Nested risk computations through non parametric Regression with Markovian design

Gersende Fort

LTCI, CNRS and Telecom ParisTech
Paris, France

Joint work with

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

Goal

Monte Carlo for the approximation of intractable quantities of the form

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

when

Goal

Monte Carlo for the approximation of intractable quantities of the form

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

when

1. the conditional expectation is unknown

Goal

Monte Carlo for the approximation of intractable quantities of the form

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

when

1. the conditional expectation is unknown
2. the event $\{\mathbf{Y} \in \mathcal{A}\}$ is rare.

Goal

Monte Carlo for the approximation of intractable quantities of the form

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

when

1. the conditional expectation is unknown
2. the event $\{\mathbf{Y} \in \mathcal{A}\}$ is rare.
3. the function f is known, with explicit evaluation.

Goal

Monte Carlo for the approximation of intractable quantities of the form

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

when

1. the conditional expectation is unknown
2. the event $\{\mathbf{Y} \in \mathcal{A}\}$ is rare.
3. the function f is known, with explicit evaluation.
4. the distribution of $\mathbf{Y}$ is known (explicit) and we can sample under the conditional distribution of $\mathbf{R}$ given $\mathbf{Y}$.

Hereafter, set

$$\phi_\star(\mathbf{X}) = \mathbb{E}\left[\mathbf{R}|\mathbf{X}\right].$$

The nested Monte Carlo approximation (1/3)

- Step 1:

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

where $\mathbf{X}\sim$ the conditional distribution of $\mathbf{Y}$ given $\{\mathbf{Y}\in\mathcal{A}\}$.

The nested Monte Carlo approximation (1/3)

- Step 1:

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

where $\mathbf{X}\sim$ the conditional distribution of $\mathbf{Y}$ given $\{\mathbf{Y}\in\mathcal{A}\}$.

- Step 2: an outer stage for Monte Carlo sampling

$$\mathbb{E}\left[f\left(\mathbf{X},\phi_\star(\mathbf{X})\right)\right]\approx\frac{1}{M}\sum_{m=1}^{M}f\left(X^{(m)},\phi_\star(X^{(m)})\right)$$

`Difficulties:` efficient sampling from the distribution of $\mathbf{Y}$ given the rare event $\{\mathbf{Y}\in\mathcal{A}\}$.

`Solution:` a MCMC approach i.e. samples from a Markov chain with kernel P.

The nested Monte Carlo approximation (1/3)

- Step 1:
$$\mathbb{E}\left[f\left(\mathbf{Y},\phi_\star(\mathbf{Y})\right)|\mathbf{Y}\in\mathcal{A}\right]=\mathbb{E}\left[f\left(\mathbf{X},\phi_\star(\mathbf{X})\right)\right]$$
where $\mathbf{X}\sim$ the conditional distribution of $\mathbf{Y}$ given $\{\mathbf{Y}\in\mathcal{A}\}$.

- Step 2: an outer stage for Monte Carlo sampling
$$\mathbb{E}\left[f\left(\mathbf{X},\phi_\star(\mathbf{X})\right)\right]\approx\frac{1}{M}\sum_{m=1}^{M}f\left(X^{(m)},\phi_\star(X^{(m)})\right)$$

`Difficulties:` efficient sampling from the distribution of $\mathbf{Y}$ given the rare event $\{\mathbf{Y}\in\mathcal{A}\}$.

`Solution:` a MCMC approach i.e. samples from a Markov chain with kernel P.

The nested Monte Carlo approximation (2/3)

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

- Step 3: an inner stage : for each $m = 1, \cdots, M$, approximation of $\phi_\star(X^{(m)})$ through regression

  fix L basis functions $\phi_1, \cdots, \phi_L$.

  for each $X^{(m)}$, a **single draw** $R^{(m)}$ under the conditional distribution of $\mathbf{R}$ given $\mathbf{X}$.

  least square regression of the samples $\{R^{(m)}, m = 1 : M\}$ on the regressors $\{\phi_\ell(X^{(m)}), \ell = 1 : L, m = 1 : M\}$

The nested Monte Carlo approximation (3/3)

Goal: $\mathbb{E}\left[f\left(\mathbf{Y},\mathbb{E}\left[\mathbf{R}|\mathbf{Y}\right]\right)|\mathbf{Y}\in\mathcal{A}\right]=\mathbb{E}\left[f\left(\mathbf{X},\phi_\star(\mathbf{X})\right)\right]$

Algorithm:

Init.: $X^{(0)}\sim\xi$ *where* ξ *is a distribution on* $\mathcal{A}$.

For $m=1:M$, *do*

$X^{(m)}\sim\mathsf{P}(X^{(m-1)},\cdot)$

$R^{(m)}\sim\mathsf{Q}(X^{(m)},\cdot)$ - *the conditional dist. of* $\mathbf{R}$ *given* $\mathbf{X}$

Choose $\hat\alpha_M$ *solving*

$$\mathrm{argmin}_{\alpha\in\mathbb{R}^L}\sum_{m=1}^{M}\left(R^{(m)}-\sum_{\ell=1}^{L}\alpha_\ell\,\phi_\ell(X^{(m)})\right)^2$$

and set $\hat\phi_M(x)=\sum_{\ell=1}^{L}\hat\alpha_{M,\ell}\;\phi_\ell(x)$.

Return

$$\frac{1}{M}\sum_{m=1}^{M}f\left(X^{(m)},\hat\phi_M(X^{(m)})\right)$$

Comparison to the literature

See e.g. Broadie et al. (2011) and refs therein

- i.i.d. samples $X^{(m)}$ / Markovian samples.
- N draws $R^{(1,m)}, \cdots, R^{(N,m)}$ draws per $X^{(m)}$ / A single draw $R^{(m)}$ per $X^{(m)}$.
- Weighted linear regression / (trivial extension).
- Full rank and orthogonal regressors / (no conditions).

Outline

The algorithm

How to sample from a distribution restricted to a rare event ?
- The method
- Application: toy example

How to estimate the conditional expectation ?

Convergence results

Goal

An efficient algorithm to sample from the distribution

$$\mu \, \mathrm{d}\lambda \equiv \text{the conditional dist. of } \mathbf{Y} \text{ given } \{\mathbf{Y} \in \mathcal{A}\}$$

The reject algorithm

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

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

Markov chain Monte Carlo sampling with target $\mu\,\mathrm{d}\lambda$

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

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

MCMC sampler (Gobet and Liu, 2015)

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

For $m = 1 : M$, *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 $X^{(m)}, m = 0 : M$.

Toy example

- 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 $s_\star$

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

In this toy example,

- The rare event probability $\mathbb{P}(S_T \leq s_\star)$ is explicit
- The conditional expectation $\phi_\star(s)$ is explicit

Toy example: how to sample $X^{(m)}$?

The target distribution:

- the distribution of S_T given $\{S_T \leq s_\star\}$:

$$S_T = S_0 \ \exp\left(\{r - \frac{1}{2}\sigma^2\}T + \sigma W_T\right)$$

- Equivalently: the distribution of a standard Gaussian distribution W restricted to $\{W \leq w_\star\}$.

In that case, we can choose

$$\widetilde{X} = \rho x + \sqrt{1-\rho^2}\mathcal{N}(0,1).$$

where $\rho \in (0,1)$.

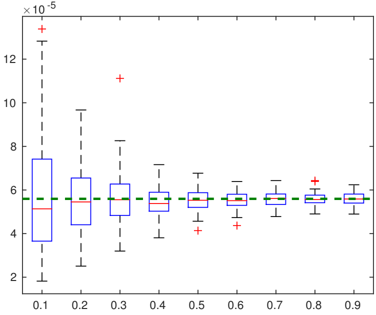
Toy example: on the design parameters

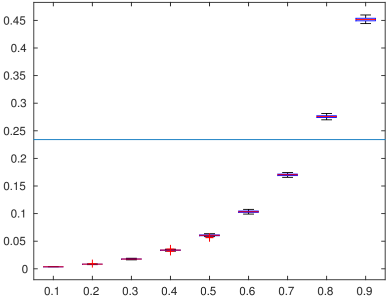
In this example: $\mathbb{P}(\mathbf{Y} \in \mathcal{A}) = 5.6e-5$. We have

$$\mathbb{P}(W \leq w_\star) = \prod_{j=1}^{J} \mathbb{P}(W \leq w_j | W \leq w_{j-1}) \qquad w_0 < w_1 < \cdots < w_J = w_\star$$

Displayed

- for different values of $\rho \in (0,1)$
- the boxplot of 100 independent realizations of the estimator $\prod_{j=1}^{J} \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}_{W_m^{(j)} \leq w_j}$,
- the boxplot of 100 independent realizations of the acceptance-rejection rate.





Outline

Goal

Wanted: An approximation of the function $\phi_\star$ given by

$$\phi_\star(\mathbf{X}) = \mathbb{E}\left[\mathbf{R}|\mathbf{X}\right].$$

Available:

- Samples $X^{(m)}, m = 1 : M$, approximating the distribution of $\mathbf{X}$.
- A transition kernel Q (easy to sample from) for the conditional distribution of $\mathbf{R}$ given $\mathbf{X}$.
- Basis functions $\phi_\ell, \ell = 1 : M$ chosen by the user.

Estimation through regression

The problem:

- Approximation by a function of the form $\sum_{\ell=1}^{L} \alpha_\ell \, \phi_\ell$
- Given an approximation $R^{(m)} \sim \mathsf{Q}(X^{(m)}, \cdot)$ of $\phi_\star(X^{(m)})$.

The approach:

$$\operatorname{argmin}_{\alpha \in \mathbb{R}^L} \sum_{m=1}^{M} \| R^{(m)} - \sum_{\ell=1}^{L} \alpha_\ell \, \phi_\ell(X^{(m)}) \|^2.$$

Estimation through regression

The problem:

- Approximation by a function of the form $\sum_{\ell=1}^{L} \alpha_\ell \, \phi_\ell$
- Given an approximation $R^{(m)} \sim \mathsf{Q}(X^{(m)}, \cdot)$ of $\phi_\star(X^{(m)})$.

The approach:

$$\mathrm{argmin}_{\alpha \in \mathbb{R}^L} \sum_{m=1}^{M} \| R^{(m)} - \sum_{\ell=1}^{L} \alpha_\ell \, \phi_\ell(X^{(m)}) \|^2 .$$

The solution: Compute a solution $\hat{\alpha}_M$ of

$$(\mathbf{A}^T \mathbf{A}) \alpha = \mathbf{A}^T \underline{\mathbf{R}}$$

so that

$$\left[\hat{\phi}_M (X^{(m)}) \right]_{m=1:M} = \mathbf{A} \hat{\alpha}_M$$

$$\underline{\mathbf{R}} \overset{\text{def}}{=} \begin{bmatrix} R^{(1)} \\ \cdots \\ R^{(M)} \end{bmatrix}, \qquad \mathbf{A} \overset{\text{def}}{=} \begin{bmatrix} \phi_1(X^{(1)}) & \cdots & \phi_L(X^{(1)}) \\ \cdots & \cdots & \cdots \\ \phi_1(X^{(M)}) & \cdots & \phi_L(X^{(M)}) \end{bmatrix}.$$

On a toy example (1/4)

Goal

$$\phi_\star(S_T) = \mathbb{E}\left[(K - S_{T'})_+ \ |S_T\right]$$

Sampling $R^{(m)}$ given $X^{(m)}$ Here,

$$R^{(m)} = (K - S_{T'}^{(m)})_+$$

where $S_{T'}^{(m)}$ is sampled from the conditional distribution of $S_{T'}$ given $S_T = X^{(m)}$.

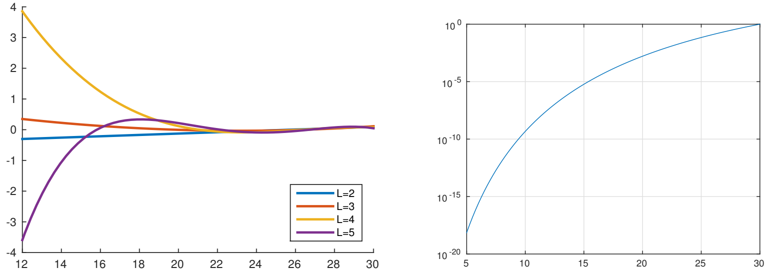
Basis functions

$$\phi_\ell(x) = x^{\ell-1}$$

On a toy example (2/4)

Displayed

- (left) for different values of L, a realization of the estimator $x \mapsto \hat{\phi}_M(x) - \phi_\star(x)$
- (right) the cdf of S_T given $S_T \leq s_\star$.

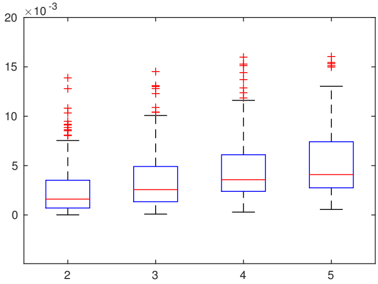


On a toy example (3/4)

Displayed

- for different values of L
- boxplot of 100 ind. realizations of the mean squared error

$$\frac{1}{M}\sum_{m=1}^{M}\left(\hat{\phi}_M(X^{(m)}) - \phi_\star(X^{(m)})\right)^2$$



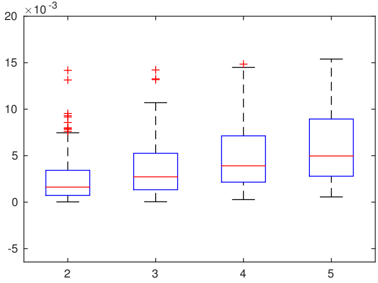
On a toy example (4/4)

Displayed

- for different values of L
- boxplot of 100 ind. realizations of

$$\frac{1}{N}\sum_{n=1}^{N}\left(\hat{\phi}_M(Z^{(n)}) - \phi_\star(Z^{(n)})\right)^2 \approx \mathbb{E}\left[\left(\hat{\phi}_M(\mathbf{X}) - \phi_\star(\mathbf{X})\right)^2\right]$$

where $Z^{(n)}, n = 1 : N$ is a Markov chain independent of $X^{(m)}, m = 1 : M$.



Outline

Convergence result on $\hat{\phi}_M$

Theorem (F.,Gobet,Moulines (2016))

Let ψ *s.t.* $|\psi - \phi_\star|_{L_2(\mu)} = \min_{\phi \in \mathrm{Span}(\phi_\ell, \ell=1:L)} |\phi - \phi_\star|^2_{L_2(\mu)}$.
Assume that

(i) *the transition kernel* P *and the initial distribution* ξ *satisfy: there exists a constant* C *and a rate sequence* $\{\rho(m), m \geq 0\}$ *such that for any* $m \geq 0$,

$$\left| \xi \mathsf{P}^m [(\psi - \phi_\star)^2] - \int (\psi - \phi_\star)^2 \, \mu \, \mathsf{d}\lambda \right| \leq C \rho(m).$$

(ii) *the conditional distribution* Q *satisfies*

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

Then,

$$\mathbb{E}\left[\frac{1}{M} \sum_{m=1}^{M} \left(\hat{\phi}_M(X^{(m)}) - \phi_\star(X^{(m)}) \right)^2 \right] \leq \frac{\sigma^2 K}{M} + \frac{C}{M} \sum_{m=1}^{M} \rho(m) + |\psi - \phi_\star|^2_{L_2(\mu)}.$$

Comments on the convergence result

1. In the case of i.i.d. sampling, $C = 0$ (same bound in e.g. Gyorfi et al. 2002)
2. When f is Lipschitz, first step for the control of the error

$$\frac{1}{M}\sum_{m=1}^{M} f\left(X^{(m)}, \hat{\phi}_M(X^{(m)})\right) - \mathbb{E}\left[f\left(\mathbf{X}, \phi_\star(\mathbf{X})\right)\right]$$

Comments on the convergence result

1. In the case of i.i.d. sampling, $C = 0$ (same bound in e.g. Gyorfi et al. 2002)
2. When f is Lipschitz, first step for the control of the error

$$\frac{1}{M}\sum_{m=1}^{M} f\left(X^{(m)}, \hat{\phi}_M(X^{(m)})\right) - \mathbb{E}\left[f\left(\mathbf{X}, \phi_\star(\mathbf{X})\right)\right]$$

3. The proof is a bias/variance decomposition:

$$\frac{1}{M}\sum_{m=1}^{M}\left(\hat{\phi}_M(X^{(m)}) - \phi_\star(X^{(m)})\right)^2$$

$$= \frac{1}{M}\sum_{m=1}^{M}\left(\hat{\phi}_M(X^{(m)}) - \mathbb{E}\left[\hat{\phi}_M(X^{(m)})|X^{(1:M)}\right]\right)^2 \qquad \text{controled by } \sigma^2 K/M$$

$$+ \frac{1}{M}\sum_{m=1}^{M}\left(\mathbb{E}\left[\hat{\phi}_M(X^{(m)})|X^{(1:M)}\right] - \phi_\star(X^{(m)})\right)^2 \qquad \text{ergodicity of the chain} + \text{the norm } \|\psi - \phi_\star\|_{L_2(\mu)}$$

Ergodicity of the MCMC sampler for rare event

Proposition (F.,Gobet,Moulines (2016))

Assume that

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

(ii) *the functions* $z \mapsto \mu(z)$ *and* $(x,z) \mapsto q(x,z)$ *are continuous for all* $x, z \in \mathcal{A}$.

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

(iv) *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*

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

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

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

$\hookrightarrow$ When the target is a truncated Gaussian distribution: satisfied with $V(x) = \exp(\beta\|x\|)$ and the proposal $\tilde{X} \sim \mathcal{N}_d(\rho x, (1-\rho^2)I_d)$.