Integral approximation by kernel smoothing

François Portier

Université catholique de Louvain - ISBA

August, 29 2014

In collaboration with Bernard Delyon

Topic of the talk: Given $\varphi : \mathbb{R}^d \to \mathbb{R}$, estimation of

$$I(\varphi) = \int \varphi(x)dx$$

Monte-Carlo

$(X_1, \ldots, X_n)$ i.i.d. with law f

$$n^{1/2}\left(n^{-1}\sum_{i=1}^{n}\frac{\varphi(X_i)}{f(X_i)} - I(\varphi)\right) = O_{\mathbb{P}}(1)$$

Importance sampling

Optimal sampler f^*, $(X_1, \ldots, X_n)$ i.i.d. with law f^*

$$n^{1/2}\left(n^{-1}\sum_{i=1}^{n}\frac{\varphi(X_i)}{f^*(X_i)} - I(\varphi)\right) = o_{\mathbb{P}}(1)$$

Adaptive to φ

Evans and Schwartz (2000, book), Zhang (1996, JASA)

Main result

$(X_1, \ldots, X_n)$ i.i.d. with law f, $\widehat{f}$ is a kernel estimator of f

$$n^{1/2}\left(n^{-1}\sum_{i=1}^{n}\frac{\varphi(X_i)}{\widehat{f}(X_i)} - I(\varphi)\right) = o_{\mathbb{P}}(1)$$

Adaptive to the design

φ is only known at the points X_i

Purpose

- Rates of convergence, asymptotic behaviour
- Regularity of f and φ with respect to the dimension, the bandwidth
- In practice: kernel, bandwidth...
- Application to regression modelling

Definition of the estimators

K a d-dimensional kernel

$$\widehat{f}^{(i)}(x) = (nh^d)^{-1} \sum_{j \neq i}^{n} K(h^{-1}(x - X_j))$$

$$\widehat{v}^{(i)}(x) = ((n-1)(n-2))^{-1} \sum_{j \neq i}^{n} (h^{-d} K(h^{-1}(x - X_j)) - \widehat{f}^{(i)}(x))^2$$

2 estimators of $I(\varphi)$

$$\widehat{I}(\varphi) = n^{-1} \sum_{i=1}^{n} \frac{\varphi(X_i)}{\widehat{f}^{(i)}(X_i)}$$

$$\widehat{I}_c(\varphi) = n^{-1} \sum_{i=1}^{n} \frac{\varphi(X_i)}{\widehat{f}^{(i)}(X_i)} \left(1 - \frac{\widehat{v}^{(i)}(X_i)}{\widehat{f}^{(i)}(X_i)^2}\right)$$

Assumptions

Nikol'ski class $\mathcal{H}_s$, $s = k + \alpha$, $k \in \mathbb{N}$, $0 < \alpha \leq 1$

$$\int (\varphi^{(l)}(x+u) - \varphi^{(l)}(x))^2 dx \leq C|u|^{2\alpha} \qquad l = (l_1, \ldots l_d), \ \sum l_i \leq k$$

$$(\Rightarrow \psi \text{ is } \alpha\text{-Hölder inside } Q \quad \Rightarrow \quad s = \min(1/2, \alpha))$$

Tsybakov (2009, book)

(A1) $\varphi \in \mathcal{H}_s$ on $\mathbb{R}^d$ and has compact support Q

(A2) The r-th order derivatives of f are bounded

(A3) For every $x \in Q$, $\quad f(x) \geq b > 0$

(A4) K symmetric with order r and $K(x) \leq C_1 \exp(-C_2\|x\|)$

Theorem

Assume (A1-A4), we have

$$n^{1/2}\left(\widehat{I}(\varphi) - I(\varphi)\right) = O_{\mathbb{P}}\left(h^s + n^{1/2}h^r + n^{-1/2}h^{-d}\right) \tag{1}$$

if the $O_{\mathbb{P}} \overset{n\to+\infty}{\to} 0$

Remarks

- Curse of dimensionality: $r > d$
- For r, s large, $\quad h_{opt} \propto n^{-\frac{1}{r+d}}, \quad$ the rate $= n^{-\frac{r-d}{2(r+d)}}$
- f is undersmooth because $h_{opt} < n^{-\frac{1}{2r+d}}$ Stone (1980, AoS)
- Regularity of φ is not crucial
- Trimming method ? Härdle and Stocker (1989, JASA)

Theorem

Assume (A1-A4), we have

$$n^{1/2}\left(\widehat{I}_c(\varphi) - I(\varphi)\right) = O_{\mathbb{P}}\left(h^s + n^{1/2}h^r + n^{-1/2}h^{-d/2} + n^{-1}h^{-3d/2}\right)$$

$$\textit{instead of } O_{\mathbb{P}}\left(h^s + n^{1/2}h^r + n^{-1/2}h^{-d}\right)$$

if the $O_{\mathbb{P}} \overset{n\to+\infty}{\to} 0$

Remarks

- Curse of dimensionality : $r > 3d/4$
- For r, s large, $h_{opt} \propto n^{-\frac{1}{r+d/2}}$, the optimal rate $= n^{-\frac{r-d/2}{2(r+d/2)}}$
- Leave-one out better than the classical

Rates of convergence

Asymptotic behaviour

Simulations

Conclusion (Application to regression modelling)

$$\widehat{I}(\varphi) - I(\varphi) = \widetilde{B}_n + M_n + \text{neglectable}$$
$$\widehat{I}_c(\varphi) - I(\varphi) = B_n + M_n + U_n + \text{neglectable}$$

with B_n and $\widetilde{B}_n$ non-random, M_n martingale, U_n U-stat

- If φ is very smooth: $M_n = o_{\mathbb{P}}(U_n)$
- If φ is not regular: $U_n = o_{\mathbb{P}}(M_n)$

Hall (1984, JMVA), Hall and Heyde (1980, book)

Regular case

Theorem

Under (A1) to (A4), if $nh^{2d} \to +\infty$, $nh^{r+d/2} \to 0$ *and* $nh^{2s+d} \to 0$,

$$nh^{d/2}(\widehat{I_c}(\varphi) - I(\varphi))$$

is asymptotically normally distributed with zero-mean and variance given by

$$\int \left(\int (K(u+v) - K(v))K(u)du \right)^2 dv \int \varphi(x)^2 f(x)^{-2} dx$$

A non smooth example

(B1) For some $s > 1/2$ the function φ belongs to $\mathcal{H}_s$ on Q and is bounded, with compact support Q.

(B2) The set Q is compact with C^2 boundary.

$$L_Q(x) = \iint \min(\langle z, u(x)\rangle, \langle z', u(x)\rangle)_+ K(z)K(z')dzdz'$$

$u(x)$ the normal outer vector of Q at the point x

Theorem

Under the assumptions (A2) to (A4), (B1) and (B2), if $nh^{(3d+1)/2} \to 0$ and $nh^{2r-1} \to 0$

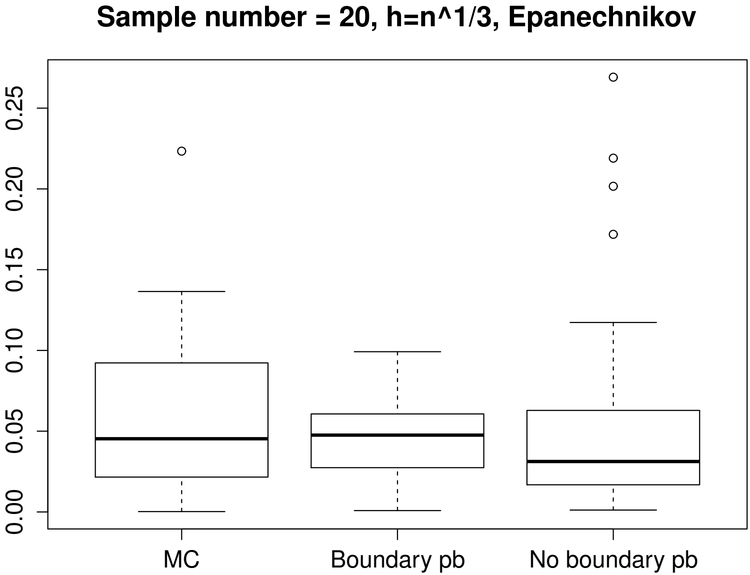
$$(nh^{-1})^{1/2}(\widehat{I}_c(\varphi) - I(\varphi))$$

is asymptotically normally distributed with zero-mean and variance given by

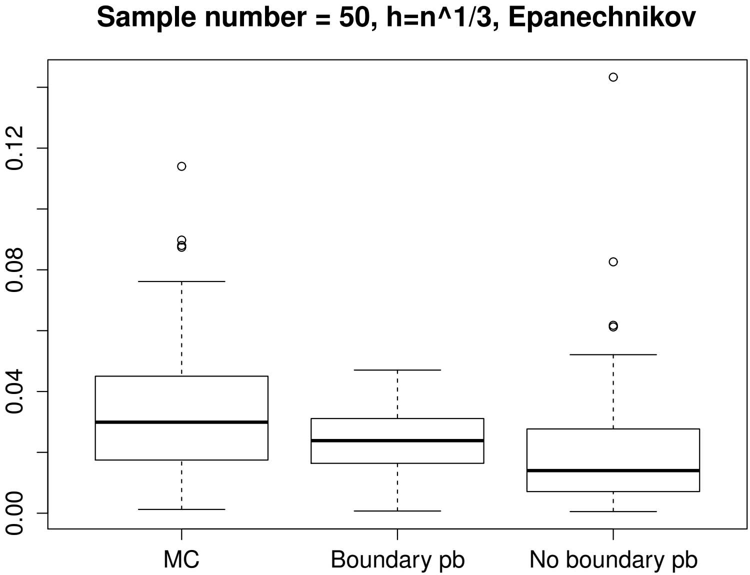
$$\int_{\partial Q} L_Q(x)\varphi(x)^2 d\mathcal{H}^{p-1}(x),$$

where $\mathcal{H}^{p-1}$ stands for the $p-1$ dimensional Hausdorff measure.

In practice

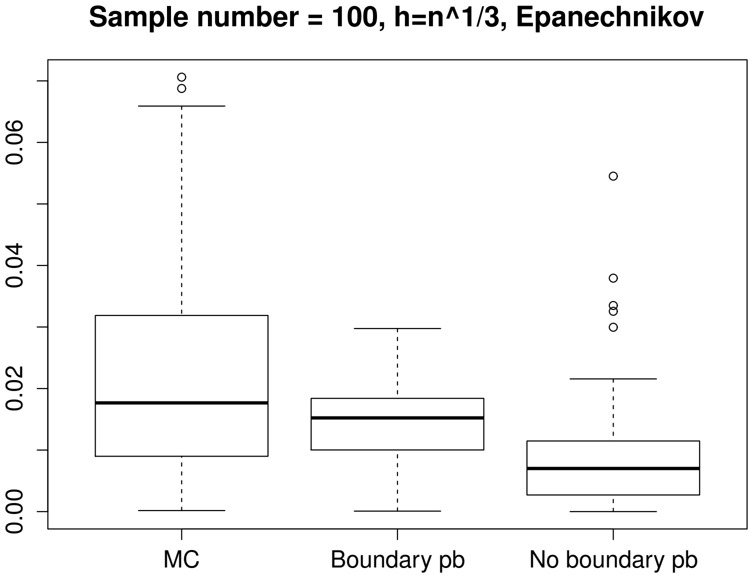

Sample number = 20, h=n^1/3, Epanechnikov
0.00
0.05
0.10
0.15
0.20
0.25
MC
Boundary pb
No boundary pb

In practice



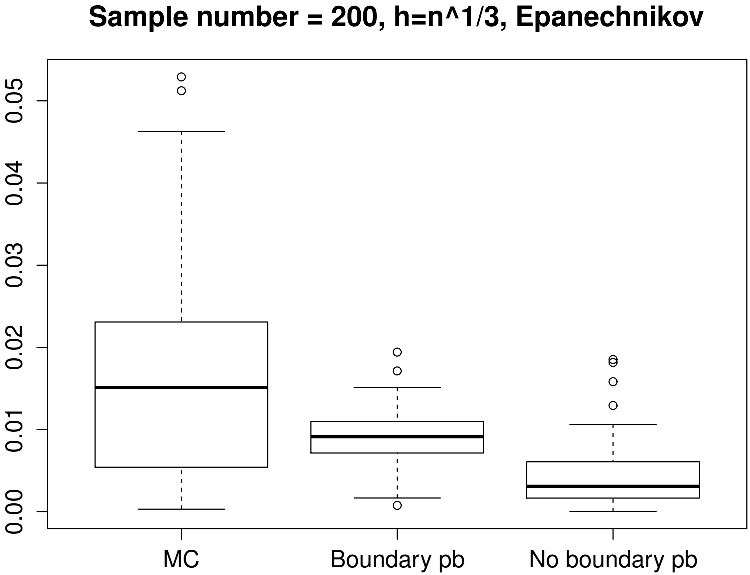
Sample number = 50, h=n^1/3, Epanechnikov
0.00
0.04
0.08
0.12
MC
Boundary pb
No boundary pb

In practice

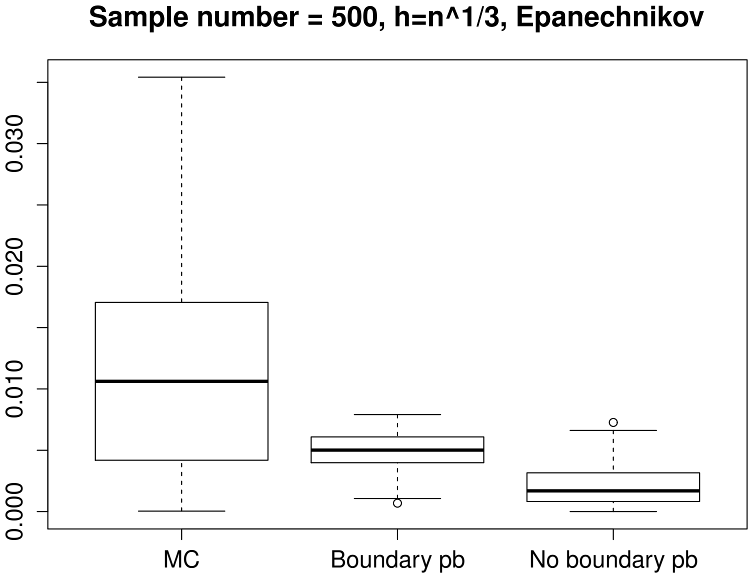


Sample number = 100, h=n^1/3, Epanechnikov
0.00
0.02
0.04
0.06
MC
Boundary pb
No boundary pb

In practice


Sample number = 200, h=n^1/3, Epanechnikov
0.00
0.01
0.02
0.03
0.04
0.05
MC
Boundary pb
No boundary pb

In practice


Sample number = 500, h=n^1/3, Epanechnikov
0.000
0.010
0.020
0.030
MC
Boundary pb
No boundary pb

Bandwidth choice

- Plug-in, e.g. Härdle, Marron and Tsybakov (1992, JASA)
- Simulation-validation

$$\widetilde{\varphi}(x) = n^{-1} \sum_{i=1}^{n} \frac{\varphi(X_i)}{\widehat{f}(X_i)} h_0^{-d} \widetilde{K}\left(\frac{x - X_i}{h_0}\right),$$

- $\widetilde{\varphi}$ looks like φ (convolution estimator)
- $I(\widetilde{\varphi})$ is known

$$\widehat{h} = \operatorname{argmin}_h |\widehat{I}_c(\widetilde{\varphi}) - I(\widetilde{\varphi})|$$

Kernel

$$K(x) \propto (d+2-(d+3)|x|)1_{|x|<1}$$

Design

$$\begin{array}{ll} \text{Model 1} & X_i \sim \mathcal{N}(\frac{1}{2}, \frac{1}{4} Id) \\ \text{Model 2} & X_i \sim \mathcal{U}([0,1]^d) \end{array}$$

$$\varphi(x) = \prod_{k=1}^{d} 2\sin(x_k)^2 1_{0 \le x_k \le 1}.$$

Model 1

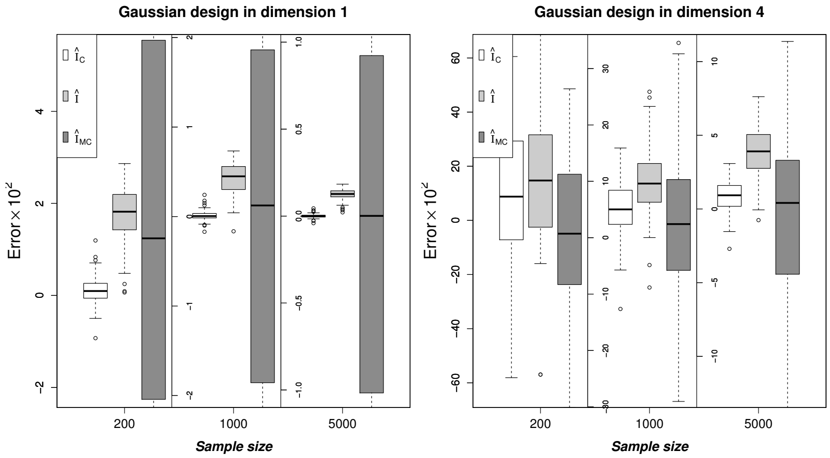


Figure : 100 estimates $\widehat{I_c}(\varphi)$, $\widehat{I}(\varphi)$ and Monte-Carlo method noted $\widehat{I_{MC}}$

Model 2

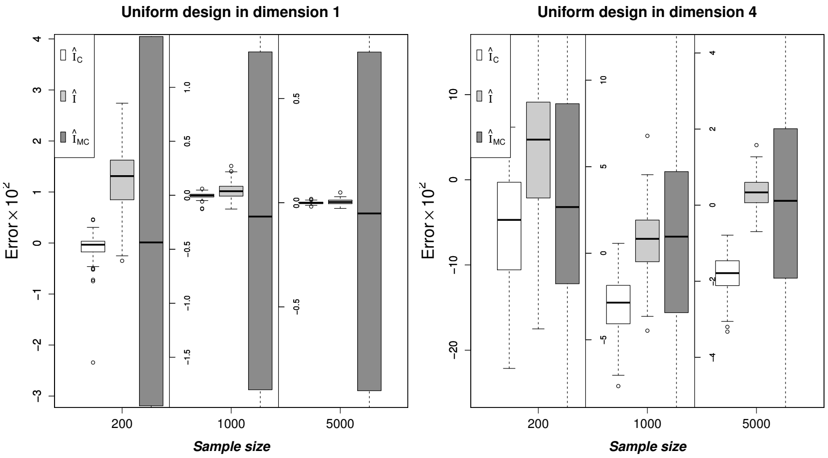


Figure : 100 estimates $\widehat{I_c}(\varphi)$, $\widehat{I}(\varphi)$ and Monte-Carlo method noted $\widehat{I_{MC}}$

Rates of convergence

Asymptotic behaviour

Simulations

Conclusion (Application to regression modelling)

Regression model

$$Y_i = g(X_i) + \sigma(X_i)e_i$$

- (X_i) **random** i.i.d. with density f
- $(X_i) \perp (e_i)$
- The functions g and σ are unknown

Let $Q \subset \mathbb{R}^d$ bounded and $L_2(Q) = \{\psi \ : \ \int_Q \psi(x)^2 dx < +\infty\}$

Purpose

Estimate $$c = <g, \psi> = \int_Q g(x)\psi(x)dx$$

(nonrandom design case treated by Donoho)

Plug-in estimates

Plug-in of g is difficult

Let $\widehat{g}$ such that

$$a_n(\widehat{g}(x) - g(x)) \xrightarrow{\text{d}} \text{Gaussian variable} \qquad \text{(e.g. NW, NN...)}$$

$a_n = o(\sqrt{n})$, but not tight, then

$$\sqrt{n}(< \widehat{g}, \psi > - < g, \psi >) = \sqrt{n} < \widehat{g} - g, \psi > \xrightarrow{\text{d}} \text{Gaussian variable}$$

is difficult to handle.

Plug-in of f may be better

$$c = < g, \psi > = \mathbb{E}\left[\frac{Y\psi(X)}{f(X)}\right] \qquad \widehat{c} = n^{-1}\sum_{i=1}^{n}\frac{Y_i\psi(X_i)}{\widehat{f}(X_i)}$$

Assumptions

(A2) The r-th order derivatives of f are bounded

(A3) For every $x \in Q, \quad f(x) \geq b > 0$

(A4) K symmetric with order r and $K(x) \leq C_1 \exp(-C_2 \|x\|)$

Assumptions

(A2) The r-th order derivatives of f are bounded

(A3) For every $x \in Q, \quad f(x) \geq b > 0$

(A4) K symmetric with order r and $K(x) \leq C_1 \exp(-C_2\|x\|)$

(A5) ψ is Hölder on its support $Q \subset \mathbb{R}^d$ nonempty bounded and convex

(A6) g is Hölder on Q and σ is bounded

(A7) $n^{1/2}h^r \overset{n\to+\infty}{\to} 0$ and $n^{1/2}h^d \overset{n\to+\infty}{\to} +\infty$

Theorem

Assume (A1-A7) we have

$$n^{1/2}(\widehat{c} - c) \xrightarrow{d} \mathcal{N}(0, v)$$

where v is the variance of the random variable $\frac{Y_1 - g(X_1)}{f(X_1)}\psi(X_1)$

Remarks

- Rates in root n
- The variance is smaller than when $\widehat{f} = f$ is known
- Trimming method ? (Härdle and Stoker (1989, JASA))