

Économetrie non paramétrique
I. Estimation d'une densité

Stéphane Adjemian

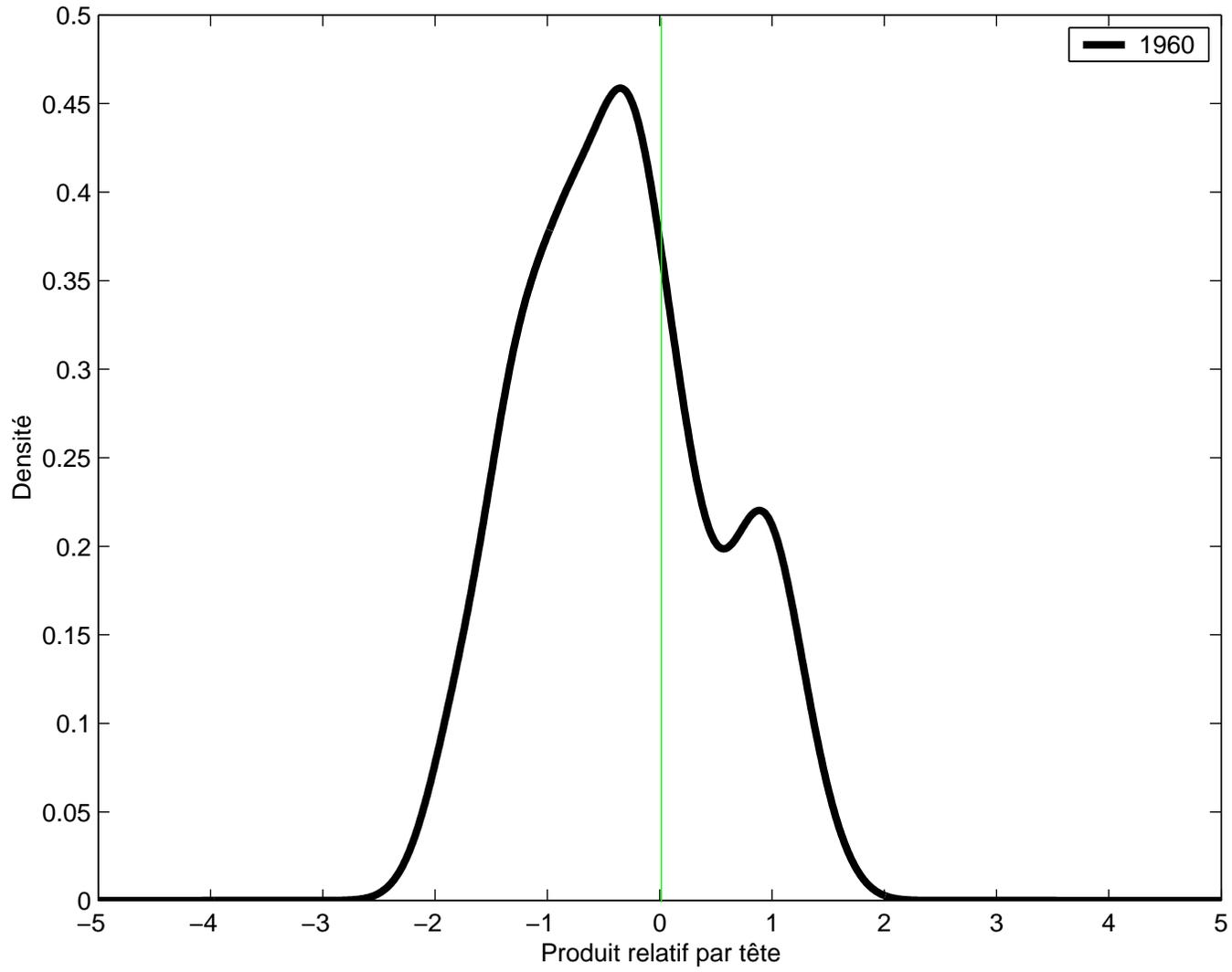
Université d'Évry

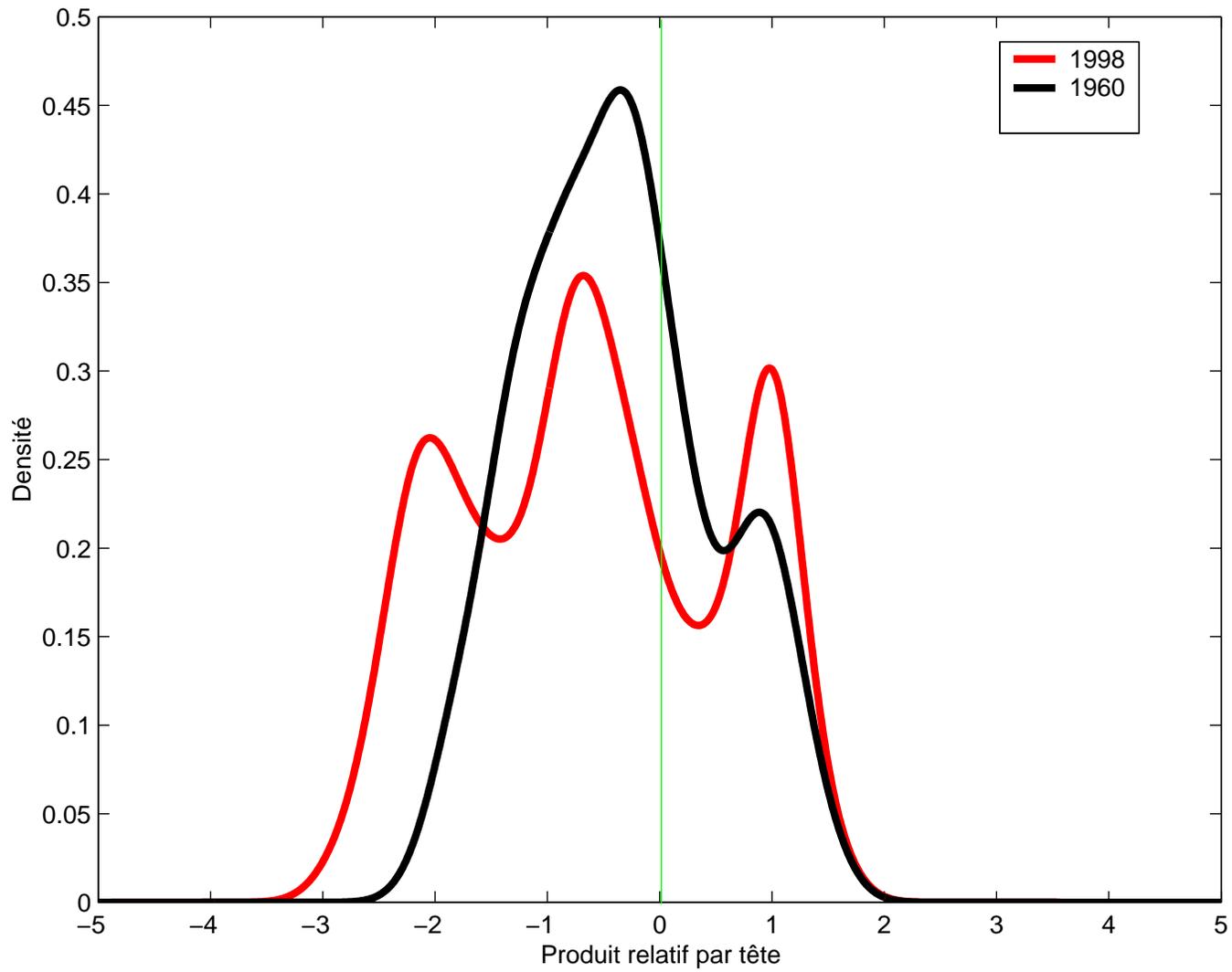
Janvier 2004

1 Introduction

1.1 Pourquoi estimer une densité ?

- Étudier la distribution des richesses...
- Proposer un résumé d'un exercice de monte-carlo...
- La distribution asymptotique d'une statistique peut dépendre d'un paramètre de nuisance lié à une distribution inconnue...





TEST DE DICKEY – FULLER

Soit le processus générateur des données suivant :

$$y_t = y_{t-1} + \varepsilon_t \quad (1)$$

avec $\varepsilon_t \sim N(0, 1)$. On estime le modèle :

$$y_t = \rho y_{t-1} + u_t, \text{ pour } t = 1, \dots, T \quad (2)$$

La statistique de Dickey – Fuller (pour tester $\rho = 1$ contre $|\rho| < 1$) est $Z(\hat{\rho}_T) = T(\hat{\rho}_T - 1)$. On connaît sa distribution asymptotique :

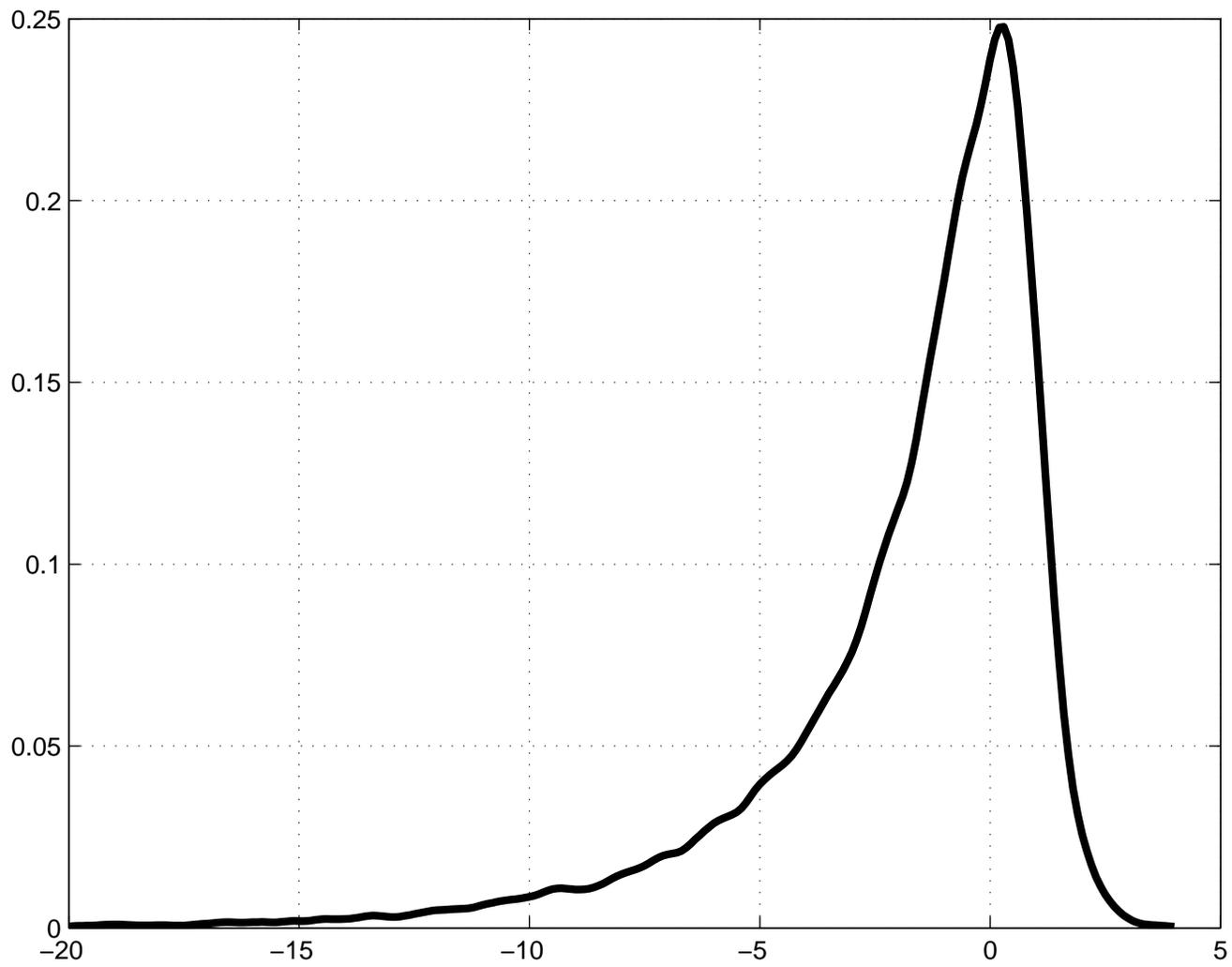
$$Z(\hat{\rho}_T) \xrightarrow{T \rightarrow \infty} \frac{1}{2} \frac{W(1)^2 - 1}{\int_0^1 W(r)^2 dr} \quad (3)$$

à quoi peut bien ressembler cette distribution ?...

MONTE – CARLO

IDÉE : Simuler le modèle sous l'hypothèse nulle, puis estimer (pour chaque série temporelle simulée) la statistique de test.

```
B = 50000;
T = 20000;
statistics = zeros(B,1);
for(i=0;i<B;i++)
{
    epsilon = rann(T,1);
    y = sumc(epsilon);
    rho = y[0:T-2][1]'*y[1:T-1][1]/
        (y[0:T-2][1]'*y[0:T-2][1]);
    statistics[i][1] = T*(rho-1);
}
```



- La variance de l'estimateur de la médiane d'une variable aléatoire X est égale à $(1/4)n^{-1}f^{-2}(0)$, où $f(\bullet)$ est la densité (inconnue) de la variable aléatoire X .

1.2 Comment estimer une densité ?

1.2.1 Approche paramétrique

Supposons que l'on observe la réalisation d'une suite de variables aléatoires indépendantes et identiques en loi (x_1, x_2, \dots, x_n) . Si on a une idée précise du processus générateur des données, on peut alors estimer la densité en utilisant un estimateur du maximum de vraisemblance.

EXEMPLE : Supposons que l'on ait $X_i \sim N(\mu, 1)$, sa densité est alors

donnée par :

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2}$$

Estimer la densité se ramène alors à estimer un unique paramètre (l'espérance μ). Limites évidentes...

1.2.2 Approche non-paramétrique

- L'HISTOGRAMME. Il s'agit de l'estimateur le plus simple. Cet estimateur n'est pas conçu pour les variables aléatoires continues mais discrètes... Cela peut poser des problèmes (par exemple lorsque l'on étudie la distribution mondiale des richesses).
- L'ESTIMATEUR À NOYAU. Il s'agit de l'estimateur le plus populaire. Il est adapté aux variables aléatoires continues (mais surtout pas aux *v.a.* discrètes).

2 L'Histogramme

- Soit X une variable aléatoire continue. On note f sa densité de probabilité. On dispose d'un échantillon de n réalisations (X_1, \dots, X_n) tirées dans cette distribution inconnue.
- En sélectionnant une origine x_0 on peut construire des intervalles:

$$B_j = [x_0 + (j - 1)h, x_0 + jh] \text{ pour } j \text{ dans } \mathbb{Z}$$

où $h > 0$ est le paramètre de *binwidth* qui définit la taille des intervalles.

- L'histogramme estime la densité f au point x de la façon suivante :

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n \sum_j \mathbb{I}_{X_i \in B_j} \mathbb{I}_{x \in B_j} \quad (4)$$

2.1 Construction de l'histogramme

2.2 Histogramme et Maximum de Vraisemblance

2.3 Biais

Supposons que $x_0 = 0$ ($\rightarrow B_j = [(j-1)h, jh[$) et que $x \in B_j$.

$$\text{Biais} \left\{ \widehat{f}_h(x) \right\} = \frac{1}{nh} \sum_{i=1}^n \mathbb{E} \left[I_{\{X_i \in B_j\}}(x) \right] - f(x)$$

$$\Leftrightarrow \text{Biais} \left\{ \widehat{f}_h(x) \right\} = f'(m_j(h)) (m_j(h) - x) + o(h)$$

$$\Leftrightarrow \text{Biais} \left\{ \widehat{f}_h(x) \right\} \approx f'(m_j(h)) (m_j(h) - x)$$

où $m_j(h) = (j - \frac{1}{2})h$ est le centre de l'intervalle B_j . On voit que le biais est généralement non nul, sauf si la densité est une fonction en escalier...

2.4 Variance

$$\mathbb{V} \left[\hat{f}_h(x) \right] = \frac{1}{nh} f(x) + o \left(\frac{1}{nh} \right)$$

Une diminution du paramètre h induit une augmentation de la variance et une augmentation du biais. \rightarrow Comment choisir le paramètre de *binwidth*?

2.5 Mean Squared Error

$$MSE \left\{ \hat{f}_h(x) \right\} = \mathbb{E} \left[\left(\hat{f}_h(x) - f(x) \right)^2 \right]$$

$$\Leftrightarrow MSE \left\{ \hat{f}_h(x) \right\} = \mathbb{V} \left[\hat{f}_h(x) \right] + \text{Biais} \left\{ \hat{f}_h(x) \right\}^2$$

$$\Leftrightarrow MSE \left\{ \hat{f}_h(x) \right\} = \frac{1}{nh} f(x) + f'(m_j(h))^2 (m_j(h) - x)^2 + o(h) + o\left(\frac{1}{nh}\right)$$

On note que l'erreur quadratique moyenne tend vers zéro lorsque h tend vers zéro et nh tend vers l'infini. Le paramètre h doit tendre moins vite vers zéro que la taille de l'échantillon (n) tend vers l'infini... Quelle vitesse relative ?... Le comportement de la MSE implique la convergence au sens de L_2 de l'histogramme vers $f(x)$ et donc la convergence en probabilité.

Pb La MSE est une mesure locale, on obtient une mesure globale en intégrant la MSE sur le support de la variable aléatoire d'intérêt.

$$MISE \left\{ \hat{f}_h(x) \right\} = \int_{-\infty}^{\infty} MSE \left\{ \hat{f}_h(x) \right\} dx$$

$$MISE \left\{ \hat{f}_h(x) \right\} \approx \frac{1}{nh} + \frac{h^2}{12} \int f'(x)dx \text{ pour } h \rightarrow 0$$

Asymptotiquement, la MISE est donnée par :

$$AMISE \left\{ \hat{f}_h(x) \right\} = \frac{1}{nh} + \frac{h^2}{12} \|f'\|_2^2$$

2.6 Paramètre de *binwidth* optimal

On va choisir le paramètre de *binwidth* de façon à minimiser l'AMISE.

$$h^* = \left(\frac{6}{n \|f'\|_2^2} \right)^{1/3}$$

- Ce résultat nous donne la vitesse de convergence de h vers zéro lorsque la taille de l'échantillon n tend vers l'infini, ie $h^*(n) \sim n^{-\frac{1}{3}}$.

- Pour avoir une idée plus précise du h optimal il faut évaluer $\|f'\|_2^2 \dots$ **Pb** On ne connaît pas $f(\bullet) \dots$
- En première approximation, on peut supposer que le DGP est gaussien, *ie* $f(x) = (2\pi)^{-1/2} e^{-\frac{1}{2}x^2}$. On obtient alors :

$$h^* = \left(\frac{24\sqrt{\pi}}{n} \right)^{\frac{1}{3}}$$

- Cette approximation est valable si la vraie distribution n'est pas trop éloignée d'une distribution gaussienne.

2.7 Limites de l'histogramme...

3 L'estimateur à noyau

3.1 Un histogramme local

Estimer une densité en un point x à l'aide d'un histogramme consiste essentiellement à compter le nombre de réalisations appartenant à un intervalle contenant x . L'estimateur à noyau de la densité en un point x est construit en comptant le nombre d'observations dans un intervalle autour de x . Plus formellement :

$$\hat{f}_h(x) = \frac{1}{2nh} \text{Card} \{X_i \in [x - h, x + h]\}$$

ou de façon équivalente :

$$\hat{f}_h(x) = \frac{1}{2nh} \sum_{i=1}^N \mathbb{I}_{\left\{ \left| \frac{x - X_i}{h} \right| \leq 1 \right\}}(x)$$

Plus généralement, on adopte les notations suivantes :

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

avec

$$K(u) = \frac{1}{2} \mathbb{I}_{\{|u| \leq 1\}}$$

où $K(\bullet)$ est le noyau (ou la fenêtre) uniforme. Pour ce noyau, les observations sont identiquement pondérées dans la détermination de $\hat{f}_h(x)$. Le choix du noyau est arbitraire... On pourrait attribuer plus de poids aux observations les plus proches de x .

Fenêtre	$K(u)$
Uniform	$\frac{1}{2} \mathbb{I}_{ u \leq 1}$
Bartlett	$(1 - u) \mathbb{I}_{ u \leq 1}$
Epanechnikov	$\frac{3}{4} (1 - u^2) \mathbb{I}_{ u \leq 1}$
Quartic	$\frac{15}{16} (1 - u^2)^2 \mathbb{I}_{ u \leq 1}$
Triweight	$\frac{35}{32} (1 - u^2)^3 \mathbb{I}_{ u \leq 1}$
Gaussian	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$
Cosinus	$\frac{\pi}{4} \cos\left(\frac{\pi}{2}u\right) \mathbb{I}_{ u \leq 1}$

Table 1: Quelques Noyaux

On supposera que le noyau satisfait les hypothèses suivantes:

(i) $K(\bullet)$ est une fonction symétrique autour de zéro.

$$(ii) \int K(u)du = 1$$

$$(iii) \int K(u)^2 du = \|K\|_2^2 < \infty$$

$$(iv) \int u^2 K(u)du = \mu_2(K) \neq 0$$

3.2 Propriétés statistiques

3.2.1 Le biais

$$\text{Biais} \left\{ \widehat{f}_h(x) \right\} = \frac{h^2 f''(x)}{2} \mu_2(K) + o(h^2)$$

Pour réduire le biais il faut choisir un petit paramètre de lissage...

Rapport à la convexité de la densité $f(x)$...

3.2.2 La variance

$$\mathbb{V} \left[\widehat{f}_h(x) \right] = \frac{1}{nh} \|K\|_2^2 f(x) + o\left(\frac{1}{nh}\right) \text{ lorsque } nh \rightarrow \infty$$

La variance est d'autant plus grande que le paramètre de lissage est faible. On retrouve l'arbitrage entre biais et variance...

3.2.3 Erreur quadratique moyenne – MSE

$$MSE \left\{ \widehat{f}_h(x) \right\} = \frac{h^4}{4} f''(x)^2 \mu_2^2(K) + \frac{1}{nh} \|K\|_2^2 f(x) + o(h^4) + o\left(\frac{1}{nh}\right)$$

L'erreur quadratique moyenne tend vers zéro lorsque h tend vers zéro et nh tend vers l'infini \rightarrow convergence en probabilité de $\widehat{f}_h(x)$ vers la densité $f(x)$. Il s'agit d'une mesure locale...

3.2.4 MISE

On obtient une mesure globale de la distance entre l'estimateur à noyau et la densité en intégrant la MSE:

$$MISE \left\{ \hat{f}_h \right\} = \frac{1}{nh} \|K\|_2^2 + \frac{h^4}{4} \mu_2^2(K) \|f''\|_2^2 + o\left(\frac{1}{nh}\right) + o(h^4)$$

En ignorant les termes d'ordre supérieurs, on obtient l'approximation suivante :

$$AMISE \left\{ \hat{f}_h \right\} = \frac{1}{nh} \|K\|_2^2 + \frac{h^4}{4} \mu_2^2(K) \|f''\|_2^2$$

La valeur de h qui minimise l'AMISE est alors :

$$h^* = \left[\frac{\|K\|_2^2}{\mu_2^2(K) \|f''\|_2^2 n} \right]^{\frac{1}{5}}$$

L'expression du paramètre de lissage optimal h^* dépend de la dérivée seconde de la densité (inconnue) mais elle nous donne la vitesse de convergence optimale de h vers zéro relativement à la divergence de la taille de l'échantillon : $h^* \propto n^{-\frac{1}{5}}$.

En substituant h^* dans l'expression de l'AMISE on montre que pour l'estimateur à noyau :

$$n^{\frac{4}{5}} AMISE \left\{ \hat{f}_{h^*} \right\} = O(1)$$

Alors que pour l'histogramme on a :

$$n^{\frac{2}{3}} AMISE \left\{ \hat{f}_{h^*} \right\} = O(1)$$

On voit ainsi que la vitesse de convergence de l'estimateur à noyau est supérieure à celle de l'histogramme.

3.3 Choix du paramètre de lissage

3.3.1 La règle du pouce (Silverman)

On veut choisir le paramètre de lissage qui minimise l'AMISE, mais on ne connaît pas la norme de la dérivée seconde de la densité que nous cherchons à estimer. En première approximation, pour obtenir h^* , on va supposer que f est la densité d'une v.a. gaussienne d'espérance μ et de variance σ^2 . On a alors :

$$\|f''\|_2^2 = \frac{3}{8\sigma^5\sqrt{\pi}} \approx 0,212\sigma^{-5}$$

et donc, si on utilise un noyau gaussien,

$$\hat{h}^* \approx 1,06\hat{\sigma}n^{-\frac{1}{5}}$$

Si la fenêtre n'est pas gaussienne on doit changer le chiffre devant l'écart type empirique...

- Ce choix du paramètre de lissage peut se révéler désastreux si la vraie distribution est éloignée du cas gaussien.
- On peut facilement proposer un critère modifié pour tenir compte de la présence éventuelle de points aberrants. Plutôt que d'utiliser l'écart type pour évaluer la dispersion de la distribution on utilise l'écart interquartile qui a l'avantage de ne pas être sensible aux points aberrants.

$$\hat{h}^* \approx 1,06 \min \left\{ \hat{\sigma}, \frac{\hat{R}}{1,34} \right\} n^{-\frac{1}{5}}$$

où \hat{R} est l'écart interquartile empirique.

- Cette correction est insuffisante dans de nombreux cas... Par exemple si la vraie densité est multi-modale.

3.3.2 Des “Plug-in methods” améliorées

Plutôt que de calculer la norme de la dérivée seconde en paramétrisant la vraie distribution, on va directement l'estimer.

- Park et Marron (JASA, 1990)
- Sheather et Jones (JRSS, 1991)

Pour estimer la dérivée s -ième de la densité f on utilise l'estimateur à noyau suivant :

$$\widehat{f}_h^{(s)}(x) = \frac{(-1)^s}{nh^{s+1}} \sum_{i=1}^n K^{(s)}\left(\frac{x - X_i}{h}\right)$$

En particulier pour $s = 2$, on a :

$$\widehat{f}_g^{(2)}(x) = \frac{1}{ng^3} \sum_{i=1}^n K^{(2)}\left(\frac{x - X_i}{g}\right)$$

Problème : Comment doit-on choisir le paramètre de lissage pour estimer la dérivée seconde de f ? On va une nouvelle fois utiliser la règle du pouce en supposant que la densité est gaussienne. Notons que nous formulons cette hypothèse plus en amont (il s'agit de déterminer le paramètre de lissage de l'estimateur de f''). Des simulations permettent de montrer que cette hypothèse devient alors, en pratique, moins gênante.

PJ90 montrent que l'on doit avoir :

$$g = \left[\frac{18 \|K^{(4)}\|_2 \mu_2^4(K)}{\mu_2^2(K \circ K) \|K\|_2^2} \frac{\|f\|_2 \|f''\|_2^2}{\|f'''\|_2^2} \right]^{\frac{1}{13}} h^{\frac{10}{13}}$$

si on choisit g de façon à minimiser l'AMISE (de \hat{f}_g''), où h est le paramètre de lissage de l'estimateur de f . On détermine les fonctionnelles de f en faisant une hypothèse paramétrique.

Finalement, PM1990 obtiennent le paramètre de lissage associé à l'estimateur à noyau de la densité comme un point fixe :

$$\hat{h}_{PM}^* = \left(\frac{\|K\|_2^2}{\|f''_{g(\hat{h}_{PM}^*)}\|_2^2 \mu_2^2(K)n} \right)^{\frac{1}{5}}$$

Sheather et Jones 1991...

3.3.3 Cross-Validation

On a obtenu un paramètre de lissage optimal en minimisant une approximation asymptotique de

$$\int \mathbb{E} \left[(\hat{f}_h(x) - f(x))^2 \right] dx \equiv AMISE$$

L'AMISE est une mesure globale de la distance entre l'estimateur à noyau et la vraie densité. Dans cette section on considère une mesure alternative de cette distance, l'erreur quadratique intégrée,

$$ISE \left\{ \hat{f}_h \right\} = \int (\hat{f}_h(x) - f(x))^2 dx$$

Contrairement à la MISE, cette mesure est une variable aléatoire. En développant, il vient :

$$ISE \left\{ \hat{f}_h \right\} = \int \hat{f}_h^2(x) dx - 2 \int \hat{f}_h(x) f(x) dx + \int f^2(x) dx$$

Par la suite on omet le dernier terme puisqu'il ne dépend pas de h .

Le terme croisé s'interprète comme une espérance :

$$\int \hat{f}_h(x) f(x) dx \equiv \mathbb{E} \left[\hat{f}_h(X) \right]$$

Un estimateur sans biais (*leave one out estimator*) de l'espérance de $\widehat{f}_h(X)$ est donné par :

$$\mathbb{E} \left[\widehat{f}_h(X) \right] = \frac{1}{n} \sum_{i=1}^n \widehat{f}_{h,-i}(X_i)$$

avec

$$\widehat{f}_{h,-i}(x) = \frac{1}{n-1} \sum_{j=1, j \neq i}^n K \left(\frac{x - X_j}{h} \right)$$

Par ailleurs, on montre que :

$$\int \widehat{f}_h(x)^2 dx = \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1}^n K \star K \left(\frac{X_i - X_j}{h} \right)$$

où $K \star K$ est la convolution du noyau par lui même. Finalement, on

obtient un paramètre h en minimisant le critère suivant :

$$CV(h) = \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1}^n K \star K \left(\frac{X_i - X_j}{h} \right) - 2 * \frac{1}{n} \sum_{i=1}^n \hat{f}_{h,-i}(X_i)$$

Remarque : Si $K(\bullet)$ est le noyau gaussien, alors on a :

$$K \star K(u) = \int_{-\infty}^{\infty} K(x)K(u - x)dx = \frac{1}{2\sqrt{\pi}} e^{-\frac{1}{4}u^2}$$

la densité d'une loi normale d'espérance nulle et de variance 2.

3.4 Retour sur les propriétés statistiques

Proposition Sous les hypothèses (i) (x_1, x_2, \dots, x_n) un échantillon *iid*, (ii) le noyau est une fonction symétrique autour de zéro vérifiant $\int K(u)du = 0$, $\int u^2 K(u)du = 0 = \mu_2 \neq 0$ et $\int K(u)^2 du < \infty$, (iii) les dérivées secondes de f sont continues et bornées dans un voisinage de x , (iv) h tend vers zéro lorsque n tend vers l'infini et nh tend vers l'infini lorsque n tend vers l'infini, (v) $\exists \delta > 0$ tel que $\int K(u)^{2+\delta} < \infty$, alors on a le résultat asymptotique suivant :

$$(nh)^{\frac{1}{2}} \left\{ \hat{f}_h(x) - \mathbb{E} \left[\hat{f}_h(x) \right] \right\} \xrightarrow[n \rightarrow \infty]{} \mathcal{N} \left(0, f(x) \|K\|_2^2 \right) \quad (5)$$

→ En pratique on s'intéresse plutôt à la distribution asymptotique de $(nh)^{\frac{1}{2}} \left\{ \hat{f}_h(x) - f(x) \right\}$ pour construire un intervalle de confiance...

$$(nh)^{\frac{1}{2}} \left\{ \widehat{f}_h(x) - f(x) \right\} = (nh)^{\frac{1}{2}} \left\{ \widehat{f}_h(x) - \mathbb{E} \left[\widehat{f}_h(x) \right] \right\} \\ + (nh)^{\frac{1}{2}} \left\{ \mathbb{E} \left[\widehat{f}_h(x) \right] - f(x) \right\}$$

Le membre de gauche sera asymptotiquement centré en zéro si et seulement si le deuxième terme (le biais de l'estimateur à noyau) du membre de droite tend vers zéro lorsque n tend vers l'infini. Sachant que le biais est un $O(h^2)$, une condition suffisante est:

$$(nh)^{\frac{1}{2}} h^2 \xrightarrow[n \rightarrow \infty]{} 0$$

Si cette hypothèse est vérifiée, alors :

$$(nh)^{\frac{1}{2}} \left\{ \widehat{f}_h(x) - f(x) \right\} \xrightarrow[n \rightarrow \infty]{} \mathcal{N} \left(0, f(x) \|K\|_2^2 \right)$$

D'où, sous cette hypothèse, l'intervalle de confiance à 95% :

$$\widehat{f}_h(x) \pm 1,96 (nh)^{-\frac{1}{2}} \left[f(x) \|K\|_2^2 \right]^{\frac{1}{2}}$$

Problème : L'hypothèse sur h , qui nous permet de construire cet intervalle de confiance, est incompatible avec un paramètre de lissage obtenu en minimisant l'AMISE.

Posons $h = c \times n^{-\frac{1}{5} + \alpha}$. Pour $\alpha = 0$ on retrouve le paramètre de lissage optimal. Pour $\alpha > 0$ (resp. $\alpha < 0$) on obtient un cas de sur-lissage (resp. sous-lissage). En substituant cette expression dans la condition suffisante qui nous permet de construire l'intervalle de confiance, on voit cette condition est vérifiée seulement dans le cas où $\alpha < 0$. \rightarrow Le biais disparaît asymptotiquement dans le cas du sous-lissage.

Si h est optimal, au sens où il minimise l'AMISE, la condition suffisante n'est pas vérifiée. Utiliser l'intervalle de confiance défini plus haut revient à supposer que le biais est négligeable (*ie* la dérivée seconde de la densité est petite).

3.5 Un estimateur alternatif de la densité

Soit X une variable aléatoire et $f(\bullet)$ sa densité sur l'intervalle $[0, 1]$ (on pourrait considérer un support plus large).

On suppose que la fonction de densité admet une représentation de Fourier (décomposition comme la somme de fonctions sinus et cosinus), c'est-à-dire que l'on peut écrire :

$$f(x) = \sum_{i=1}^{\infty} a_j \phi_j(x)$$

où $a_j = \int_0^1 f(x) \phi_j(x) dx \equiv \mathbb{E}[\phi_j(X)]$ et

$$\phi_j(x) = \begin{cases} \sqrt{2} \cos(\pi(j+1)x), & \text{si } j \text{ pair,} \\ \sqrt{2} \sin(\pi j x), & \text{sinon.} \end{cases}$$

Un estimateur convergent du coefficient a_j est donné par :

$$\hat{a}_j = n^{-1} \sum_{i=1}^n \phi_j(x_i)$$

Un estimateur de la densité est alors :

$$\hat{f}_m(x) = \sum_{j=1}^m \hat{a}_j \phi_j(x)$$

où le paramètre m de troncation joue un rôle analogue au paramètre de lissage dans l'estimateur à noyau.

On pourrait considérer d'autres séries orthogonales que les séries de Fourier (par exemple les polynômes hermitien)

3.6 Le cas multivarié

Soit un vecteur aléatoire de dimension d : $\mathbb{X} = (X_1, X_2, \dots, X_d)'$, où X_i est une variable aléatoire univariée. On dispose d'un échantillon *iid* de dimension n : $(\mathbb{X}_1, \mathbb{X}_2, \dots, \mathbb{X}_n)$.

Cet échantillon est *iid* mais $X_{1,i}$ n'est pas nécessairement indépendant de $X_{2,i}$ (pour $i = 1, \dots, n$).

On note $f(x_1, \dots, x_d)$ la densité jointe de cette variable aléatoire multivariée.

On généralise l'estimateur à noyau de la façon suivante :

$$f_H(\mathbf{x}) = \frac{1}{n|H|} \sum_{i=1}^n \mathcal{K}(H^{-1}(\mathbb{X}_i - \mathbf{x}))$$

avec $\mathbf{x} = (x_1, \dots, x_d)'$, où H est la matrice de bandwidth et $\mathcal{K}(\mathbf{u})$ est une fenêtre multivariée.

La fenêtre multivariée vérifie les hypothèses usuelles : $\int \mathcal{K}(\mathbf{u})d\mathbf{u} = 1$,
 $\int \mathbf{u}\mathcal{K}(\mathbf{u})d\mathbf{u} = \vec{0}$ (symétrie), $\int \mathbf{u}\mathbf{u}'\mathcal{K}(\mathbf{u})d\mathbf{u} = \mu_2(\mathcal{K})\mathbf{I}_d > 0$ et
 $\int \mathcal{K}(\mathbf{u})^2d\mathbf{u} = \|\mathcal{K}\|_2^2 < \infty$.

Alternatives pour la fenêtre multivariée :

- Utiliser une densité multivariée, comme, par exemple une densité normale de dimension d : $(2\pi)^{-\frac{d}{2}}e^{-\frac{1}{2}\mathbf{u}'\mathbf{u}}$. De la même façon on peut généraliser la fenêtre d'Epanechnikov :

$$\mathcal{K}(\mathbf{u}) = \frac{1}{2}c_d^{-1}(d+2)(1-\mathbf{u}'\mathbf{u})\mathbb{I}_{\{\mathbf{u}'\mathbf{u}\leq 1\}}(\mathbf{u})$$

avec $c_1 = 2$, $c_2 = \pi$, $c_3 = 4\pi/3$. Où c_d est le volume de la sphère unité de dimension d .

- Utiliser un produit de fenêtres univariées :

$$\mathcal{K}(\mathbf{u}) = \prod_{j=1}^d K_j(u_j)$$

Par exemple dans le cas d'une fenêtre d'Epanechnikov, en supposant que **la matrice de lissage est diagonale**, l'estimateur à noyau aurait la forme suivante :

$$\hat{f}_{h_1, \dots, h_d} = \frac{1}{nh_1 \dots h_d} \left(\frac{3}{4}\right)^d \times \sum_{i=1}^n \prod_{j=1}^d \left\{ 1 - \left(\frac{x_j - \mathbb{X}_{j,i}}{h_j}\right)^2 \right\} \mathbb{I}_{\left|\frac{x_j - \mathbb{X}_{j,i}}{h_j}\right| \leq 1}$$

Remarque : Dans le cas de la fenêtre gaussienne on a équivalence entre les deux spécifications.

La matrice de lissage H est définie positive.

Rappel : L'approximation à l'ordre deux d'une fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$ dans \mathcal{C}_2 est donnée par :

$$f(\mathbf{x} + \mathbf{s}) = f(\mathbf{x}) + \mathbf{s}' \nabla_f(\mathbf{x}) + \frac{1}{2} \mathbf{s}' \mathcal{H}_f(\mathbf{x}) \mathbf{s} + o(\|\mathbf{s}\|^2)$$

où $\nabla_f(\mathbf{x})$ est le gradient de f évalué en \mathbf{x} et $\mathcal{H}_f(\mathbf{x})$ est le hessien de f évalué en \mathbf{x} .

3.6.1 Propriétés statistiques

Le biais :

$$\begin{aligned}\mathbb{E} \left[\widehat{f}_H(\mathbf{x}) \right] &= \mathbb{E}_{\mathbb{X}} [\mathcal{K}_H(\mathbf{x} - \mathbb{X})] \\ &= \int \mathcal{K}_H(\mathbf{u} - \mathbf{x}) f(\mathbf{u}) d\mathbf{u} \\ &= \int \mathcal{K}(\mathbf{s}) f(\mathbf{x} + H\mathbf{s}) d\mathbf{s} \\ &\approx \int \mathcal{K}(\mathbf{s}) \left[f(\mathbf{x}) + \mathbf{s}' \nabla f(\mathbf{x}) + \frac{1}{2} \mathbf{s}' \mathcal{H}_f(\mathbf{x}) \mathbf{s} \right] d\mathbf{s} \\ &= f(\mathbf{x}) + \frac{1}{2} \mu_2(\mathcal{K}) \text{trace} \{ H' \mathcal{H}_f(\mathbf{x}) H \}\end{aligned}$$

D'où le biais au asymptotique intégré :

$$AIBiais\{\widehat{f}_H\} = \frac{1}{2} \mu_2(\mathcal{K}) \int \text{trace} \{ H' \mathcal{H}_f(\mathbf{x}) H \} d\mathbf{x}$$

La variance : On procède de la même façon avec une approximation à l'ordre un et on obtient la variance asymptotique intégrée :

$$AIV\{\hat{f}_H\} = \frac{\|\mathcal{K}\|_2^2}{n|H|}$$

Finalement, l'AMISE est donnée par :

$$AMISE\{\hat{f}_H\} = \frac{1}{4}\mu_2^2(\mathcal{K}) \int \text{trace}\{H'\mathcal{H}_f(\mathbf{x})H\}^2 d\mathbf{x} + \frac{\|\mathcal{K}\|_2^2}{n|H|}$$

Afin d'évaluer la vitesse de convergence de l'estimateur à noyau multivarié, on va supposer que la matrice de lissage est $H = h\mathbf{I}_d$.

L'AMISE s'écrit alors de la façon suivante :

$$AMISE\{\hat{f}_h\} = \frac{h^4}{4}\mu_2^2(\mathcal{K}) \int \text{trace}\{\mathcal{H}_f(\mathbf{x})\}^2 d\mathbf{x} + \frac{\|\mathcal{K}\|_2^2}{nh^d}$$

La matrice qui minimise l'AMISE vérifie alors:

$$h^* = O\left(n^{-\frac{1}{4+d}}\right)$$

et on a :

$$AMISE\{\hat{f}_{h^*}\} = O\left(n^{-\frac{4}{4+d}}\right)$$

La vitesse de convergence de l'estimateur à noyau dépend de la dimension de la variable aléatoire. Elle est d'autant plus faible que la dimension est grande (*curse of dimensionality*).

3.6.2 Choix du paramètre de lissage

Rule of Thumb : On suppose que le processus générateur des données est gaussien, d'espérance μ et de matrice de variance-covariance Σ . Si la fenêtre est gaussienne, alors on a $\mu_2(\mathcal{K}) = 1$ et $\|\mathcal{K}\|_2^2 = 2^{-d}\pi^{-\frac{d}{2}}$. Si la matrice Σ est diagonale, on peut

alors montrer que la matrice de lissage qui minimise l'AMISE est diagonale et vérifie :

$$h_j^* = \left(\frac{4}{d+2} \right)^{\frac{1}{d+4}} n^{-\frac{1}{d+4}} \sigma_j$$

On ne peut obtenir la matrice de lissage optimale dans le cas plus général où la matrice Σ est non diagonale. Mais le résultat précédent suggère d'utiliser :

$$H^* = n^{-\frac{1}{d+4}} \widehat{\Sigma}^{\frac{1}{2}}$$

→ Cela revient à appliquer une transformation de Mahalanobis aux observations.

Cross-Validation : On généralise le critère obtenu dans le cadre univarié de la façon suivante :

$$cv(H) = \frac{1}{n^2 |H|} \sum_{i=1}^n \sum_{j=1}^n \mathcal{K} \star \mathcal{K} (H^{-1}(\mathbb{X}_j - \mathbb{X}_i)) \\ - \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n K_H (\mathbb{X}_j - \mathbb{X}_i)$$

Il reste alors à minimiser $cv(H)$ para rapport à $d(d+1)/2$ paramètres...

BIBLIOGRAPHIE

Trois ouvrages utilisés pour préparer ce chapitre :

- Adrian PAGAN et Aman ULLAH (1999), “Nonparametric Econometrics”, Cambridge University Press.
- Wolfgang HÄRDLE, Marlene MÜLLER, Stefan SPERLICH et Axel WERWATZ (1999), “A course on Non- and Semiparametric Modelling”, Humboldt-Universität zu Berlin.
- B.W. SILVERMAN (1986), “Density estimation for Statistics and Data Analysis”, Chapman and Hall.