Integral approximation by kernel smoothing

François Portier

Télécom ParisTech

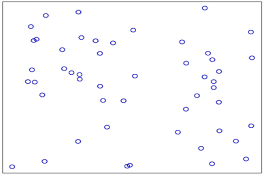
February, 2 2017

In collaboration with Romain Azais and Bernard Delyon

Stochastic integration problem

- $x_1, \ldots, x_n$ random points
- Observe $(x_1, \varphi(x_1)), \ldots, (x_n, \varphi(x_n))$
- Goal : Evaluate $\int \varphi(x)dx$

$x_1, \ldots, x_n$ in $[0,1]^2$ with uniform law

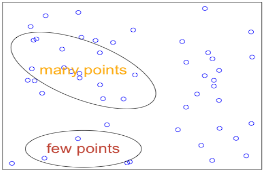


Monte-Carlo: $n^{-1} \sum_{i=1}^{n} \varphi(x_i)$

Stochastic integration problem

- $x_1, \ldots, x_n$ random points
- Observe $(x_1, \varphi(x_1)), \ldots, (x_n, \varphi(x_n))$
- Goal : Evaluate $\int \varphi(x)dx$

$x_1, \ldots, x_n$ in $[0,1]^2$ with uniform law

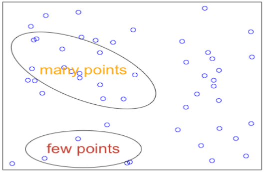


Monte-Carlo: $n^{-1} \sum_{i=1}^{n} \varphi(x_i)$

Stochastic integration problem

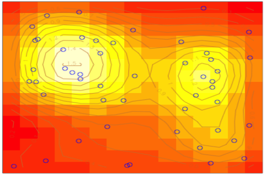
- $x_1, \dots, x_n$ random points
- Observe $(x_1, \varphi(x_1)), \dots, (x_n, \varphi(x_n))$
- Goal : Evaluate $\int \varphi(x)dx$

$x_1, \dots, x_n$ in $[0,1]^2$ with uniform law



Monte-Carlo: $n^{-1} \sum_{i=1}^n \varphi(x_i)$

$\widehat{\pi}(x) = n^{-1} \sum_{i=1}^n K(x - x_i)$



Our proposal: $n^{-1} \sum_{i=1}^n \frac{\varphi(x_i)}{\widehat{\pi}(x_i)}$

Adaptive to the design

Advantages

- one-sample based procedure
- the design distribution is not use
- fast rates
- robust to dependent design

Difficulties

- computational time
- choice of the bandwidth
- dimension curse

Applications

- Real data
- Stochastic algorithms

Advantages

- one-sample based procedure
- the design distribution is not use
- fast rates
- robust to dependent design

Difficulties

- computational time
- choice of the bandwidth
- dimension curse

Applications

- Real data ✓
- Stochastic algorithms

Advantages

- one-sample based procedure
- the design distribution is not use
- fast rates
- robust to dependent design

Difficulties

- computational time
- choice of the bandwidth
- dimension curse

Applications

- Real data ✓
- Stochastic algorithms ✓ (when φ or X_i expensive)

Independent design

- Asymptotic behaviour

Markovian design

Simulations

Applications

- Regression
- Stochastic integration
- Estimation of the temperature of oceans

Definition of the estimator

context

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with density π w.r.t. Lebesgue measure

$$\widehat{\pi}_i = ((n-1)h_n^d)^{-1} \sum_{j \neq i}^{n} K(h_n^{-1}(X_i - X_j))$$

$K : \mathbb{R}^d \to \mathbb{R}$ is called kernel, h_n is called bandwidth

estimator of $I(\varphi)$

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

Definitions

Definition 1

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

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

($\Rightarrow \psi$ is α-Hölder inside Q $\Rightarrow$ $s = \min(1/2, \alpha)$)
Tsybakov (2009, book)

Definition 2

A function K is said to be a kernel with order r if

$$\int K(x)dx = 1, \qquad \int x_1^{l_1} \dots x_d^{l_d} K(x) dx = 0, \qquad 0 < \sum_{k=1}^{d} l_k \leqslant r - 1,$$

Assumptions

(A1) $\varphi \in \mathcal{H}_s$ and has compact support Q

(A2) $\pi \in \mathcal{H}_r$ is continuous on Q

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

(A4) K has order (strictly) greater than r and s and

$$(i) \quad K(x) = K^{(0)}(|x|), \qquad \text{or} \qquad (ii) \quad K(x) = \prod_{k=1}^{d} K^{(0)}(x_k),$$

where $K^{(0)}$ a bounded real function of bounded variation with bounded support

(A5) $h_n \to 0$, $\frac{nh_n^d}{\log(n)} \to +\infty$

Theorem

Assume (A1-A5), we have

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

Remarks

- Fast rates if $nh_n^{2r} \to 0$ and $nh_n^{2d} \to +\infty$
- Curse of dimensionality: $r > d$
- For r, s large, $h_{opt} \propto n^{-\frac{1}{r+d}}$, the rate $= n^{-1/2} n^{-\frac{r-d}{2(r+d)}}$
- $\widehat{\pi}$ is undersmooth because $h_{opt} < n^{-\frac{1}{2r+d}}$ Stone (1980, AoS)
- Regularity of φ is not crucial

Theorem

Assume (A1-A5), we have

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

Remarks

- Fast rates if $nh_n^{2r} \to 0$ and $nh_n^{2d} \to +\infty$
- Curse of dimensionality: $r > d$
- For r, s large, $h_{opt} \propto n^{-\frac{1}{r+d}}$, the rate $= n^{-1/2}n^{-\frac{r-d}{2(r+d)}}$
- $\widehat{\pi}$ is undersmooth because $h_{opt} < n^{-\frac{1}{2r+d}}$ Stone (1980, AoS)
- Regularity of φ is not crucial

Sketch of the proof (kernel semiparametric litterature, e.g., C. Vial (2003, thesis))

Correction of the estimator

estimator of $I(\varphi)$

$$\widehat{v}^i(x) = ((n-1)(n-2))^{-1} \sum_{j \neq i}^{n} (h_n^{-d} K(h_n^{-1}(X_i - X_j)) - \widehat{\pi}_i)^2$$

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

Theorem

Assume (A1-A5), we have

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

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

Remarks

- Fast rates if $nh_n^{2r} \to 0$ and $nh_n^{3d/2} \to +\infty$
- Curse of dimensionality : $r > 3d/4$
- For r, s large, $h_{opt} \propto n^{-\frac{1}{r+d/2}}$, the optimal rate $= n^{-1/2}n^{-\frac{r-d/2}{2(r+d/2)}}$ ($\widehat{\pi}$ is undersmooth)
- Regularity of φ is not crucial
- Leave-one out better than the classical

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

with B_n non-random, M_n martingale, U_n U-stat

- If φ is very smooth $(s > d/2)$: $M_n << U_n$
- If φ is not regular $(s < d/2)$: $U_n << M_n$

Regular case

Theorem

Under (A1) to (A5), if $nh_n^{2d} \to +\infty$, $nh_n^{r+d/2} \to 0$ and $nh_n^{2s+d} \to 0$,

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

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

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

(based on Hall (1984, JMVA))

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

Theorem

Under the assumptions (A2) to (A5), (B1) and (B2), if $nh_n^{(3d+1)/2} \to +\infty$ *and* $nh_n^{2r-1} \to 0$

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

is asymptotically normally distributed with zero-mean

(based on Hall and Heyde (1980, book), Evans and Gariepy (1992))

Delyon, B., & Portier, F. (2014). Integral approximation by kernel smoothing. Bernoulli. arXiv:1409.0733.

Independent design

- Asymptotic behaviour

Markovian design

Simulations

Applications

- Regression
- Stochastic integration
- Estimation of the temperature of oceans

Markovian context

$(X_i)_{i \geqslant 0}$ a time homogeneous Markov chain with state space $\mathcal{E}$ and transition $P(x, dy)$

Renewal approach

there exists A, such that $\tau_A = \min\{i \geqslant 1 : X_i \in A\}$ satisfies

$$\forall x \in \mathcal{E}, P_x(\tau_A < \infty) = 1$$
$$\sup_{x \in A} E_x[\tau_A] < \infty$$

there exists a probability measure ψ, and some $\lambda > 0$, such that

$$\forall x \in A,\ \forall B \text{ measurable}, \quad P(x, B) \geqslant \lambda \psi(B)$$

Block decomposition

There exists a *split* chain $Z_i = (X_i, Y_i),\ i = 1, 2, \ldots$ for which $\tilde{A} = A \times \{1\}$ is such that

main feature

the blocks

$$B_k = (Z_{\tau_{\tilde{A}}(k)+1}, \ldots Z_{\tau_{\tilde{A}}(k+1)}),\ k = 1, 2, \ldots \qquad \text{form an i.i.d. sequence}$$

- first and last block treated separately
- the total number of blocks is random
- the length of the blocks $\tau_A(k+1) - \tau_A(k)$ is random

Nummelin (1978), Athreya and Ney (1978), Leventhal (1988), Bertail and Clemençon (2011)

Assumptions

In addition to the Doeblin's condition, there exists $p_0 > 3$ such that

(A1)

$$\sup_{x \in A} E_x[\tau_A^{p_0}] < +\infty,$$

(A2) $h_n \to 0, \qquad \frac{n h_n^{dp_0/(p_0-1)}}{\log(n)} \to +\infty$

(A2) $\varphi \in \mathcal{H}_s$ on $\mathbb{R}^d$ and has compact support Q

(A3) $\pi \in \mathcal{H}_r$ is continuous on Q

(A4) For every $x \in Q, \quad \pi(x) \geq b > 0$

(A5) K has order (strictly) greater than r and s and

$$(i) \quad K(x) = K^{(0)}(|x|), \qquad \text{or} \qquad (ii) \quad K(x) = \prod_{k=1}^{d} K^{(0)}(x_k),$$

where $K^{(0)}$ a bounded real function of bounded variation with bounded support

Theorem

Assume (A1-A5), we have

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

Remarks

- Fast rates if $n^{1/2}h_n^r \to 0$ and $nh_n^{2d} \to +\infty$
- Curse of dimensionality: $r > d$
- For r, s large, $h_{opt} \propto n^{-\frac{1}{r+d}}$, the rate $= n^{-1/2}n^{-\frac{r-d}{2(r+d)}}$
- f is undersmooth because $h_{opt} < n^{-\frac{1}{2r+d}}$ Stone (1980)
- Regularity of φ is not crucial

Theorem

Assume (A1-A5), we have

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

Remarks

- Fast rates if $n^{1/2}h_n^r \to 0$ and $nh_n^{2d} \to +\infty$
- Curse of dimensionality: $r > d$
- For r, s large, $\quad h_{opt} \propto n^{-\frac{1}{r+d}}, \quad$ the rate $= n^{-1/2}n^{-\frac{r-d}{2(r+d)}}$
- f is undersmooth because $h_{opt} < n^{-\frac{1}{2r+d}}$ Stone (1980)
- Regularity of φ is not crucial

Idea of the proof

$K_h(x) = h^{-d} K(x/h)$

Theorem

Under the previous conditions, we have

$$\sup_{y \in \mathbb{R}^d} |\widehat{\pi}(y) - (\pi \star K_h)(y)| \longrightarrow 0, \quad \text{in } \mathbb{P}_\pi\text{-probability.}$$

Based on Einmahl and Mason (2005): control of

$$\mathbb{E}\left[\sup_{f \in \mathcal{F}} \left| \sum_{i=1}^{n} f(\xi_i) - \mathbb{E} f(\xi_1) \right| \right],$$

for $(\xi_i)_{i \in \mathbb{N}^*}$ i.i.d. and $\mathcal{F}$ a VC class

Bandwidth choice

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

Let $(t_k)_{k=1,\dots,K}$ be a regular grid over Q

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

where h_0 is selected by cross-validation

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

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

Bandwidth choice

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

implemented in the R package `ks` (see Chacon and Duong (2010))

$$Q = [0,1]^d$$
$$K(x) \propto (1 - \|x\|^2) 1_{\|x\| \leq 1}$$

- $\mathcal{M}_1$: $\varphi(x_1, \ldots, x_d) = \prod_{i=1}^d \left[2 \sin(\pi x_i)^2 1_{[0,1]}(x_i) \right]$;
- $\mathcal{M}_2$: $\varphi(x_1, \ldots, x_d) = \prod_{i=1}^d \left[\frac{1+\pi^2}{\pi(1+\exp(1))} \sin(\pi x_i) \exp(x_i) 1_{[0,1]}(x_i) \right]$;
- $\mathcal{M}_3$: $\varphi(x_1, \ldots, x_d) = \prod_{i=1}^d \left[\frac{\pi}{2} \sin(\pi x_i)(1 + \cos(5\pi x_i)) 1_{[0,1]}(x_i) \right]$.

- independent design: uniform distribution over Q, $\mathcal{U}_Q$
- Markovian design : Metropolis-Hastings algorithm with proposition kernel

$$P(x, dy) = \mathcal{U}_{[x-\varepsilon, x+\varepsilon]^d}(dy),$$

with $\varepsilon = 0.2$ and target measure $\mathcal{U}_Q$

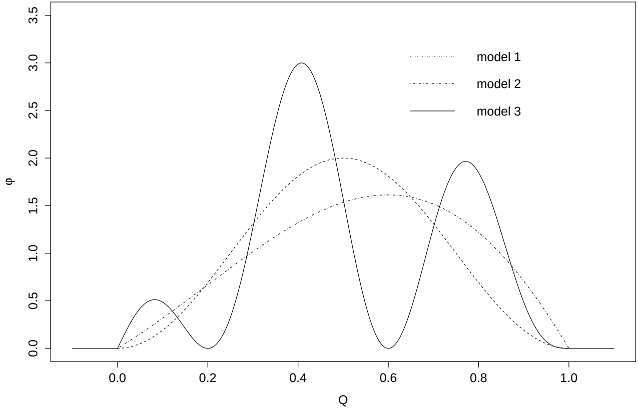


Figure: Shape of function φ for each model $\mathcal{M}_i$, $1 \leq i \leq 3$, in dimension $d = 1$.

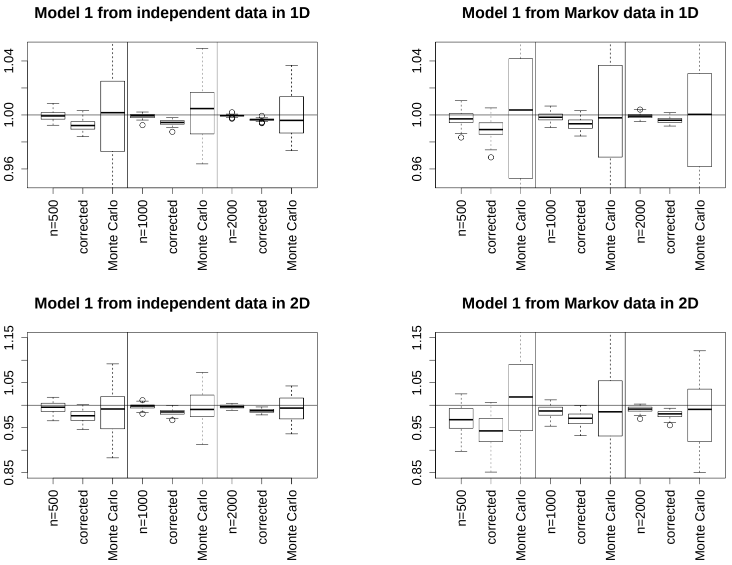


Figure: Boxplots of $\widehat{I_{ks}}$, $\widehat{I_{ks}^c}$ and $\widehat{I_{mc}}$ computed from 50 replicates for model $\mathcal{M}_1$ in dimension $d = 1$ (top), $d = 2$ (bottom) from independent data (left) and Markov data (right).

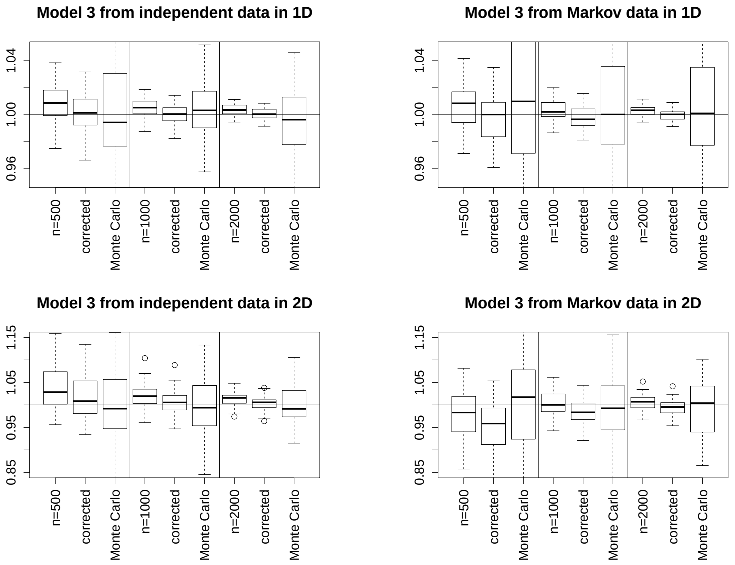


Figure: Boxplots of $\widehat{I_{ks}}$, $\widehat{I^c_{ks}}$ and $\widehat{I_{mc}}$ computed from 50 replicates for model $\mathcal{M}_3$ in dimension $d = 1$ (top), $d = 2$ (bottom) from independent data (left) and Markov data (right).

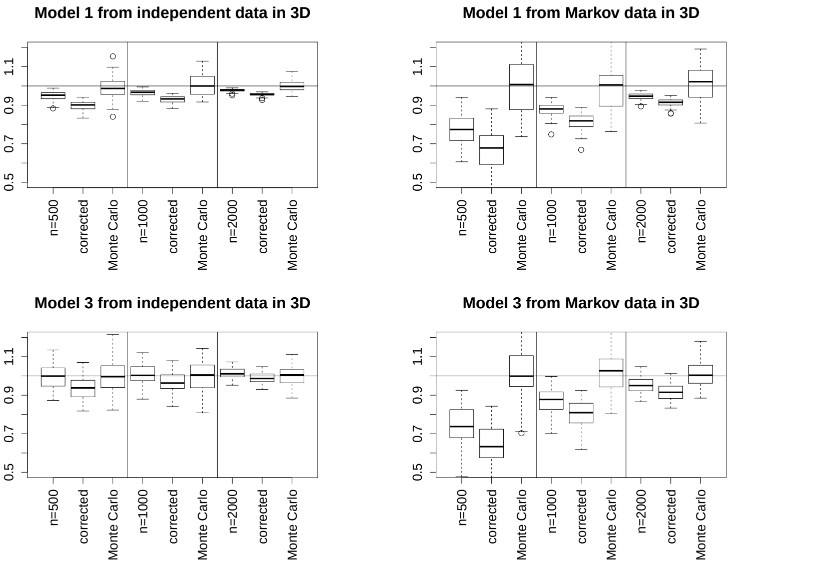


Figure: Boxplots of $\widehat{I_{\mathrm{ks}}}$, $\widehat{I^c_{\mathrm{ks}}}$ and $\widehat{I_{\mathrm{mc}}}$ computed from 50 replicates for model $\mathcal{M}_1$ (top) and $\mathcal{M}_2$ (bottom) in dimension $d = 3$ from independent data (left) and Markov data (right).

Independent design

- Asymptotic behaviour

Markovian design

Simulations

Applications

- Regression
- Stochastic integration
- Estimation of the temperature of oceans

Regression modelling

$$g \text{ **regression**}; \qquad c = \langle g, \psi \rangle = \int_Q g(x)\psi(x)dx$$

Curve estimation

- Orthogonal series or wavelet $\quad \sum_{k=1}^{K} \langle g, \psi_k \rangle \psi_k(y) \underset{approx.th.}{\overset{K \to +\infty}{\longrightarrow}} g(y)$
- Kernel smoothing $\quad \langle g, K_h(\cdot - y) \rangle \underset{regular.th.}{\overset{h \to 0}{\longrightarrow}} g(y)$

(Book of Härdle)

Semiparametric estimation

- Dimension reduction $\quad g(x) = g_0(\beta^T x)$

$$\langle g, \nabla\psi \rangle = -\langle \nabla g, \psi \rangle \propto \beta$$

  (Vial, Juditsky, Spokoiny, Härdle...)
- Estimation of a location parameter in a regression (Vimond, Bercu)

Stochastic algorithm for integral approximation

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

$$\text{MC ERROR BOUND} = \sup_{\varphi \in \Psi} \left| n^{-1} \sum_{i=1}^{n} \frac{\varphi(X_i)}{f(X_i)} - I(\varphi) \right| = O_{\mathbb{P}}(n^{-1/2})$$

Deterministic Methods:

$$\text{OPTIMAL ERROR BOUND} = O(n^{-s/d})$$ Novak (2014, review)

deterministic algorithm are

- badly affected by large d
- badly affected by small s

Stochastic algorithm for integral approximation

Importance sampling

find f^* such that, $(X_1, \ldots, X_n)$ i.i.d. with law f^*

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

Adaptive to φ (Evans and Schwartz (2000, book))

(parametric approach in $n^{-1/2}$ if misspecification)

Nonparametric Imp. Sampling Zhang (NIS) (1996, JASA)

- $(X_1^{(0)}, \ldots, X_n^{(0)})$ according to f_0
- $(X_1, \ldots, X_n)$ according to

$$\hat{f}^*(x) \propto n^{-1} \sum_{i=1}^{n} \frac{\varphi(X_i^{(0)}) K_h(X_i^{(0)} - x)}{f_0(X_i^{(0)})}$$

Stochastic algorithm for integral approximation

Summary of Zhang (1996)

- ERROR BOUND $\mathrm{NIS} = n^{-1/2} n^{-\frac{2}{d+4}}$
- heavy computationally

KS in a stochastic algorithm

Draw $(X_1, \dots, X_n)$ by Hastings-Metropolis with a "nice stationary law" ($\propto |\varphi|^{2/3}$), and choose K as the transition probability ($B_n = 0$). Then

$$\text{ERROR BOUND KS} = n^{-1/2} n^{-\frac{s}{2(d+s)}}.$$

- ERROR BOUND KS $<<$ ERROR BOUND NIS iff $s > 4$
- computational time OK as soon as φ is expensive

Application: Evaluate average temperature of oceans

- Data collected from drifting buoys (dependency between the locations)
- Too few points at some places

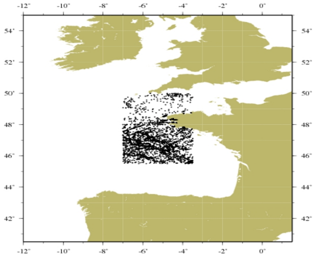
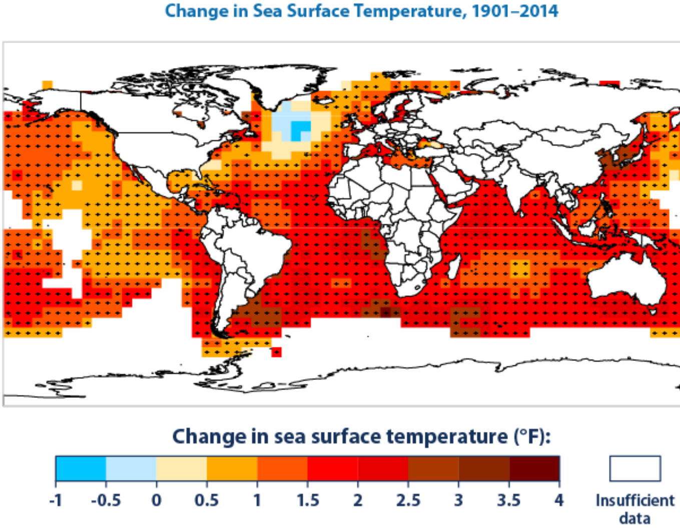


Figure: Data from the National Oceanographic Data Center

Temperature of Oceans

US Environmental Protection Agency (EPA)

“Changes in sea surface temperature can alter marine ecosystems [...] sea surface temperature can also have profound effects on global climate”



Database

From National Oceanic and Atmospheric Administration (NOAA)

- sea surface temperatures obtained all around the world between 2005 and 2015
- from profiling floats
- about 1.3M observations

Total 1 343 094					
Ocean					
Pacific 727 135		Atlantic 336 180		Indian 279 779	
Year		Year		Year	
2005	2015	2005	2015	2005	2015
35 773	86 961	16 242	45 488	14 134	33 049

Table: Size of the sub-datasets extracted from PFL dataset between 2005 and 2015.

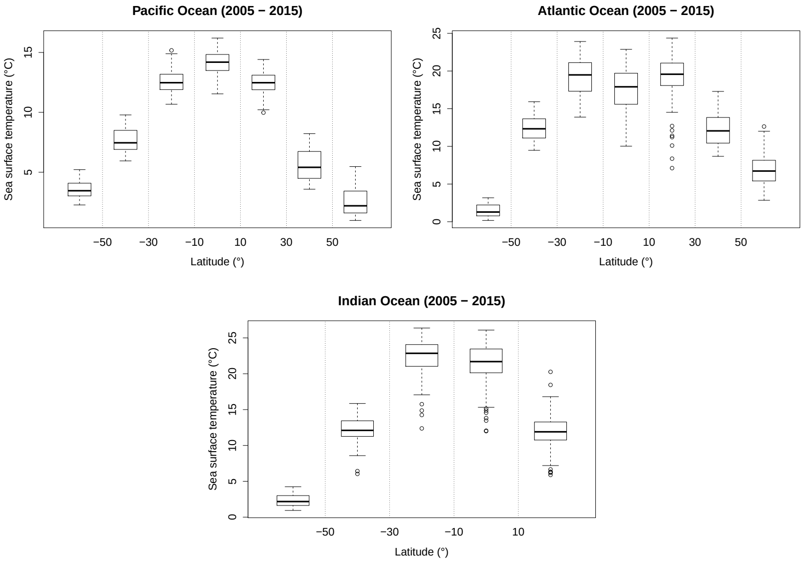


Figure: Average sea surface temperatures according to the latitude of the considered area for the 3 major oceans. Each boxplot has been computed from $11 \times 12 = 132$ estimates of the average temperature for each month of each year between 2005 and 2015.

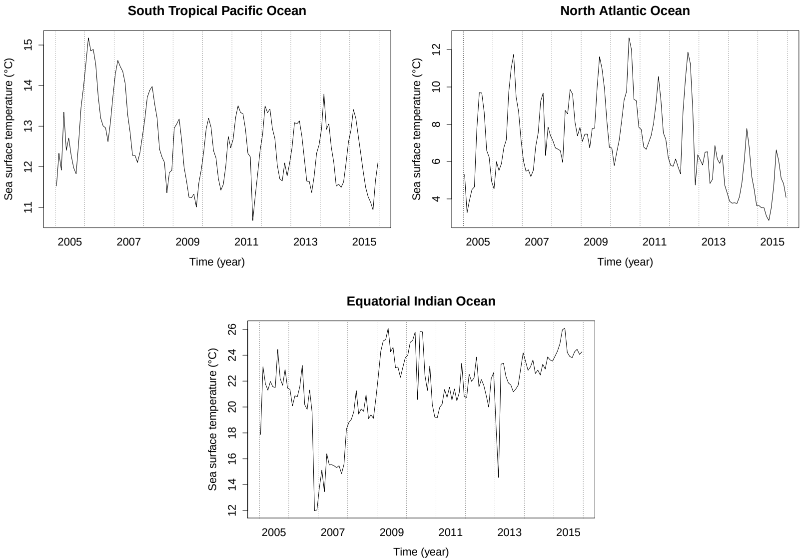


Figure: Times series of sea surface temperature in some specific areas of the 3 major oceans between 2005 and 2015. Latitude between -30° and -10° for South Tropical Pacific Ocean (top), 50° and 60° for North Atlantic Ocean (middle), and -10° and 10° for Equatorial Indian Ocean (bottom). See Kosaka and Xie (2013) Rahmstorf et al. (2015) Roxy et al. (2014)

What's next?

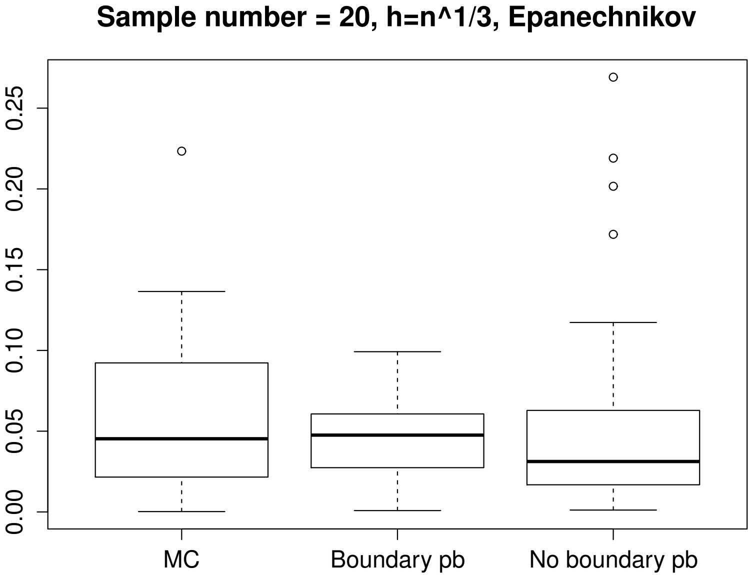
Theoretical results

- Trimming method ? Härdle and Stocker (1989, JASA)
- Local polynomial for boundary problem

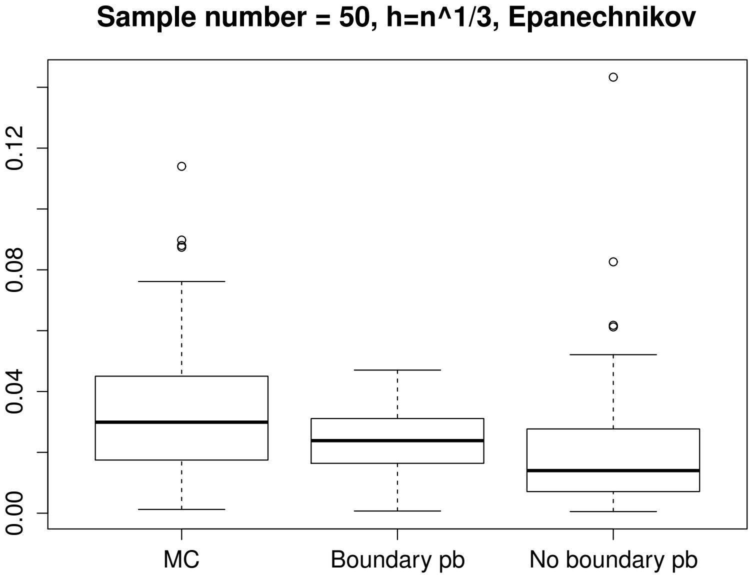
Computational

- Beat the curse of dimensionality
- Stochastic algorithm for integration

In practice

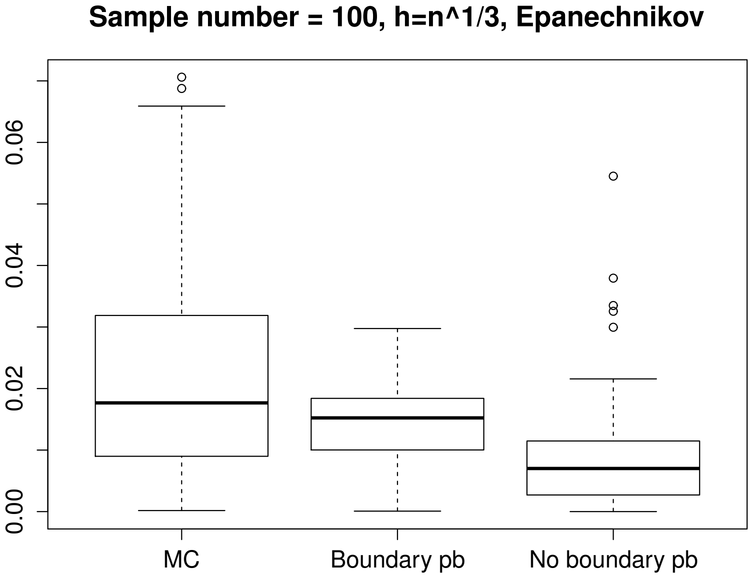

Sample number = 20, h=n^1/3, Epanechnikov
0.00
0.05
0.10
0.15
0.20
0.25
MC
Boundary pb
No boundary pb

In practice

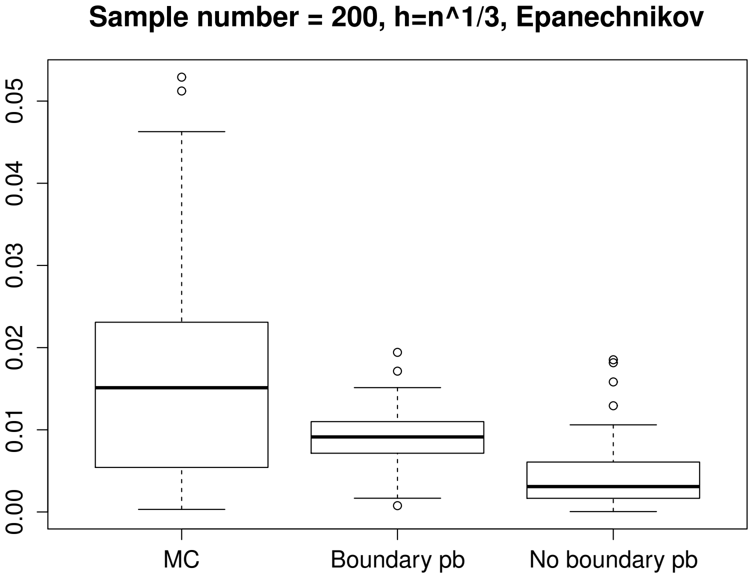


Sample number = 50, h=n^1/3, Epanechnikov
0.00
0.04
0.08
0.12
MC
Boundary pb
No boundary pb

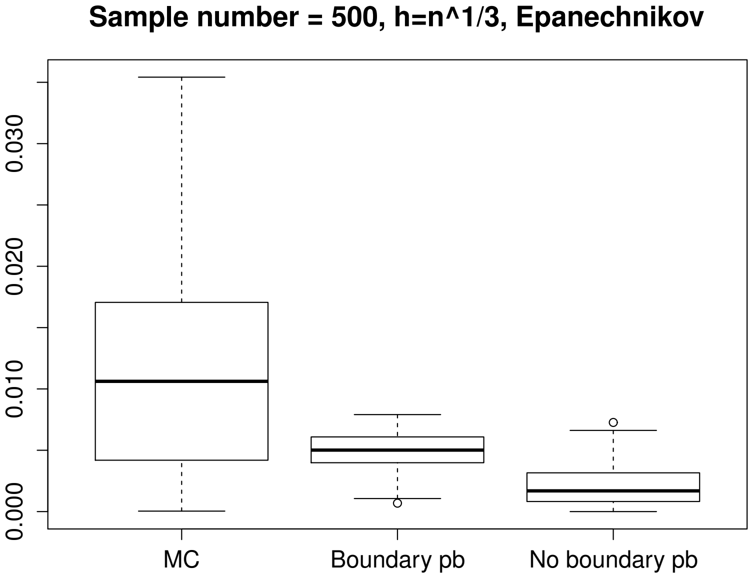
In practice


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

Simulations in dimension 1

Functions to integrate over $[0, 1]$

$$\varphi_1(x) = 1$$
$$\varphi_2(x) = 2\sin(\pi x)^2$$
$$\varphi_3(x) = \pi x \sin(\pi x)$$
$$\varphi_4(x) = \frac{\pi}{2}\sin(\pi x)(1 + \cos(5\pi x))$$

Uniform design

$$X_i \sim \mathcal{U}[0, 1]$$

Simulations in dimension 1

Functions to integrate over $[0, 1]$

$$\varphi_1(x) = 1$$
$$\varphi_2(x) = 2\sin(\pi x)^2$$
$$\varphi_3(x) = \pi x \sin(\pi x)$$
$$\varphi_4(x) = \frac{\pi}{2}\sin(\pi x)(1 + \cos(5\pi x))$$

Uniform design

$$X_i \sim \mathcal{U}[0, 1].$$

φ_2

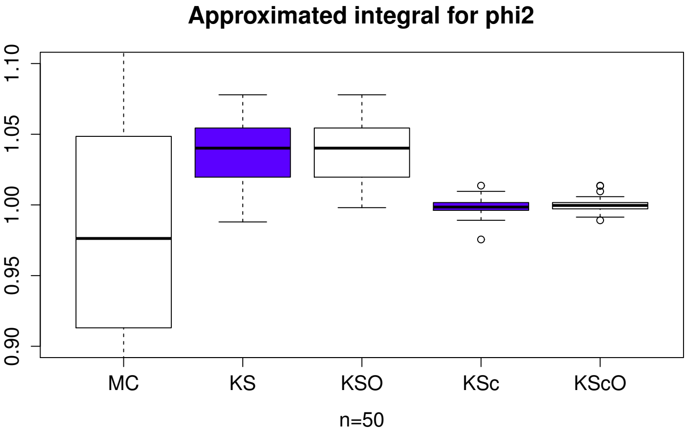


Figure: Boxplot based on 50 estimates

φ_2

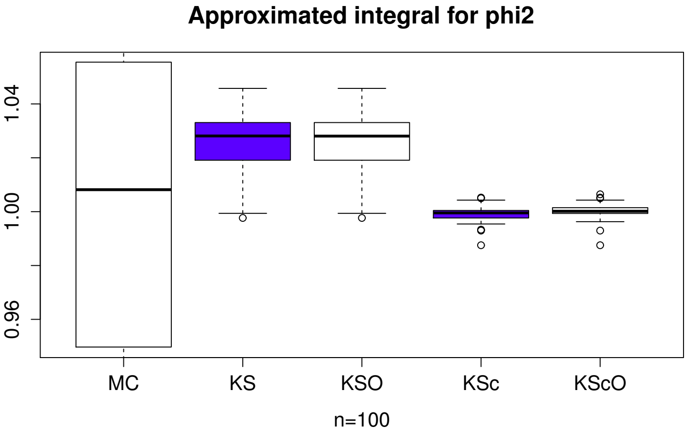


Figure: Boxplot based on 50 estimates

φ_2

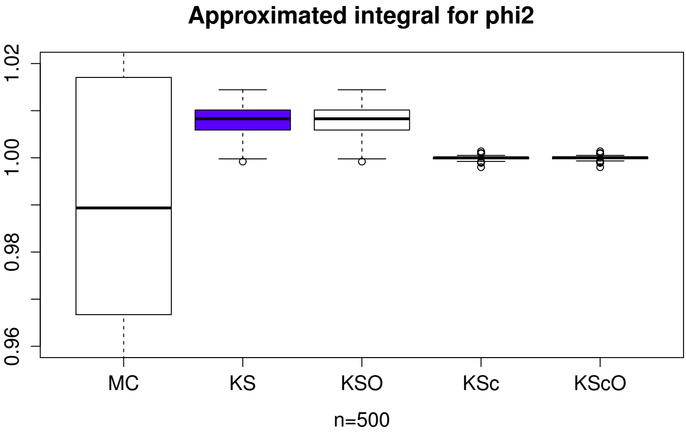


Figure: Boxplot based on 50 estimates

φ_4

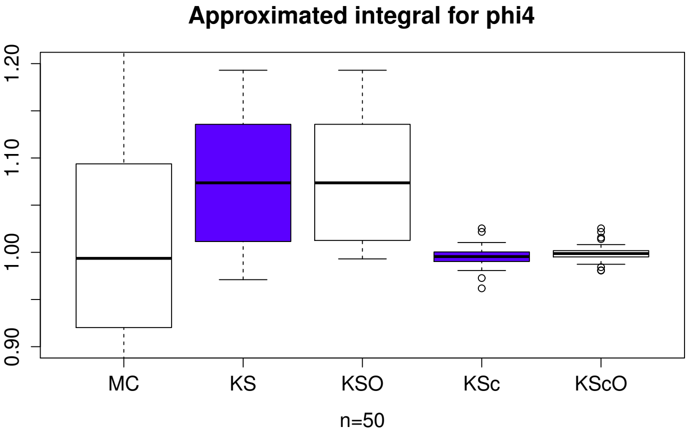


Figure: Boxplot based on 50 estimates

φ_4

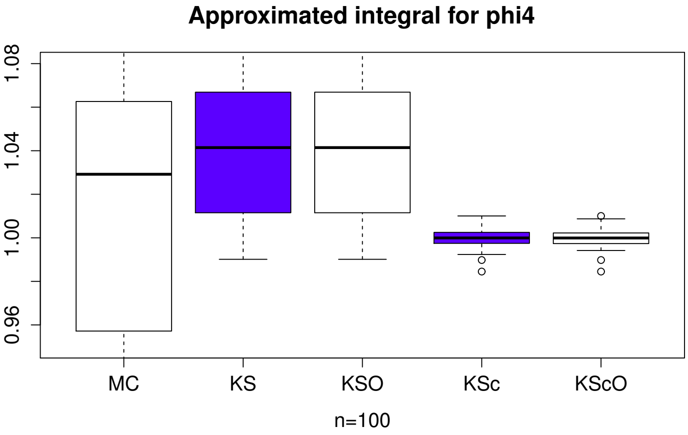


Figure: Boxplot based on 50 estimates

φ_4

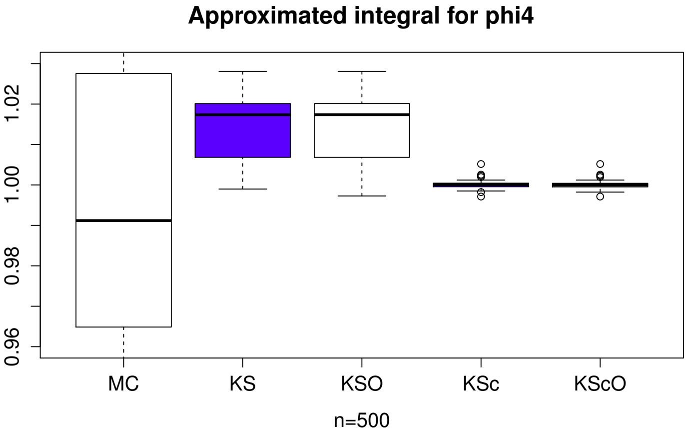


Figure: Boxplot based on 50 estimates