Efficient Simulation of Light-Tailed Sums

Brennan Hall

UC Santa Barbara

March 19, 2019

Overview

Goal

We aim to present an importance sampling based methodology for rare-event simulation that provides an estimator with bounded relative error.

Rare Event Simulation

- Let $\{X_i\}_{i=1}^n$ be a sequence of iid d-dimensional random variables distributed under measure $\mathbf{P}$ and

$$S_n = X_1 + \ldots + X_n$$

- Of interest is estimating

$$\alpha_n = \mathbf{P}\left(S_n/n \in A\right) = \mathbf{E}[Y]$$

for $Y = \mathbb{1}_{\{S_n/n \in A\}}$ for a closed and convex set A.

- One issue for finding an estimator of α_n is determining whether its accuracy deteriorates as $\alpha_n \downarrow 0$.

Shortfalls of Standard Methods

- If the event $\{S_n/n \in A\}$ is rare under $\mathbf{P}$, the variance of the estimator

$$Q_K = (Y_1 + \ldots + Y_K)/K$$

is large compared to $\mathbf{E}[Q_K]$.

- Large number of Monte Carlo simulations are required to estimate α_M to a given relative accuracy, eg. coefficient of variation

$$\frac{\sigma/\sqrt{M}}{\alpha_n} = \frac{\sqrt{\alpha_n(1-\alpha_n)}}{M\sqrt{\alpha_n}} \approx \frac{1}{\sqrt{M\alpha_n}}$$

where $\alpha_n \downarrow 0$.

Importance Sampling

- Used to reduce variance of Q_K by a change of measure so that the number of MC simulations needed is reduced.
- Sample $\{\bar{Y}_i\}_{i=1}^{n}$ from the distributiuon defined by the Radon-Nikodym derivative, $\mathbf{f} = \frac{d\mathbf{P}}{d\mathbf{P}_\theta}$ for $\mathbf{P} \ll \mathbf{P}_\theta$.
- Instead of Q_K, consider

$$\bar{Q}_K = \frac{1}{K} \sum_{i=1}^{K} \bar{Y}_i \mathbf{f}(\bar{Y}_i)$$

Note: $\mathbf{f}$ is restricted to a parametrized family of alternative sampling distribtutions since some distributions are inappropriate to use.

Optimal Exponential Tiliting

- Let $F(\cdot) = \mathbf{P}(X \leq \cdot)$ and ψ be the log-MGF of X , then

$$\mathrm{d}F_\theta = \exp\left\{\theta x - \psi(\theta)\right\}\mathrm{d}F$$

  is said to be *exponentially tilted* by the parameter θ.

- $Z = \frac{\mathrm{d}F}{\mathrm{d}F_\theta}$ is an exponential martingale with $\mathbf{E}[Z] = 1$ guarantees existence of F_θ [Girsanov 1960].

Examples:

- $\mathcal{N}(\mu, \sigma^2) \to \mathcal{N}(\mu + \theta\sigma^2, \sigma^2)$
- $Gamma(\alpha, \beta) \to Gamma(\alpha, \beta - \theta)$

OET Efficiency

Definition

An estimator Q is logarithmically efficient if

$$\liminf_{n\uparrow\infty} \frac{\log \mathbf{E}\left[Q^2\right]}{\log \mathbf{E}\left[Q\right]^2} = 1$$

Proposition ([Asmussen, Glynn 2007])

OET is the only iid importance sampling algorithm that provides at least logarithmic efficiency.

Motivation for improving on OET

Consider in the $d = 1$-dimensional case

- Set $A = (\beta, \infty)$, $\beta > 0$ so $Y = \mathbb{1}_{\{S_n > n\beta\}}$.
- Under OET, we get LE for unbiased estimator $\bar{Q}_K$ with $\mathbf{Var}(\bar{Q}_K) \to 0$ but the squared coefficient of variation,

$$\frac{\mathbf{Var}(\bar{Q}_K)}{\mathbf{E}[\bar{Q}_K]^2} \to \infty, \quad \text{as } n \uparrow \infty$$

  unless K increases exponentially with n.
- OET adds bias to increments when S_n near $n\beta$.
- Use conditional knowledge of current state – Optimal State-Dependent Exponential Tilting – to obtain bounded relative error.

OSDET

Dynamically update OET at each step based on the current step.

- Update parameter θ_w to sample from F_{θ_w} through the Legendre transform

$$J(w) = \sup_{\theta \geq 0} \left\{ \theta w - \psi(\theta) \right\}$$

  with

$$J'(w) = \psi'^{-1}(w) = \theta_w$$

- $J(\cdot)$ is convex along $\mathbb{R}$ and is twice continuously differentiable (besides $w = 0$).
- Under certain conditions before random walk completes, apply OET for remaining time.

Algorithm

Set $w = \beta > n^{-1/2}$, $L = 1$, $s = 0$, $\bar{s} = 0$, $k = 0$, $\lambda > 2\beta$.
Repeat Step (1) until $n = k$ or $w \leq (n-k)^{-1/2}$ or $w > \lambda$.

1. Sample X from F_{θ_w}

$$\begin{aligned} L &\leftarrow \exp\{-\theta_w X + \psi(\theta_w)\} L \\ s &\leftarrow s + X \\ k &\leftarrow k + 1 \\ w &\leftarrow (n\beta - s)/(n-k) \end{aligned}$$

2. If $k < n$, sample $X_{k+1}, \ldots, X_n$ from F_{θ_w}

$$\begin{aligned} \bar{s} &\leftarrow X_{k+1} + \ldots + X_n \\ L &\leftarrow \exp\left\{-\theta_w \bar{s} + (n-k)\psi(\theta_w)\right\} L \end{aligned}$$

3. Output $Y_n = L \times \mathbb{1}_{\{s+\bar{s}>n\beta\}}$

Estimator

The algorithm provides an unbiased, BRE estimator:

$$Y_n = \mathbb{1}_{\{S_n > n\beta\}} \prod_{j=1}^{n} \exp\{-\theta_j X_j + \psi(\theta_j)\}$$

- Unbiased:

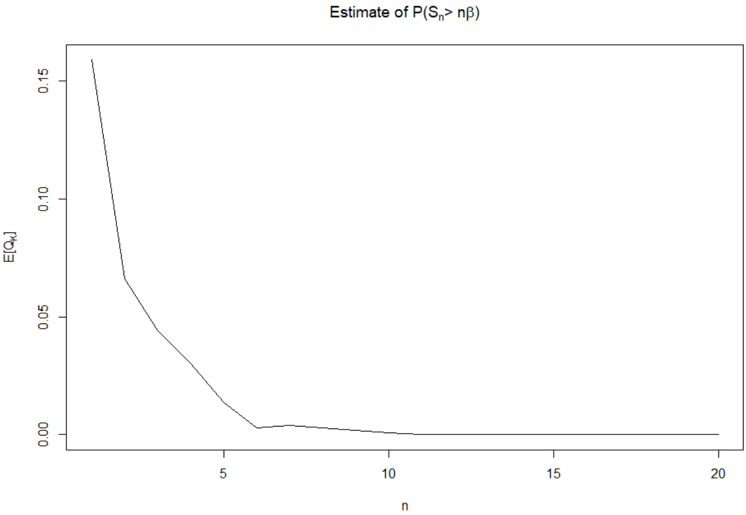
$$\begin{aligned}\mathbf{E}[Y_n] &= \mathbf{E}\left[\mathbb{1}_{\{S_n > n\beta\}} \prod_{j=1}^{n} \exp\{-\theta_j X_j + \psi(\theta_j)\}\right] \\ &= \prod_{j=1}^{n} \int_{\omega:\,\{S_n > n\beta\}} \frac{1}{Z_j} \mathrm{d}F_{\theta_j} = \mathbf{P}(S_n > n\beta)\end{aligned}$$

- BRE:

$$\sup_{n \geq 1} \frac{\tilde{\mathbf{E}}[Y_n^p]}{\mathbf{P}(S_n > n\beta)^p} < \infty$$

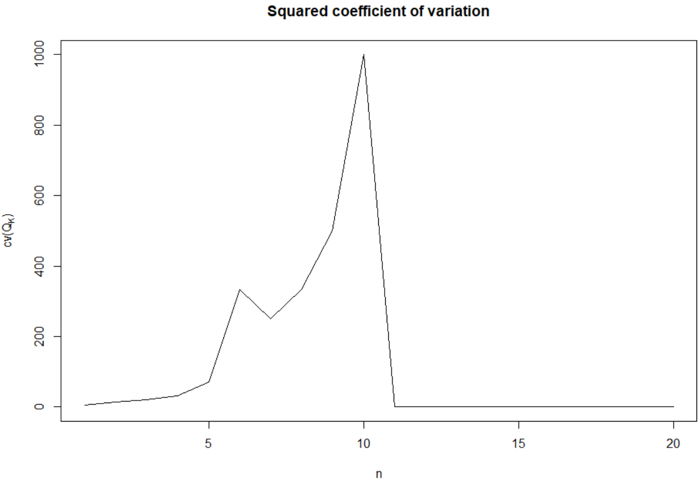
$\mathbf{P}(S_n > n\beta)$ with $X \sim \mathcal{N}(0,1)$, $\alpha_n \to 0$.

Fix number of MC trials, $K = 1000$, and $\beta = 1$.



$\mathbf{P}(S_n > n\beta)$ with $X \sim \mathcal{N}(0,1)$, $cv(Q_K)$ bounded

Fix number of MC trials, $K = 1000$, and $\beta = 1$.



References

Asmussen, S. and Glynn, P.W., 2007. Stochastic simulation: algorithms and analysis (Vol. 57). Springer Science & Business Media.

Blanchet, J.H., Leder, K. and Glynn, P.W., 2009. "Efficient Simulation of Light-Tailed Sums: an Old-Folk Song Sung to a Faster New Tune..." In Monte Carlo and Quasi-Monte Carlo Methods 2008 (pp. 227-248). Springer, Berlin, Heidelberg.

Girsanov, I.V., 1960. "On transforming a certain class of stochastic processes by absolutely continuous substitution of measures". Theory of Probability & Its Applications, 5(3), pp.285-301.

Dupuis, P. and Wang, H., 2004. "Importance sampling, large deviations, and differential games". Stochastics: An International Journal of Probability and Stochastic Processes, 76(6), pp.481-508.