Title
Improving the Normalized Importance Sampling Estimator

Authors
Ghamami, Samim
Ross, Sheldon M.

Publication Date
2012-07-01

COLEMAN FUNG
RISK MANAGEMENT
RESEARCH CENTER

Working Paper # 2012 - 08

Improving the Normalized Importance Sampling Estimator

Samim Ghamami,University of Southern California,Los Angeles
Sheldon M. Ross,University of Southern California,Los Angeles

July, 2012

University of California Berkeley

Improving the Normalized Importance Sampling Estimator *

Samim Ghamami and Sheldon M. Ross

Department of Industrial and Systems Engineering
University of Southern California, Los Angeles, CA 90089.
ghamami@usc.edu
smross@usc.edu

Abstract

The normalized importance sampling estimator allows the target density f to be known only up to a multiplicative constant. We indicate how it can be derived by a delta method based approximation of a Rao-Blackwellized acceptance rejection estimator. Using additional terms in the delta method then results on a new estimator that also only requires f to be known only up to a multiplicative constant. Numerical examples indicate that the new estimator usually outperforms the normalized importance sampling estimator in terms of mean square error.

1 Introduction

Consider the problem of estimating

$$\theta = E[h(X)] = \int h(x)f(x)dx,$$

where X is a random element of R^d with probability density f and h is a function from R^d to R, with the $n-$simulation-run importance sampling estimator,

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} h(X_i)\frac{f(X_i)}{g(X_i)}$$

where $X_1, ..., X_n$ are drawn from g, the importance sampling density, which is any other probability density on R^d satisfying $f(x) > 0 \Rightarrow g(x) > 0$ for all $x \in R^d$, (see [3] or [1] for more on importance sampling). The normalized importance sampling estimator, which divides the standard importance sampling estimator by the Monte Carlo average of the likelihood ratios,

*This material is based upon work supported by the U. S. Army Research Laboratory and the U. S. Army Research Office under grant number W911NF-11-1-0115.

$$\hat{\theta}_2 = \frac{\sum_{i=1}^{n} h(X_i)\frac{f(X_i)}{g(X_i)}}{\sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)}}$$

is a well-known alternative of the standard importance sampling estimator whose advantage over $\hat{\theta}_1$ is that in using $\hat{\theta}_2$ one needs to know the target density, $f(x)$, only up to a multiplicative constant, a practical constraint arising quite often in sequential importance sampling, Markov chain Monte Carlo, and Bayesian statistics, (see [4]).

In this paper (section 2.1), we show how both $\hat{\theta}_2$, and our new estimator

$$\hat{\theta}_3 = \frac{\sum_{i=1}^{n} h(X_i)\frac{f(X_i)}{g(X_i)}}{\sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)}} + \frac{\sum_{i=1}^{n} h(X_i)\frac{f^2(X_i)}{g^2(X_i)}}{(\sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)})^2} - \frac{(\sum_{i=1}^{n} h(X_i)\frac{f(X_i)}{g(X_i)})(\sum_{i=1}^{n} \frac{f^2(X_i)}{g^2(X_i)})}{(\sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)})^3} \tag{1}$$

can be derived by approximating a conditional expectation of an acceptance-rejection sampling estimator of θ. Numerical evidence indicates that our new estimator usually improves upon $\hat{\theta}_2$ in terms of mean square error.

2 Obtaining The Estimators by Approximating a Conditional Expectation of an Acceptance-Rejection Based Estimator

The acceptance-rejection method generates the value of a random variable having density function $f(x)$ by first generating the value of a random variable having density function $g(x)$. If the generated value is x, then that value is accepted with probability $\frac{f(x)}{Cg(x)}$, where C is such that $\frac{f(x)}{g(x)} \leq C$ for all x; if the value is not accepted then the process is repeated. Let $X_1, \ldots, X_n$ be the first n generated values using the density g, and consider the following estimator of $\theta = E_f[h(X)]$ based on these values.

$$\hat{\theta} = \frac{\sum_{i=1}^{n} h(X_i)\mathbf{1}\{U_i \leq \frac{f(X_i)}{Cg(X_i)}\}}{\sum_{i=1}^{n} \mathbf{1}\{U_i \leq \frac{f(X_i)}{Cg(X_i)}\}},$$

where U_i's are independent $[0, 1]$ uniform random variables. Now, suppose that we would like to use conditional expectation to derive a new estimator of θ which improves upon $\hat{\theta}$ by conditioning on the simulated values of $X_1, X_2, ..., X_n$ under g, i.e.,

$$E[\frac{\sum_{i=1}^{n} h(X_i)\mathbf{1}\{U_i \leq \frac{f(X_i)}{Cg(X_i)}\}}{\sum_{i=1}^{n} \mathbf{1}\{U_i \leq \frac{f(X_i)}{Cg(X_i)}\}}|X_1, X_2, ..., X_n]. \tag{2}$$

An approximation of (2) by the delta method using the first three terms in the Taylor series expansion of the function $f(X, Y) = \frac{X}{Y}$ around $f(E(X), E(Y))$ results in

$$\hat{\theta}_2 = \frac{E[\sum_{i=1}^n h(X_i)\mathbf{1}\{U_i \leq \frac{f(X_i)}{Cg(X_i)}\}|X_1, ..., X_2]}{E[\sum_{i=1}^n \mathbf{1}\{U_i \leq \frac{f(X_i)}{Cg(X_i)}\}|X_1, ..., X_n]} = \frac{\sum_{i=1}^n h(X_i)\frac{f(X_i)}{g(X_i)}}{\sum_{i=1}^n \frac{f(X_i)}{g(X_i)}},$$

the *normalized importance sampling* estimator. It is a biased but consistent estimator of θ whose MSE, compared to the classical importance sampling estimator, $\hat{\theta}_1$, has been discussed in section 2.5.3 of [4].

Considering additional terms in the delta method used to approximate (2) now yields the new estimator. More particularly, using the delta method with the first six terms in the Taylor series expansion of the function $f(X, Y) = \frac{X}{Y}$ around $f(E(X), E(Y))$ yields

$$E[\frac{X}{Y}] \approx \frac{E[X]}{E[Y]} + \frac{\text{Var}(Y)E[X]}{(E[Y])^3} - \frac{\text{cov(X,Y)}}{(E[Y])^2}, \tag{3}$$

The proof of the following Lemma is in the Appendix.

Lemma 1 *Approximating (2) by using the first six terms yields the estimator*

$$\hat{\theta}_3 = \frac{\sum_{i=1}^n h(X_i)\frac{f(X_i)}{g(X_i)}}{\sum_{i=1}^n \frac{f(X_i)}{g(X_i)}} + \frac{\sum_{i=1}^n h(X_i)\frac{f^2(X_i)}{g^2(X_i)}}{(\sum_{i=1}^n \frac{f(X_i)}{g(X_i)})^2} - \frac{(\sum_{i=1}^n h(X_i)\frac{f(X_i)}{g(X_i)})(\sum_{i=1}^n \frac{f^2(X_i)}{g^2(X_i)})}{(\sum_{i=1}^n \frac{f(X_i)}{g(X_i)})^3} \tag{4}$$

Remark: Note that $\hat{\theta}_3$ does not depend on C and also allows the target density to be known only up to a multiplicative constant. In addition, like the normalized importance sampling estimator, it is a biased but consistent estimator of θ.

2.1 A Numerical Example

In the following example we compare $\hat{\theta}_3$, with the normalized importance sampling estimator, $\hat{\theta}_2$, in terms of mean square error (MSE).

Example: Consider estimating $P(Y > x)$ where Y is a Weibull random variable with density $f(x) = \beta x^{\beta-1}e^{-x^\beta}$, g is the Weibull density $g(x) = \theta\beta x^{\beta-1}e^{-\theta x^\beta}$, where we use the result from [2] and set $\theta = 1/x^\beta$. Table 1 below shows the estimates of $\text{Var}(\hat{\theta}_i)$ and $\text{MSE}(\hat{\theta}_i),\ i = 2, 3$. These estimates are based on 1000 simulation runs.

MSE($\hat{\theta}_2$)	Var($\hat{\theta}_2$)	MSE($\hat{\theta}_3$)	Var($\hat{\theta}_3$)	β	x	$P(Y > x)$	n
6.592×10^{-4}	6.500×10^{-4}	4.397×10^{-4}	4.179×10^{-4}	.5	.1	.7289	5000
1.900×10^{-4}	1.891×10^{-4}	1.390×10^{-4}	1.386×10^{-4}	.3	.05	.6656	10^4
8.98×10^{-6}	8.95×10^{-6}	8.78×10^{-6}	8.762×10^{-6}	.2	20	.1619	10^4

Table 1: Comparing MSE of $\hat{\theta}_2$ and $\hat{\theta}_3$ for $\theta = P(Y > x)$ in the Weibull case

Appendix

Proof of Lemma 1

Set $X \equiv \sum_{i=1}^{n} h(X_i)\mathbf{1}\{U_i \leq \frac{f(X_i)}{Cg(X_i)}\}$, $Y \equiv \sum_{i=1}^{n} \mathbf{1}\{U_i \leq \frac{f(X_i)}{Cg(X_i)}\}$ and $Z \equiv (X_1, ..., X_n)$. Recall the following identity from section 2,

$$E[\frac{X}{Y}] \approx \frac{E[X]}{E[Y]} + \frac{\text{Var}(Y)E[X]}{(E[Y])^3} - \frac{\text{cov(X,Y)}}{(E[Y])^2},$$

And that we derive $\hat{\theta}_3$ by conditioning all the expectations on the right hand side of the above approximation on Z, i.e., we are to show that

$$\frac{E[X|Z]}{E[Y|Z]} + \frac{\text{Var}(Y|Z)E[X|Z]}{(E[Y|Z])^3} - \frac{\text{cov}(X,Y|Z)}{(E[Y|Z])^2}$$

is equal to $\hat{\theta}_3$, which is,

$$\hat{\theta}_3 = \frac{\sum_{i=1}^{n} h(X_i)\frac{f(X_i)}{g(X_i)}}{\sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)}} + \frac{\sum_{i=1}^{n} h(X_i)\frac{f^2(X_i)}{g^2(X_i)}}{(\sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)})^2} - \frac{(\sum_{i=1}^{n} h(X_i)\frac{f(X_i)}{g(X_i)})(\sum_{i=1}^{n} \frac{f^2(X_i)}{g^2(X_i)})}{(\sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)})^3}$$

It is then fairly straight forward to arrive at the above mentioned equality merely by noting that,

$$\text{Var}(Y|Z) = \sum_{i=1}^{n} \frac{f(X_i)}{Cg(X_i)} - \sum_{i=1}^{n} \frac{f^2(X_i)}{C^2g^2(X_i)} \quad , \quad \text{cov}(X,Y|Z) = \sum_{i=1}^{n} h(X_i)\frac{f(X_i)}{Cg(X_i)} - \sum_{i=1}^{n} h(X_i)\frac{f^2(X_i)}{C^2g^2(X_i)}$$

$$E[X|Z] = \sum_{i=1}^{n} h(X_i)\frac{f(X_i)}{Cg(X_i)} \quad , \quad (E[Y|Z])^j = \frac{1}{C^j}(\sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)})^j$$

where $j = 1, 2, 3$ are used in the derivation of $\hat{\theta}_3$.

References

[1] Asmussen, S., and Glynn, P. W., "Stochastic Simulation", Springer-Verlag, (2007)

[2] Asmussen, S., Kroese, D.P., Rubinstein, R.Y, "Heavy Tails, Importance Sampling and Cross-Entropy", *Stochastic Models* 21 (1), 57-76, (2005)

[3] Glasserman, P., "Monte Carlo Methods in Financial Engineering", Springer-Verlag, (2004)

[4] Liu, S. J., "Monte Carlo Strategies in Scientific Computing", Springer-Verlag, (2004)