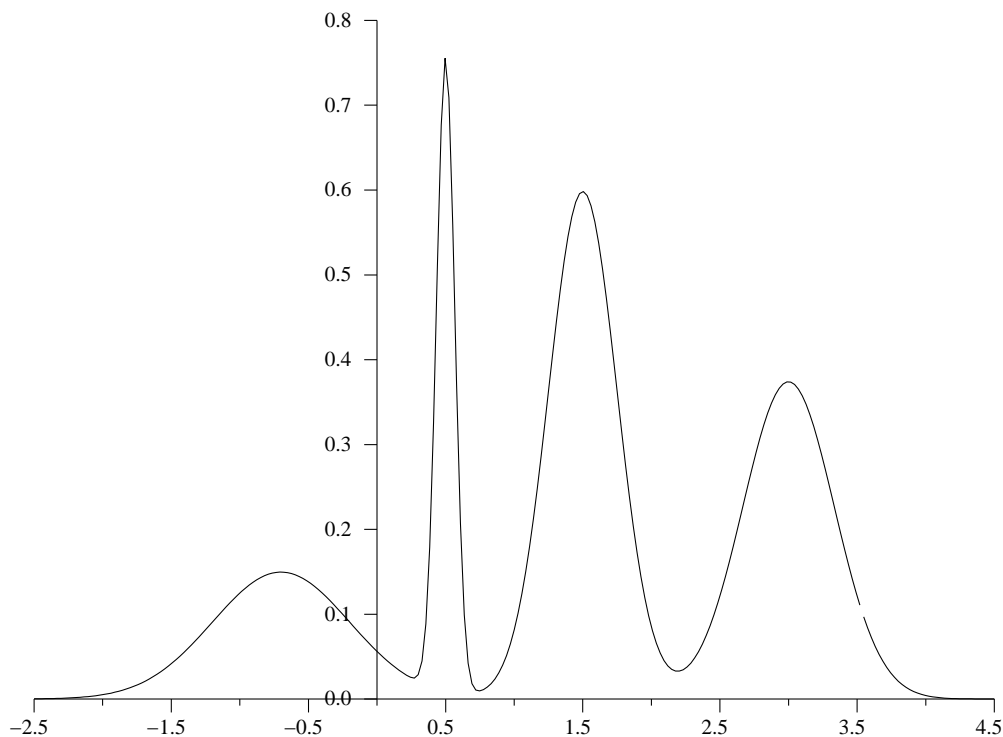


Mélanges de lois normales

La densité qui suit est la combinaison linéaire des densités des lois normales $\mathcal{N}(-0.7, \frac{1}{4})$, $\mathcal{N}(0.5, \frac{1}{225})$, $\mathcal{N}(1.5, \frac{1}{16})$ et $\mathcal{N}(3, \frac{1}{9})$ affectées respectivement des coefficients $\frac{3}{16}$, $\frac{1}{8}$, $\frac{3}{8}$, $\frac{5}{16}$. Plus généralement, on appelle *mélange de lois normales* toute loi de probabilité à densité de la forme

$$(5) \quad x \mapsto \sum_{k=1}^n c_k \cdot \frac{1}{\sqrt{2\pi} \cdot \sigma_k} \cdot \exp\left(-\frac{(x - m_k)^2}{2\sigma_k^2}\right)$$

où c_1, \dots, c_n sont des réels > 0 de somme 1, m_1, \dots, m_k des réels quelconques et $\sigma_1, \dots, \sigma_n$ des réels > 0 .



On verra que tous les estimateurs de la densité à noyau gaussien sont des densités de mélanges de lois normales.

3 Histogrammes

3.1 Histogrammes de la loi $\mathcal{U}(0, 1)$

On se place ici dans le cas où l'ensemble \mathcal{F} des densités possibles est réduit à la densité $f = \mathbf{1}_{[0,1]}$ de la loi $\mathcal{U}(0, 1)$ ¹¹. Evidemment, dans ce cas, il n'est pas utile d'estimer la densité f puisqu'elle est connue! Néanmoins, cette étude exploratoire nous montrera divers défauts et qualités de l'histogramme, d'où son intérêt.

Nous partons donc d'un échantillon $(x_n, n \geq 1)$ de la loi $\mathcal{U}(0, 1)$. Quelle que soit la taille n de l'échantillon (x_1, \dots, x_n) , les nombres x_1, \dots, x_n appartiennent à l'intervalle $]0, 1[$ et sont tous différents¹² parce que la probabilité qu'il n'en soit pas ainsi est nulle.

¹¹c'est à dire, en clair, que l'on suppose que l'échantillon a été tiré suivant la loi $\mathcal{U}(0, 1)$.

¹²En fait, comme les calculettes ou les ordinateurs ne travaillent qu'avec un nombre fini (très grand) de décimales dont très peu sont affichées, la probabilité d'obtenir deux nombres égaux dans l'échantillon n'est pas nulle. Si on utilise des échantillons de très grande taille, cette probabilité peut même devenir

Nous allons ci-dessous calculer l'histogramme à partir d'un échantillon de taille $n = 10, 100, 1000, 10000, 50000$ et 1000000 ¹³ successivement lorsque l'intervalle $[0, 1]$ est divisé en 50 classes : $[0, 0.02], \dots,]0.98, 1]$.

Ces classes sont notées $I_j, j = 1, \dots, N$ dans les formules ci-dessous. l désigne leur longueur commune ($l = 0.02$).

n et N ayant été choisis, on rappelle que l'histogramme correspondant est constant sur chacune des classes I_j où il vaut

$$(6) \quad \bar{f}_{n,N,j} = \frac{1}{nl} \sum_{i=1}^n \mathbf{1}_{I_j}(x_i)$$

(il s'agit de l'effectif de la classe I_j divisés par le produit nl). L'histogramme est la fonction de \mathbf{R} dans \mathbf{R}^+ définie par¹⁴ :

$$(7) \quad x \mapsto \bar{f}_{n,N}(x) = \sum_{j=1}^N \bar{f}_{n,N,j} \cdot \mathbf{1}_{I_j}(x),$$

ou, si l'on n'écrit pas la variable x

$$(8) \quad \boxed{\bar{f}_{n,N} = \sum_{j=1}^N \bar{f}_{n,N,j} \cdot \mathbf{1}_{I_j}.}$$

En fait, les notations (6), (7) et (8) sont trop imprécises car elles ne montrent pas que l'histogramme est calculé à partir de l'échantillon (x_1, \dots, x_n) , autrement dit, qu'il dépend du hasard. En fait, $\bar{f}_{n,N}(x)$ aurait dû être noté

$$(9) \quad \bar{\bar{f}}_{n,N}(x_1, \dots, x_n; x) = \sum_{j=1}^N \frac{1}{nl} \left(\sum_{i=1}^n \mathbf{1}_{I_j}(x_i) \right) \cdot \mathbf{1}_{I_j}(x)$$

ou, en utilisant (1),

$$(10) \quad \check{\check{f}}_{n,N}(\omega, x) = \sum_{j=1}^N \left(\frac{1}{nl} \sum_{i=1}^n \mathbf{1}_{I_j}(X_i(\omega)) \right) \cdot \mathbf{1}_{I_j}(x),$$

autrement dit, en n'écrivant pas la variable ω ,

$$(11) \quad \boxed{\check{\check{f}}_{n,N}(x) = \sum_{j=1}^N \left(\frac{1}{nl} \sum_{i=1}^n \mathbf{1}_{I_j}(X_i) \right) \cdot \mathbf{1}_{I_j}(x) \quad \text{ou} \quad \check{\check{f}}_{n,N} = \sum_{j=1}^N \left(\frac{1}{nl} \sum_{i=1}^n \mathbf{1}_{I_j}(X_i) \right) \cdot \mathbf{1}_{I_j}.}$$

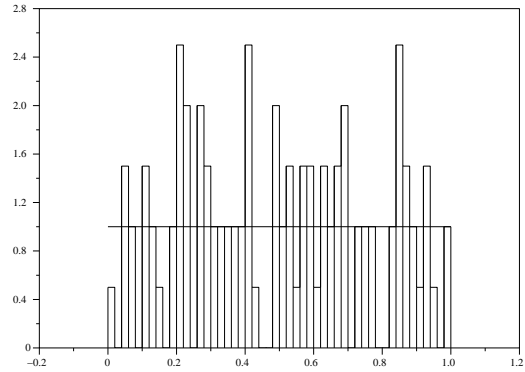
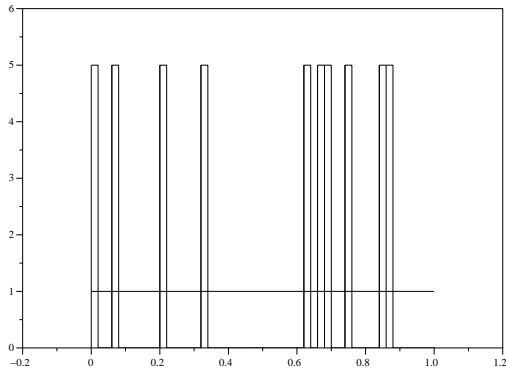
grande. Mais on sait aussi que la taille des échantillons que l'on peut engendrer valablement avec un générateur donné est limitée, à cause de la périodicité de ce générateur! Il y a là tout une série de problèmes redoutables que nous préférons ignorer.

¹³Chaque échantillon a été obtenu en ajoutant à l'échantillon précédent de nouveaux nombres tirés au hasard suivant la loi $\mathcal{U}(0, 1)$, c'est à dire calculés par l'instruction "rand" ou "grand" de scilab, en donnant toujours la même valeur initiale ("seed") au générateur.

¹⁴On dit que les histogrammes sont des fonctions en escalier.

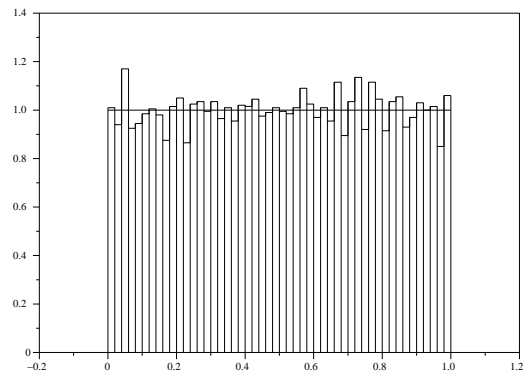
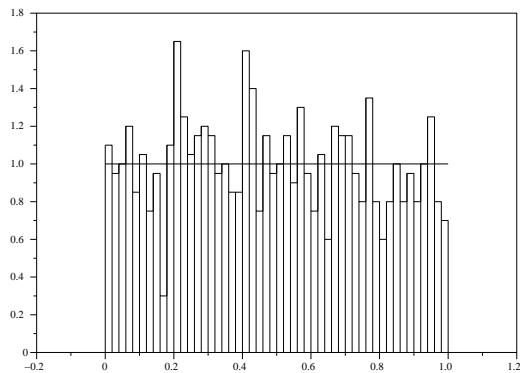
**On voit bien maintenant le caractère aléatoire des histogrammes.
Remarquons aussi que ce sont des densités de probabilité.**

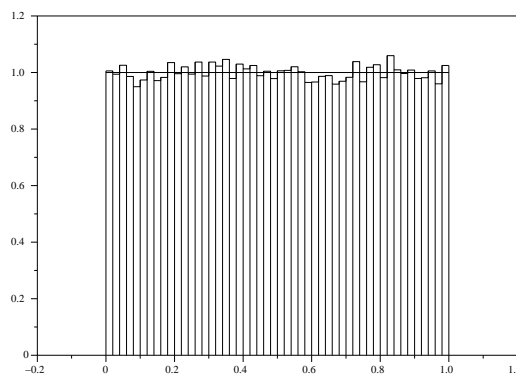
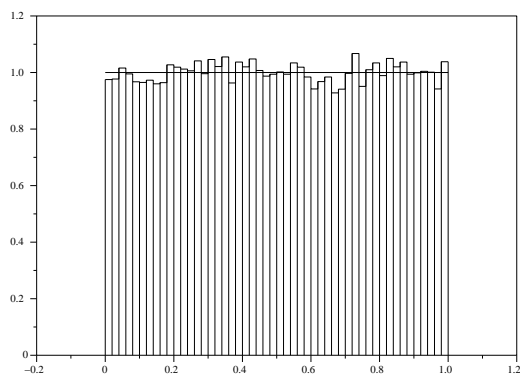
Voici maintenant les graphes des 6 histogrammes annoncés, complétés par le graphe de f .



Premier graphe : $nl = 10 \times 0.02 = 0.2$. Dans une classe qui contient 1 point, la valeur de l'histogramme est $\frac{1}{0.2} = 5$. Il y a 10 classes de cette sorte ci-dessus.

Deuxième graphe : $nl = 100 \times 0.02 = 2$. Il y a des classes contenant 0,1,2,3,4 et 5 points.





On constate que ces histogrammes se rapprochent de plus en plus du graphe de f au sens de la convergence uniforme quand la taille n de l'échantillon grandit. La convergence semble lente.

Justification des résultats précédents

Etudier la convergence uniforme vers f de la suite des histogrammes $(\check{f}_{n,N}, n \geq 1)$ quand n tend vers $+\infty$, N étant fixé, revient à étudier la convergence vers 0 de la suite des variables aléatoires

$$(12) \quad d_{sup}(f, \check{f}_{n,N}) = \sup_{0 \leq x \leq 1} |f(x) - \check{f}_{n,N}(x)|,$$

parce que les fonctions f et $\check{f}_{n,N}$ sont bornées et nulles en dehors de l'intervalle $[0, 1]$. Pour $j = 1, \dots, N$, comme $(\mathbf{1}_{I_j}(X_i), i \geq 1)$ est une suite de variables aléatoires indépendantes¹⁵, ne prenant que les valeurs 1 et 0, avec les probabilités l et $1 - l$ ¹⁶, d'après la loi forte des grands nombres¹⁷,

$$(13) \quad \frac{1}{nl} \sum_{i=1}^n \mathbf{1}_{I_j}(X_i) - 1 \xrightarrow[n \rightarrow +\infty]{p.s.} 0.$$

Comme $d_{sup}(f, \check{f}_{n,N}) = \sup_{j=1, \dots, N} \left| \frac{1}{nl} \sum_{i=1}^n \mathbf{1}_{I_j}(X_i) - 1 \right| \leq \sum_{j=1}^N \left| \frac{1}{nl} \sum_{i=1}^n \mathbf{1}_{I_j}(X_i) - 1 \right|$, la suite majorante converge vers 0 presque sûrement, donc

$$(14) \quad d_{sup}(f, \check{f}_{n,N}) \xrightarrow[n \rightarrow +\infty]{p.s.} 0,$$

¹⁵ parce que les variables aléatoires $X_i, i \geq 1$ sont indépendantes

¹⁶ Elles suivent donc la même loi de Bernoulli.

¹⁷ Si $(X_n, n \geq 1)$ est une suite de variables aléatoires indépendantes et de même loi dont l'espérance m existe,

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow +\infty]{p.s.} m.$$

ce qu'il fallait démontrer. Les résultats précédents auraient été conservés si nous avions choisi un partage quelconque de $[0, 1]$ en classes fixes, la démonstration restant inchangée. Nous en déduisons pratiquement que, quel que soit l'échantillon $(x_n, n \geq 1)$, la suite des histogrammes converge uniformément vers la densité de la loi $\mathcal{U}(0, 1)$, à savoir $\mathbf{1}_{[0,1]}$. Par contre, on ne peut rien dire sur la vitesse de convergence, qui semble lente. De toute façon, on ne peut espérer que des résultats du genre :

si la taille de l'échantillon est assez grande, la distance $d_{sup}(f, \check{f}_{n,N})$ devient plus petite que tant, avec au moins telle probabilité.

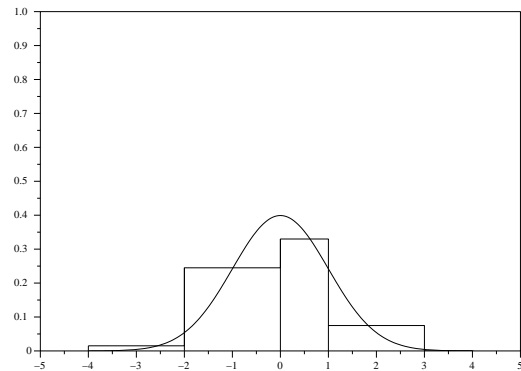
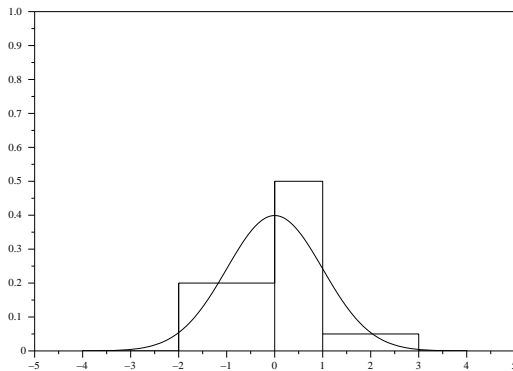
Rappelons encore une fois que l'intérêt de ces considérations est uniquement exploratoire : il n'y a pas ici de problème d'estimation de la densité f parce que f est connue.

3.2 Histogrammes de la loi $\mathcal{N}(0, 1)$

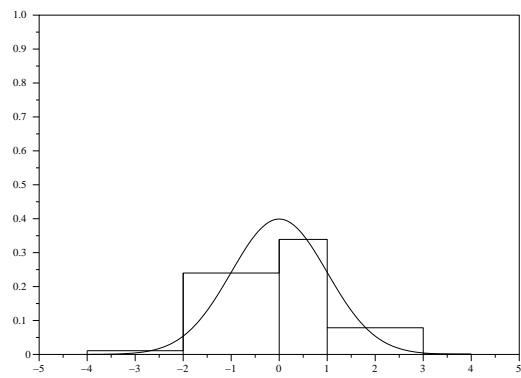
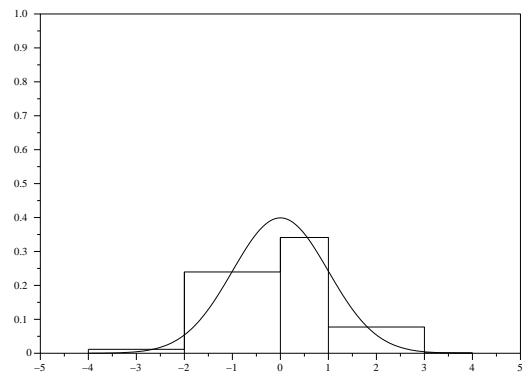
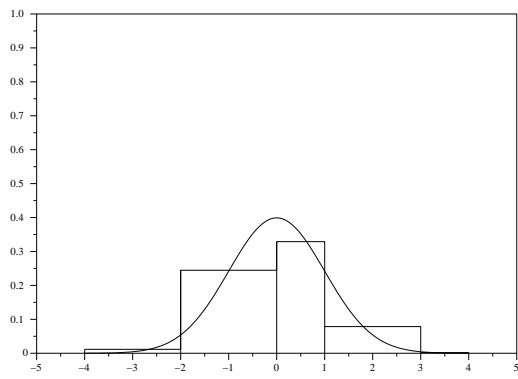
Nous supposons maintenant que l'ensemble \mathcal{F} des densités possibles est réduit à la densité $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$ de la loi normale centrée réduite ¹⁸, toujours dans un but exploratoire, le cas précédemment étudié de la loi $\mathcal{U}[0, 1]$, trop simple, masquant les problèmes de convergence des histogrammes.

Nous utilisons donc dans cette division des échantillons de $\mathcal{N}(0, 1)$ et traçons le graphe de l'histogramme ainsi que le graphe de la densité de cette loi en donnant successivement à la taille de l'échantillon les valeurs 10, 100, 1000, 10000 et 100000, l'histogramme étant construit sur l'intervalle $[-4, 4]$ partagé en $[-4, -2]$, $[-2, 0]$, $[0, 1]$, $[1, 3]$ et $[3, 4]$ dans le premier cas, en 20 intervalles de même longueur dans le second.

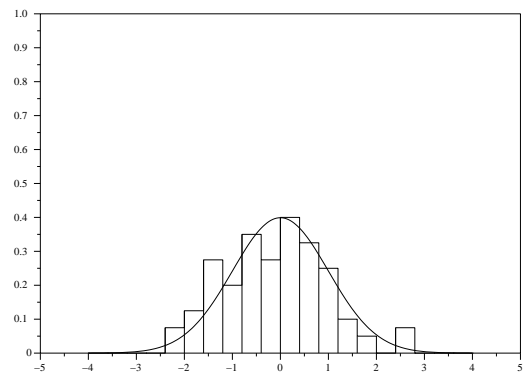
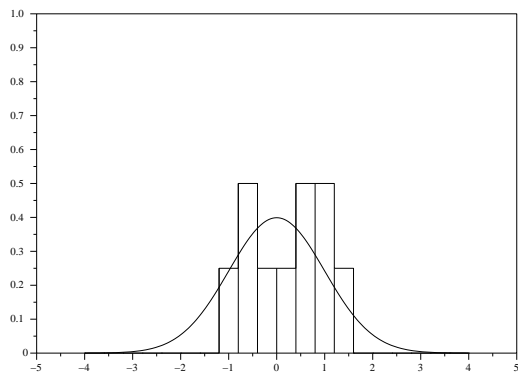
Premier partage de $[-4, 4]$

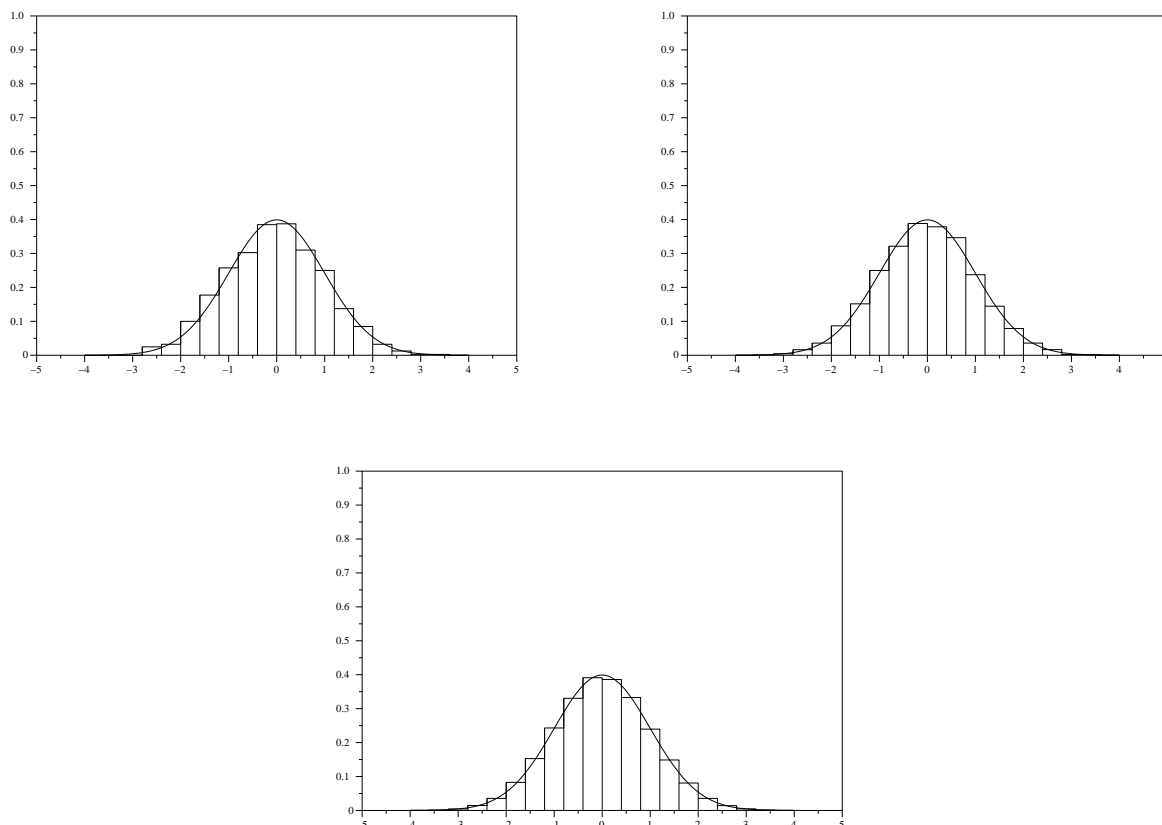


¹⁸c'est à dire que nous supposons que l'échantillon a été tiré suivant la loi $\mathcal{N}(0, 1)$.



Second partage de $[-4, 4]$





On constate qu'il n'y a apparemment aucun espoir que l'histogramme converge vers f . En fait, les deux suites d'histogrammes semblent avoir une limite qui est une fonction en escalier. Effectivement, il est facile de démontrer, de nouveau à l'aide de la loi forte des grands nombres, que l'histogramme converge uniformément sur $[-4, 4]$, au sens de la convergence presque sûre, vers la fonction en escalier

$$x \mapsto \varphi(x) = \sum_{j=1}^N \frac{\alpha_j}{l_j} \cdot \mathbf{1}_{I_j}(x)$$

où (I_1, \dots, I_N) désigne le partage en classes de l'intervalle $[-4, 4]$ et où, pour $j = 1, \dots, N$, α_j désigne la probabilité pour qu'une variable aléatoire X suivant la loi $\mathcal{N}(0, 1)$ prenne ses valeurs dans l'intervalle I_j , dont la longueur est notée l_j , soit

$$\forall j \in \{1, \dots, N\}, \quad \alpha_j = P(X \in I_j) = \int_{I_j} \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2}{2}\right) dx.$$

On remarquera ¹⁹ que φ n'est pas une densité de probabilité, son intégrale valant

$$\sum_{j=1}^N \alpha_j = P(X \in [-4, 4]) < 1.$$

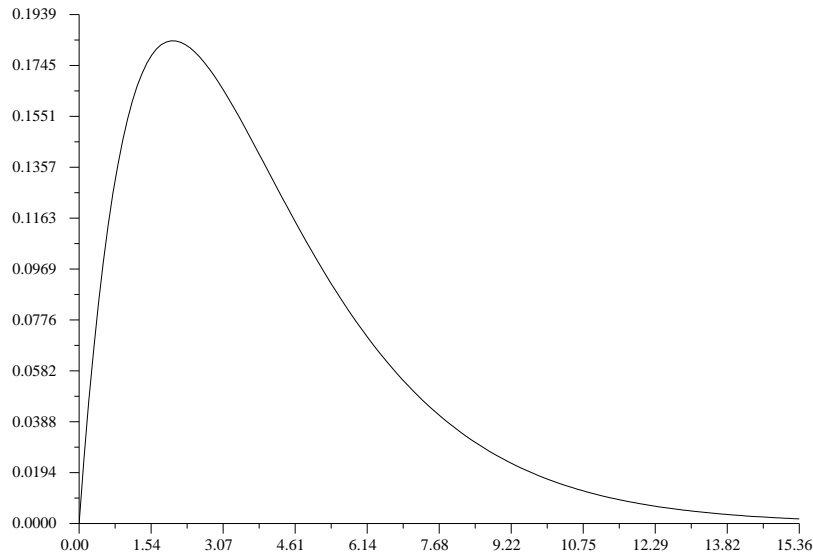
¹⁹On a un peu oublié, dans cette partie, que l'on désirait étudier la convergence de l'histogramme vers f sur \mathbf{R} et non sur $[-4, 4]$!

3.3 Un résultat théorique sur la convergence des histogrammes

Suivant le programme annoncé, nous allons simuler suivant une loi de densité connue des échantillons de taille de plus en plus grande, calculer les histogrammes correspondants et voir si les suites d'histogrammes obtenues semblent converger ou non vers cette densité. Nous avons choisi de tirer les échantillons suivant la loi du χ^2 à 4 degrés de liberté (ddl). Rappelons pour cela que la loi du χ^2 à N degrés de liberté (ddl) est la loi de densité

$$x \mapsto \frac{e^{-\frac{x}{2}} \cdot x^{\frac{N}{2}-1}}{\sqrt{2}^N \cdot \Gamma(\frac{N}{2})} \cdot \mathbf{1}_{]0,+\infty[}(x), \quad \text{où } \forall a > 0, \quad \Gamma(a) = \int_0^{+\infty} e^{-x} x^{a-1} dx,$$

dont voici le graphe, dans le cas $N = 4$.



Cette densité a pour dérivée à droite à l'origine le nombre $\frac{1}{4}$, ce qui n'apparaît pas sur le graphe parce que le repère n'est pas orthonormé. On en déduit facilement qu'elle est uniformément continue sur \mathbf{R} , ce qui n'est pas le cas de la densité de la loi du χ^2 à 3 ddl, dont la dérivée à droite à l'origine est $+\infty$, voir plus loin. Nous allons préciser la définition des histogrammes utilisés, qui seront tous des densités de probabilité. Ils seront construits sur les partitions de \mathbf{R} de la forme

$$(15) \quad (]kh_n, (k+1)h_n]; k \in \mathbf{Z})$$

où h_n est un réel > 0 que nous appellerons *largeur de classe*²⁰. L'effectif de l'échantillon (x_1, \dots, x_n) dans l'intervalle $]kh_n, (k+1)h_n]$ sera

$$\sum_{j=1}^n \mathbf{1}_{]kh_n, (k+1)h_n]}(x_j).$$

Par définition, la valeur de l'histogramme, noté \hat{H}_n , dans l'intervalle $]kh_n, (k+1)h_n]$ sera

$$\frac{1}{nh_n} \cdot \sum_{j=1}^n \mathbf{1}_{]kh_n, (k+1)h_n]}(x_j);$$

si bien que la valeur de \hat{H}_n en un point quelconque x sera

$$(16) \quad \hat{H}_n(x) = \sum_{k \in \mathbf{Z}} \left(\frac{1}{nh_n} \cdot \sum_{j=1}^n \mathbf{1}_{]kh_n, (k+1)h_n]}(x_j) \right) \mathbf{1}_{]kh_n, (k+1)h_n]}(x).$$

Compte tenu de (1), si on veut faire apparaître l'échantillon théorique, on écrira

$$(17) \quad \hat{H}_n(x) = \sum_{k \in \mathbf{Z}} \left(\frac{1}{nh_n} \cdot \sum_{j=1}^n \mathbf{1}_{]kh_n, (k+1)h_n]}(X_j(\omega)) \right) \mathbf{1}_{]kh_n, (k+1)h_n]}(x)$$

ou, si on veut faire apparaître $\hat{H}_n(x)$ comme une variable aléatoire

$$(18) \quad \hat{H}_n(x) = \sum_{k \in \mathbf{Z}} \left(\frac{1}{nh_n} \cdot \sum_{j=1}^n \mathbf{1}_{]kh_n, (k+1)h_n]}(X_j) \right) \mathbf{1}_{]kh_n, (k+1)h_n]}(x).$$

Remarquons que les formules ci-dessus ne font intervenir que des sommes d'au plus n termes et non des séries. En effet, il y a au plus n intervalles de la partition (15) qui contiennent au moins un point de l'échantillon (x_1, \dots, x_n) . Les autres intervalles n'interviennent pas dans la sommation. En dehors de ces intervalles, $\hat{H}_n(x)$ vaut donc 0. Plus précisément²¹,

$$kh_n < u \leq (k+1)h_n \iff k = -\left[\frac{-u}{h_n}\right] - 1$$

si bien que si l'on pose

$$(19) \quad \alpha_n = \min(x_1, \dots, x_n) \quad \text{et} \quad \beta_n = \max(x_1, \dots, x_n)$$

ainsi que

$$(20) \quad k_n = -\left[\frac{-\alpha_n}{h_n}\right] - 1 \quad \text{et} \quad K_n = -\left[\frac{-\beta_n}{h_n}\right] - 1,$$

²⁰De nouveau, nous simplifions. En fait, les classes d'une même partition n'ont pas besoin d'avoir toutes la même longueur.

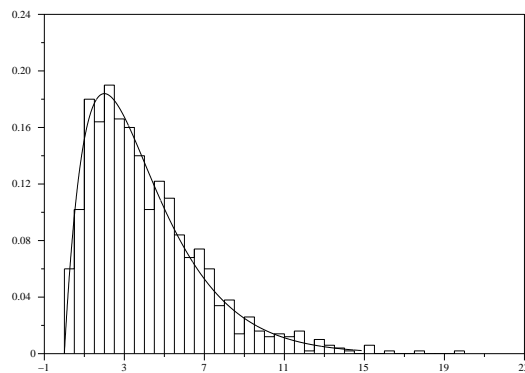
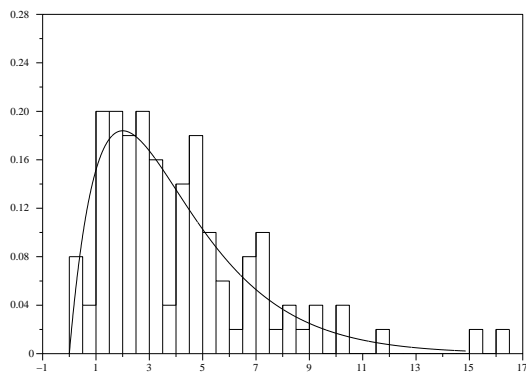
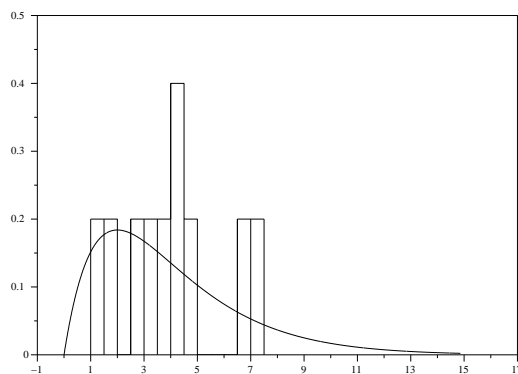
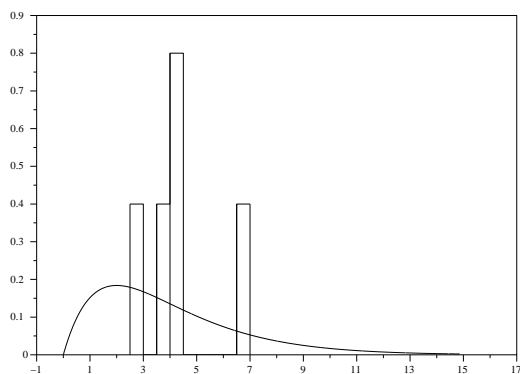
²¹Dans l'égalité qui suit, $[]$ désigne la fonction partie entière.

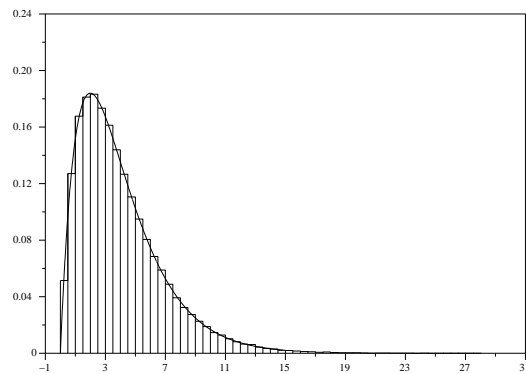
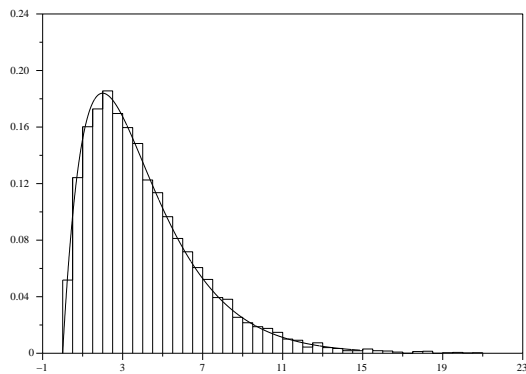
l'égalité (16) (et de même les égalités (17) et (18)) s'écrit

$$(21) \quad \hat{H}_n(x) = \frac{1}{nh_n} \cdot \sum_{k=k_n}^{K_n} \left(\sum_{j=1}^n \mathbf{1}_{]kh_n, (k+1)h_n]}(x_j) \right) \mathbf{1}_{]kh_n, (k+1)h_n]}(x).$$

Première suite de simulations

Nous laissons constante et égale à 0.5 la largeur de classe, notée provisoirement h au lieu de h_n . La taille n de l'échantillon prend successivement les valeurs 5, 10, 100, 1000, 10000 et 100000.





Notre conclusion sera que

la suite des histogrammes à largeur de classe constante ne semble pas converger.

En fait, on démontrerait facilement, à l'aide de la loi forte des grands nombres comme précédemment, que les histogrammes convergent dans ce cas vers la fonction qui est constante sur chaque intervalle de la forme $]kh, (k + 1)h]$ où elle vaut

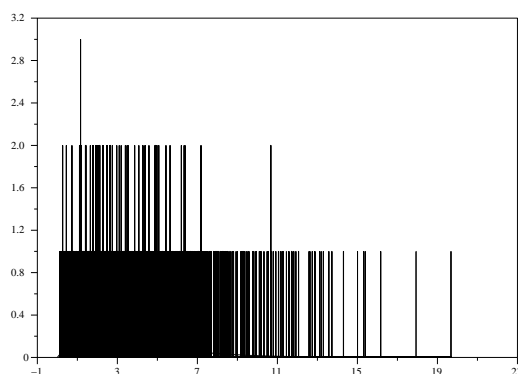
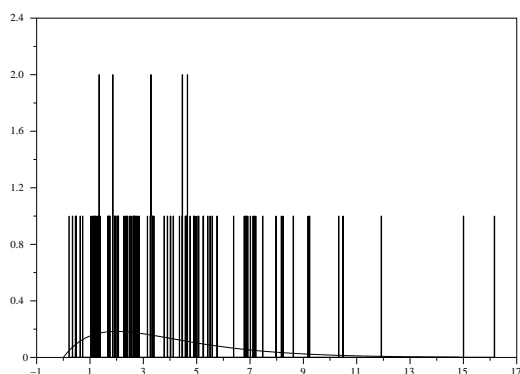
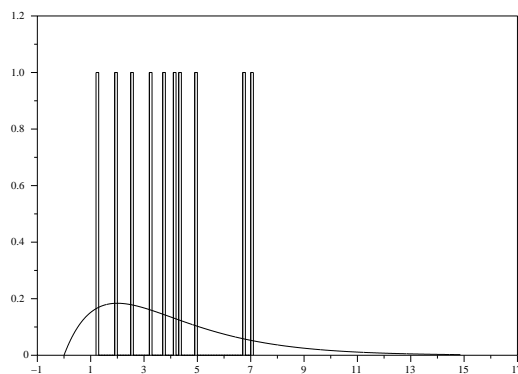
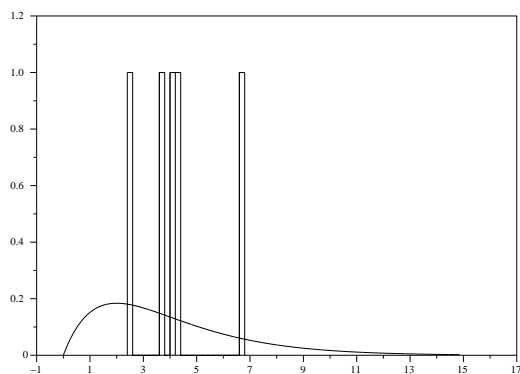
$$\frac{1}{h} \cdot P(kh < X \leq (k + 1)h),$$

X désignant une variable aléatoire qui suit la loi du χ^2 à 4 ddl. Il est donc établi qu'il n'y a pas convergence dans ce cas.

Deuxième série de simulations

Nous reprenons maintenant les simulations précédentes en définissant la largeur de classe par $h_n = \frac{1}{n}$. Les choses étant claires dans ce cas où la convergence vers 0 de la largeur de classe est assez rapide²², nous avons donné à la taille n de l'échantillon seulement les valeurs 5, 10, 100 et 1000.

²²Une convergence vers 0 en $\frac{1}{n}$ n'est pas très rapide : penser par exemple à la vitesse de convergence de $\exp(-n)$. Mais on va voir que c'est déjà trop rapide!



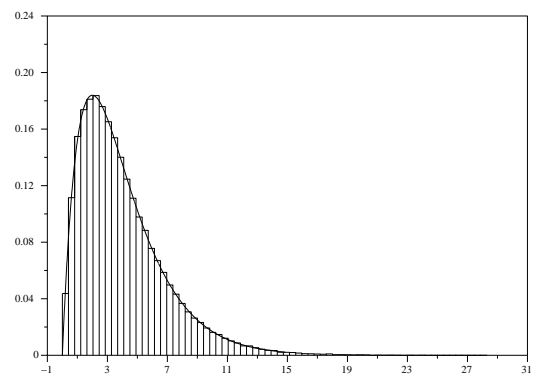
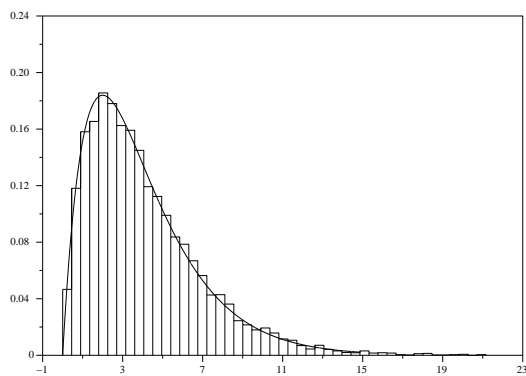
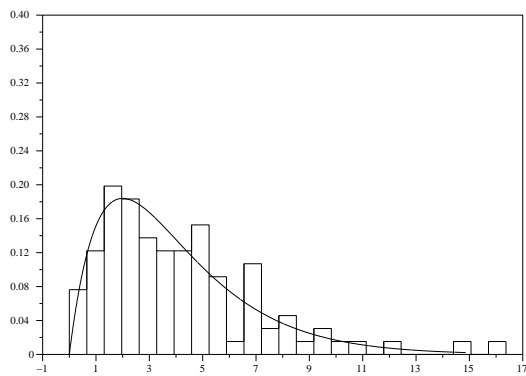
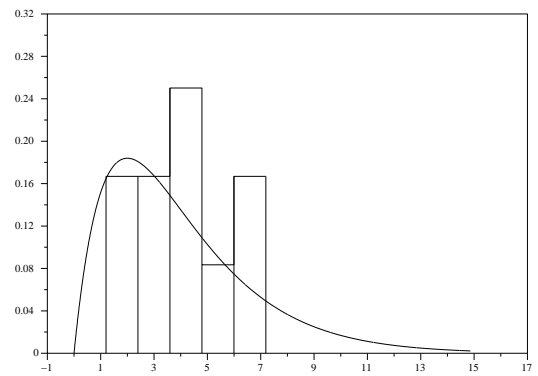
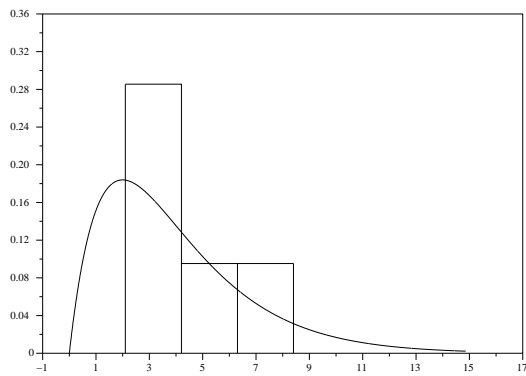
Le produit nh_n valant 1, les valeurs des histogrammes seront exactement les effectifs des classes. Dans le dernier graphique, par exemple, l'histogramme ne prend que les valeurs 0, 1, 2 ou 3. Bien entendu, une suite de fonctions à valeurs entières ne peut converger uniformément vers la densité de la loi du χ^2 à 4 ddl.

Il n'y a donc pas convergence dans ce cas et cela semble bien dû au fait que les classes sont trop étroites, c'est à dire au fait que la largeur de classe tend trop vite vers 0.

Troisième série de simulations

Nous posons pour terminer $h_n = \frac{1}{\log \log(n)}$ ²³ et donnons à n successivement les valeurs 5, 10, 100, 1000, 10000 et 100000.

²³Il s'agit d'une suite convergeant très lentement vers 0.



Cette fois, il semble bien y avoir convergence uniforme des histogrammes vers la densité de la loi du χ^2 à 4 ddl, de manière assez lente.

En fait, c'est exact, la convergence vers 0 de la largeur de classe se faisant suffisamment lentement, d'après le théorème suivant :

Théorème 3.1 *Pour tout échantillon théorique $(X_n, n \geq 1)$ suivant une loi inconnue*

de densité f uniformément continue sur \mathbf{R} ²⁴, la suite des histogrammes $(\hat{H}_n, n \geq 1)$ définis par (18) converge presque sûrement au sens de la convergence uniforme, c'est à dire que

$$(22) \quad d_{sup}(\hat{H}_n, f) \xrightarrow[n \rightarrow +\infty]{p.s.} 0$$

si et seulement si

$$(23) \quad h_n \xrightarrow[n \rightarrow +\infty]{} 0 \quad \text{et} \quad \frac{\log n}{n \cdot h_n} \xrightarrow[n \rightarrow +\infty]{} 0. \quad \square$$

4 Estimateurs à noyau

4.1 Définition des estimateurs à noyau

Supposons que l'on ne dispose que d'un échantillon observé de taille 1, soit (x_1) (sous l'hypothèse qu'il provient d'une loi à densité f inconnue). Si K désigne une densité de probabilité qui a un maximum en 0, la fonction

$$x \mapsto K(x - x_1),$$

qui est encore une densité de probabilité, admet un maximum en x_1 . Comme il est raisonnable de penser que la densité inconnue f prend de grandes valeurs (c'est à dire des valeurs voisines de son maximum²⁵) au voisinage de x_1 puisque x_1 a été tiré au hasard suivant la loi de densité f , on pourra décider de choisir $x \mapsto K(x - x_1)$ comme estimateur de f . En effet, la loi de densité $x \mapsto K(x - x_1)$ a au moins la propriété de rendre les plus probables les valeurs autour de x_1 . Cette raison n'est pas très convaincante, mais il ne faut pas oublier que pour choisir cet estimateur de f , on n'a pratiquement pas d'information !

On peut de plus donner de la souplesse à l'estimateur en lui adjoignant un paramètre supplémentaire $h > 0$ qui s'appelle la *largeur de la fenêtre* et en choisissant comme nouvel estimateur de f la fonction un peu plus sophistiquée suivante, qui est toujours une densité de probabilité²⁶ :

$$(24) \quad x \mapsto \frac{1}{h} \cdot K\left(\frac{x - x_1}{h}\right).$$

Le choix de h permet d'accorder plus ou moins d'importance à l'observation : en effet, plus h est petit, plus le maximum de l'estimateur (24) est grand. On verra sur des

²⁴La classe \mathcal{F} des densités possibles est donc la classe des densités uniformément continues sur \mathbf{R} . Toute densité uniformément continue sur \mathbf{R} est bornée et tend vers 0 à l'infini.

²⁵si maximum il y a, s'il est unique, etc : ces imprécisions sont sans importance puisqu'il s'agit seulement de montrer comment est née une théorie

²⁶Si U est une variable aléatoire qui suit la loi de densité K , la variable aléatoire $V = hU + x_1$ suit la loi de densité $x \mapsto \frac{1}{h} \cdot K\left(\frac{x - x_1}{h}\right)$.

simulations que le choix de h va se révéler primordial.

On peut alors continuer dans cet esprit : si l'échantillon observé est de taille 2, soit (x_1, x_2) , et comme l'on n'a aucune raison d'accorder plus d'importance à x_1 qu'à x_2 , on choisira comme estimateur de f la fonction

$$x \mapsto \frac{1}{2h} \cdot \left(K\left(\frac{x-x_1}{h}\right) + K\left(\frac{x-x_2}{h}\right) \right),$$

et plus généralement, on associera à l'échantillon observé (x_1, \dots, x_n) , de taille n , l'estimateur

$$(25) \quad \boxed{x \mapsto \hat{f}_n(x) = \frac{1}{nh_n} \cdot \left(K\left(\frac{x-x_1}{h_n}\right) + \dots + K\left(\frac{x-x_n}{h_n}\right) \right),}$$

qui est encore une densité de probabilité ²⁷. La formule (25) permet de surcroît de faire varier la largeur de la fenêtre avec n , parce qu'elle a été notée h_n et non h . Finalement, compte tenu de (1), autrement dit en faisant apparaître l'échantillon théorique, on peut l'écrire, au lieu de (25), sous la forme

$$(26) \quad x \mapsto \hat{f}_n(x) = \frac{1}{nh_n} \cdot \left(K\left(\frac{x-X_1(\omega)}{h_n}\right) + \dots + K\left(\frac{x-X_n(\omega)}{h_n}\right) \right),$$

ou, si l'on indique pas la variable ω ,

$$(27) \quad \boxed{x \mapsto \check{f}_n(x) = \frac{1}{nh_n} \cdot \left(K\left(\frac{x-X_1}{h_n}\right) + \dots + K\left(\frac{x-X_n}{h_n}\right) \right).}$$

(25) ²⁸ est appelé *estimateur à noyau* de la densité inconnue f . On voit qu'il dépend uniquement des choix du noyau K et de la largeur de la fenêtre h_n et qu'il se calcule à partir de l'échantillon observé (x_1, \dots, x_n) (tandis que (27) se calcule à partir de l'échantillon théorique).

4.2 Convergence des estimateurs à noyau gaussien

Nous allons commencer par simuler des estimateurs à noyau lorsque K est le noyau gaussien, c'est à dire la densité de la loi $\mathcal{N}(0, 1)$, soit

$$(28) \quad x \mapsto K(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2}{2}\right).$$

Il en résulte que les estimateurs de ce type sont des densités de mélanges de lois gaussiennes.

Nous regarderons comment ces estimateurs se comportent quand la taille de l'échantillon

²⁷ parce que cet estimateur de f est la moyenne arithmétique de n densités de probabilité.

²⁸ ou (27)

grandit (successivement $n = 5, 10, 100, 1000, 10000, 100000$) et essaierons d'en tirer des conclusions.

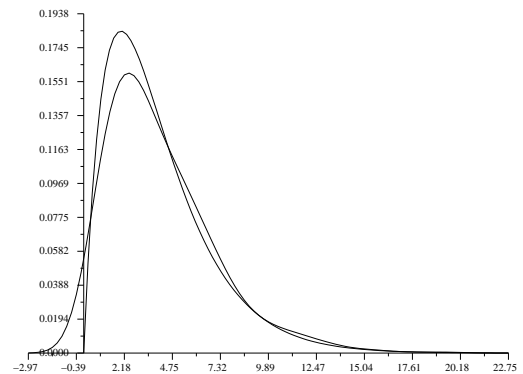
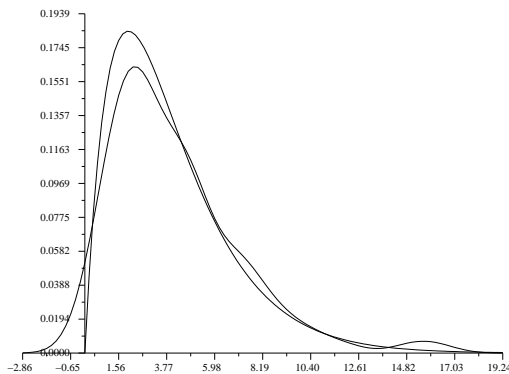
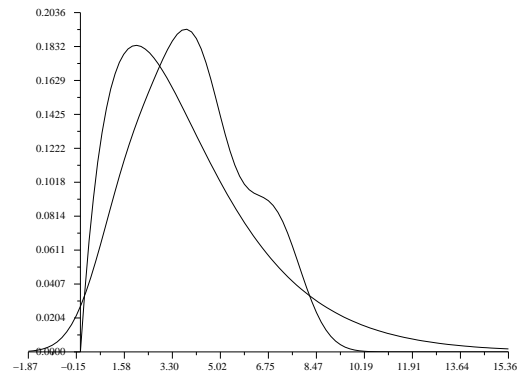
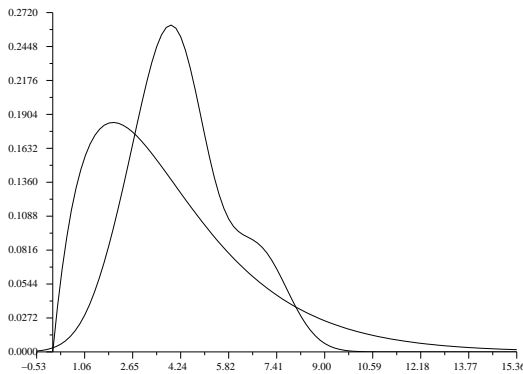
Nous commencerons par laisser constante la fenêtre h_n de l'estimateur à noyau, puis nous la ferons décroître vers 0 assez rapidement ; enfin, nous la ferons décroître vers 0 à vitesse intermédiaire.

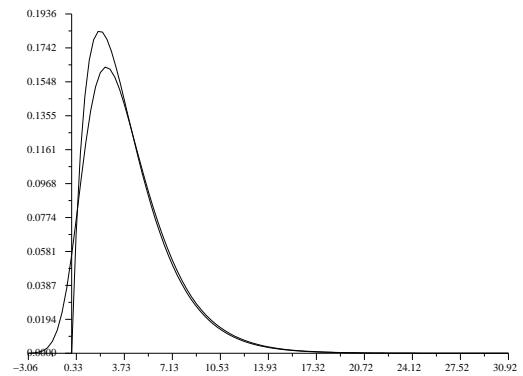
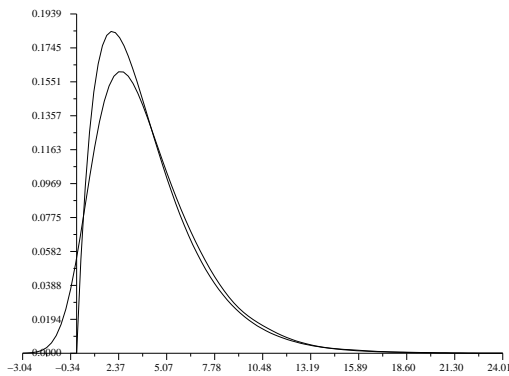
Comme pour l'histogramme, nous utiliserons un échantillon dont la loi est connue, afin de pouvoir porter un jugement sur la convergence des estimateurs (25) considérés. Nous avons choisi la loi du χ^2 à 4 ddl.

Chaque dessin ci-dessous est composé du graphe de l'estimateur considéré et du graphe de cette densité.

Estimateurs à largeur de fenêtre fixée

Les estimateurs ci-dessous sont calculés lorsque K est le noyau gaussien, $h = 1$, n valant successivement 5, 10, 100, 1000, 10000, 100000.



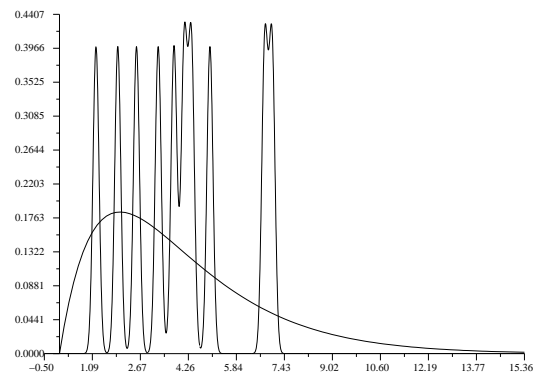
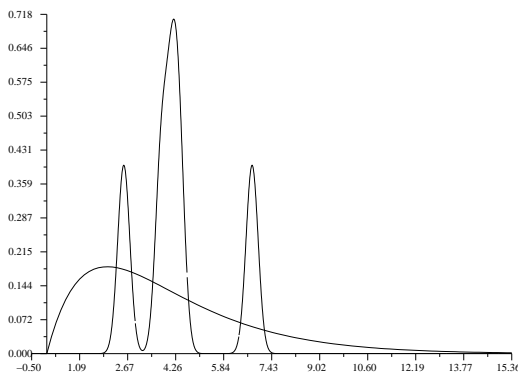


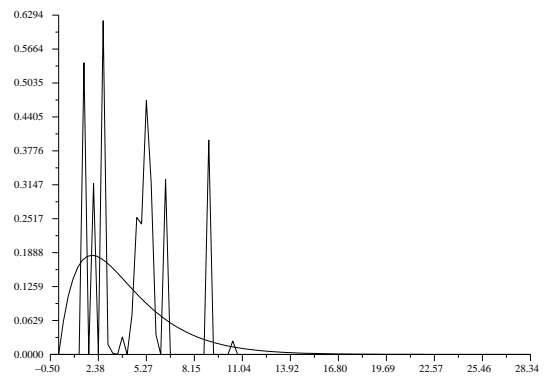
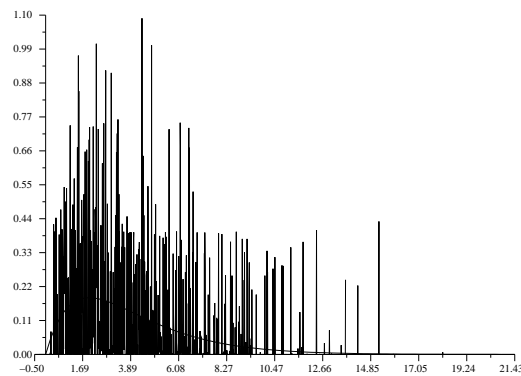
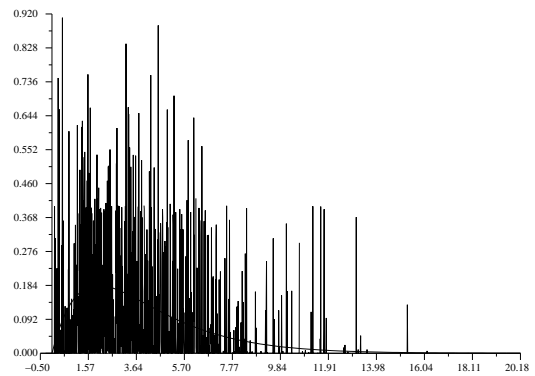
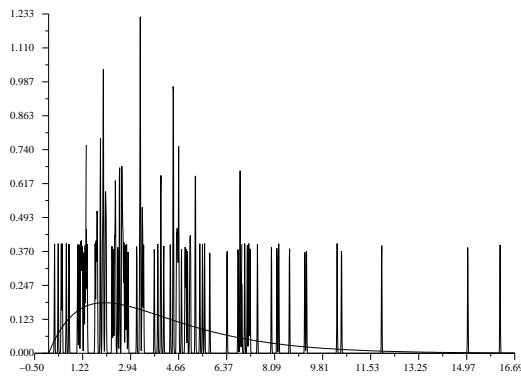
Ces essais conduisent à la conclusion suivante :

Les estimateurs considérés, dont la largeur de fenêtre est constante, ne semblent pas appartenir à une suite d'estimateurs convergeant uniformément vers la densité de la loi du χ^2 à 4 ddl.

Estimateurs à largeur de fenêtre décroissant rapidement

La seule modification par rapport à l'étude précédente est que l'on pose maintenant $h_n = \frac{1}{n}$, les échantillons et les autres choix demeurant inchangés.





Notons que l'on cumule ici toutes les difficultés des calculs précédents, par exemple :

1 - notre générateur de nombres au hasard est-il un bon générateur de la loi du χ^2 à 4 ddl ?

2 - quelle est l'influence sur les résultats des erreurs dues à la représentation des nombres et aux calculs approchés par les ordinateurs ?

3 - le 6^{ième} graphique est très différent des précédents parce que, contraint d'économiser du temps de calcul et de l'espace mémoire, on a calculé 101 points du graphe de l'estimateur au lieu de 2001 comme dans les 5 graphiques précédents ²⁹. Il est clair que le nombre de points calculés pour le dernier graphe est trop petit. Il en résulte que ce graphe paraît beaucoup moins chaotique que les graphes 3, 4, 5. En fait, il ne ressemble pas du tout au vrai graphe de l'estimateur ³⁰.

Tous ces graphes, qui représentent des densités de mélanges gaussiens, ont un aspect chaotique parce qu'il s'agit de lois gaussiennes d'écart-type $\frac{1}{n}$. Par exemple, pour $n =$

²⁹On sait que pour tracer le graphe d'une fonction, scilab - pour l'option choisie dans ce papier - calcule un certain nombre de points de ce graphe et les rejoint ensuite par des segments, si bien que tous les graphes scilab sont en fait des lignes polygonales, ce qui ne se voit pas si suffisamment de points ont été calculés.

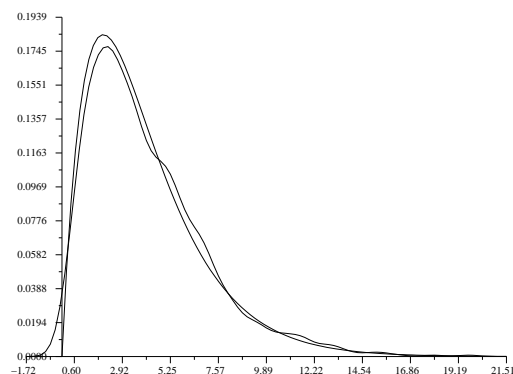
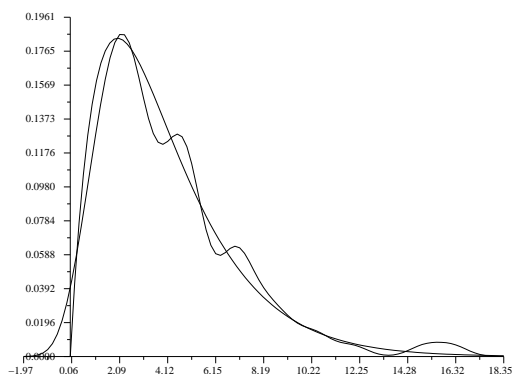
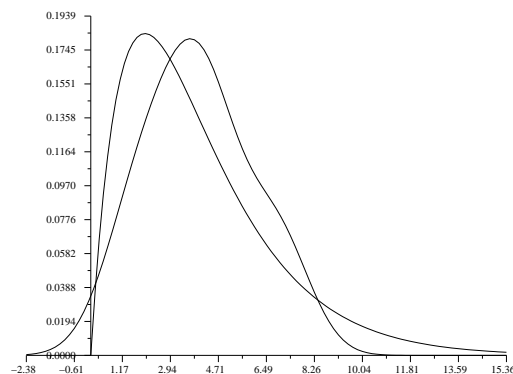
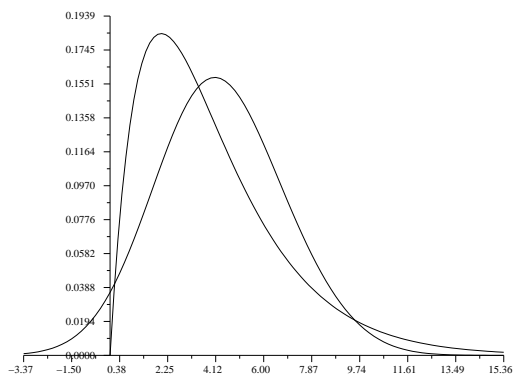
³⁰Il aurait été plus adroit d'augmenter le nombre de points calculés avec la taille de l'échantillon observé. On voit ainsi qu'une question qui paraît secondaire est en fait très importante.

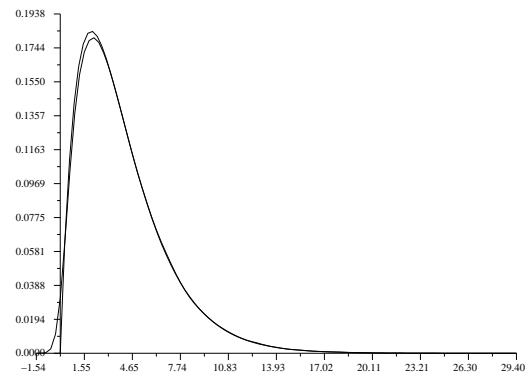
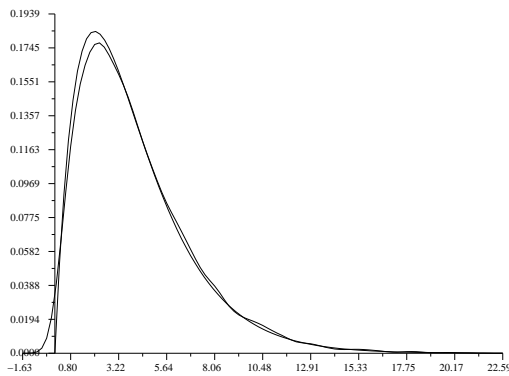
10000, il s'agit de lois gaussiennes dont 99% de la probabilité est concentrée sur un intervalle de longueur inférieure à $5, 152 \cdot 10^{-4}$. Pratiquement, le graphe d'une telle densité se réduit à l'axe des abscisses et à un segment vertical. Nous en déduisons que

Ces estimateurs ne semblent pas appartenir à une suite convergente vers f .

Estimateurs à largeur de fenêtre décroissant convenablement

Nous recommençons une nouvelle fois l'étude ci-dessus en posant $h_n = \frac{1}{\log(\log n)}$.





On peut conclure cette fois que

Ces estimateurs semblent appartenir à une suite convergent vers f assez lentement.

4.3 Convergence des estimateurs de la fenêtre mobile

Nous allons reprendre l'étude de la division précédente en remplaçant le noyau gaussien par un noyau appelé *noyau de la fenêtre mobile*, encore noté K ³¹ et défini par

$$(29) \quad K = \mathbf{1}_{[-\frac{1}{2}, \frac{1}{2}]}.$$

Les estimateurs de la densité construits à partir du noyau de la fenêtre mobile s'appellent *estimateurs de la fenêtre mobile*.

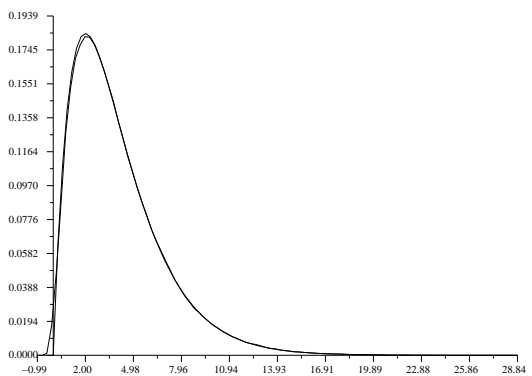
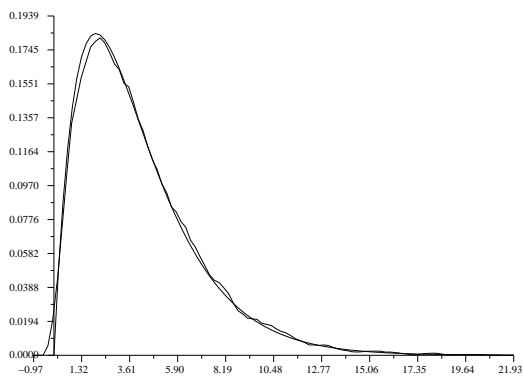
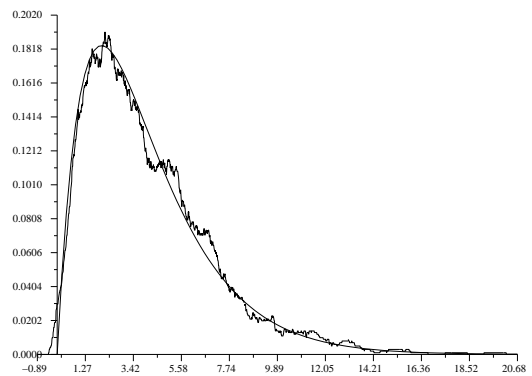
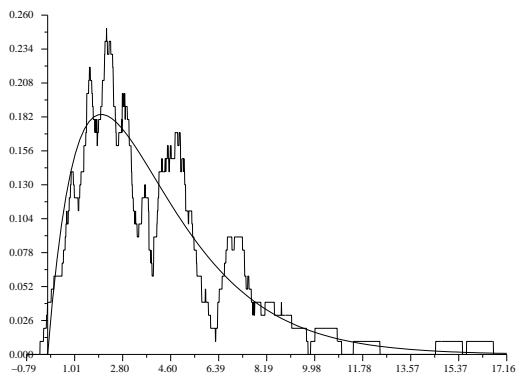
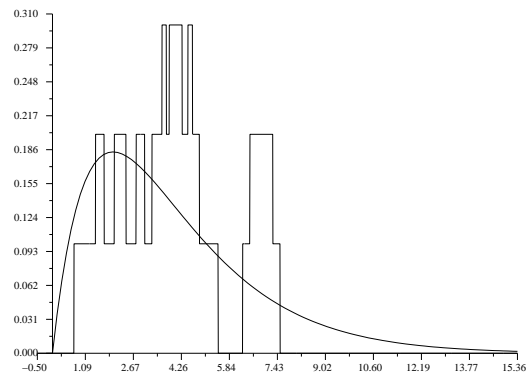
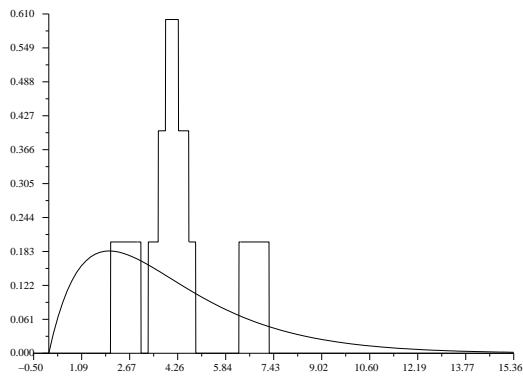
Dans ce qui suit, nous simulons des estimateurs de la fenêtre mobile calculés sur des échantillons de la loi du χ^2 à 4 ddl, de taille égale successivement à 5, 10, 100, 1000, 10000, 100000 lorsque la largeur de la fenêtre reste constante, puis décroît rapidement vers 0 et enfin décroît vers 0 suffisamment lentement, les autres éléments de ces 3 séries de simulations demeurant inchangés. Tous les estimateurs de la fenêtre mobile sont des densités de probabilité puisque le noyau $K = \mathbf{1}_{[-\frac{1}{2}, \frac{1}{2}]}$ est lui-même une densité de probabilité.

Chaque dessin ci-dessous est composé du graphe de l'estimateur considéré et du graphe de la densité f de la loi du χ^2 à 4 ddl.

Estimateurs à largeur de fenêtre fixée

Les estimateurs de la fenêtre mobile ci-dessous sont calculés lorsque $h = 1$, n valant successivement 5, 10, 100, 1000, 10000, 100000.

³¹ K est la densité de la loi uniforme sur l'intervalle $[-\frac{1}{2}, \frac{1}{2}]$.

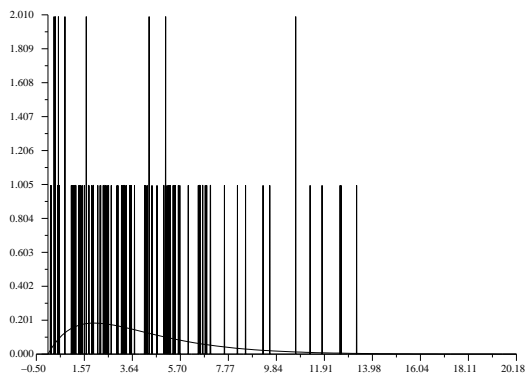
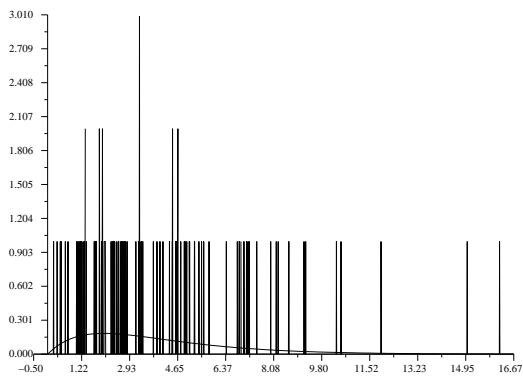
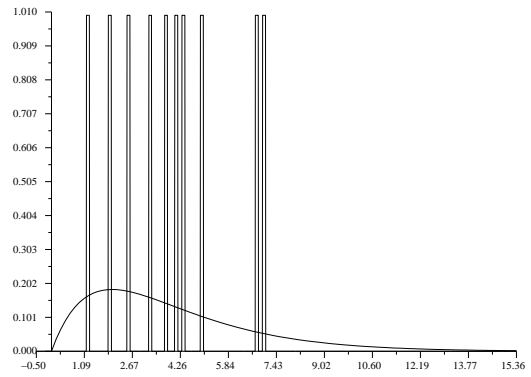
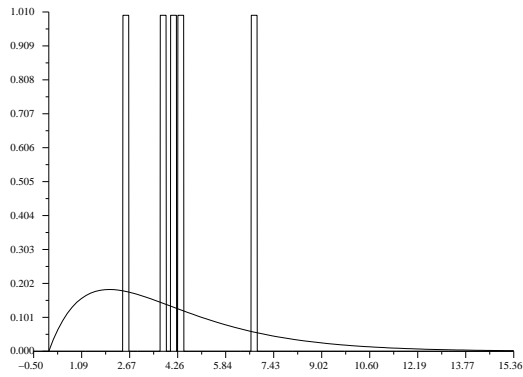


Ces essais conduisent à la même conclusion que lorsque le noyau est gaussien (observer le voisinage de 0).

Estimateurs à largeur de fenêtre décroissant rapidement

La seule modification par rapport à l'étude précédente est que l'on pose maintenant $h_n = \frac{1}{n}$, les échantillons et les autres choix demeurant inchangés. Nous n'avons pas utilisé les échantillons de taille 10000 et 100000, faute de pouvoir calculer un nombre

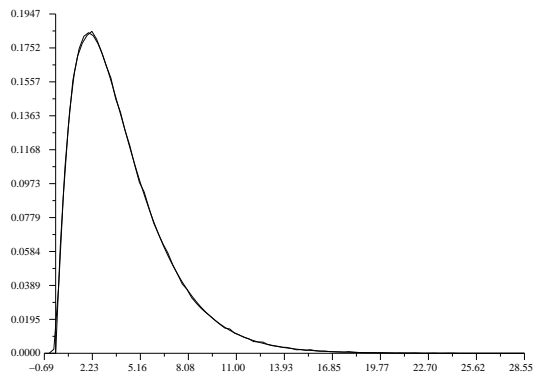
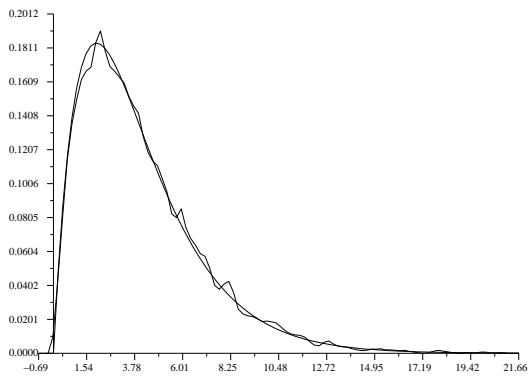
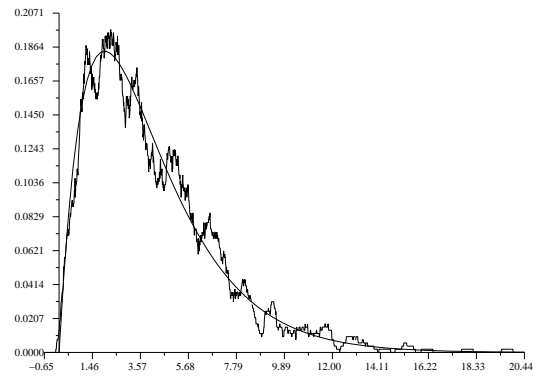
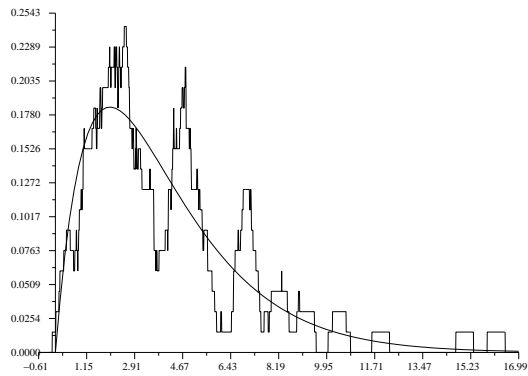
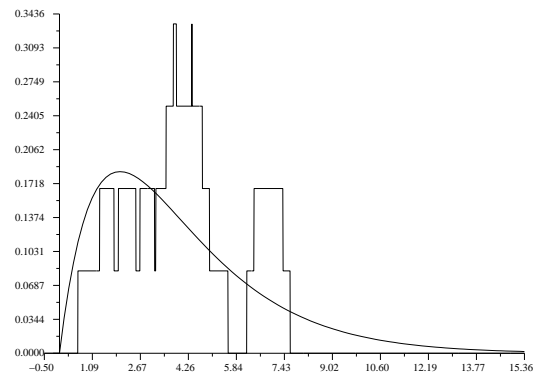
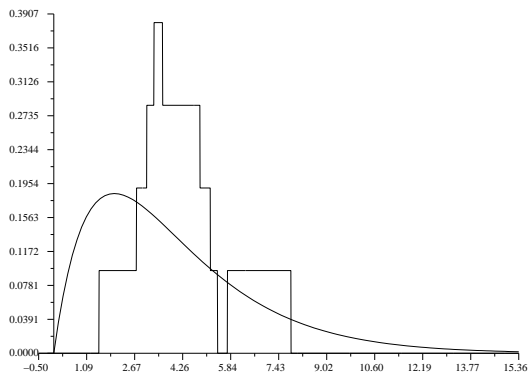
suffisant de points du graphe qui donne un tracé significatif.



Comme dans le cas analogue précédent, ces estimateurs de la densité ne prennent que des valeurs entières. Ils ne peuvent donc pas converger uniformément vers f .

Estimateurs à largeur de fenêtre décroissant convenablement

Nous recommençons une nouvelle fois l'étude ci-dessus en posant maintenant $h_n = \frac{1}{\log(\log n)}$.



On peut conclure finalement que

Les estimateurs de la fenêtre mobile et à noyau gaussien se comportent de la même manière vis à vis de la convergence uniforme vers f quand la largeur de la fenêtre reste constante ou décroît vers 0 en $\frac{1}{n}$ ou en $\frac{1}{\log(\log n)}$. Il semble qu'il n'y ait convergence que dans le dernier cas.

Il y a de bonnes raisons à cela que nous allons maintenant exposer.

4.4 Un résultat théorique sur la convergence des estimateurs à noyau

Appelons noyau de Geffroy ³² toute densité de probabilité bornée K dont l'ensemble des points de discontinuité est de mesure de Lebesgue nulle ³³ et telle que l'application

$$(30) \quad x \mapsto \sup (|K(u)|; x-1 < u < x+1)$$

soit intégrable sur \mathbf{R} (toujours par rapport à la mesure de Lebesgue ou pour l'intégrale de Riemann impropre). Alors (cf. [1], p. 65),

Théorème 4.1 *Pour tout échantillon théorique $(X_n, n \geq 1)$ suivant une loi inconnue de densité f uniformément continue sur \mathbf{R} , la suite $(\check{f}_n, n \geq 1)$ des estimateurs à noyau de Geffroy K définis par (27) converge presque sûrement au sens de la convergence uniforme, c'est à dire que*

$$(31) \quad d_{sup}(\check{f}_n, f) \xrightarrow[n \rightarrow +\infty]{p.s.} 0$$

si et seulement si

$$(32) \quad h_n \xrightarrow[n \rightarrow +\infty]{} 0 \quad \text{et} \quad \frac{\log n}{n \cdot h_n} \xrightarrow[n \rightarrow +\infty]{} 0. \quad \square$$

Il est facile de vérifier que le noyau gaussien et le noyau de la fenêtre mobile sont des noyaux de Geffroy. De plus, la densité de la loi du χ^2 à 4 ddl est uniformément continue sur \mathbf{R} . Par conséquent, le théorème 4.1 s'applique et ce qui était suggéré par les simulations ci-dessus est vrai :

Il n'y a convergence uniforme presque sûre dans le cas où h_n tend vers 0 à vitesse intermédiaire.

Notons aussi qu'il est tout à fait remarquable d'obtenir pour la *largeur de classe* dans le cas des histogrammes et pour la *largeur de fenêtre* dans le cas des estimateurs à noyau de Geffroy exactement la même condition nécessaire et suffisante de convergence uniforme presque sûre.

³²Jean Geffroy, ancien professeur à l'Université Pierre et Marie Curie, maintenant retraité, a eu plus d'une centaine d'élèves dont le rédacteur de ce papier, qui rend ici hommage à ses grandes qualités d'homme, de chercheur et de statisticien.

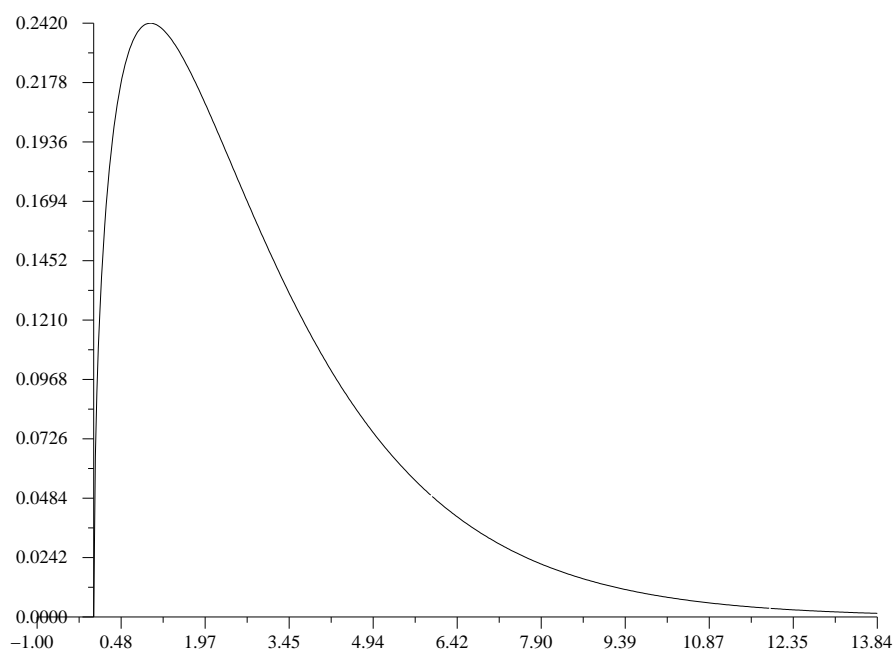
³³Il s'agit des fonctions f de \mathbf{R} dans \mathbf{R}^+ , bornées, dont l'ensemble des points de discontinuité peut être recouvert, quel que soit $\varepsilon > 0$, par une réunion d'intervalles bornés $I_{n,\varepsilon}$, $n \geq 1$, dont la somme des longueurs : $\sum_{n \geq 1} l(I_{n,\varepsilon})$ est inférieure à ε et dont l'intégrale de Riemann impropre (ou l'intégrale de Lebesgue) sur \mathbf{R} vaut 1.

5 Estimation avec hypothèses défailantes

Il est toujours intéressant d'étudier des cas où le théorème 4.1 ne s'applique pas et où les suites d'estimateurs à noyau peuvent ne pas converger.

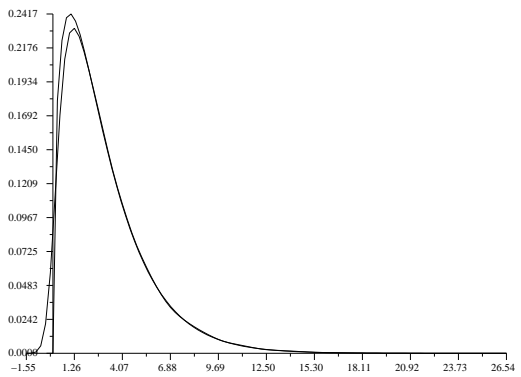
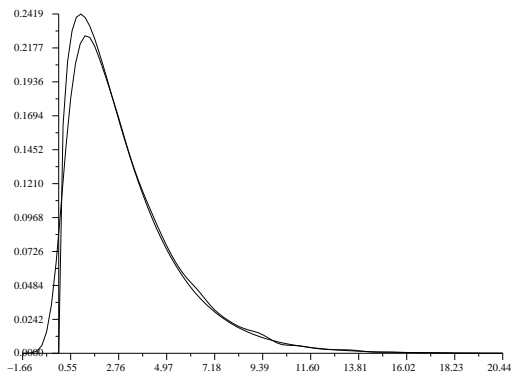
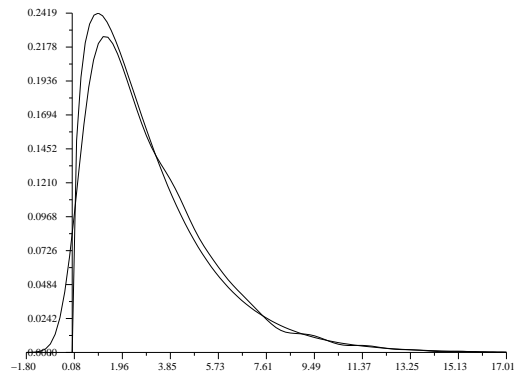
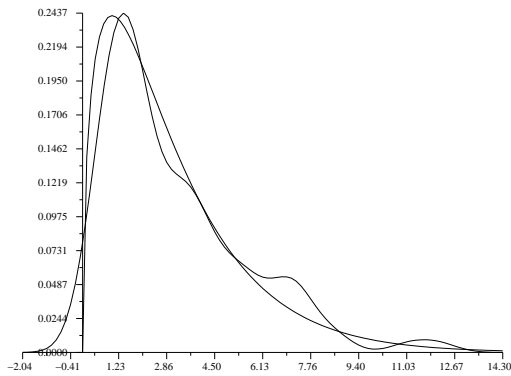
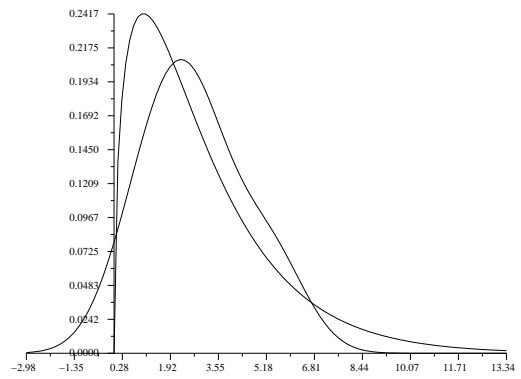
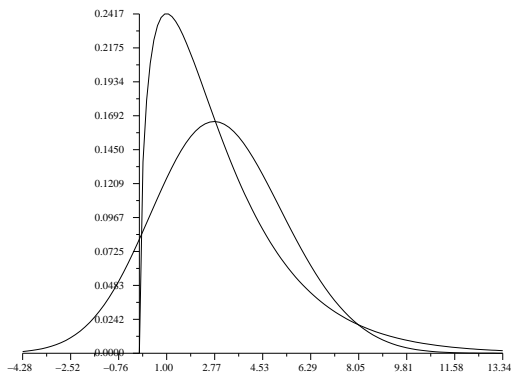
5.1 Cas d'une densité à estimer non uniformément continue

Nous recommençons une nouvelle fois l'étude ci-dessus lorsque K est le noyau gaussien, lorsque $h_n = \frac{1}{\log(\log n)}$ décroît convenablement vers 0³⁴, lorsque la taille n de l'échantillon prend successivement les valeurs 5, 10, 100, 1000, 10000, 100000 et enfin lorsque l'échantillon est tiré suivant la loi du χ^2 à 3 ddl. Le graphe de la densité de cette loi



a une demi-tangente à droite en 0 qui est verticale. Il en résulte que la densité de la loi du χ^2 à 3 ddl n'est pas uniformément continue sur \mathbf{R} , ce qui fait que le théorème 4.1 ne s'applique pas. Voyons comment se comportent dans ce cas les estimateurs de la densité.

³⁴L'adverbe convenablement fait évidemment référence au théorème 4.1.

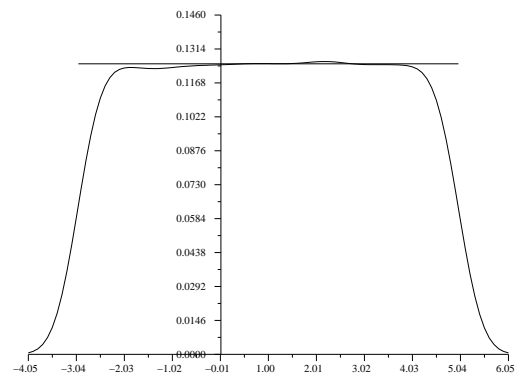
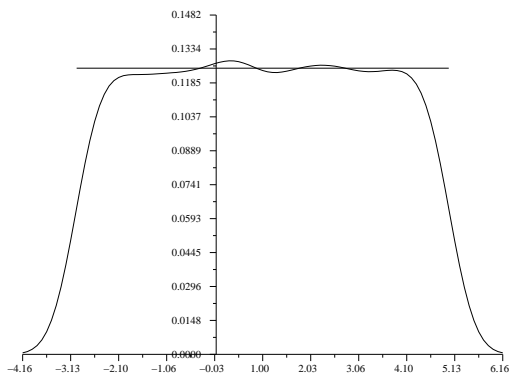
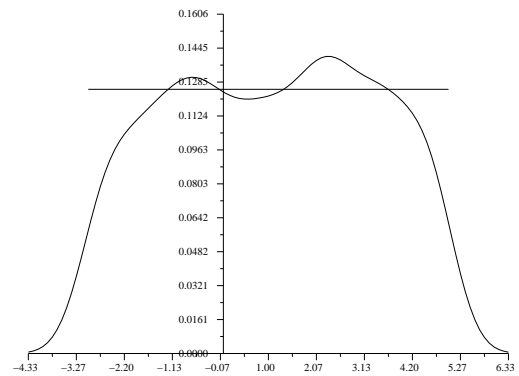
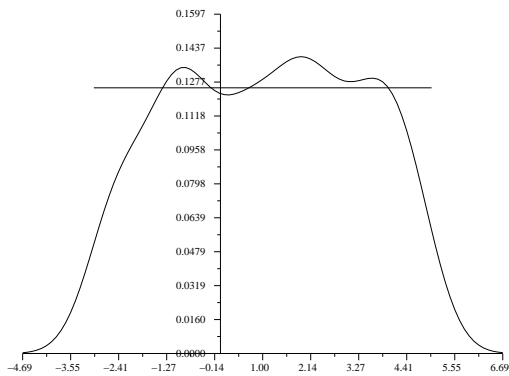
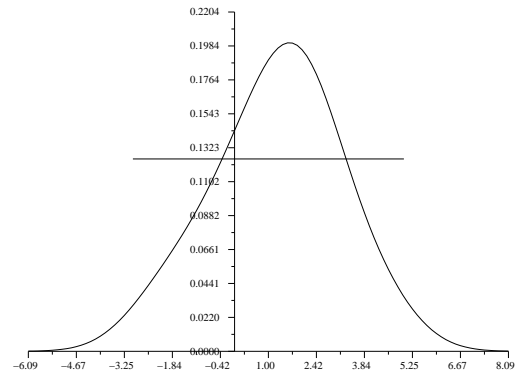
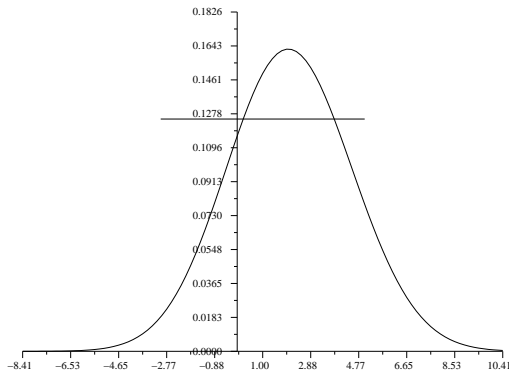


On peut conclure cette fois que

Ces estimateurs semblent appartenir à une suite ne convergeant pas vers f , puisqu'il y a manifestement un problème de convergence au voisinage de 0, à moins qu'il s'agisse d'une convergence extrêmement lente.

5.2 Cas de l'estimation d'une densité discontinue

Nous partons maintenant d'un échantillon de la loi $\mathcal{U}(a, b)$, où $a = -3$ et $b = 5$. Le théorème 4.1 ne s'applique pas ici parce que la densité à estimer a 2 points de discontinuité, à savoir -3 et 5. Les autres caractéristiques des simulations sont inchangées.



Il est clair cette fois que

Il ne peut y avoir convergence uniforme parce que l'on sait que toute limite uniforme sur \mathbf{R} d'une suite de fonctions continues est elle-même une fonction continue sur \mathbf{R} . Il semble y avoir convergence en dehors des points de discontinuité.

5.3 Cas de l'estimation de la densité d'une loi mixte

Il s'agit des lois de probabilité qui ont une partie discrète et une partie à densité. Plus précisément, nous avons engendré des échantillons suivant la loi de probabilité d'une variable aléatoire X qui prend toutes ses valeurs dans l'intervalle $[-3, 5]$ avec les probabilités suivantes, deux valeurs α et β ($-3 < \alpha < \beta < 5$) et deux probabilités m_α et m_β ($0 < m_\alpha < m_\alpha + m_\beta < 1$) étant données : pour tout sous-intervalle $[a, b]$ de $[-3, 5]$,

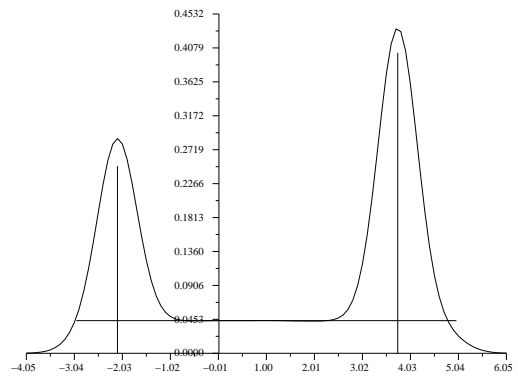
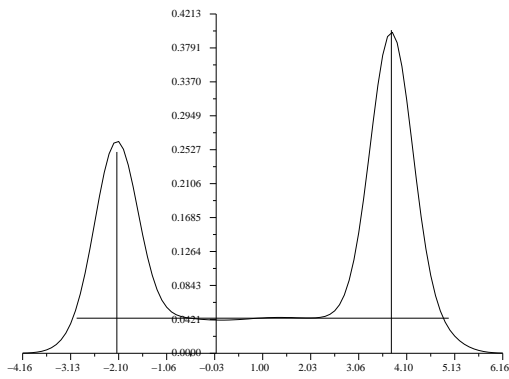
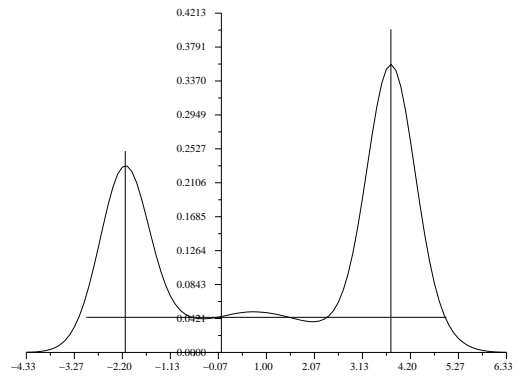
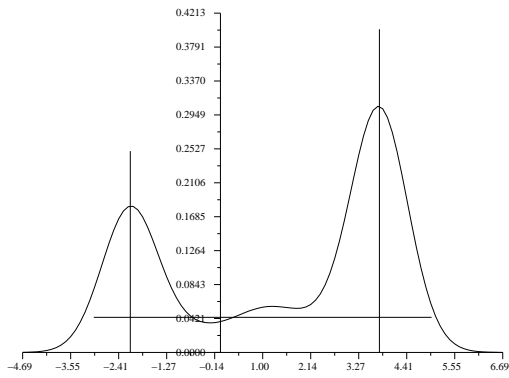
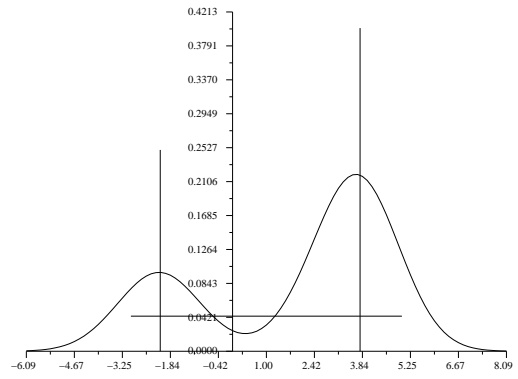
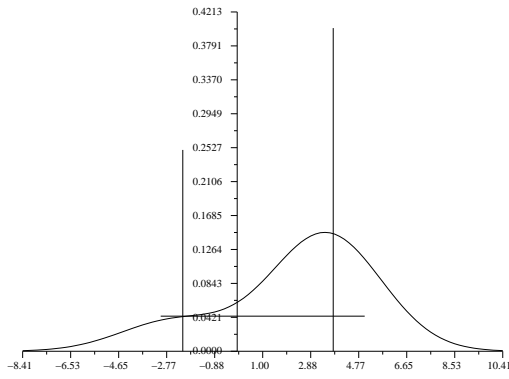
$$P(X \in [a, b]) = \begin{cases} \frac{1-m_\alpha-m_\beta}{8} \int_a^b \mathbf{1}_{[-3,5]}(x) dx = \frac{(b-a)(1-m_\alpha-m_\beta)}{8} & \text{si } \alpha, \beta \notin [a, b] \\ \frac{(b-a)(1-m_\alpha-m_\beta)}{8} + m_\alpha & \text{si } \alpha \in [a, b], \beta \notin [a, b] \\ \frac{(b-a)(1-m_\alpha-m_\beta)}{8} + m_\beta & \text{si } \alpha \notin [a, b], \beta \in [a, b] \\ \frac{(b-a)(1-m_\alpha-m_\beta)}{8} + m_\alpha + m_\beta & \text{si } \alpha, \beta \in [a, b] \end{cases}$$

Ce type de loi ne figure dans aucun programme des lycées, mais reste simple ³⁵. Il y a bien pire !

Dans les simulations ci-dessous, $\alpha = -2.1352769$, $\beta = 3.7651438$, $m_\alpha = 0.25$ et $m_\beta = 0.40125$ ³⁶. Nous avons représenté graphiquement cette loi mixte par 2 bâtons verticaux d'abscisses α et β et de hauteurs m_α et m_β et par le graphe de la densité de la partie à densité de cette loi mixte qui est la fonction constante et égale à $\frac{1-m_\alpha-m_\beta}{8}$ sur $[-3, 5]$ et à 0 ailleurs. Le reste du programme de simulations est inchangé.

³⁵Naturellement, on dira que sa partie discrète est la loi d'une variable aléatoire qui prend seulement les valeurs α et β avec les probabilités $\frac{m_\alpha}{m_\alpha+m_\beta}$ et $\frac{m_\beta}{m_\alpha+m_\beta}$ et que sa partie à densité est la loi uniforme sur l'intervalle $[-3, 5]$.

³⁶Cet exemple pose de nombreuses questions. Si nous avons choisi des probabilités m_α et m_β beaucoup plus petites, l'examen des premiers termes de l'échantillon engendré ne permettrait pas de voir que les valeurs α et β se répètent et par conséquent ne permettrait pas d'exclure une loi purement à densité. Avec les valeurs assez grandes ici choisies pour m_α et m_β , on voit que α et plus encore β apparaissent souvent (en moyenne β apparaît 4 fois sur 10 termes de l'échantillon). On voit donc tout de suite que la loi de l'échantillon a une composante discrète, mais on n'a pas envie de dire que c'est une loi discrète parce que il prend trop de valeurs différentes. Pourtant, on peut se tromper : il n'est pas possible par exemple, même avec beaucoup de métier, de distinguer sur des échantillons la loi d'une variable aléatoire prenant comme valeurs tous les nombres rationnels avec des probabilités très petites et la loi d'une variable aléatoire à densité.



Nous dirons seulement qu' ³⁷

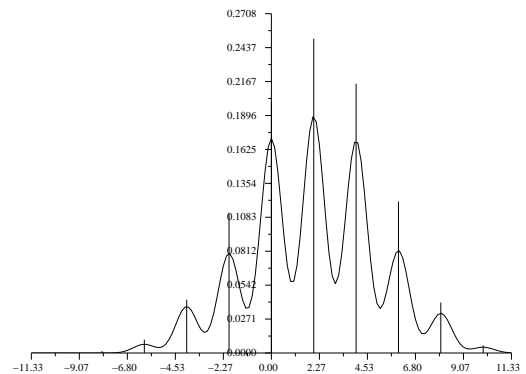
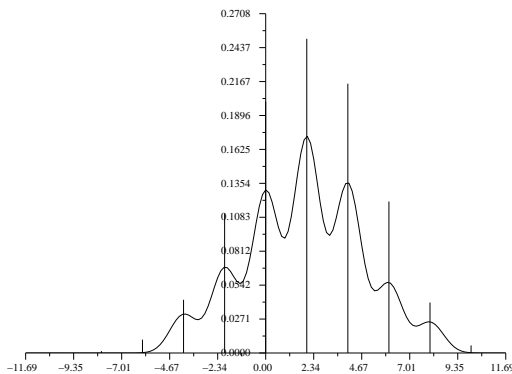
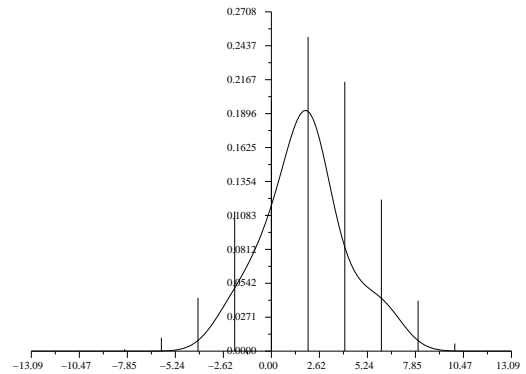
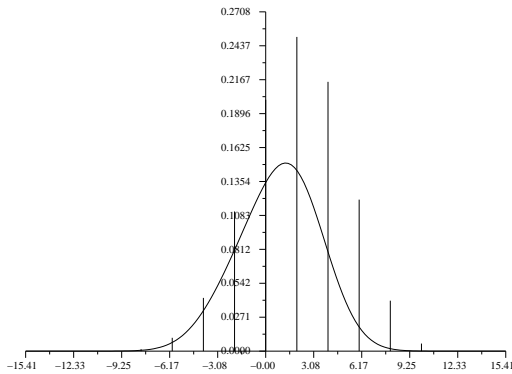
³⁷La convergence vers $+\infty$ se démontre simplement comme application de la loi forte des grands nombres : supposons que l'on parte d'un échantillon théorique infini $(X_n, n \geq 1)$ de la loi d'une variable aléatoire X qui prend une certaine valeur a avec la probabilité $m > 0$. Alors la suite $(\hat{f}_n(a), n \geq 1)$ des valeurs en a des estimateurs à noyau gaussien construits sur l'échantillon théorique donné tend vers $+\infty$. En effet,

$$\hat{f}_n(a) = \frac{1}{n} \sum_{j=1}^n \frac{1}{h_n} K\left(\frac{a-X_j}{h_n}\right) \geq \frac{1}{n} \sum_{X_j=a}^n \frac{1}{h_n} K(0) = \frac{K(0)}{h_n} \left(\frac{1}{n} \sum_{j=1}^n \mathbf{1}_{(X_j=a)}\right).$$

Il semble y avoir une estimation correcte non uniforme de la densité sauf aux points α et β où la suite des estimateurs semble tendre vers $+\infty$ (avec beaucoup d'imagination).

5.4 Cas d'une loi purement discrète

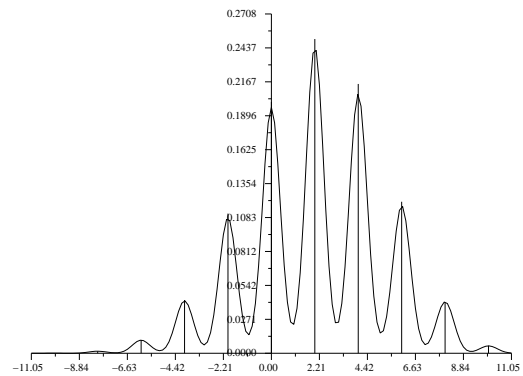
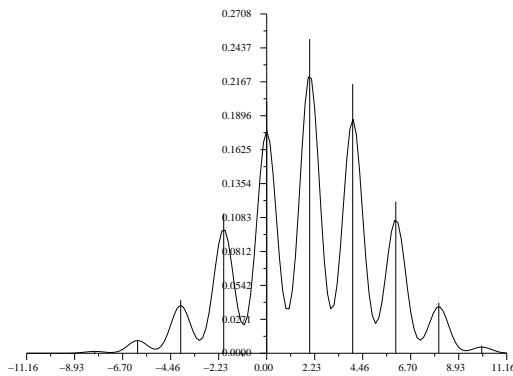
Nous appliquons enfin l'estimation de la densité par les estimateurs à noyau de Gauss au cas où la loi de l'échantillon est la loi de la variable aléatoire $2X - N$, X suivant la loi $\mathcal{B}(N, p)$, avec $N = 10$, $p = 0.6$. C'est donc un cas purement discret, c'est à dire un cas où il n'y a aucune composante à densité ³⁸.



D'après la loi forte des grands nombres, $\frac{1}{n} \sum_{j=1}^n \mathbf{1}_{(X_j=a)} \xrightarrow[n \rightarrow +\infty]{p.s.} m$, donc $\check{f}_n(a) \xrightarrow[n \rightarrow +\infty]{p.s.} +\infty$ dès que

$$h_n \xrightarrow[n \rightarrow +\infty]{} 0.$$

³⁸Ce cas n'est pas raisonnable du point de vue de la pratique statistique, parce que X suivant la loi $\mathcal{B}(10, 0.6)$ prendra des valeurs entières comprises entre 0 et 10, donc l'échantillon sera formé d'entiers compris entre $-N$ et N . Par conséquent, on pensera immédiatement à une loi discrète!



On constate qu'entre les entiers $-N, -N+1, \dots, N$, les estimateurs à noyau semblent converger vers 0, très lentement et certainement pas uniformément, ce qui confirme la convergence vers la densité - nulle - de la partie à densité ³⁹ tandis qu'il y a convergence vers $+\infty$ aux points chargés par la loi discrète de l'échantillon, ce qui ne se voit pas, mais a été démontré précédemment. Nous avons ainsi des convergences très lentes, ce qui est aussi intéressant : tout ne se passe pas toujours très bien en mathématiques.

6 Intérêt de l'estimation de la densité

Il n'est pas évident. En effet, au lycée, les densités servent à calculer des probabilités d'événements et pourraient servir à calculer des espérances mathématiques. Si X est une variable aléatoire suivant une loi de densité inconnue f , on sait que pour tous couples de réels a et b tels que $a < b$, la probabilité $p = P(a \leq X \leq b)$ est donnée par la formule (4). Si on dispose d'un échantillon (x_1, \dots, x_n) de taille n de la loi de X , il sera très naturel d'estimer la quantité inconnue p par

$$(33) \quad \hat{p} = \int_a^b \hat{f}_n(x) dx$$

si l'on utilise, par exemple, l'estimateur de la densité donné par la formule (25). Pour l'espérance mathématique inconnue m de X , si elle existe, c'est à dire si l'intégrale de Riemann impropre

$$(34) \quad m = \int_{-\infty}^{+\infty} x \cdot f(x) dx$$

est absolument convergente (ou si l'intégrale de Lebesgue qui s'écrit de la même façon est finie), on estimera de même m par

$$(35) \quad \hat{m} = \int_{-\infty}^{+\infty} x \cdot \hat{f}_n(x) dx$$

³⁹si l'on convient que la loi de X est une loi mixte dont la partie à densité est nulle, pour se rattacher au cas précédent !

en supposant que cette intégrale de Riemann impropre ait un sens. Mais en fait, on n'utilisera jamais les estimateurs (33) ou (35) parce qu'il existe des estimateurs bien connus de p et de m qui sont directs, donc probablement beaucoup plus efficaces au sens commun du terme : il s'agit d'une part

$$(36) \quad \check{m}_n = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{(X_j \in [a,b])}$$

où $(X_n, n \geq 1)$ désigne l'échantillon théorique, ce qui fait que \check{m}_n est la fréquence relative de réalisation de l'événement "la variable aléatoire prend ses valeurs dans l'intervalle $[a, b]$ ", qui converge presque sûrement vers p d'après la loi forte des grands nombres, d'autre part de la moyenne empirique $\frac{X_1 + \dots + X_n}{n}$ qui converge de même vers m .

L'intérêt d'estimer la densité inconnue est pourtant évident dans certaines conditions : quand il y a convergence uniforme presque sûre (ce qui est le cas des deux théorèmes que nous avons cités) et si nous disposons d'un échantillon de taille assez grande, nous aurons une bonne idée des zones où la densité est grande, qui seront des zones où toute variable aléatoire suivant la loi inconnue aura de bonnes chances de prendre ses valeurs. L'estimateur permettra aussi de faire de meilleures hypothèses pour la modélisation du phénomène aléatoire considéré : en clair, l'intérêt essentiel de l'estimation de la densité est l'aide à la modélisation. Mais il ne faut pas oublier que ces questions sont délicates : on peut être facilement induit en erreur (voir les simulations concernant des lois mixtes⁴⁰) et on ne sait jamais d'avance si les hypothèses de tel ou tel théorème sont vérifiées. Tout modèle est entaché d'incertitude que l'on essaie ensuite de lever par divers procédés de validation, ce qui est un autre problème.

Références

- [1] **Bosq Denis et Lecoutre Jean-Pierre** *Théorie de l'estimation fonctionnelle*, Economica, Paris (1987).
- [2] **Delecroix Michel** *Histogrammes et estimation de la densité*, Que-sais-je? numéro 2055, P.U.F., Paris (1983).
- [3] **Ycart B.** *Démarrer en Scilab*,
<http://www.math-info.univ-paris5.fr/Enseignements/>

Table des matières

1	Position du problème	169
1.1	Distances entre des courbes	170
1.2	Convergences stochastiques	171
2	Rappels sur les densités de probabilité	172

⁴⁰L'argument est faible. Les lois à densité et les lois discrètes ne se mélangent pas dans la pratique.

3	Histogrammes	176
3.1	Histogrammes de la loi $\mathcal{U}(0, 1)$	176
3.2	Histogrammes de la loi $\mathcal{N}(0, 1)$	180
3.3	Un résultat théorique sur la convergence des histogrammes	183
4	Estimateurs à noyau	189
4.1	Définition des estimateurs à noyau	189
4.2	Convergence des estimateurs à noyau gaussien	190
4.3	Convergence des estimateurs de la fenêtre mobile	195
4.4	Un résultat théorique sur la convergence des estimateurs à noyau	199
5	Estimation avec hypothèses défailantes	200
5.1	Cas d'une densité à estimer non uniformément continue	200
5.2	Cas de l'estimation d'une densité discontinue	202
5.3	Cas de l'estimation de la densité d'une loi mixte	203
5.4	Cas d'une loi purement discrète	205
6	Intérêt de l'estimation de la densité	206