Variance reduction for nonlinear Monte Carlo

D. Belomestny[1]

[1]University of Duisburg Essen

PARIS, 2016

Nested Expectations

Let $(X, Y) \in \mathbb{R}^d \times \mathbb{R}^d$ be a random vector such that one can generate samples from the conditional distribution $Y|X$.

Problem

The problem is to compute the quantity

$$F = \mathrm{E}[f(\mathrm{E}[g(X, Y)|X])],$$

where $g : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}^m$ and $f : \mathbb{R}^m \mapsto \mathbb{R}$.

Importance

This problem can be frequently encountered in risk management, mathematical finance and engeneering.

Examples

McKean-Vlasov Equation

Let $(X_t)_{t\in[0,T]}$ satisfy

$$dX_t = A_t(X_t)\,dt + \sigma^\top dW_t,$$

where $A_t(x) := \mathrm{E}\left[a(X_t, X_t')\,|\,X_t = x\right]$, $a : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}^d$, $\sigma \in \mathbb{R}^d \times \mathbb{R}^d$ is a constant matrix, W_t is a d-dimensional Brownian motion and X_t' is an independent copy of X_t. Suppose that we want to compute

$$\mathrm{E}\left[F(X_t)\right]$$

for some $t > 0$ and some function $F : \mathbb{R}^d \mapsto \mathbb{R}$. The Euler scheme yields

$$X_{\Delta,j\Delta} = X_{\Delta,(j-1)\Delta} + A_{(j-1)\Delta}(X_{\Delta,(j-1)\Delta})\,\Delta + \sigma^\top \sqrt{\Delta}\,\xi_j, \quad j = 1,\ldots,J,$$

with $\xi_j \sim \mathcal{N}(0, I_d)$.

Examples

McKean-Vlasov Equation

Suppose that we can sample from $X_{\Delta,(j-1)\Delta}$, then we can compute

$$\mathrm{E}\left[F\left(X_{\Delta,j\Delta}\right)\right] = \mathrm{E}\left[f\left(\mathrm{E}\left[g(X_{\Delta,(j-1)\Delta}, X'_{\Delta,(j-1)\Delta})\middle| X_{\Delta,(j-1)\Delta}\right]\right)\right],$$

where $g(X,X') = X + a(X,X')\Delta$ and $f(y) = \mathrm{E}_{\xi}\left[F\left(y + \sigma^{\top}\sqrt{\Delta}\,\xi\right)\right]$.

Remark

In general, we need to take into account an error in the distribution of $X_{\Delta,(j-1)\Delta}$ originating from approximations in the previous steps.

Examples

Optimal stopping

Consider a discrete time optimal stopping problem

$$V_j(x) = \sup_{\tau \in \mathscr{T}[j,\dots,T]} \mathrm{E}\left[G_\tau(X_\tau) \middle| X_j = x \right],$$

where $(X_j)_{j=0}^T$ is a d-dimensional Markov chain and $G_j : \mathbb{R}^d \to \mathbb{R}$ and $\mathscr{T}[j,\dots,T]$ is a set of stopping times with values in $\{j,\dots,T\}$.

Dynamic programming principle

The following relations hold for $j = 1, \dots, T-1$

$$C_j(x) = \mathrm{E}\left[\max\{G_{j+1}(X_{j+1}), C_{j+1}(X_{j+1})\} \middle| X_j = x \right],$$

where $C_j(x) = \mathrm{E}[V_{j+1}(X_{j+1}) | X_j = x]$ with $C_0 \equiv 0$ by definition.

Examples

Optimal stopping

Consider a simplest two-step stopping problem with $C_2(x) \equiv 0$ and $X_0 = x_0$. In this case

$$
\begin{aligned}
C_1(x) &= \mathrm{E}[\, G_2(X_2) | X_1 = x] \\
C_0(x_0) &= \mathrm{E}[\max\{G_1(X_1), C_1(X_1)\} | X_0 = x_0] \\
&= \mathrm{E}_{x_0}[f(\mathrm{E}[g(X_1, X_2) | X_1])]
\end{aligned}
$$

with $g(X_1, X_2) = (G_1(X_1), G_2(X_2))^\top$ and $f(x, y) = \max\{x, y\}$.

Examples

Dual approach for optimal stopping

It holds

$$
\begin{aligned}
V_j(X_j) &= \inf_{M \in \mathscr{M}} \mathrm{E}\left[\max_{t=j,\ldots,T}(G_t(X_t) - M_t + M_j)\right] \\
&= \mathrm{E}\left[\max_{t=j,\ldots,T}(G_t(X_t) - M_t^* + M_j^*)\right]
\end{aligned}
$$

with

$$
M_j^* := \sum_{i=1}^{j} (V_i(X_i) - \mathrm{E}[V_i(X_i)|X_{i-1}]).
$$

Examples

Dual approach for optimal stopping

Suppose that some approximation $\hat{V}$ of the value process V is available, then one construct an upper bound $\tilde{V}$ for V via

$$\tilde{V}_j = \mathrm{E}\left[\max_{t=j,\dots,T}(G_t(X_t) - M_t + M_j)\right],$$

where

$$M_j = \sum_{i=1}^{j}(\hat{V}_i(X_i) - \mathrm{E}[\hat{V}_i(X_i)|X_{i-1}]).$$

Nested Monte Carlo approach

Idea

Suppose we want to compute

$$F = \mathrm{E}\left[f\left(\mathrm{E}\left[g(X,Y)|X\right]\right)\right].$$

Approximate

$$F_{N,K} := \frac{1}{N}\sum_{n=1}^{N} f\left(\frac{1}{K}\sum_{k=1}^{K} g(X^{(n)}, Y^{(n,k)})\right),$$

where $\left(Y^{(n,k)}, k = 1, \ldots, K\right)$ is a sample from the conditional distribution of Y given $X = X^{(n)}$.

Nested Monte Carlo approach

Error estimates

If the function f is Lipschitz continuous on $\mathbb{R}^m$, we derive by the Jensen inequality

$$\mathrm{E}\left[\left|F_{N,K} - F\right|^2\right] \leq \underbrace{\frac{L_f^2}{K}\mathrm{E}_X\left\{\sum_{l=1}^{m} \mathrm{Var}\left[g_l(X,Y)|X\right]\right\}}_{Bias^2} + \underbrace{\frac{1}{N}\mathrm{Var}_X\left[f\left(\mathrm{E}\left[g(X,Y)|X\right]\right)\right]}_{Variance}$$

Observation

Note that the variances $\mathrm{Var}\left[g_l(X,Y)|X\right]$ appear in the bias of $F_{N,K}$, a common feature of many nonlinear Monte Carlo problems.

Nested Monte Carlo approach

Complexity

In order to get $\mathrm{E}\left[\left|F_{N,K}-F\right|^2\right] \leq \varepsilon^2$, we can take

$$\begin{aligned} K &= 2L_f^2\varepsilon^{-2}\mathrm{E}_X\left\{\sum_{l=1}^{m}\mathrm{Var}\left[g_l(X,Y)|X\right]\right\}, \\ N &= 2\varepsilon^{-2}\mathrm{Var}_X\left[f\left(\mathrm{E}\left[g(X,Y)|X\right]\right)\right], \end{aligned}$$

yielding the complexity of order

$$\mathscr{C}_{NMC} = KN = O\left(\varepsilon^{-4}\right).$$

Question

Can this complexity order be improved ?

Nested Monte Carlo approach

Observation

In order to reduce the bias of the estimate $F_{N,K}$, we need to reduce the variances $\mathrm{Var}[g_l(X,Y)|X]$.

Separation assumption

To simplify the analysis assume that

$$Y = \Phi(X,\xi),$$

where the random vector $\xi \in \mathbb{R}^p$ is independent of X. This assumption can be verified in many SDE related applications including the case of McKean-Vlasov Equations. Under this assumption, the original problem becomes

$$F = \mathrm{E}\left[f\left(\mathrm{E}_\xi\left[\tilde{g}(X,\xi)\right]\right)\right]$$

with $\tilde{g}(X,\xi) = g(X,\Phi(X,\xi))$.

Nested Monte Carlo approach

Projection based variance reduction

Suppose, for simplicity, that $m = 1$. Let ϕ_k, $k = 0, 1, \ldots$ with $\phi_0 \equiv 1$ be a complete orthonormal system in $L_2(\mathrm{P}_\xi)$, i.e.,

$$\mathrm{E}[\phi_k(\xi)\phi_l(\xi)] = \delta_{kl},$$

then it holds

$$\tilde{g}(x,\xi) = \mathrm{E}[\tilde{g}(x,\xi)] + \sum_{k=1}^{\infty} a_k(x)\phi_k(\xi),$$

where $a_k(x) := \mathrm{E}[\tilde{g}(x,\xi)\phi_k(\xi)]$, provided $\mathrm{E}\left[(\tilde{g}(x,\xi))^2\right] < \infty$ for any $x \in \mathbb{R}^d$.

Nested Monte Carlo approach

Projection based variance reduction

Suppose that for some $\beta > 0$

$$\sum_{k=1}^{\infty} k^{\beta} \mathrm{E}[a_k^2(X)] \leq C_a, \tag{1}$$

then the control variate $M_L(x,\xi) := \sum_{k=1}^{L} a_k(x)\phi_k(\xi)$ satisfies

$$\mathrm{E}_X \mathrm{Var}[\tilde{g}(X,\xi) - M_L(X,\xi)] \leq C_a L^{-\beta}.$$

The assumption (1) means some kind of smoothness of $\tilde{g}(\cdot,x)$.

Nested Monte Carlo approach

Example

Suppose that $p = 1$ and $\xi \sim \mathcal{N}(0,1)$, then we can take $\phi_k = H_k$ for $k \in \mathbb{N}_0$, where H_k stands for the (normalised) k-th Hermite polynomial, i.e.

$$H_k(x) \doteq \frac{(-1)^k}{\sqrt{k!}} e^{\frac{x^2}{2}} \frac{d^k}{dx^k} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$

We require that for any fixed $x > 0$, $\tilde{g}(x,s)$ admits derivatives up to order $\beta \in \mathbb{N}$ which satisfy

$$\int s^{2(\beta-\ell)} \mathrm{E}\left[\tilde{g}_s^{(\ell)}(X,s)\right]^2 ds \leq C, \quad \ell = 0, \ldots, \beta - 1$$

for some constant $C > 0$.

Discussion

In SDE applications p can be large, but one can significantly reduce the number of basis functions using the structure of the underlying discretisation scheme, see Belomestny et al, 2016.

Nested Monte Carlo approach

Estimation of coefficients

Each coefficient a_k in a fixed point x can be estimated via

$$a_{k,n}(x) := \frac{1}{n}\sum_{j=1}^{n} \tilde{g}(x, \xi^{(j)}).$$

Set $M_{L,n}(x,\xi) := \sum_{k=1}^{L} a_{k,n}(x)\phi_k(\xi)$, then by the Jensen's inequality

$$\begin{aligned} \mathrm{E}_X\left\{\mathrm{Var}[\tilde{g}(X,\xi) - M_{L,n}(X,\xi)]\right\} &= \mathrm{E}_X\left\{\mathrm{Var}[\tilde{g}(X,\xi) - M_L(X,\xi)]\right\} \\ &\quad + \mathrm{E}\left[\left|M_L(X,\xi) - M_{L,n}(X,\xi)\right|^2\right] \\ &\leq C_a L^{-\beta} + \sqrt{C_a}\frac{L}{n}. \end{aligned}$$

Nested Monte Carlo approach

Variance reduction

A new variance reduced nested Monte Carlo estimate

$$F_{N,K,L,n} = \frac{1}{N}\sum_{i=1}^{N} f\left(\frac{1}{K}\sum_{k=1}^{K}\left\{\tilde{g}(X^{(i)},\xi^{(k)}) - M_{L,n}(X^{(i)},\xi^{(k)})\right\}\right)$$

has MSE error of the form

$$\begin{aligned}\mathrm{E}\left[\left|F_{N,K,L}-F\right|^2\right] \quad &\leq \quad \frac{L_f^2}{K}\left(C_a L^{-\beta} + \sqrt{C_a}\,\frac{L}{n}\right) \\ &\quad + \frac{1}{N}\,\mathrm{Var}_X\left[f\left(\mathrm{E}_\xi\left[\tilde{g}(X,\xi)\right]\right)\right]\end{aligned}$$

while the cost of computing $F_{N,K,L,n}$ is of order $O(NnL + NKL)$. The resulting complexity of $F_{N,K,L,n}$ can be bounded as

$$\mathscr{C}_{VRNMC}(\varepsilon) = O\left(\varepsilon^{-\frac{3\beta}{\beta+1/2}}\right).$$

Nested Monte Carlo approach

Question

Can we further improve the complexity ?

Multilevel Monte Carlo

Set

$$U_{K,L,n}(X) := \frac{1}{K}\sum_{k=1}^{K}\left\{\tilde{g}(X,\xi^{(k)}) - M_{L,n}(X,\xi^{(k)})\right\}$$

and define a MLMC estimates $F_{\mathbf{N},\mathbf{K},\mathbf{L},\mathbf{n}}$ via

$$\frac{1}{N_0}\sum_{i=1}^{N_0} U_{K_0,L_0,n_0}(X^{(i)}) + \sum_{r=1}^{R}\frac{1}{N_r}\sum_{i=1}^{N_r}\left\{U_{K_r,L_r,n_r}(X^{(i)}) - U_{K_{r-1},L_{r-1},n_{r-1}}(X^{(i)})\right\},$$

where $\mathbf{N},\mathbf{K},\mathbf{L},\mathbf{n} \in \mathbb{R}^{R+1}$.

Nested Monte Carlo approach

Complexity

Using the estimate

$$\mathrm{E}_X\left\{\mathrm{Var}[U_{K,L,n}(X)|X]\right\} \quad \leq \quad \frac{1}{K}\left[C_a L^{-\beta} + \sqrt{C_a}\,\frac{L}{n}\right]$$

and the fact that the cost of computing $U_{K,L,n}(x)$ for a fixed x is of order $O(nL+KL)$, we derive

$$\mathscr{C}_{F_{\mathbf{N},\mathbf{K},\mathbf{L},\mathbf{n}}}(\varepsilon) \lesssim \begin{cases} \varepsilon^{-2}, & \beta > 1, \\ \varepsilon^{-2}\log^2(\varepsilon), & 0 \leq \beta \leq 1, \end{cases}$$

provided $\mathbf{N}, \mathbf{K}, \mathbf{L}, \mathbf{n}$ are chosen appropriately.

Observation

If $L = 1$, we recover the standard MLMC for nested simulations (see, Belomestny and Schoenmakers (2013), Lemaire and Pagés, (2016)).

Nested Monte Carlo approach

Formal complexity result

Let $Q = (f, \tilde{g}, \xi, X) \in \mathscr{G}(\beta, C_a, L_f)$ for some $\beta > 1$, $C_a, L_f > 0$, where $\mathscr{G}(\beta, C_a, L_f)$ is a class of separable nested models such that

$$\sum_{k=1}^{\infty} k^{\beta}\, \mathrm{E}[a_k^2(X)] \le C_a \text{ with } a_k(x) = \mathrm{E}[\tilde{g}(x, \xi)\phi_k(\xi)]$$

and

$$|f(x) - f(y)| \le L_f \|x - y\|, \quad x, y \in \mathbb{R}^d.$$

Then

$$A\varepsilon^{-2} \le \sup_{Q \in \mathscr{G}(\beta, C_a, L_f)} \inf_{\hat{F}} \left\{ \mathrm{Cost}(\hat{F}) : \mathrm{E}_Q[|\hat{F} - F|^2] \le \varepsilon^2 \right\} \le B\varepsilon^{-2},$$

where infimum is taken over the set of all measurable functions of the finite samples from the distributions P_{ξ} and P_X, and the constant s A and B depend on C_a and L_f only.

Nested Monte Carlo approach

Further complexity reduction

Let ψ_k, $k = 0, 1, \ldots$ with $\psi_0 \equiv 1$ be a complete orthonormal system in $L_2(\mathrm{P}_X)$, i.e.,

$$\mathrm{E}\left[\psi_k(X)\psi_l(X)\right] = \delta_{kl},$$

then it holds

$$H(X) = \mathrm{E}[H(X)] + \sum_{k=1}^{\infty} b_k \psi_k(X),$$

where $b_k(x) := \mathrm{E}\left[H(X)\phi_k(X)\right]$, provided $\mathrm{E}\left[H^2(X)\right] < \infty$.

Nested Monte Carlo approach

Proposition

Define a new outer control variate via

$$M_J(X) := \sum_{j=1}^{J} b_{K,L,n,j}\, \psi_j(X),$$

where

$$b_j(x) := \mathrm{E}\left[f\left(\mathrm{E}_\xi\left[\tilde{g}(X,\xi)\right]\right)\phi_j(X)\right], \quad j = 1, \ldots, J.$$

If the function $x \mapsto f\left(\mathrm{E}_\xi\left[\tilde{g}(x,\xi)\right]\right)$ is smooth, then the estimate

$$F_{N,K,L,n,J} = \frac{1}{N} \sum_{i=1}^{N} \left[U_{K,L,n}(X^{(i)}) - M_J(X^{(i)})\right]$$

has (under a proper choice of K, L, n, J) the complexity order of $\varepsilon^{-2+\delta}$ for some $\delta \in [0, 0.5)$.

Regression approach

We approximate

$$G(x) = \mathrm{E}[g(X,Y) | X = x] \approx \sum_{j=0}^{K} a_j \psi_j(x).$$

The coefficients (a_j), $j = 1, \ldots, K$, can be estimated based on the data $D_n = (X_i, Y_i)_{i=1}^n$, where $(X_i, Y_i)_{i=1}^n$ is an i.i.d. sample from the distibution (X, Y). Define an estimate

$$(a_{0,n}, \ldots, a_{K,n}) = \operatorname*{argmin}_{a_0,\ldots,a_K} \sum_{i=1}^{n} \left(g(X_i, Y_i) - \sum_{j=0}^{K} a_j \psi_j(X_i) \right)^2$$

and set

$$G_{K,n}(x) = \sum_{j=0}^{K} a_{j,n} \psi_j(x).$$

Regression approach

Now we estimate the quantity F via

$$F_{N,K,n} = \frac{1}{N}\sum_{j=1}^{N} f\left(G_{K,n}(X^{(j)})\right),$$

where $X^{(1)},\ldots,X^{(N)}$ is an iid sample from P_X.

Convergence

Suppose that f is Lipschitz continuous, then it holds

$$\mathrm{E}\left[\left|F_{N,K,n} - F\right|^2\right] \leq L_f^2\left[\mathrm{E}\left|G_{K,n}(X) - G(X)\right|^2\right] + \frac{1}{N}\mathrm{Var}\left[f\left(G_{K,n}(X)\right)\right].$$

Regression approach

Convergence

Suppose that

$$\sigma^2 = \sup_x \mathrm{Var}\left[g(X,Y)|X=x\right] < \infty$$

and

$$\|G\|_\infty \leq M,$$

then

$$\begin{aligned} \mathrm{E}\left|\widehat{G}_{K,n}(X) - G(X)\right|^2 &\leq c\max\left\{\sigma^2, M\right\}\frac{(\log(n)+1)\cdot K}{n} \\ &\quad +8\inf_{\Psi\in\mathrm{Span}(\psi_0,\ldots,\psi_K)}\mathrm{E}\left|\Psi(X)-G(X)\right|^2. \end{aligned}$$

Regression estimate

Cost of regression

The cost of constructing the least-squares estimate $G_{K,n}(x)$ for one fixed x is of order nK^2, so that the overall computational cost of the regression-based MC approach is proportional to NnK^2.

Complexity

Set

$$\rho_K := \inf_{\Psi \in \mathrm{Span}(\psi_0,\ldots,\psi_K)} \mathrm{E}\left|\Psi(X) - G(X)\right|^2$$

then the complexity of the estimate $G_{K,n}(x)$ is given by

$$\mathscr{C}_{RMC}(\varepsilon) \lesssim \varepsilon^{-3}\rho_K^{-}(\varepsilon/\sqrt{3}).$$

By assuming $\rho_K = K^{-\alpha} l(K)$ for some $\alpha > 0$ and some slow varying function l, we derive $\mathscr{C}_{RMC}(\varepsilon) \lesssim \varepsilon^{-3-1/\alpha}$.

Discussion

- While RMC requires a rather strong uniform bound (in x) for the variance $\mathrm{Var}[g(X,Y)|X=x]$ and for the function $G(x)$, VRMC works under weaker assumptions (in the case of normal distribution)

$$\int \mathrm{E}\left[\partial_s \tilde{g}(X,s)\right]^2 ds < \infty.$$

- Any reduction of the variance $\mathrm{Var}[g(X,Y)|X=x]$ will have no effect on the complexity of the regression estimate because of the term $\max\left\{\sigma^2, M\right\}$. This is intrinsic problem of the global regression !