Operations Research Letters 45 (2017) 63–67

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

Computing the variance of a conditional expectation via non-nested Monte Carlo

Takashi Goda

Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

ARTICLE INFO

Article history:
Received 18 May 2016
Received in revised form
27 September 2016
Accepted 1 December 2016
Available online 10 December 2016

Keywords:
Variance of a conditional expectation
Monte Carlo estimation
Pick-freeze scheme
Bias correction

ABSTRACT

Computing the variance of a conditional expectation has often been of importance in uncertainty quantification. Sun et al. has introduced an unbiased nested Monte Carlo estimator, which they call $1\frac{1}{2}$-level simulation since the optimal inner-level sample size is bounded as the computational budget increases. In this letter, we construct unbiased non-nested Monte Carlo estimators based on the so-called pick-freeze scheme due to Sobol'. An extension of our approach to compute higher order moments of a conditional expectation is also discussed.

1. Introduction

Let X be a random variable with probability density function p_X defined on Ω_X, and let $f\colon \Omega_X \to \mathbb{R}$ be a function. For another random variable Y which is correlated with X, we are interested in computing the *variance of a conditional expectation*

$$\mathrm{Var}_Y\left[\mathbb{E}_{X|Y}[f]\right] := \int_{\Omega_Y}\left(\int_{\Omega_X} f(x)p_{X|Y=y}(x)\,\mathrm{d}x-\mu\right)^2 p_Y(y)\,\mathrm{d}y \quad (1)$$

$$= \int_{\Omega_Y}\left(\int_{\Omega_X} f(x)p_{X|Y=y}(x)\,\mathrm{d}x\right)^2 p_Y(y)\,\mathrm{d}y-\mu^2, \quad (2)$$

where p_Y and $p_{X|Y=y}$ denote the probability density function (defined on Ω_Y) of Y and the conditional probability density function of X given $Y=y$, respectively, and further μ is defined by

$$\mu := \int_{\Omega_Y}\int_{\Omega_X} f(x)p_{X|Y=y}(x)p_Y(y)\,\mathrm{d}x\,\mathrm{d}y = \int_{\Omega_X} f(x)p_X(x)\,\mathrm{d}x.$$

It follows from the well-known variance decomposition formula that

$$\mathrm{Var}_Y\left[\mathbb{E}_{X|Y}[f]\right] = \mathrm{Var}_X[f]-\mathbb{E}_Y\left[\mathrm{Var}_{X|Y}[f]\right], \quad (3)$$

where each term on the right-hand side is defined similarly.

The quantity $\mathrm{Var}_Y\left[\mathbb{E}_{X|Y}[f]\right]$ has been used in the area of uncertainty quantification. For instance, in [11], Zouaoui and Wilson used $\mathrm{Var}_Y\left[\mathbb{E}_{X|Y}[f]\right]$ as a quality measure of uncertainty on the mean time in a single-server queueing system due to uncertainty on the parameters. Another usage of $\mathrm{Var}_Y\left[\mathbb{E}_{X|Y}[f]\right]$ which we have in mind is as follows: Assume that X denotes a set of uncertain simulation inputs and Y does a sample observation data. In the absence of the data Y, the prior variance $\mathrm{Var}_X[f]$ measures the uncertainty of a simulation output f. If the data Y is available, the uncertainty of f after knowing $Y=y$ is represented by the posterior variance $\mathrm{Var}_{X|Y=y}[f]$. Thus, the uncertainty of f is expected to be reduced to $\mathbb{E}_Y\left[\mathrm{Var}_{X|Y}[f]\right]$ by obtaining the data Y. It can be seen from the identity (3) that $\mathrm{Var}_Y\left[\mathbb{E}_{X|Y}[f]\right]$ quantifies how much the uncertainty of f can be reduced before and after obtaining the data Y.

By the definition (1), it seems natural to use a nested Monte Carlo estimator for $\mathrm{Var}_Y\left[\mathbb{E}_{X|Y}[f]\right]$. Recently, in [10], Sun et al. have introduced the following unbiased nested Monte Carlo estimator: For positive integers K and $n_1,\ldots,n_K$, let $y_1,\ldots,y_K$ be sampled independently from p_Y, and for $k=1,\ldots,K$, let $x_{1k},\ldots,x_{n_k k}$ be sampled independently but conditionally from $p_{X|Y=y_k}$. Moreover, let $C=n_1+\cdots+n_K$,

$$\bar{f}_k=\frac{1}{n_k}\sum_{j=1}^{n_k} f(x_{jk}) \quad\text{and}\quad \bar{\bar{f}}=\frac{1}{C}\sum_{k=1}^{K}\sum_{j=1}^{n_k} f(x_{jk}).$$

Then, the quantity of interest $\mathrm{Var}_Y\left[\mathbb{E}_{X|Y}[f]\right]$ is estimated by

$$W := \frac{SS_\tau-\frac{K-1}{C-K}SS_\epsilon}{C-\sum_{k=1}^{K} n_k^2/C}, \quad (4)$$

E-mail address: goda@frcer.t.u-tokyo.ac.jp.

http://dx.doi.org/10.1016/j.orl.2016.12.002
0167-6377/

where

$$SS_\tau = \sum_{k=1}^{K} n_k(\bar{f}_k - \bar{\bar{f}})^2 \quad \text{and} \quad SS_\epsilon = \sum_{k=1}^{K}\sum_{j=1}^{n_k}(f(x_{jk}) - \bar{f}_k)^2.$$

The interesting property of the estimator W is that the optimal inner-level sample sizes $n_1, \ldots, n_K$ remain bounded above as the total computational budget C increases. That is, there is no need to increase both the inner- and outer-level sample sizes simultaneously for making the approximation error converge to zero, and thus, it can be inferred from [10, Equation (10)] that the estimator W achieves the Monte Carlo root mean square error (RMSE) of order $C^{-1/2}$ asymptotically. Because of this nice property, Sun et al. have referred to their estimator as $1\frac{1}{2}$-level simulation. However, it may require a precomputation step to choose proper inner-level sample sizes depending on a problem at hand and a given cost C.

In this letter, by assuming that i.i.d. samplings from p_Y and $p_{X|Y=y}$ for any $y \in \Omega_Y$ are possible, we construct several unbiased *non-nested* Monte Carlo estimators for $\mathrm{Var}_Y\left[\mathbb{E}_{X|Y}[f]\right]$. We also show that our approach can be extended in a straightforward way to compute higher order moments of a conditional expectation. We note that our assumption is same as that considered in [10]. Since our estimators are no longer of the nested form, we do not need to take care of a proper choice of inner-level sample sizes, and our estimators are naturally expected to achieve the Monte Carlo RMSE of order $C^{-1/2}$. Our idea for constructing non-nested estimators stems from the pick-freeze scheme due to Sobol' [7,8], which was originally introduced for computing variance-based sensitivity indices and has been thoroughly studied in the context of global sensitivity analysis by Saltelli [6], Owen [4], Janon et al. [1], and Owen et al. [5] to list just a few. In fact, in that context, the quantity $\mathrm{Var}_Y\left[\mathbb{E}_{X|Y}[f]\right]$ corresponds to the so-called first order sensitivity index, if Y denotes a subset of uncertain simulation inputs contained in X. Thus, our result of this letter can be regarded as a generalization of the known results on variance-based sensitivity analysis.

The remainder of this letter is organized as follows. In the next section, we introduce four straightforward non-nested Monte Carlo estimators; one based on Mauntz [3] and Kucherenko et al. [2] is unbiased whereas the other three essentially based on Janon et al. [1] is biased. In the third section, we give bias corrections of the latter estimators. In the fourth section, we discuss an extension of our approach to compute higher order moments. We conclude this letter with numerical experiments in the last section.

2. Non-nested Monte Carlo

The key ingredient of the pick-freeze scheme lies in how to deal with the square appearing in the first and second terms of (2). It is easy to see from Fubini's theorem that we can rewrite the first term of (2) into

$$\begin{aligned}\int_{\Omega_Y}&\left(\int_{\Omega_X} f(x)p_{X|Y=y}(x)\,\mathrm{d}x\right)\\ &\times\left(\int_{\Omega_X} f(x')p_{X|Y=y}(x')\,\mathrm{d}x'\right)p_Y(y)\,\mathrm{d}y\\ =&\int_{\Omega_Y}\int_{\Omega_X}\int_{\Omega_X} f(x)f(x')p_{X|Y=y}\\ &\times (x)p_{X|Y=y}(x')p_Y(y)\,\mathrm{d}x\,\mathrm{d}x'\,\mathrm{d}y. \qquad (5)\end{aligned}$$

The second term of (2), i.e., the squared expectation μ^2, can be simply rewritten into

$$\begin{aligned}\mu^2 &= \left(\int_{\Omega_Y}\int_{\Omega_X} f(x)p_{X|Y=y}(x)p_Y(y)\,\mathrm{d}x\,\mathrm{d}y\right)\\ &\quad\times\left(\int_{\Omega_X} f(x'')p_X(x'')\,\mathrm{d}x''\right)\\ &=\int_{\Omega_Y}\int_{\Omega_X}\int_{\Omega_X} f(x)f(x'')p_{X|Y=y}\\ &\quad\times (x)p_X(x'')p_Y(y)\,\mathrm{d}x\,\mathrm{d}x''\,\mathrm{d}y,\end{aligned}$$

where we used Fubini's theorem in the second equality. Thus, the quantity $\mathrm{Var}_Y\left[\mathbb{E}_{X|Y}[f]\right]$ is given by

$$\begin{aligned}\mathrm{Var}_Y\left[\mathbb{E}_{X|Y}[f]\right] &= \int_{\Omega_Y}\int_{\Omega_X}\int_{\Omega_X}\int_{\Omega_X} f(x)\left(f(x')-f(x'')\right)\\ &\times p_{X|Y=y}(x)p_{X|Y=y}(x')p_X(x'')p_Y(y)\,\mathrm{d}x\,\mathrm{d}x'\,\mathrm{d}x''\,\mathrm{d}y.\end{aligned}$$

Therefore, our first Monte Carlo estimator, which has some similarity to that in [3,2] for variance-based sensitivity analysis, can be constructed as

$$U := \frac{1}{N}\sum_{n=1}^{N} f(x_n)\left(f(x'_n)-f(x''_n)\right),$$

where, for each n, we first sample y_n randomly from p_Y and then sample x_n and x'_n independently and randomly from $p_{X|Y=y_n}$. Further, we sample x''_n randomly from p_X, or we first sample y''_n randomly from p_Y (independently of y_n) and then sample x''_n randomly from $p_{X|Y=y''_n}$. It is obvious that $\mathbb{E}[U] = \mathrm{Var}_Y\left[\mathbb{E}_{X|Y}[f]\right]$, meaning that the estimator U is unbiased.

Since the estimator U requires three function evaluations for each n, the total computational budget C equals $3N$. It is further possible to construct Monte Carlo estimators which require two function evaluations for each n, i.e., $C = 2N$. Let us consider an approximation of μ instead of μ^2. This can be done by using the samples x_n's and x'_n's commonly as either

$$\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} f(x_n) \quad \text{or} \quad \hat{\mu}' = \frac{1}{N}\sum_{n=1}^{N} f(x'_n) \quad \text{or} \quad \frac{\hat{\mu}+\hat{\mu}'}{2}.$$

Using these estimators for μ, we can introduce the following Monte Carlo estimators for $\mathrm{Var}_Y\left[\mathbb{E}_{X|Y}[f]\right]$:

$$V_1 := \frac{1}{N}\sum_{n=1}^{N} f(x_n)f(x'_n) - \hat{\mu}^2,$$

$$\begin{aligned}V_2 &:= \frac{1}{N}\sum_{n=1}^{N} f(x_n)f(x'_n) - \left(\frac{\hat{\mu}+\hat{\mu}'}{2}\right)^2\\ &= \frac{1}{N}\sum_{n=1}^{N}\left(f(x_n)-\frac{\hat{\mu}+\hat{\mu}'}{2}\right)\left(f(x'_n)-\frac{\hat{\mu}+\hat{\mu}'}{2}\right),\end{aligned}$$

$$\begin{aligned}V_3 &:= \frac{1}{N}\sum_{n=1}^{N} f(x_n)f(x'_n) - \hat{\mu}\hat{\mu}'\\ &= \frac{1}{N}\sum_{n=1}^{N}\left(f(x_n)-\hat{\mu}\right)\left(f(x'_n)-\hat{\mu}'\right).\end{aligned}$$

Note that the last two estimators are exactly of the same forms as those in [1] for variance-based sensitivity analysis. These estimators are actually the special cases of a generalized estimator

$$V := \frac{1}{N}\sum_{n=1}^{N} f(x_n)f(x'_n) - \left(w_1\hat{\mu}^2 + w_2\hat{\mu}'^2 + w_3\hat{\mu}\hat{\mu}'\right), \qquad (6)$$