Nested risk computations through non parametric Regression, with Markovian design

Gersende Fort

Institut de Mathématiques de Toulouse
CNRS and Univ. de Toulouse
France

Joint work with

- Emmanuel Gobet (Ecole Polytechnique, France)
- Eric Moulines (Ecole Polytechnique, France)

Talk based on the paper

G. Fort, E. Gobet and E. Moulines. *MCMC design-based non-parametric regression for rare event. Application to nested risk computation.*, `Monte Carlo Methods and Applications`, 23(1):21–42, 2017.

Numerical method for the approximation of quantities of the form

$$\mathbb{E}\Big[\ f\Big(\mathbf{Y}, \phi_\star(\mathbf{Y})\Big)\Big|\mathbf{Y} \in \mathcal{A}\Big]$$

when (outer expectation)

- the integration w.r.t. to $\mathcal{L}(\mathbf{Y}|\mathbf{Y} \in \mathcal{A})$ is intractable
- the event $\{Y \in \mathcal{A}\}$ is rare

and (inner expectation)

- The function $\phi_\star$ is unknown, and is assumed of the form

$$\phi_\star(\mathbf{Y}) = \mathbb{E}\left[\mathbf{R}|\mathbf{Y}\right] \qquad \text{a.s.}$$

  with exact sampling from the conditional distribution $\mathcal{L}(\mathbf{R}|\mathbf{Y})$.

but

- For all (y, r), the quantity $f(y, r)$ can be explicitly computed.

$$\mathbb{E}\Big[\, f\Big(\mathbf{Y}, \mathbb{E}\left[\mathbf{R}|\mathbf{Y}\right]\Big)\Big|\mathbf{Y} \in \mathcal{A}\Big]$$

- Solving dynamical programming equations for stochastic control and optimal stopping problems - see the plenary talk by E. Gobet (Tsitsiklis and Van Roy, 2001; Egloff, 2005; Lemor et al. 2006; Belomestny et al. 2010)

- Financial and Actuarial Management (Mc Neil et al., 2005)

  ex. : risk management of portfolios written with derivative options (Gordy and Juneja, 2010) where

  $\mathbf{Y}$ is the underlying asset or financial variables at time T
  $\mathbf{R}$ aggregated cashflows of derivatives at time $T' > T$
  $\mathbb{E}[\mathbf{R}|\mathbf{Y}]$ is the portfolio value at time T given a scenario $\mathbf{Y}$

  and the aim is to compute the extreme exposure of the portfolio (VaR, CVaR).

A solution based on nested Monte Carlo (1/2)

- Step 1: An outer Monte Carlo step

$$\mathbb{E}\Big[\, f\Big(\mathbf{Y}, \phi_\star(\mathbf{Y})\Big)\Big|\mathbf{Y} \in \mathcal{A}\Big] \approx \frac{1}{N_{\text{out}}} \sum_{m=1}^{N_{\text{out}}} f\Big(\mathbf{X}^{(m)}, \phi_\star\big(\mathbf{X}^{(m)}\big)\Big)$$

How to draw the points $\mathbf{X}^{(m)}$?

A solution based on nested Monte Carlo (1/2)

- Step 1: An outer Monte Carlo step

$$\mathbb{E}\Big[\, f\Big(\mathbf{Y}, \phi_\star(\mathbf{Y})\Big)\Big|\mathbf{Y} \in \mathcal{A}\Big] \approx \frac{1}{N_{\text{out}}} \sum_{m=1}^{N_{\text{out}}} f\Big(\mathbf{X}^{(m)}, \phi_\star\big(\mathbf{X}^{(m)}\big)\Big)$$

How to draw the points $\mathbf{X}^{(m)}$?

- Rejection algorithm i.e. exact sampling under $\mathcal{L}(\mathbf{Y}|\mathbf{Y} \in \mathcal{A})$

  Repeat

  Draw independently, samples $Y^{(m)}$ with distribution $\mathbf{Y}$

  until

  $Y^{(m)} \in \mathcal{A}$

  $\hookrightarrow$ inefficient in the rare event setting: the mean number of loops to accept one sample is $1/\mathbb{P}(\mathbf{Y} \in \mathcal{A})$.

A solution based on nested Monte Carlo (1/2)

- Step 1: An outer Monte Carlo step

$$\mathbb{E}\Big[\, f\Big(\mathbf{Y}, \phi_\star(\mathbf{Y})\Big)\Big|\mathbf{Y} \in \mathcal{A}\Big] \approx \frac{1}{N_{\text{out}}} \sum_{m=1}^{N_{\text{out}}} f\Big(\mathbf{X}^{(m)}, \phi_\star\big(\mathbf{X}^{(m)}\big)\Big)$$

How to draw the points $\mathbf{X}^{(m)}$?

- Rejection algorithm i.e. exact sampling under $\mathcal{L}(\mathbf{Y}|\mathbf{Y} \in \mathcal{A})$
- Importance sampling (Rubinstein and Kroese, 2008; Blanchet and Lam, 2012)

  - efficient in small dimension, fails to deal with larger dimensions
  - relies heavily on particular types of models for $\mathbf{Y}$, and on suitable information about the problem.

A solution based on nested Monte Carlo (1/2)

- Step 1: An outer Monte Carlo step

$$\mathbb{E}\left[\, f\Big(\mathbf{Y}, \phi_\star(\mathbf{Y})\Big)\Big|\mathbf{Y} \in \mathcal{A}\right] \approx \frac{1}{N_{\text{out}}} \sum_{m=1}^{N_{\text{out}}} f\Big(\mathbf{X}^{(m)}, \phi_\star\big(\mathbf{X}^{(m)}\big)\Big)$$

How to draw the points $\mathbf{X}^{(m)}$?

- Rejection algorithm i.e. exact sampling under $\mathcal{L}(\mathbf{Y}|\mathbf{Y} \in \mathcal{A})$
- Importance sampling (Rubinstein and Kroese, 2008; Blanchet and Lam, 2012)
- MCMC approach: $\{\mathbf{X}^{(1)}, \cdots, \mathbf{X}^{(m)}, \cdots\}$ is a Markov chain having the conditional distribution $\mathcal{L}(\mathbf{Y}|\mathbf{Y} \in \mathcal{A})$ as the unique invariant distribution.

A solution based on nested Monte Carlo (2/2)

- Step 2: An inner Monte Carlo step

$$\mathbb{E}\Big[\, f\Big(\mathbf{Y}, \phi_\star(\mathbf{Y})\Big)\Big|\mathbf{Y} \in \mathcal{A}\Big] \approx \frac{1}{N_{\mathrm{out}}} \sum_{m=1}^{N_{\mathrm{out}}} f\Big(\mathbf{X}^{(m)}, \widehat{\phi}_\star^{(m)}\Big) \qquad \widehat{\phi}_\star^{(m)} \approx \phi_\star(\mathbf{X}^{(m)})$$

$\phi_\star(\mathbf{Y}) = \mathbb{E}[\mathbf{R}|\mathbf{Y}]$, exact sampling from $\mathcal{L}(\mathbf{R}|\mathbf{Y})$: available

- Crude Monte Carlo.

A solution based on nested Monte Carlo (2/2)

- Step 2: An inner Monte Carlo step

$$\mathbb{E}\Big[f\Big(\mathbf{Y}, \phi_\star(\mathbf{Y})\Big)\Big|\mathbf{Y} \in \mathcal{A}\Big] \approx \frac{1}{N_{\text{out}}} \sum_{m=1}^{N_{\text{out}}} f\Big(\mathbf{X}^{(m)}, \widehat{\phi}_\star^{(m)}\Big) \qquad \widehat{\phi}_\star^{(m)} \approx \phi_\star(\mathbf{X}^{(m)})$$

$\phi_\star(\mathbf{Y}) = \mathbb{E}[\mathbf{R}|\mathbf{Y}]$, exact sampling from $\mathcal{L}(\mathbf{R}|\mathbf{Y})$: available

- Crude Monte Carlo. cost: $N_{\text{in}} \times N_{\text{out}}$ draws
  for each sample $X^{(m)}$,
  draw $\{\mathbf{R}^{(m,1)}, \mathbf{R}^{(m,2)}, \cdots, \mathbf{R}^{(m,N_{\text{in}})}\} \overset{i.i.d.}{\sim} \mathcal{L}(\mathbf{R}|\mathbf{Y} = \mathbf{X}^{(m)})$
  set

$$\widehat{\phi}_\star^{(m)} \overset{\text{def}}{=} \frac{1}{N_{\text{in}}} \sum_{n=1}^{N_{\text{in}}} \mathbf{R}^{(m,n)}$$

- Regression.

A solution based on nested Monte Carlo (2/2)

- Step 2: An inner Monte Carlo step

$$\mathbb{E}\Big[f\Big(\mathbf{Y}, \phi_\star(\mathbf{Y})\Big)\Big|\mathbf{Y} \in \mathcal{A}\Big] \approx \frac{1}{N_{\text{out}}} \sum_{m=1}^{N_{\text{out}}} f\Big(\mathbf{X}^{(m)}, \widehat{\phi}_\star^{(m)}\Big) \qquad \widehat{\phi}_\star^{(m)} \approx \phi_\star(\mathbf{X}^{(m)})$$

$\phi_\star(\mathbf{Y}) = \mathbb{E}[\mathbf{R}|\mathbf{Y}]$, exact sampling from $\mathcal{L}(\mathbf{R}|\mathbf{Y})$: available

- Crude Monte Carlo. cost: $N_{\text{in}} \times N_{\text{out}}$ draws
- Regression. cost: N_{out} draws; Take into account cross-information between points $\mathbf{X}^{(m)}$

  for each sample $\mathbf{X}^{(m)}$,

  draw a single $R^{(m)} \sim \mathcal{L}(\mathbf{R}|\mathbf{Y} = \mathbf{X}^{(m)})$

  set $\widehat{\phi}_\star^{(m)} \overset{\text{def}}{=} \widehat{\phi}_\star(\mathbf{X}^{(m)})$ where

$$\widehat{\phi}_\star(x) \overset{\text{def}}{=} \text{argmin}_{\phi \in \mathcal{F}} \frac{1}{N_{\text{out}}} \sum_{m=1}^{N_{\text{out}}} \|\mathbf{R}^{(m)} - \phi(\mathbf{X}^{(m)})\|^2.$$

MCMC combined with Regression

$$\mathbb{E}\Big[\, f\Big(\mathbf{Y}, \phi_\star(\mathbf{Y})\Big)\Big|\mathbf{Y} \in \mathcal{A}\Big] \approx \frac{1}{N_{\text{out}}} \sum_{m=1}^{N_{\text{out}}} f\Big(\mathbf{X}^{(m)}, \widehat{\phi}_\star(\mathbf{X}^{(m)})\Big)$$

(I) samples points $\mathbf{X}^{(1)}, \cdots, \mathbf{X}^{(N_{\text{out}})}$ from a MCMC targeting $\mathcal{L}(\mathbf{Y}|\mathbf{Y} \in \mathcal{A})$

(II) choose L basis functions $\phi_1, \cdots, \phi_L$ and set

$$\widehat{\phi}_\star = \widehat{\alpha}_1 \phi_1 + \cdots + \widehat{\alpha}_L \phi_L$$

where

$$(\widehat{\alpha}_1, \cdots, \widehat{\alpha}_L) = \text{argmin}_{(\alpha_1, \cdots, \alpha_L) \in \mathbb{R}^L} \sum_{m=1}^{N_{\text{out}}} \|\mathbf{R}^{(m)} - \sum_{\ell=1}^{L} \alpha_\ell\, \phi_\ell(\mathbf{X}^{(m)})\|^2.$$

For alternatives to this regression approach, see e.g.: kernel estimators (Hong and Juneja, 2009); kriging techniques (Liu and Staum, 2010)

For alternatives to this MCMC design, see e.g. (Broadie et al., 2015) with a weighted regression

1. Convergence analysis when the outer Monte Carlo step relies on (non stationary) MCMC samples

   - existing results on the regression error address the case of a i.i.d. or a stationary design $\mathbf{X}^{(1)}, \cdots, \mathbf{X}^{(m)}, \cdots$ (Gyorfi et al. 2002, Ren and Mojirsheibani 2010, Delattre and Gaïffas, 2011)

   - consistent numerical method under weaker conditions on the basis functions $\phi_1, \cdots, \phi_L$ and on the distribution of the design (Broadie et al. 2015)

2. Ergodic properties of a MCMC sampler, designed to sample distributions restricted to a rare event.

On the MCMC step

An efficient algorithm to sample from the distribution

$$\pi \ \mathsf{d}\lambda \equiv \mathcal{L}(\mathbf{Y} | \mathbf{Y} \in \mathcal{A})$$

Choose a proposal kernel $q(x,z)\mathrm{d}\lambda(z)$ such that for all $x, z \in \mathcal{A}$

$$q(x,z)\pi(z) = \pi(x)q(x,z) \qquad \text{(reversible w.r.t. } \pi\text{)}$$

MCMC sampler (Gobet and Liu, 2015)

Init: $X^{(0)} \sim \xi$ *- a distribution on* $\mathcal{A}$

For $m = 1 : N_{\text{out}}$, *repeat:*

Draw a candidate $\widetilde{X}^{(m)} \sim q(X^{(m)}, z)\mathrm{d}\lambda(z)$

Update the chain: set

$$X^{(m+1)} = \begin{cases} \widetilde{X}^{(m)} & \textit{if } \widetilde{X}^{(m)} \in \mathcal{A} \\ X^{(m)} & \textit{otherwise} \end{cases}$$

Return $X^{(m)}, m = 0 : N_{\text{out}}$.

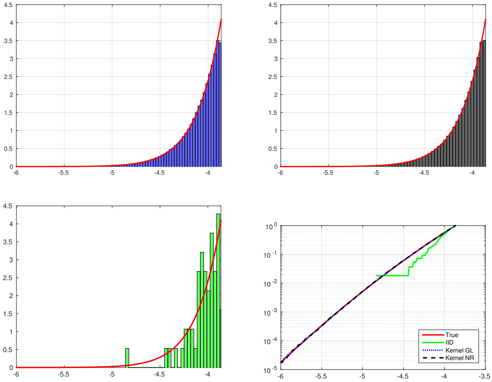
Application: sampling a $\mathcal{N}(0,1)$ in the left tail (1/3)

- Goal: $\mathbb{P}(\mathbf{Y} \in \cdot | \mathbf{Y} \leq y_\star)$, $\quad \mathbf{Y} \sim \mathcal{N}(0,1)$.
- Displayed: The histogram of the draws obtained by rejection (bottom left), by the MCMC sampler GL (top left) and the MCMC sampler NR (top right). The associated empirical cdf's (bottom right).

$$\mathrm{GL:}\ q(x,\cdot) = \rho x + \sqrt{1-\rho^2}\mathcal{N}(0,1) \qquad \text{(reversible)}$$

$$\mathrm{NR:}\ q(x,\cdot) = \rho x + (1-\rho)y_\star + \sqrt{1-\rho^2}\mathcal{N}(0,1) \qquad \text{(non reversible)}$$

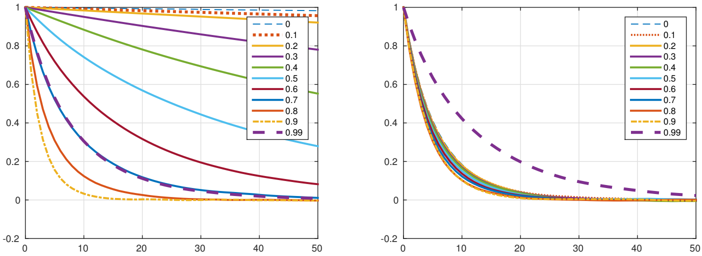
- Num. Appl.: $1e6$ draws for each algorithm ($\Rightarrow 50-60$ accepted draws for the rejection algorithm). $\rho = 0.85$. $\mathbb{P}(\mathbf{Y} \leq y_\star) = 5.6e-5$



- Goal: Role of the design parameter ρ in the efficiency of the samplers.
- Displayed: The autocorrelation function, averaged over 100 estimations with lag 0 to 50. For different values of $\rho \in (0,1)$. For the MCMC sampler GL (left) and NR (right).

$$\mathrm{GL:}\ q(x,\cdot) = \rho x + \sqrt{1-\rho^2}\mathcal{N}(0,1) \qquad \text{(reversible)}$$

$$\mathrm{NR:}\ q(x,\cdot) = \rho x + (1-\rho)y_\star + \sqrt{1-\rho^2}\mathcal{N}(0,1) \qquad \text{(non reversible)}$$

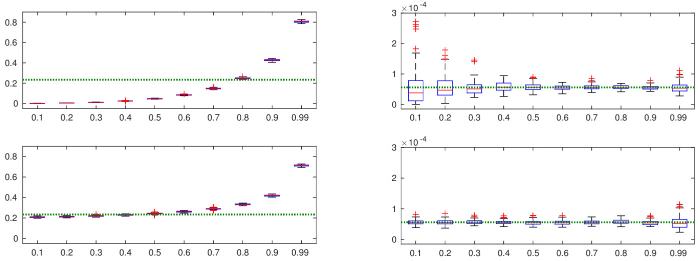


Application: sampling a $\mathcal{N}(0,1)$ in the (left) tail (3/3)

- Goal: Role of the design parameter ρ on the efficiency of the samplers.
- Displayed: [left] Boxplot of the mean acceptance rate computed along a path of length $N_{\text{out}} = 1e4$; 50 independent runs of the algorithms. [right] estimation of $\mathbb{P}(\mathbf{Y} \in \mathcal{A})$ for different values of $\rho \in (0,1)$; for the MCMC sampler GL (top) and NR (bottom).

$$\text{GL: } q(x,\cdot) = \rho x + \sqrt{1-\rho^2}\mathcal{N}(0,1) \qquad \text{(reversible)}$$

$$\text{NR: } q(x,\cdot) = \rho x + (1-\rho)y_\star + \sqrt{1-\rho^2}\mathcal{N}(0,1) \qquad \text{(non reversible)}$$



Proposition (F., Gobet, Moulines (2017))

Assume that

(i) *for all* $x \in \mathcal{A}$, $\pi(z) > 0 \Longrightarrow q(x,z) > 0$.

(ii) *there exists* $\delta_1 \in (0,1)$ *such that* $\sup_{x \in \mathcal{A}} \int_{\mathcal{A}^c} q(x,z) \mathrm{d}\lambda(z) \leq \delta_1$.

(iii) *there exist a measurable set* $\mathcal{C}$ *in* $\mathcal{A}$, $\delta_2 \in (\delta_1, 1)$ *and an unbounded off compact set measurable function* $V : \mathcal{A} \to [1, +\infty)$ *such that*

$$b \stackrel{\text{def}}{=} \sup_{x \in \mathcal{C}} \int_{\mathcal{A}} V(z)\, q(x,z) \mathrm{d}\lambda(z) < \infty, \quad \sup_{x \in \mathcal{C}^c} V^{-1}(x) \int_{\mathcal{A}} V(z)\, q(x,z) \mathrm{d}\lambda(z) \leq \delta_2 - \delta_1.$$

(iv) *For some* $v_\star > b/(1-\delta_2)$, *the level set* $\mathcal{C}_\star \stackrel{\text{def}}{=} \{V \leq v_\star\}$ *is such that*

$$\inf_{(x,z) \in \mathcal{C}_\star^2} \left(\frac{q(x,z) \mathbb{1}_{\pi(z) \neq 0}}{\pi(z)} \right) > 0, \qquad \int_{\mathcal{C}_\star} \pi \mathrm{d}\lambda > 0.$$

Then there exist $\kappa \in (0,1)$ *and* $C < \infty$ *such that for any function* $f : \mathcal{A} \to \mathbf{R}$,

$$\left| \mathsf{P}^m f(x) - \int f(z)\, \pi(z)\, \mathsf{d}\lambda(z) \right| \leq C \left(\sup_{\mathcal{A}} \frac{|f|}{V} \right) \kappa^m V(x), \qquad \forall x \in \mathcal{A}.$$

Comments on the theorem

- When π is a truncated Gaussian distribution:

$$V(x) = \exp(\beta\|x\|), \qquad q(x, \cdot) = \mathcal{N}_d(\rho x, (1-\rho^2)I_d).$$

- Sketch of the proof:

  Irreducibility, Aperiodicity
  The level sets of V are *small sets*
  Drift inequality: $\mathrm{P}V(x) \leq \delta V(x) + C \qquad$ for some $\delta \in (0,1]$.
  Then, standard results on Markov chains (Meyn and Tweedie, 1993)

- Ergodicity at a polyomial rate

  weaker assumptions for a weaker rate of convergence (Fort and Moulines, 2003; Douc et al. 2004)

Control of the regression approximation

- The unknown quantities $\phi_\star$ given by

$$\phi_\star(\mathbf{Y}) = \mathbb{E}\left[\mathbf{R}|\mathbf{Y}\right] \; a.s.$$

- Available

  $\mathbf{X}^{(1)}, \cdots, \mathbf{X}^{(N_{\text{out}})}$ from a Markov chain with stationary dist $\mathcal{L}(\mathbf{Y}|\mathbf{Y} \in \mathcal{A})$
  $\mathbf{R}^{(1)}, \cdots, \mathbf{R}^{(N_{\text{out}})}$ s.t. $\mathbf{R}^{(i)} \sim \mathcal{L}(\mathbf{R}|\mathbf{Y} = \mathbf{X}^{(i)})$.

- Estimation

$$\widehat{\phi}_\star(x) \stackrel{\text{def}}{=} \sum_{\ell=1}^{L} \widehat{\alpha}_\ell \; \phi_\ell(x)$$

where

$\phi_1, \cdots, \phi_L$ are basis functions chosen by the user,
the $\widehat{\alpha}_i' s$ are explicit solutions of

$$\text{argmin}_{(\alpha_1, \cdots, \alpha_L) \in \mathbb{R}^L} \sum_{m=1}^{N_{\text{out}}} \left(\mathbf{R}^{(m)} - \sum_{\ell=1}^{L} \alpha_\ell \, \phi_\ell(\mathbf{X}^{(m)}) \right)^2$$

Explicit control

Denote

$\mathsf{Q}(x, \mathsf{d}r)$ the cond. distribution of $\mathbf{R}|\mathbf{Y} = x$

$\psi_\star \stackrel{\text{def}}{=} \operatorname{argmin}_{\phi \in \mathcal{F}} \int (\phi - \phi_\star)^2 \, \pi \, \mathsf{d}\lambda$, where $\pi \, \mathsf{d}\lambda \equiv \mathcal{L}(\mathbf{Y}|\mathbf{Y} \in \mathcal{A})$.

Mean squared error along the design (F., Gobet, Moulines (2017))

Assume that

(i) *the MCMC kernel* P *and the initial distribution* ξ *satisfy: there exists a constant* C_{P} *and a rate sequence* $\{\rho(m), m \geq 1\}$ *such that for any* $m \geq 1$,

$$\left| \xi \mathsf{P}^m [(\psi_\star - \phi_\star)^2] - \int (\psi_\star - \phi_\star)^2 \, \pi \, \mathsf{d}\lambda \right| \leq C_{\mathsf{P}} \, \rho(m). \tag{1}$$

(ii) *the transition kernel* Q *of the cond. distribution* $\mathcal{L}(\mathbf{R}|\mathbf{Y})$ *satisfies*

$$\sigma^2 \stackrel{\text{def}}{=} \sup_{x \in \mathcal{A}} \left\{ \int r^2 \, \mathsf{Q}(x, \mathsf{d}r) - \left(\int r \, \mathsf{Q}(x, \mathsf{d}r) \right)^2 \right\} < \infty. \tag{2}$$

Then,

$$\mathbb{E}\left[\frac{1}{N_{\text{out}}} \sum_{m=1}^{N_{\text{out}}} \left(\widehat{\phi}_\star(\mathbf{X}^{(m)}) - \phi_\star(\mathbf{X}^{(m)}) \right)^2 \right] \leq \frac{\sigma^2 L}{N_{\text{out}}} + |\psi_\star - \phi_\star|^2_{L_2(\pi)} + \frac{C_{\mathsf{P}}}{N_{\text{out}}} \sum_{m=1}^{N_{\text{out}}} \rho(m).$$

The proof is a bias/variance decomposition:

$$\frac{1}{N_{\text{out}}}\sum_{m=1}^{N_{\text{out}}}\left(\widehat{\phi}_\star(\mathbf{X}^{(m)})-\phi_\star(\mathbf{X}^{(m)})\right)^2$$

$$=\frac{1}{N_{\text{out}}}\sum_{m=1}^{N_{\text{out}}}\left(\widehat{\phi}_\star(\mathbf{X}^{(m)})-\mathbb{E}\left[\widehat{\phi}_\star(\mathbf{X}^{(m)})|\mathbf{X}^{(1:N_{\text{out}})}\right]\right)^2$$

controled by $\sigma^2 L/N_{\text{out}}$

$$+\frac{1}{N_{\text{out}}}\sum_{m=1}^{N_{\text{out}}}\left(\mathbb{E}\left[\widehat{\phi}_\star(\mathbf{X}^{(m)})|\mathbf{X}^{(1:N_{\text{out}})}\right]-\phi_\star(\mathbf{X}^{(m)})\right)^2$$

ergodicity of the chain
$+$ the norm $|\psi_\star-\phi_\star|^2_{L_2(\pi)}$

Toy example - description (1/2)

- A stock price $\{S_t, t \geq 0\}$, modeled as a 1-D geometric Brownian motion
- A put option $(K - S_{T'})_+$ with strike K and maturity T'
- The owner of the contract aims at valuing the excess of the put price at time $T < T'$ above the threshold $p_\star$, conditionally to a stock value S_T lower that $s_\star$

$$\mathbb{E}\left[\left(\underbrace{\mathbb{E}\left[(K - S_{T'})_+ \; | S_T \right]}_{\text{put price at time } T;\ \phi_\star(S_T)} - p_\star \right)_+ \middle| \underbrace{S_T \leq s_\star}_{\text{rare event}} \right]$$

Of the form

$$\mathbb{E}\left[f\left(\mathbf{Y}, \mathbb{E}\left[\mathbf{R} | \mathbf{Y} \right] \right) | \mathbf{Y} \leq y_\star \right]$$

where

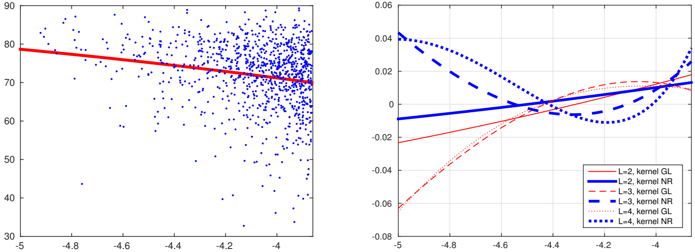
- $\mathbf{Y} \sim \mathcal{N}(O, 1)$,
- $\mathbf{R} = \Xi(\mathbf{Y}, Z)$ with $Z \sim \mathcal{N}(0, 1)$ and indep. of $\mathbf{Y}$.

Toy example - Estimation of $\phi_\star$ (2/2)

▶ Goal: In this example, $\phi_\star$ is explicit $\to$ the error $\widehat{\phi}_\star - \phi_\star$ can be displayed.

▶ Displayed: [left] The N_{out} points $(\mathbf{X}^{(m)}, \mathbf{R}^{(m)})$ when the $\mathbf{X}^{(m)}$'s are sampled from the MCMC sampler GL; and the function $x \mapsto \phi_\star(x)$ in red.
[right] Six realizations of $\widehat{\phi}$ resp. obtained with $L = 2, 3, 4$ and a design $\mathbf{X}^{(m)}$'s sampled from the kernel GL (red) and NR (blue). The basis functions are

$$\phi_\ell(x) = S_0^{\ell-1} \ \exp\Big((\ell-1)(-0.5\sigma^2 T + \sigma\sqrt{T} x) \Big) \, ;$$



Error of the numerical method

$$\mathcal{I} - \widehat{\mathcal{I}}_{N_{\text{out}}} \stackrel{\text{def}}{=} \mathbb{E}\Big[\, f\Big(\mathbf{Y}, \phi_\star(\mathbf{Y})\Big)\Big|\mathbf{Y} \in \mathcal{A}\Big] - \frac{1}{N_{\text{out}}} \sum_{m=1}^{N_{\text{out}}} f\Big(\mathbf{X}^{(m)}, \widehat{\phi}_\star(\mathbf{X}^{(m)})\Big)$$

Error of the numerical method

$$
\begin{aligned}
\mathcal{I} - \widehat{\mathcal{I}}_{N_{\text{out}}} &\stackrel{\text{def}}{=} \mathbb{E}\Big[\, f\Big(\mathbf{Y}, \phi_\star(\mathbf{Y})\Big)\Big|\mathbf{Y} \in \mathcal{A}\Big] - \frac{1}{N_{\text{out}}} \sum_{m=1}^{N_{\text{out}}} f\Big(\mathbf{X}^{(m)}, \widehat{\phi}_\star(\mathbf{X}^{(m)})\Big) \\
&= \mathbb{E}\Big[\, f\Big(\mathbf{Y}, \phi_\star(\mathbf{Y})\Big)\Big|\mathbf{Y} \in \mathcal{A}\Big] - \frac{1}{N_{\text{out}}} \sum_{m=1}^{N_{\text{out}}} f\Big(\mathbf{X}^{(m)}, \phi_\star(\mathbf{X}^{(m)})\Big) \\
&+ \frac{1}{N_{\text{out}}} \sum_{m=1}^{N_{\text{out}}} f\Big(\mathbf{X}^{(m)}, \phi_\star(\mathbf{X}^{(m)})\Big) - \frac{1}{N_{\text{out}}} \sum_{m=1}^{N_{\text{out}}} f\Big(\mathbf{X}^{(m)}, \widehat{\phi}_\star(\mathbf{X}^{(m)})\Big) \qquad \leftarrow
\end{aligned}
$$

(F., Gobet, Moulines (2017))

Assume

(i) $f : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ *is globally Lipschitz in the second variable: there exists a finite constant* C_f *such that for any* $(r_1, r_2, y) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^d$,

$$|f(y, r_1) - f(y, r_2)| \leq C_f \; |r_1 - r_2| \, .$$

(ii) *There exists a finite constant* C *such that for any* N_{out}

$$\mathbb{E}\left[\left({N_{\text{out}}}^{-1} \sum_{m=1}^{N_{\text{out}}} f\left(\mathbf{X}^{(m)}, \phi_\star(\mathbf{X}^{(m)})\right) - \int f(x, \phi_\star(x))\pi(x) \, \mathsf{d}\lambda(x)\right)^2\right] \leq \frac{C}{N_{\text{out}}}.$$

Then

$$\left(\mathbb{E}\left[\left|\mathcal{I} - \widehat{\mathcal{I}}_{N_{\text{out}}}\right|^2\right]\right)^{1/2} \leq C_f \, \sqrt{\Delta_{N_{\text{out}}}} + \sqrt{\frac{C}{N_{\text{out}}}},$$

where (for the rate, see the slide on the regression error)

$$\Delta_{N_{\text{out}}} \overset{\text{def}}{=} \mathbb{E}\left[\frac{1}{N_{\text{out}}} \sum_{m=1}^{N_{\text{out}}} \left(\widehat{\phi}_\star(\mathbf{X}^{(m)}) - \phi_\star(\mathbf{X}^{(m)})\right)^2\right] = O\left(\frac{1}{N_{\text{out}}}\right).$$