Introduction to Rare Event Simulation for Processes with Light Tailed Increments

Thomas Dean

Signal Processing and Communications Laboratory
Department of Engineering
University of Cambridge
Email: tad36@cam.ac.uk

October 2009

Problem Description

Assume that a stochastic process $\{X_0, X_1, \ldots\}$ taking values in $\mathbb{R}$ is given.

Want to estimate probabilities of the form

$$P \triangleq P\left(\{X_0, X_1, \ldots\} \in A\right)$$

for some $A \in \mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R}) \cdots$ when

$$P\left(\{X_0, X_1, \ldots\} \in A\right) \ll 1.$$

Chemists often use models of the form $dZ_t = \nabla b(Z_t) + \epsilon dW_t$ to analyse chemical reactions.

Let $\{X_0, X_1, \ldots\}$ be a discrete approximation to Z_t

$$X_{i+1} = X_i + \nabla b(X_i) + \epsilon W_{i+1}$$

where $\{W_1, \ldots\}$ is a sequence of i.i.d. $\mathcal{N}(0,1)$ random variables.

Assume that $b(\cdot)$ is a double well potential, that A and B are neighbourhoods of the two local minima and that $X_0 \in A$.

Two probabilities of interest are

$$P(X_T \in B)$$

and

$$P\left(\bigcup_{i=1}^{T} X_i \in B\right).$$

When ϵ is small these probabilities are very (exponentially) small!

The simplest way to estimate a probability of the form P is to generate an i.i.d. sequence of samples $\{X_0^1, X_1^1, \dots\}, \{X_0^2, X_1^2, \dots\}, \dots, \{X_0^N, X_1^N, \dots\}$ such that $\{X_0^k, X_1^k, \dots\} \sim \{X_0, X_1, \dots\}$ and to estimate P by

$$P \approx \frac{1}{N} \sum_{k=1}^{N} 1_{\left\{\left\{X_0^k, X_1^k, \dots\right\} \in A\right\}}.$$

The variance of this estimator is equal to $\left(P - P^2\right)/N$ and so the relative error is equal to

$$\sqrt{\frac{P - P^2}{N}}.\frac{1}{P} \approx \frac{1}{\sqrt{PN}}.$$

It follows that the amount of work required to estimate a probability P is of the order $O(\frac{1}{P})$!

Theory of Large Deviations

Let $Y_1, Y_2, \ldots$ be a sequence of i.i.d. random variables such that $E\left[Y_k^2\right] < \infty$.

Strong Law of Large Numbers: $\frac{1}{N}\sum_{k=1}^{N} Y_k \xrightarrow{a.s.} E\left[Y_1\right]$.

Central Limit Theorem: $\sqrt{N}\frac{\left(\frac{1}{N}\sum_{k=1}^{N} Y_k - E[Y_1]\right)}{\sqrt{E\left[Y_1^2\right]}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,1\right)$.

This suggests that for any $\gamma > 0$, for N large enough

$$\log P\left(\left|\frac{1}{N}\sum_{k=1}^{N} Y_k - E\left[Y_1\right]\right| \geq \gamma\right) = O\left(-N\right).$$

What can we say about the asymptotic decay rate of $P\left(\frac{1}{N}\sum_{k=1}^{N} Y_k \geq \gamma\right)$ for large N?

Let $Y_1, Y_2, \ldots$ be a sequence of centered $\mathbb{R}$ valued i.i.d. light tailed random variables, i.e. such that $E\left[Y_k = 0\right]$ and $E\left[e^{\theta Y_k}\right] < \infty$ for all $\theta \in \mathbb{R}$.

For each θ let $H(\theta) \triangleq \log E\left[e^{\theta Y_k}\right]$ and define $L(\alpha)$ by

$$L(\alpha) = \sup_{\theta} \left\{\theta\alpha - H(\theta)\right\}$$

for all α.

Cramer's Theorem

For any $\gamma > 0$

$$\lim_{N\to\infty} -\frac{1}{N} \log P\left(\frac{1}{N}\sum_{k=1}^{N} Y_k \geq \gamma\right) = L(\gamma)$$

Upper bound: For each $\theta \geq 0$ let $Y_1^\theta, Y_2^\theta, \ldots$ be a sequence of i.i.d. random variables with probability law given by $\frac{dP^{Y_k^\theta}}{dP^{Y_k}} = e^{\theta Y - H(\theta)}$.

$$
\begin{aligned}
P\left(\frac{1}{N}\sum_{k=1}^{N} Y_k \geq \gamma\right) &= E\left[1_{\left\{\frac{1}{N}\sum_{k=1}^{N} Y_k^\theta \geq \gamma\right\}} e^{\left(NH(\theta) - \sum_{k=1}^{N} \theta Y_k^\theta\right)}\right] \\
&\leq e^{N(H(\theta)-\theta\gamma)}
\end{aligned}
$$

Thus $-\frac{1}{N}\log P\left(\frac{1}{N}\sum_{k=1}^{N} Y_k \geq \gamma\right) \geq \sup_{\theta\geq 0}\{\theta\gamma - H(\theta)\}$. It is easy to show that $H'(0) = 0$ and that $H(.)$ is strictly convex. Hence

$$
\lim_{N\to\infty} -\frac{1}{N}\log P\left(\frac{1}{N}\sum_{k=1}^{N} Y_k \geq \gamma\right) \geq \sup_{\theta}\{\theta\gamma - H(\theta)\} \triangleq L(\gamma).
$$

Lower bound: Let θ_γ be such that $L(\gamma) = \theta_\gamma \gamma - H(\theta_\gamma)$. Note that then we have

$$\gamma = H'(\theta_\gamma) = E\left[Ye^{\theta_\gamma Y - H(\theta_\gamma)}\right] = E\left[Y^{\theta_\gamma}\right].$$

Thus for any $\delta > 0$

$$P\left(\frac{1}{N}\sum_{k=1}^{N} Y_k \geq \gamma\right) = E\left[1_{\left\{\frac{1}{N}\sum_{k=1}^{N} Y_k^{\theta_\gamma} > \gamma\right\}} e^{\left(NH(\theta_\gamma) - \sum_{k=1}^{N} \theta_\gamma Y_k^{\theta_\gamma}\right)}\right]$$
$$\geq e^{N(H(\theta_\gamma) - \theta_\gamma(\gamma+\delta))} P\left(1_{\left\{\gamma+\delta > \frac{1}{N}\sum_{k=1}^{N} Y_k^{\theta_\gamma} > \gamma\right\}}\right).$$

Since δ is arbitrary it follows that

$$\lim_{N\to\infty} -\frac{1}{N} \log P\left(\frac{1}{N}\sum_{k=1}^{N} Y_k \geq \gamma\right) \leq \theta_\gamma \gamma - H(\theta_\gamma) = L(\gamma).$$

Using exactly the same proof techniques as above one can show that for any x and $t \in [0,1]$

$$\lim_{N\to\infty} -\frac{1}{N} \log P\left(\frac{1}{N}\sum_{k=1}^{N} Y_k \geq \gamma \,\middle|\, \frac{1}{N}\sum_{k=1}^{\lfloor tN \rfloor} Y_k = x \right) = V(x,t)$$

where

$$V(x,t) = \begin{cases} (1-t)\, L\left(\frac{\gamma - x}{1-t}\right) & \text{if } x < \gamma \\ 0 & \text{otherwise} \end{cases} .$$

By definition $L(\cdot)$ is strictly convex. Thus it follows that for all x, t

$$V(x,t) = \inf_{\psi:\psi(t)=x,\psi(1)\geq\gamma} \left\{ \int_t^1 L\left(\dot{\psi}(s)\right) ds \right\}.$$

In particular $V(\cdot,\cdot)$ is a solution to the HJB equation

$$0 = V_t - \mathbb{H}(-V_x)$$

where $\mathbb{H}(\beta) = \sup_\alpha \{\alpha\beta - L(\alpha)\} = H(\beta)$.

Simulating Rare Events

Assume an i.i.d. $Y_1, Y_2, \ldots$ of $\mathbb{R}$ valued, centered and light tailed random variables is given.

Given N define the "sample mean" process $\{X_0^N, \ldots\}$ by $X_i^N = \frac{1}{N}\sum_{k=1}^{i} Y_k$ for all $i \in \{0, 1, \ldots\}$.

Consider the problem of estimating the probabilities

$$P\left(\frac{1}{N}\sum_{k=1}^{N} Y_k \geq \gamma\right) = P\left(X_N^N \geq \gamma\right)$$

for some $\gamma > 0$.

Suppose that the probabilities $P\left(X_N^N \geq \gamma | X_i^N = x\right)$ are known for all $i \in \{0, 1, \ldots\}$ and all x.

Further suppose that we we can sample from a sequence of random variables $\tilde{Y}_1, \tilde{Y}_2, \ldots$ distributed according to the law

$$\frac{dP^{\tilde{Y}_k}}{dP^{Y_k}} = \frac{P\left(X_N^N \geq \gamma | X_k^N = \tilde{X}_{k-1}^N + \frac{1}{N}\tilde{Y}_k\right)}{P\left(X_N^N \geq \gamma | X_{k-1}^N = \tilde{X}_{k-1}^N\right)}$$

where $\left\{\tilde{X}_0^N, \ldots\right\}$ denotes the sample mean process for the random variables $\tilde{Y}_1, \tilde{Y}_2, \ldots$.

Note that the sequence $\tilde{Y}_1, \tilde{Y}_2, \ldots$ has the following properties:

- $P\left(\frac{1}{N}\sum_{k=1}^{N} \tilde{Y}_k \geq \gamma\right) = 1$.
- Given $\tilde{Y}_1, \ldots, \tilde{Y}_N$

$$\frac{dP^{Y_1,\ldots,Y_N}}{dP^{\tilde{Y}_1,\ldots,\tilde{Y}_N}} = \prod_{k=1}^{N} \frac{P\left(X_N^N \geq \gamma | X_{k-1}^N = \tilde{X}_{k-1}^N\right)}{P\left(X_N^N \geq \gamma | X_k^N = \tilde{X}_{k-1}^N + \tilde{Y}_k\right)}$$

$$= \frac{P\left(X_N^N \geq \gamma | X_0^N = 0\right)}{P\left(X_N^N \geq \gamma | X_N^N = \tilde{X}_N^N\right)}$$

$$= P\left(X_N^N \geq \gamma\right).$$

Thus if we could sample from the random variables $\tilde{Y}_1, \ldots$ the quantity $1_{\{\tilde{X}_N^N \geq \gamma\}} \frac{dP^{Y_1,\ldots,Y_N}}{dP^{\tilde{Y}_1,\ldots,\tilde{Y}_N}}$ would yield a perfect (zero variance) estimate!

Unfortunately the conditional probabilities $P(\cdot|\cdot)$ are unknown. However we do know that $P\left(X_N^N \geq \gamma | X_i^N = x\right) \approx e^{-NV(x, \frac{i}{N})}$.

This suggests sampling from the sequence $\hat{Y}_1, \ldots$ where

$$\frac{dP^{\hat{Y}_k}}{dP^{Y_k}} = \frac{e^{-NV\left(\hat{X}_{k-1}^N + \frac{1}{N}\hat{Y}_k, \frac{k}{N}\right)}}{e^{-NV\left(\hat{X}_{k-1}^N, \frac{k-1}{N}\right)}}.$$

Using elementary calculus we have the relation

$$\begin{aligned} \frac{dP^{\hat{Y}_k}}{dP^{Y_k}} &= e^{-N\left(\frac{1}{N}V_t\left(\hat{X}_{k-1}^N, \frac{k-1}{N}\right) + \frac{1}{N}\hat{Y}_k V_x\left(\hat{X}_{k-1}^N, \frac{k-1}{N}\right) + O\left(\frac{1}{N^2}\right)\right)} \\ &= e^{-\left(V_t\left(\hat{X}_{k-1}^N, \frac{k-1}{N}\right) + \hat{Y}_k V_x\left(\hat{X}_{i-1}^N, \frac{k-1}{N}\right) + O\left(\frac{1}{N}\right)\right)}. \end{aligned}$$

In practice we sample from the sequence $\bar{Y}_1, \ldots$ where

$$\frac{dP^{\bar{Y}_k}}{dP^{Y_k}} = e^{-\left(V_t\left(\bar{X}^N_{k-1}, \frac{k-1}{N}\right) + \bar{Y}_k V_x\left(\bar{X}^N_{k-1}, \frac{k-1}{N}\right)\right)}.$$

Recall that $V_t - H(-V_x) = 0$ so this does define a change of probability measure!

We calculate the variance of the estimator $1_{\{\bar{X}^N_N \geq \gamma\}} \frac{dP^{Y_1,\ldots,Y_N}}{dP^{\bar{Y}_1,\ldots,\bar{Y}_N}}$

$$E\left[1_{\{\bar{X}^N_N \geq \gamma\}} \left(\frac{dP^{Y_1,\ldots,Y_N}}{dP^{\bar{Y}_1,\ldots,\bar{Y}_N}}\right)^2\right]$$

$$= E\left[1_{\{\bar{X}^N_N \geq \gamma\}} \left(e^{N\left(V(\bar{X}^N_N,1) - V(0,0)\right) + O(1)}\right)^2\right]$$

$$= e^{-2NV(0,0)} E\left[1_{\{\bar{X}^N_N \geq \gamma\}} \left(e^{O(1)}\right)^2\right].$$

One can show that $\lim_{N\to\infty} \frac{1}{N} \log E\left[1_{\{\bar{X}_N^N \geq \gamma\}} \left(e^{O(1)}\right)^2\right] = 0$.

Thus

$$\lim_{N\to\infty} \frac{1}{N} \log \frac{\sqrt{E\left[1_{\{\bar{X}_N^N \geq \gamma\}} \left(\frac{dP^{Y_1,\ldots,Y_N}}{dP^{\bar{Y}_1,\ldots,\bar{Y}_N}}\right)^2\right]}}{P\left(\sum_{k=1}^N Y_k \geq \gamma\right)} = 0.$$

This is known as **asymptotic optimality**.

Extensions

Let a probability kernel $P(\cdot|x)$ on $\mathbb{R}$ be given. For each N define a Markov Chain $\{X_0^N, X_1^N, \ldots\}$ such that $X_0^N = 0$ and for all i

$$N\left(X_{i+1}^N - X_i^N\right) \sim P\left(\cdot|X_i^N\right).$$

We again consider the problem of estimating

$$P\left(X_N^N \geq \gamma\right)$$

for some γ.

Assume that for every $x \in \mathbb{R}$ all exponential moments of the form $E_{P(\cdot|x)}\left[e^{\theta Y}\right]$ exist.

Define

$$H(\theta, x) = \log E_{P(\cdot|x)}\left[e^{\theta Y}\right]$$

for all θ, x and

$$L(\alpha, x) = \sup_{\theta} \{\theta\alpha - H(\theta, x)\}$$

for all α, x.

Under certain conditions it can be shown that for all x and $t \in [0,1]$

$$\lim_{N\to\infty} -\frac{1}{N} \log P\left(X_N^N \geq \gamma \,\middle|\, X_{\lfloor tN \rfloor}^N = x \right) = V\left(x,t\right)$$

where

$$V\left(x,t\right) = \inf_{\psi:\psi(t)=x,\psi(1)\geq\gamma} \left\{ \int_t^1 L\left(\dot{\psi}(s), \psi(s)\right) ds \right\}.$$

In this case $V\left(\cdot,\cdot\right)$ is a solution to the HJB equation

$$0 = V_t - \mathbb{H}(-V_x, x)$$

where $\mathbb{H}(\beta,x) = \sup_\alpha \left\{\alpha\beta - L\left(\alpha,x\right)\right\} = H\left(\beta,x\right)$.

As in the i.i.d. case one can use the function $V(\cdot, \cdot)$ to define an importance sampling scheme. Further the same reasoning can be used to show that the resulting estimator is asymptotically optimal.

However in general the function $V(\cdot, \cdot)$ can be difficult to find, further the partial derivatives V_t, V_x may not even exist.

Suppose we can find a function $U(\cdot,\cdot)$ such that

$$0 \leq U_t - \mathbb{H}(-U_x)\,;\; U(x,1) \leq 0 \text{ for all } x \geq \gamma.$$

Such a function is called a **subsolution**. We could then use $U(\cdot,\cdot)$ to define a sequence $\bar{\bar{Y}}_1, \ldots$ where

$$\frac{dP^{\bar{\bar{Y}}_k}}{dP^{Y_k}} = \frac{e^{-\left(U_t\left(\bar{\bar{X}}^N_{k-1}, \frac{k-1}{N}\right) + \bar{\bar{Y}}_k U_x\left(\bar{\bar{X}}^N_{k-1}, \frac{k-1}{N}\right)\right)}}{E\left[e^{-\left(U_t\left(\bar{\bar{X}}^N_{k-1}, \frac{k-1}{N}\right) + \bar{\bar{Y}}_k U_x\left(\bar{\bar{X}}^N_{k-1}, \frac{k-1}{N}\right)\right)}\right]}$$

and use this as the change of measure for an importance sampling estimator.

We can again calculate the variance of the estimator $1_{\left\{\bar{\bar{X}}_N^N \geq \gamma\right\}} \frac{dP^{Y_1,\ldots,Y_N}}{dP^{\bar{\bar{Y}}_1,\ldots,\bar{\bar{Y}}_N}}$

$$
\begin{aligned}
& E\left[1_{\left\{\bar{\bar{X}}_N^N \geq \gamma\right\}} \left(\frac{dP^{Y_1,\ldots,Y_N}}{dP^{\bar{\bar{Y}}_1,\ldots,\bar{\bar{Y}}_N}}\right)^2\right] \\
& = E\left[1_{\left\{\bar{\bar{X}}_N^N \geq \gamma\right\}} \left(e^{N\left(U(\bar{\bar{X}}_N^N,1)-U(0,0)\right)+O(1)} \prod_{k=1}^{N} E\left[e^{-\left(U_t+\bar{\bar{Y}}_k U_x\right)}\right]\right)^2\right] \\
& \leq e^{-2NU(0,0)} E\left[1_{\left\{\bar{X}_N^N \geq \gamma\right\}} \left(e^{O(1)}\right)^2\right].
\end{aligned}
$$

As before it can be shown that

$$\lim_{N\to\infty} \frac{1}{N} \log E\left[1_{\{\bar{X}_N^N \geq \gamma\}} \left(e^{O(1)}\right)^2\right] = 0$$

and so the estimator has **asymptotic relative error** equal to

$$\lim_{N\to\infty} \frac{1}{N} \log \frac{\sqrt{E\left[1_{\{\bar{X}_N^N \geq \gamma\}} \left(\frac{dP^{Y_1,\ldots,Y_N}}{dP^{\bar{Y}_1,\ldots,\bar{Y}_N}}\right)^2\right]}}{P\left(\sum_{k=1}^N Y_k \geq \gamma\right)} = V(0,0) - U(0,0).$$

References

Large Deviations Theory

A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Jones and Bartlett, Boston, 1993.

P. Dupuis and R. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley & Sons, New York, 1997.

Importance Sampling

P. Dupuis and H. Wang. Subsolutions of an I saacs equation and efficient schemes for importance sampling, *Math. O. R.*, 32:1–35, 2007.

P. Dupuis and H. Wang. Importance Sampling, Large Deviations and Differential Games, *Stoch. and Stoch. Rep.*, 76:481–508, 2004.

Other

T. Dean and P. Dupuis. Splitting for Rare Event Simulation: A Large Deviation Approach to Design and Analysis. *Stoc. Proc. Appl.*, 119:562–587, 2009.

T. Dean and P. Dupuis. The design and analysis of a generalised RESTART/DPR algorithm for rare event simulation. *Submitted to Annals of OR*.

P. Del Moral and J. Garnier. Genealogical Particle Analysis of Rare Events. *Ann. Appl. Prob.*, 15(4):2496–2534, 2005.

H.P. Chan and T.L. Lai. A sequential Monte Carlo approach to computing tail probabilities in stochastic models. *Submitted to Ann. Appl. Prob.*.

J. Blanchet and P. Glynn. Efficient Rare-event Simulation for the Maximum of Heavy-tailed Random Variables. *Ann. Appl. Prob.*, 18:1351–1378, 2008.