Cours n°3 :Estimation non-paramétrique d'une densité de probabilité

Hamel Elhadj

Département de mathématiques
Université Hassiba Benbouali-Chlef
2021/2022
Ce cours est distiné aux étudiants Master2 mathématiques
Option :Mathématique Apliquées et statistique

Plan de la présentation

Estimation non-paramétrique d'une densité de probabilité

Introduction générale

Comment estimer non-paramétriquement la densité de probabilité f, en se basant sur les observations $X_1, ..., X_n$?
Il existe plusieurs méthodes d'estimation non-paramétrique d'une densité. La méthode la plus simple est celle de l'histogramme.
L'objectif de cette section est de décrire quelques méthodes importantes d'estimation nonparamétrique d'une densité.

l'estimation de densité (univariée)

- Densité d'une variable X :

$$f(x) = \lim_{\Delta x \to 0} \mathbb{P}[X \in [x - \Delta x, x + \Delta x]]/2\Delta x$$

- Observations : $X_1, ..., X_n$ v.a. i.i.d. réelles de fdr F et admettant une densité $f = F'$
- But : estimer (à partir des observations) f en faisant le moins d'hypothèses possibles sur cette densité.
- Typiquement, on supposera que $f \in \mathcal{F}$ espace fonctionnel et on notera $\hat{f}_n$ un estimateur de f .

- Objectif : Obtenir des informations de nature géométrique sur la distribution des variables. ex (mode ....)

Illustre avec R : f(x)<= curve(dnorm,-4,4,col=4, main="Densitée de la loi Normale centrée réduite")

Histogramme

On suppose que la densité f est définie sur un intervalle borné $[a,b] \in \mathbb{R}$ et $f \in \mathbb{L}^2([a,b])$.
Définition :

- Soit $I = (\mathcal{A}_k)_{1 \leq k \leq n}$ une partition de $[a,b]$ (i.e. intervalles disjoints dont l'union est $[a,b]$),
- On note $\nu_k = Card\{i; X_i \in \mathcal{A}_k\}$ le nombre d'observations dans la classe $\mathcal{A}$, et h la longueur de l'intervalle (classe) $\mathcal{A}$.
- L'estimateur par histogramme de f est défini par :

$$\hat{f}_H(x) = \frac{\nu_k}{nh} = \frac{1}{nh} \sum_{i=1}^{n} \mathbb{I}_{\mathcal{A}_k}(X_i)$$

- Il affecte à chaque intervalle une valeur égale à la fréquence des observations dans cet intervalle, renormalisée par la longueur de l'intervalle

Histogramme

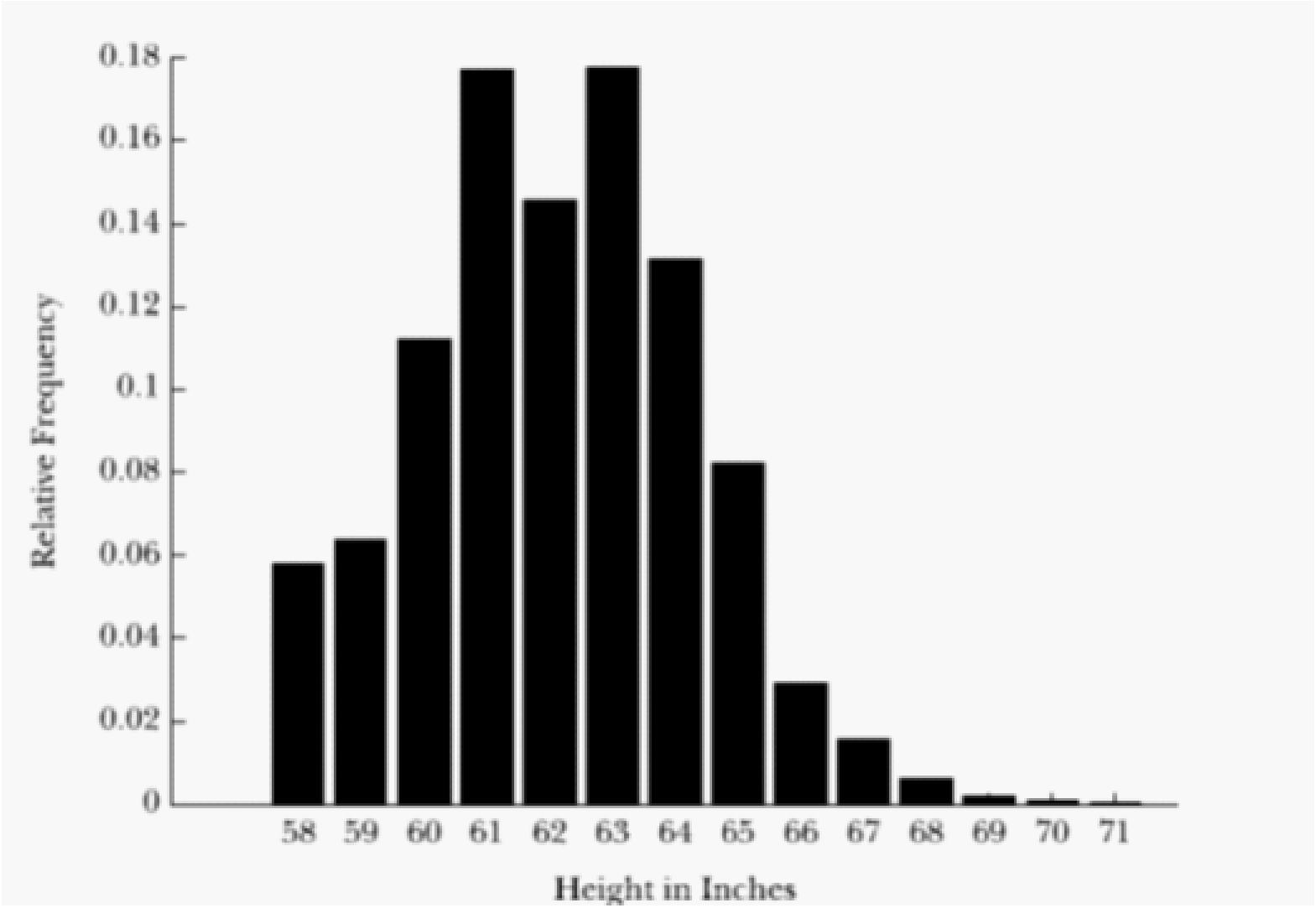
Histogrammes dits "réguliers"

- Un histogramme est dit régulier lorsque tous les intervalles (classe) $\mathcal{A}_k$ de la partition ont la même largeur.
- la longueur de $\mathcal{A}_k$, est appelé fenêtre où Bandwith, le pas .
- Un histogramme régulier prend des valeurs proportionnelles à la fréquence des observations dans chaque intervalle.

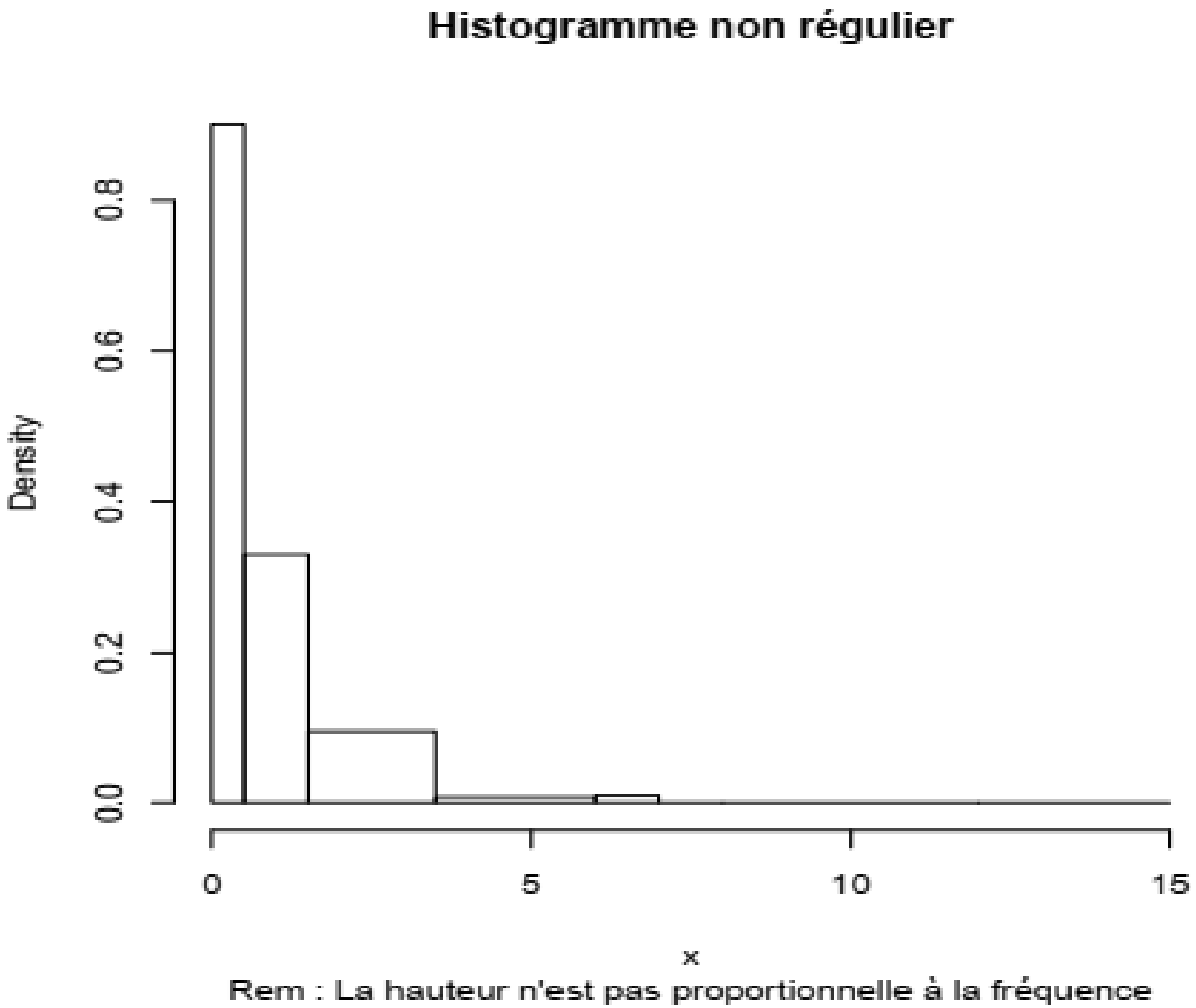
Remarque :

- L'histogramme est une fonction constante par morceaux. C'est donc une fonction très irrégulière. Cette notion de régularité n'a rien à voir avec la précédente ....

Histogramme régulière ($h_1 = h_2, ... = h_k$



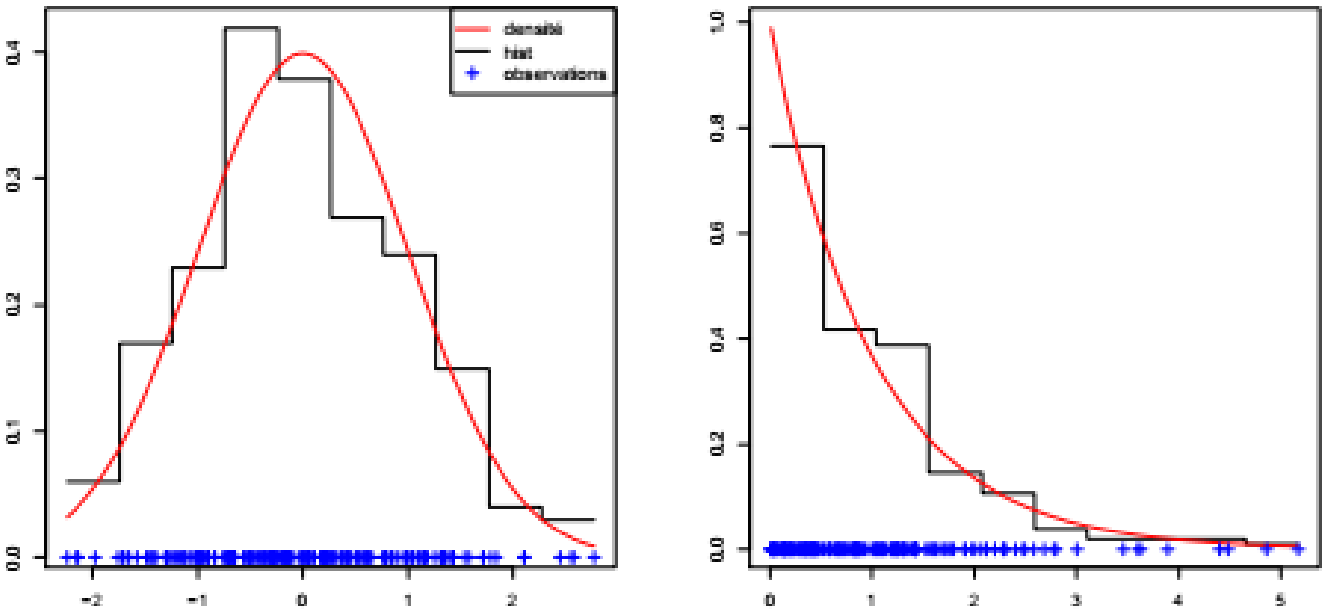
Histogramme irrégulière ($h_1, h_2, ..., h_k$



Rem : La hauteur n'est pas proportionnelle à la fréquence

Estimation par Histogramme (Exemple introductif)

- Soit $(X_i)_{i=1,\ldots,n}$ i.i.d. de densité f . On considére l'estimateur $\hat{f}_H$ par histogramme régulier $h_1 = h_2 = ... = h_k$.
- L'intervalle d'estimation est découpé en k sous-intervalles de même longueur
- Chaque classe de l'histogramme est égal à (proportion d'observations dans l'intervalle $\times$ cte).
- Exemples :



Estimation par Histogramme

- On choisi un point d'origine a_0 et une longueur de classe h ($h > 0$). Les classes sont définies par : $\mathcal{A}_k = [a_k, a_{k+1}[, k \in \mathbb{Z}$ (la k^{eme} classe) avec $a_{k+1} = a_k + h,$
- Un estimateur de f est donné par :

$$\hat{f}_H(x) = \frac{1}{nh} \# \{i : X_i \quad \text{est dans la classe qui contient x}\} .$$

- Si nous notons le nombre d'observations dans une classe $\mathcal{A}_k$ par ν_k, l'estimateur du type histogramme de densité s'écrit

$$\hat{f}_H(x) = \frac{\nu_k}{nh} = \frac{1}{nh} \sum_{i=1}^{n} \mathbb{I}_{[a_k, a_{k+1}[}(X_i)$$

Propriété asymptotique

- **Le biais**

$$\text{biais}(\hat{f}_H(x)) = \mathbb{E}[\hat{f}_H(x)] - f(x) = \frac{1}{2} f(x)\,(h - 2(x - a_k)) + \mathcal{O}(h^2).$$

où $\mathcal{O}$ est un terme résiduel et f' est la dérivée de f.
f' doit être une fonction de $L^1(\Omega)$ absolument continue [1] et de carrée intégrable [2].

- **La variance** de l'estimateur est donnée , pour tout $x \in \Omega$, par :

$$\text{Var}(\hat{f}_H(x)) = \mathbb{E}[(\hat{f}_H(x))^2] - (\mathbb{E}[\hat{f}_H(x)])^2 = \frac{f(x)}{nh} + \mathcal{O}(n^{-1}).$$

Cette variance tend vers zéro quand le produit $nh \longrightarrow \infty$ quand le nombre d'observation $n \longrightarrow \infty$.

- **Erreur quadratique moyenne (En anglais Mean Squared Error) (MSE) :** La distance la plus couramment utilisée est définie par la moyenne du carré de la valeur absolue de la différence entre l'estimateur et la densité à estimer.

conclusion

- L'histogramme de densité est un estimateur très élémentaire, mais peut quand même déjà donner une première idée assez bonne de la forme de la densité a estimé f. Par contre, si on voulait utiliser cet estimateur dans d'autres analyses statistiques (comme par exemple l'analyse discriminante, l'estimation d'un taux de hasard, etc) il vaudrait mieux démarrer avec un estimateur plus précis.
- L'histogramme de densité est une fonction étagée, et donc discontinue.
- L'estimateur $\hat{f}_H$ dépend de deux paramètres : le point d'origine a_0 et la largeur de classe h. Ces deux paramètres peuvent avoir une influence importante sur l'histogramme.

Histogramme Mobile

Rappelons que la densité de probabilité f est égale à la dérivée de la fonction de répartition F (si cette dérivée existe). On peut donc écrire

$$\begin{aligned} f(x) &= \lim_{h\to 0} \frac{F(x+h)-F(x-h)}{2h} \\ &= \lim_{h\to 0} \frac{P\{x-h<X\leq x+h\}}{2h} \end{aligned}$$

Un estimateur de $f(x)$ est alors

$$\begin{aligned} \widehat{f_{HM}}(x) &= \frac{1}{2h}\frac{\#\{i : x-h<X_i\leq x+h\}}{n} \\ &= \frac{1}{2hn}\sum_{i=1}^{n} I\{x-h<X_i\leq x+h\} \\ &= \frac{1}{2hn}\sum_{i=1}^{n} I\left\{-1\leq \frac{x-X_i}{h}<1\right\} \end{aligned}$$

Histogramme Mobile

Notons que cet estimateur peut encore s'écrire comme :

$$\hat{f}_{HM}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}\omega\left(\frac{x - X_i}{h}\right)$$

où

$$\omega(y) = \begin{cases} 1/2 & \text{si } y \in [-1,1] \\ 0 & \text{sinon} \end{cases}$$

les propriétés de l'estimateur simple $\hat{f}_{HM}(x)$

- Remarquons que

$$\hat{f}_{HM}(x) = \frac{\hat{F}_n(x+h) - \hat{F}_n(x-h)}{2h} = \lim_{h\to 0} \frac{P\{x-h < X \leq x+h\}}{2h}$$

avec $\hat{F}_n$ la fonction de répartition empirique. Le paramètre de lissage h dépend de la taille de l'échantillon n, c'est-à-dire $h = h_n$.

- Nous savons que $n\hat{F}_n(x) = \sum_{i=1}^{n} I_{\{X_i \leq x\}} \sim \mathcal{B}in(n, F(x))$
- Le biais on a :

$$2nh\hat{f}_{HM}(x) = \hat{F}_n(x+h) - \hat{F}_n(x-h) \sim \mathcal{B}in(n, F(x+h) - F(x-h))$$

$$\Longrightarrow \mathbb{E}[2nh\hat{f}_{HM}(x)] = n(F(x+hn) - F(x-hn))$$

$$\Longrightarrow \mathbb{E}[\hat{f}_{HM}(x)] = \frac{1}{2nh}[F(x+hn) - F(x-hn)]$$

les propriétés de l'estimateur simple $\hat{f}_{HM}(x)$

- Pour la variance nous trouvons :

$$Var[2nh\hat{f}_{HM}(x)] = n[F(x+h) - F(x-h)][1 - (F(x+h) - F(x-h))]$$

$$\Longrightarrow Var[\hat{f}_{HM}(x)] = \frac{1}{4nh^2} n[F(x+h) - F(x-h)][1 - (F(x+h) - F(x-h))]$$

- Remarquons que, si $n \longrightarrow \infty$ et $h_n \longrightarrow 0$, alors

$$\mathbb{E}[\hat{f}_{HM}(x)] \longrightarrow f(x) \quad , \quad nh_n Var[\hat{f}_{HM}(x)] \longrightarrow \frac{1}{2} f(x)$$

- Le risque moyen quadratique : $MSE = var[\hat{f}_{HM}(x)] + biais^2[\hat{f}_{HM}(x)]$ si $h_n \longrightarrow 0$ et $nh_n \longrightarrow \infty$ quand $n \longrightarrow \infty$, on a que

$$\mathbb{E}[\hat{f}_{HM}(x) - f(x)]^2 \longrightarrow 0$$

Conclusion

- pour tout point x. L'estimateur simple $\hat{f}_{HM}(x)$ est alors un estimateur consistent de $f(x)$.
- On n'a plus le problème du choix d'un point d'origine (un point a_0) comme dans le cas d'un histogramme de densité.
- L'estimateur

$$\hat{f}_{HM}(x) = \frac{1}{2nh} \sum_{i=1}^{n} I\{x - h < X_i \leq x + h\} = \frac{1}{2nh} \sum_{i=1}^{n} I\{X_i - h \leq x < X_i +$$

est une fonction discontinue, avec des discontinuités aux points $X_i \pm h$, et constante entre ces points.

L'estimateur à noyau

construction

Rappelons l'estimateur simple :

$$\hat{f}_{HM}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}\omega\left(\frac{x - X_i}{h}\right)$$

où

$$\omega(y) = \begin{cases} 1/2 & \text{si } y \in [-1, 1] \\ 0 & \text{sinon} \end{cases}$$

la densité de probabilité uniforme sur l'intervalle [-1, 1[. Cet estimateur peut être généralisé en remplaçant la fonction de poids $\omega(.)$ (la densité de probabilité uniforme) par une fonction de poids plus générale K (par exemple une densité de probabilité quelconque).

L'estimateur à noyau

La méthode d'estimation non paramétrique à noyau fut introduit par Rosenblatt en 1956 , puis amélioré par Parzen en 1962 .
L'estimateur à noyau est de la forme

$$\hat{f}_K(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - X_i}{h}\right)$$

- K la fonction de poids ("weight function") ou le noyau ("the kernel function") (Souvent K une densité de probabilité symétrique)
- h (le paramètre de lissage ("smoothing parameter") la fenêtre .
- Le noyau K détermine la forme des bosses, et la fenêtre h détermine la largeur des bosses
- Le paramètre de lissage h a une grande influence sur la performance de l'estimateur.

definition de noyau

Définition

Soit K : $\mathbb{R} \longrightarrow \mathbb{R}^+$ *une fonction intégrable tel que* $\int_{\mathbb{R}} K(u)du = 1$

K est dit noyau.

Un noyau K est dit ***symétrique*** *si, pour tout u dans son ensemble de définition* $K(u) = K(-u)$ *;* $\forall$ *u dans sa domaine de définition.*

ce qui implique les égalités suivantes :

$$\int_{\mathbb{R}} uK(u)du = 0, \quad \int_{\mathbb{R}} K^2(u)du \leq \infty, \quad \int_{\mathbb{R}} u^2K(u)du \leq \infty$$

La plupart des noyaux sommatifs couramment utilisés en estimation fonctionnelle sont monomodaux, symétriques et centrés .

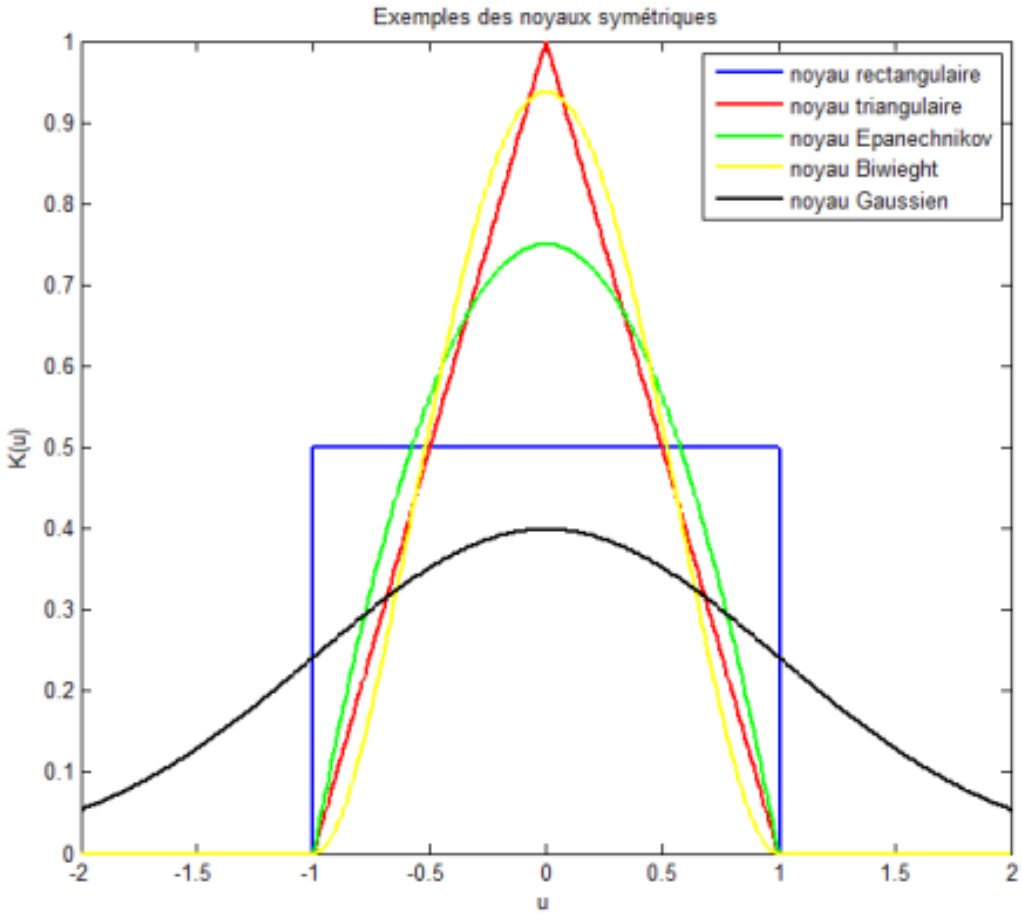
Exemples des noyaux continus symétriques

Voici quelques exemples de noyaux symétriques les plus utilisés :

Noyaux	Supports	Densités
Biwieght	$[-1,1]$	$K(u)=\frac{15}{16}\left(1-u^2\right)^2$
Epanechnikov	$[-1,1]$	$K(u)=\frac{3}{4}\left(1-u^2\right)$
Gaussien	$\mathbb{R}$	$K(u)=\frac{1}{\sqrt{2\pi}}\exp\left(-u^2/2\right)$
Rectangulaire	$[-1,1]$	$K(u)=\frac{1}{2}$
Triangulaire	$[-1,1]$	$K(u)=1-\lvert u\rvert 1\{\lvert u\rvert\leq 1\}$

(1)

Exemples des noyaux continus symétriques



Quelques propriétés de l'estimateur à noyau :

l'estimateur à noyau $\hat{f}_K(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x-X_i}{h}\right)$ possède les propriétés suivantes :

- Si K est une densité de probabilité, alors $\hat{f}_K(x)$ est aussi une densité de probabilité.
- $\hat{f}_K(x)$ a les mêmes propriétés de continuité et de différentiabilité que K :
  - Si K est continue, $\hat{f}_K(.)$ sera une fonction continue.
  - Si K est différentiable, $\hat{f}_K(.)$ sera une fonction différentiable.
  - Si K peut prendre des valeurs négatives, alors $\hat{f}_K(.)$ pourra aussi prendre des valeurs négatives.

Quelques propriétés asymptotique :

Expressions du biais et de la variance : On est besion de ces conditions :

- f fonction differentiable (f" existe), f' bornée, et $R(f) = \int [f''(x)]^2 dx < \infty$
- $h = h_n$ une suite réelle tq $h \to 0$ and $nh \to \infty$ as $n \to \infty$
- K a kernel non-negative

donc

- **Le biais :**

$$\mathbb{E}[\hat{f}_K(x)] - f(x) = \frac{h^2}{2}\mu_2(K)f''(x) + o\left(h^2\right)$$

- **La variance :**

$$V[\hat{f}_K(x)] = \frac{R(K)}{nh}f''(x) + o\left(\frac{1}{nh}\right)$$

-

$$MSE(\hat{f}_K(x)) = \frac{R(K)}{nh}f(x) + \frac{h^4}{4}\mu_2^2(K)f''(x)^2 + o\left(h^4 + \frac{1}{nh}\right) \xrightarrow[n\to\infty]{} 0$$

Quelques propriétés asymptotique :

- MISE :

$$MISE(\hat{f}_K(x)) = \frac{R(K)}{nh} + \frac{h^4}{4}\mu_2^2(K)R(f'') + o\left(h^4 + \frac{1}{nh}\right)$$

où $\mu_2 = \int u^2 K(u)du$ et $R(K) = \int K^2(u)du$, où $R(g) = \int g^2(u)du$, pour une fonction g de carré intégrable.

Remarquons que

–Si h décroît alors le $(bias^2) \searrow$ et la variance $\nearrow$

–Si h augmente alors le $(bias^2) \nearrow$ et la variance $\searrow$

Il faut donc essayer de choisir un h qui fasse un compromis entre le $(bias^2)$ et la variance.

- On note l'approximation asymptotique de la MISE par

$$AMISE(\hat{f}_K(x)) \simeq \frac{R(K)}{nh} + \frac{h^4}{4}\mu_2^2(K)R(f'')$$

Mean Squared Error

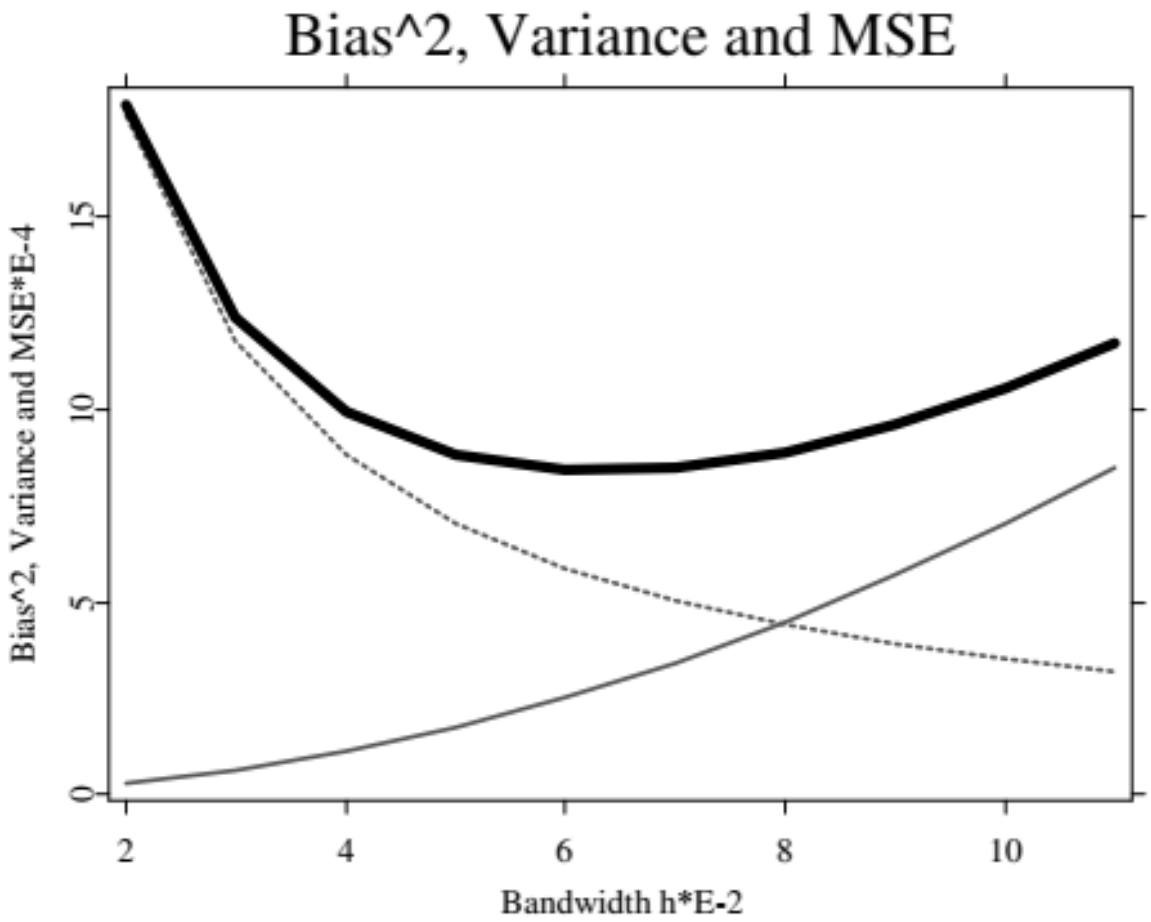


FIGURE: Squared bias part (thin solid), variance part (thin dashed) and MSE (thick solid) for kernel density estimate

Choix théoriques optimaux du paramètre de lissage

Pour le paramètre de lissage on fait la distinction entre

- h paramètre de lissage constant (ou global)
- h(x) paramètre de lissage variable (local)

Ces choix différents du paramètre de lissage résultent en les estimateurs à noyau suivants :

$$\hat{f}_K(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x-X_i}{h}\right), \quad \hat{f}_K(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h(x)}K\left(\frac{x-X_i}{h(x)}\right)$$

Le choix h(x) implique qu'un noyau différent est utilisé en chaque point. Nous allons ensuite décrire des choix théoriques optimaux des paramètres de lissage h et h(x).

Choix théoriques optimaux du paramètre de lissage

Un critère approprié pour sélectionner un paramètre de lissage constant h est la MISE.
Le paramètre de lissage optimal est la valeur de h qui minimise la MISE. Notons cette valeur par $h^\star = h_{MISE}$. Une approximation asymptotique de h_{MISE} est donnée par

$$h_{\text{AMISE}} = \left\{ \frac{R(K)}{\mu_2^2 R\left(f''\right)} \right\}^{1/5} n^{-1/5} \tag{2}$$

et

$$h_{\text{MISE}} \sim \left\{ \frac{R(K)}{\mu_2^2 R\left(f''\right)} \right\}^{1/5} n^{-1/5} \tag{3}$$

c'est-à-dire $\lim_{n\to\infty} \dfrac{h_{\text{MISE}}}{h_{\text{AMISE}}} = 1$
Remarquons que si f montre des changements rapides, alors $R(f'')$ sera grand, et h_{MISE} sera petit.

Choix théoriques optimaux du paramètre de lissage

Un critère approprié pour sélectionner un paramètre de lissage variable (local) $h(x)$ est la mesure de performance locale $\mathrm{MSE}\hat{f}_{n,L}(x)$.
Nous introduisons les notations suivantes :
$h_{\mathrm{MSE}}(x) = \arg\min_h \mathrm{MSE}\hat{f}_{n,L}(x)$ et $h_{\mathrm{AMSE}}(x) = \arg\min_h \mathrm{MSE}\hat{f}_{n,L}(x)$
nous trouvons que

$$h_{\mathrm{AMSE}}(x) \sim \left\{ \frac{f(x)R(K)}{\mu_2^2 \{f''\}^2} \right\}^{1/5} n^{-1/5} \tag{4}$$

sous condition que $f''(x) \neq 0$.
Remarque : Les choix h_{AMISE} et $h_{\mathrm{AMSE}}(x)$ sont des choix théoriques, qui ne sont pas utilisables en pratique car ils dépendent des quantités inconnues f et f''.

Choix pratiques du paramètre de lissage

Nous allons maintenant décrire quelques choix optimaux pratiques pour un paramètre de lissage constant et un paramètre de lissage variable (local)

- La règle simple de référence à une distribution normale
- La méthode de validation croisée

La règle simple de référence à une distribution normale

Rappellons l'expression pour le paramètre de lissage optimal constant :

$$h_{\text{AMISE}} = \left\{ \frac{R(K)}{\mu_2^2 R\left(f''\right)} \right\}^{1/5} n^{-1/5} \tag{5}$$

Supposons que f appartient à une famille de distributions normales $\mathcal{N}(\mu, \sigma^2)$, de moyenne μ et variance σ^2 inconnues. Alors

- $f(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$
- $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ la densité de probabilité normale réduite
- $f''(x) = \frac{1}{\sigma^3}\varphi''\left(\frac{x-\mu}{\sigma}\right)$

On cacule La quantité inconnue $R(f'') = \int (f''(x))^2 dx$

- $R(f'') = \int (f''(x))^2 dx = \frac{1}{\sigma^6} \int \left\{ \varphi''\left(\frac{x-\mu}{\sigma}\right) \right\}^2 dx$.....

La règle simple de référence à une distribution normale

Donc aprés les calculs, en faisant référence à une densité de probabilité normale, l'expression du paramètre de lissage optimal asymptotique devient :

$$\hat{h}_{\mathrm{NR}} = \left\{ \frac{8\sqrt{\pi}R(K)}{3\mu_2^2} \right\}^{1/5} \sigma n^{-1/5} \tag{6}$$

Le paramètre de lissage du type "normal reference" est défini par

$$\hat{h}_{\mathrm{NR}} = \left\{ \frac{8\sqrt{\pi}R(K)}{3\mu_2^2} \right\}^{1/5} \hat{\sigma} n^{-1/5} \tag{7}$$

où $\hat{\sigma}$ est un estimateur de σ, l'écart-type de la population X. Ce paramètre de lissage est trés simple (*"Rule- of-Thumb" bandwidth selector*).

- Comment estimer σ ?

La règle simple de référence à une distribution normale

Quelques choix possibles pour sb sont donnés ci-dessous.

- L'écart-type empirique $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$
- L'écart interquartile empirique standardisé :

$$\frac{\text{l'ecart interquartile empirique}}{\phi^{-1}(\frac{3}{4}) - \phi^{-1}(\frac{1}{4})} \simeq \frac{R}{1.349}$$

où $\phi(.)$ est la fonction de répartition d'une normale réduite.

- utiliser le mimimum entre S et $R/1.349$, c'est-à-dire d'utiliser le paramètre de lissage suivant :

$$\hat{h}_{\text{NR}} = \left\{\frac{8\sqrt{\pi}R(K)}{3\mu_2^2}\right\}^{1/5} \min\left\{S, \frac{R}{1.349}\right\} n^{-1/5} \tag{8}$$

La règle simple de référence à une distribution normale

Voici, pour quelques noyaux, l'expression de $\hat{h}_{\mathrm{NR}}$:

noyau K	paramétre de lissage pratique $\hat{h}_{\mathrm{NR}}$
densité normale réduite	$\hat{h}_{\mathrm{NR}} = 1.06 \min \left\{ S, \frac{R}{1.349} \right\} n^{-1/5}$
noyau Epanechnikov	$\hat{h}_{\mathrm{NR}} = 2.34 \min \left\{ S, \frac{R}{1.349} \right\} n^{-1/5}$
noyau biweight	$\hat{h}_{\mathrm{NR}} = 2.78 \min \left\{ S, \frac{R}{1.349} \right\} n^{-1/5}$

Excercice

Calculer R(K) et R(K") pour les noyaux Epanechnikov et biweight ?

La méthode de validation croisée- (cross-validation)

As mentioned earlier, we will focus on least squares cross-validation. To get started, consider an alternative distance measure between $\widehat{f}$ and f , the *integrated squared error* (ISE) :

$$ISE(h) = IS\widehat{Ef} = \int \{\widehat{f}_h(x) - f(x)\}^2 dx.$$

$$ISE(h) = \int \widehat{f}^2 x dx - 2\int \{\widehat{f}_h . f\}(x)dx + \int f^2(x)dx.$$

Apparently, $\int f^2(x)dx$ does not depend on h and can be ignored as far as minimization over h is concerned. Moreover, $\int \widehat{f}_h^2(x)dx$ can be calculated from the data. This leaves us with one term that depends on h and involves the unknown quantity f .

La méthode de validation croisée- (cross-validation)

If we look at this term more closely, we observe that $\int \{\widehat{f_h}.f\}(x)dx$ is the expected value of $\widehat{f_h}(X)$, where the expectation is calculated w.r.t. an independent random variable X. We can estimate this expected value by
$\widehat{\mathbb{E}\{\hat{f}_h(X)\}} = \frac{1}{n}\sum_{i=1}^{n}\hat{f}_{-i}(X_i)$, where $\hat{f}_{-i}(X_i) = \frac{1}{n-1}\sum_{j=1,i\neq j}^{n} K_h(x - X_i)$
Here $\hat{f}_{-i}(x)$ is the leave-one-out estimator. As the name of this estimator suggests the i^th observation is not used in the calculation of $\hat{f}_{-i}(X_i)$.
Let us repeat the formula of the integrated squared error (ISE), the criterion function we seek to minimize with respect to h :

$$ISE(h) = \int \hat{f}^2(x)dx - 2\mathbb{E}\{\hat{f}^2(x)\} + \int f^2(x)dx.$$

La méthode de validation croisée- (cross-validation)

Now we can reap the fruits of the work done above and plug in for estimating $\mathbb{E}\{\hat{f}^2(x)\}$. This gives the so-called *cross-validation criterion*

$$CV(h) = \int \hat{f}^2(x)dx - \frac{2}{n(n-1)} \sum_{j=1, i\neq j}^{n} K_h(X_i - X_j).$$

We have almost everything in place for an applicable formula that allows us to calculate an optimal bandwidth from a set of observations.

l'intervalle de confiance

Un résultat important issu de Silvermann.[1986] montre que si f'', la dérivée seconde de f, existe et que la fenètre $h = qn^{-1/5}$ (avec q une constante), alors, pour tout $x \in \mathbb{R}$, l'estimateur à noyau $\hat{f}_{\mathcal{K}}(x)$ est asymptotiquement normal (c'est-à-dire converge en loi vers la loi normale) :

$$n^{-1/5}\left(\hat{f}_{\mathcal{K}}(x) - f(x)\right) \xrightarrow{L} \mathcal{N}\left(\frac{q^2}{2}f''(x)\sigma_K^2, \frac{1}{q}f(x)R(K)\right) \tag{9}$$

l'intervalle de confiance

L'expression précédente est équivalente, par une transformation linéaire de la loi normale,à

$$\left(\hat{f}_{\mathcal{K}}(x) - f(x)\right) \xrightarrow{L} \mathcal{N}\left(\underbrace{\frac{h^2}{2} f''(x)\sigma_K^2}_{Biais(\hat{f})}, \underbrace{\frac{1}{nh} f(x) R(K)}_{var(\hat{f})}\right) \tag{10}$$

l'intervalle de confiance

Sous l'hypothèse que le terme de biais est négligeable par rapport au terme de la variance, l'intervalle de confiance au seuil α sur la valeur de la densité $f(x)$, pour tout $x \in \Omega$, est donné par :

$$\left[\hat{f}_{\mathcal{K}}(x) - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{f}_{\mathcal{K}}(x)R(K)}{nh}}, \hat{f}_{\mathcal{K}}(x) + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{f}_{\mathcal{K}}(x)R(K)}{nh}}\right] \tag{11}$$

où $Z_{1-\frac{\alpha}{2}}$ est le $(1-\frac{\alpha}{2})$-quantile de la loi normale centrée réduite.

Conclusion

L'estimateur de densité de Parzen-Rosenblatt $\hat{f}_{\mathcal{K}}(x)$ défini par l'expression **??**) dépend de deux paramètres : la forme fonctionnelle du noyau $\mathcal{K}$ et la largeur de bandwith h. De manière générale, il est admis le choix du noyau $\mathcal{K}$ a beaucoup moins d'importance que celui de la largeur de fenètre h. h détermine le degré de lissage de l'estimation d'une densité. Une faible largeur de fenètre implique un faible degré de lissage et résulte en une fonction de densité irrégulière. Une large valeur de h conduit à une estimation lisse.
Ils existent d'autres approches que la méthode des noyaux ont été proposées : *séries orthogonales, splines, maximum de vraisemblance pénalisé, plus proches voisins (nearest neighbour), ondelettes*,.. etc. Globalement on peut dire qu'elles ne donnent pas des résultats significativement meilleurs que la méthode des noyaux. Aucune ne peut se soustraire à l'incontournable problème du choix d'un paramètre de lissage, explicite ou non.

La fonction de régression