State–dependent Importance Sampling and Rare–event Simulation

Peter W. Glynn

Stanford University

Lunteren Conference, January 16th, 2007

Rare–event Simulation

Goal : Compute $p = P(A)$, where A is "rare"

Applications :

- Reliability modeling
- Dependability systems
- Communications networks
- Supply chains
- Finance

Outline of Talk:

Review of rare event simulation

State independent changes of measure for random walks

State dependent changes of measure

Lyapunov bounds on importance variances

Extension to expectations

Connection to optimal control

Conventional Sampling and Rare–event Simulation

Method : Generate iid copies $I_1, I_2, \ldots, I_n$ of I_A.

$$p_n = \frac{1}{n} \sum_{j=1}^{n} I_j$$

Analysis :

$$n^{1/2}(p_n - p) \Rightarrow \sqrt{p(1-p)}N(0,1)$$

$$p_n \overset{\mathcal{D}}{\approx} p + \sqrt{\frac{p(1-p)}{n}}N(0,1)$$

As p $\downarrow 0$, absolute error $\downarrow 0$

relative error $\uparrow \infty$

Importance Sampling

Change the sampling distribution from P to Q

$$\begin{aligned} p = E_p I_A &= \int_\Omega I_A(\omega) P(d\omega) \\ &= \int_\Omega I_A(\omega) L(\omega) Q(d\omega) \\ &= E_Q I_A L \end{aligned}$$

Zero–variance Change of Measure

If we select

$$Q^*(d\omega) = \frac{I_A(\omega)P(d\omega)}{P(A)} = P(d\omega|A)$$

then

$$LI_A = P(A) \qquad Q^* \text{ a.s.}$$

Zero–variance!

Moral of the Story :

Choose an easily generated Q that is close to Q^*

Theoretical Analysis

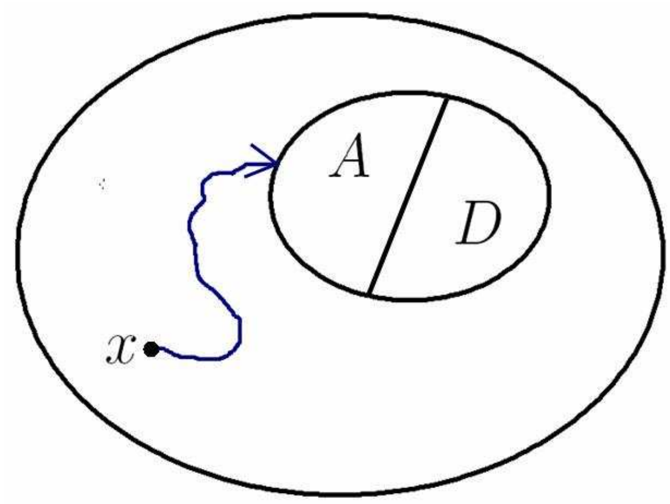
- Family of problem instances : $(P_n(A_n) : n \geq 1)$
- with rare–event property : $p_n = P_n(A_n) \to 0$ as $n \to \infty$
- Importance Sampling estimator : $p_n = \widetilde{E}_n I_{A_n} L_n$
- "Bounded Relative Variance" (strongly efficient)

$$\sup_{n \geq 1} \frac{\widetilde{\mathrm{var}}_n I_{A_n} L_n}{p_n^2} < \infty$$

- "Logarithmic efficiency"

$$\underline{\lim}_{n\to\infty} \frac{\log \widetilde{E}_n I_{A_n} L_n^2}{\log p_n} = 2$$

Computing Exit Probabilities for Markov Chains



$X = (X_n : n \geq 0)$ S–valued Markov chain

$T = \inf\{n \geq 0 : X_n \in A \cup D\}$ ("exit time")

Goal : Compute $p(x) = P_x(X_T \in A, T < \infty)$

Applications

- $P_x(T < \infty)$; ruin probabilities
- mean time to failure ($P_x(T < \tau)$)
- regenerative analysis ($E_x \int_0^\tau I(X(s) \in A)ds$)
- $P_x(X_n \in B)$ (hitting time for "space–time" chain $((i, X_i), i \geq 0)$ to hit $\{n\} \times B$)

Description of Zero–variance Change of Measure

Conditional dynamics of X given $\{X_T \in A, T < \infty\}$ are Markovian

$$P_x(X_1 \in dx_1, \ldots, X_n \in dx_n | X_T \in A, n \le T < \infty)$$
$$= Q^*_x(X_2 \in dx_2, \ldots, X_n \in dx_n)$$

where $Q^*(x, dy) = P(x, dy)p(y)/p(x)$

In general, change of measure is state-dependent

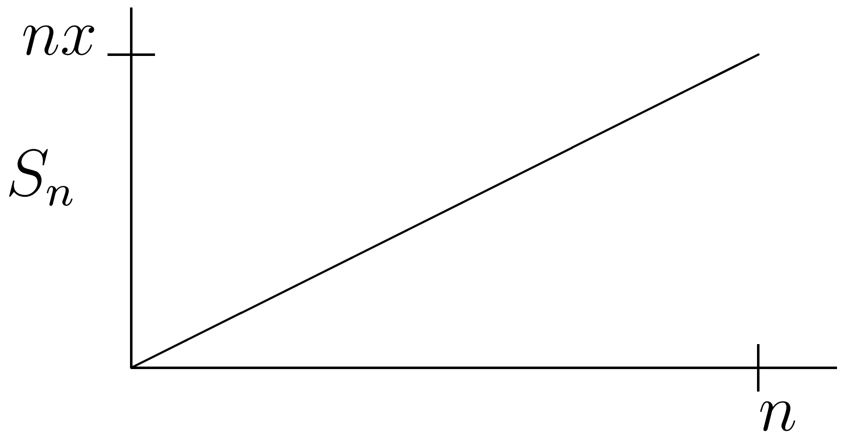
Asymptotic Description of the Conditional Distribution for Light–tailed Random Walks

$S_n = Y_1 + \ldots + Y_n$

$Y_1, Y_2, \ldots$ iid with $EY_1 = 0$

$\psi(\theta) = \log E \exp(\theta Y_i) < \infty$

Goal : Compute $p_n = P(S_n > nx) \quad (x > 0)$



- $P(Y_1 \in dy_1, \ldots, Y_k \in dy_k | S_n > nx)$

$$\to \prod_{i=1}^{k} \exp\left(\theta(x) y_i - \psi(\theta(x))\right) P(Y_i \in dy_i)$$

as $n \to \infty$ (where $\psi'(\theta(x)) = x$)

- Suggests simulating $Y_1, Y_2, \ldots Y_n$ as iid using "exponentially twisted" distribution

$$\exp\left(\theta(x)y - \psi(\theta(x))\right) P(Y \in dy)$$

- Distribution of increments is state–independent ("hard–wired" , "static" , "blind")

Level–Crossing Probability

$S_n = Y_1 + \ldots + Y_n$

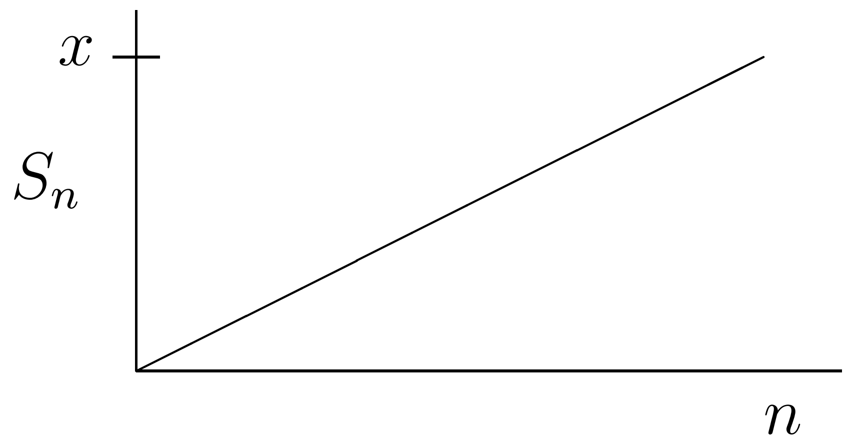
$Y_1, Y_2, \ldots$ iid with $EY_1 < 0$

$\psi(\theta) = \log E \exp(\theta Y_i) < \infty$

Goal : Compute $p(x) = P(T(x) < \infty)$ where

$$T(x) = \inf\{n \geq 0 : S_n > x\}$$

Arises in analysis of $G/G/1$ queue, insurance risk theory, sequential analysis



- $P(Y_1 \in dy_1, \ldots Y_k \in dy_k | T(x) < \infty)$
$$\to \prod_{i=1}^{k} \exp(\theta^* y_i) P(Y_i \in dy_i)$$
as $x \to \infty$ where $\theta^* > 0$ satisfies $\psi(\theta^*) = 0$

- Suggests simulating $Y_1, Y_2, \ldots$ up to $T(x)$ as iid using "exponentially twisted" distribution

$$\exp(\theta^* y) P(Y \in dy)$$

- Distribution of increments is state–independent

Asymptotic Efficiency

- $P(S_n > nx)$ logarithmic efficiency
- $P(T(x) < \infty)$ bounded relative variance

$S_n = Y_1 + \ldots + Y_n$

$Y_1, Y_2, \ldots, Y_n$ iid with $EY_1 = 0$

$E \exp(\theta Y_1) < \infty$

Goal : Compute $p_n = P(S_n < na$ or $S_n > nb)$ $(a < 0 < b)$

$$P(S_n > nb) \sim \frac{1}{\sqrt{2\pi n \sigma(\theta(b))}} \exp\left(-\theta(b)nb + n\psi(\theta(b))\right)$$

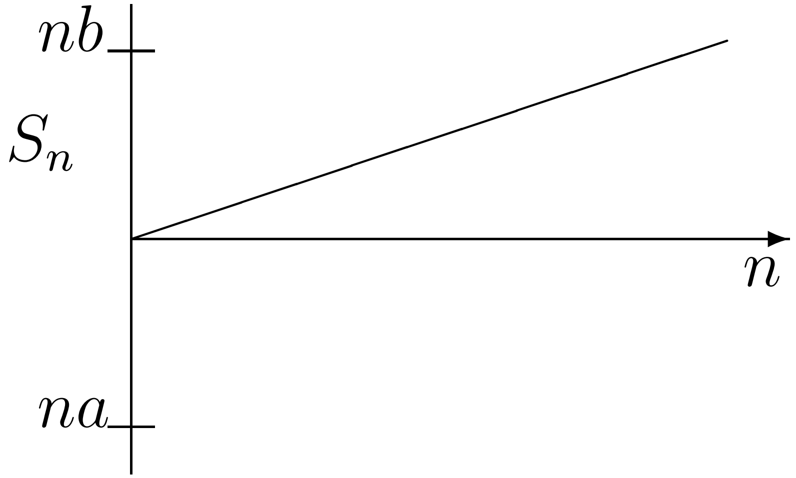
$$P(S_n < na) \sim \frac{1}{\sqrt{2\pi n \sigma(\theta(a))}} \exp\left(-\theta(a)na + n\psi(\theta(a))\right)$$

Assume $\theta(b)b - \psi(\theta(b)) < \theta(a) - \psi(\theta(a))$.

Then,

$$P(Y_1 \in dy_1, \ldots Y_k \in dy_k | S_n < na \text{ or } S_n > nb)$$

$$\to \prod_{i=1}^{k} \exp(\theta(b)y_i - \psi(\theta(x)))P(Y_i \in dy_i)$$



- Suggests simulating $Y_1, Y_2, \ldots Y_n$ as iid using "exponentially twisted" distribution

$$\exp\left(\theta(b)y - \psi(\theta(b))\right) P(Y \in dy)$$

- A disaster! Variance inflation, not variance reduction

Level Crossing Probabilities for Heavy–Tailed Random Walks

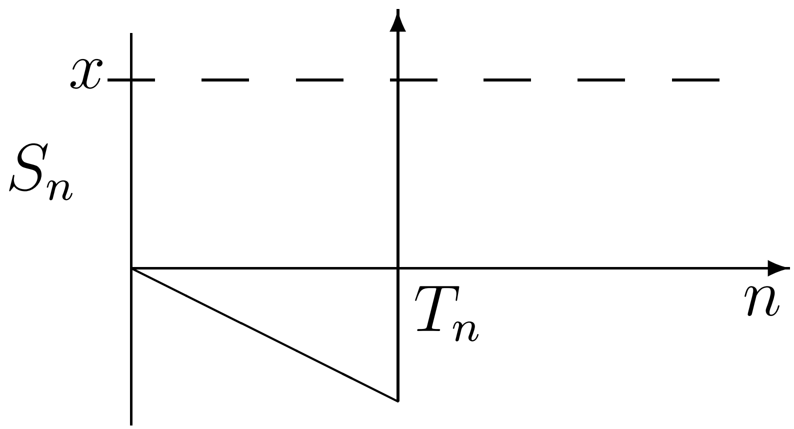
$S_n = Y_1 + \ldots + Y_n$

$Y_1, Y_2, \ldots$ iid with $EY_1 = 0$

$P(Y_1 > x) \sim cx^{-\alpha}$ as $x \to \infty$

Goal : Compute $p(x) = P(T(x) < \infty)$ where

$$T(x) = \inf\{n \geq 0 : S_n > x\}$$



Level Crossing Probabilities for Heavy–Tailed Random Walks

- $P(Y_1 \in dy_1, \ldots Y_k \in dy_k | T(x) < \infty)$
  $\to \prod_{i=1}^{k} P(Y_i \in dy_i) \quad$ as $x \to \infty$
- no obvious means of implementing this asymptotic description

Return to state–dependent (zero–variance) change of measure :

Simulate X according to

$$Q^*(x, dy) = P(x, dy)\frac{p(y)}{p(x)}$$

- Approximate Q^* by using a good approximation to $p(\cdot)$
- An alternative :
  Note that $Q^*(x, dy) = P_x(X_1 \in dy | X_T \in A, T < \infty)$
  Use conditional distribution of X_1 given $\{X_T \in A, T < \infty\}$

Return to Light–tailed Random Walk

$$p_n = P(S_n > nx) \quad (x > 0)$$

Here,

$$Q^*(Y_{i+1} \in dy | S_i = z) = P(Y \in dy) \frac{P(S_{n-i-1} > nx - z - y)}{P(S_{n-i} > nx - z)}$$

One can use the approximation suggested by (exact) large deviations :

$$P(S_j > w) \approx \frac{1}{\sqrt{2\pi j \sigma(\theta(w/j))}} \cdot \exp(-\theta(w/j)w + j\psi(\theta(w/j)))$$

Leads to change of measure that is difficult to simulate

The alternative : Simulate Y_{i+1} from "exponentially twisted" distribution

$$\exp\left(\theta((nx - z)/(n - i))y - \psi(\theta((nx - z)/(n - i))\cdot P(Y \in dy)\right)$$

i.e. Simulate Y_{i+1} with the exponential twist having drift that maximizes the likelihood of ending at level nx at time n, given that $S_i = z$.

Theorem (Blanchet and G (06)) : Under suitable regularity conditions, both the above algorithms enjoy bounded relative variance.

Remark : The state–independent algorithm has a relative variance that increases in n.

Return to Level Crossing Probability for Light–tailed Random Walk

$$p(x) = P(T(x) < \infty) \qquad (x > 0)$$

Here,

$$Q^*(Y_{i+1} \in dy | S_i = z) = P(Y \in dy) p(x - z - y) / p(x - z).$$

When Y is non-lattice and a positive root θ^* of $\psi(\theta^*) = 0$ exists, $p(w) \sim ce^{-\theta^* w}$ as $w \to \infty$. So,

$$\frac{p(x - z - y)}{p(x - z)} \to e^{\theta^* y} \qquad \text{as } x \to \infty$$

Alternative approach leads to same result : Simulate $Y_1, Y_2, \ldots$ iid according to "exponentially twisted" distribution $e^{\theta^* y} P(Y \in dy)$.

Return to Two–sided Exit Probability

$$p_n = P(S_n > nb \text{ or } S_n < na)$$

Here, $Q^*(Y_{i+1} \in dy | S_i = z)$

$$= P(Y \in dy) \frac{P(S_{n-i-1} > nb - z - y \text{ or } S_{n-i-1} < na - z - y)}{P(S_{n-i} > nb - z \text{ or } S_{n-i} < na - z)}$$

- If z is close enough to na, the preferred exit boundary is na.
- If one plugs approximations for probabilities directly in, variate generation becomes difficult.
- A strongly efficient algorithm can be developed that uses only exponential twisting.

Return to Level Crossing Probabilities for Heavy–tailed Random Walk

$S_n = Y_1 + \ldots + Y_n$

$Y_1, Y_2, \ldots$ iid with $EY_1 < 0$

$P(Y_1 > x) \sim cx^{-\alpha}$ as $x \to \infty$

$$Q^*(Y_{i+1} \in dy | S_i = z) = P(Y \in dy)p(x-y-z)/p(x-z)$$

Heavy–tail asymptotic for p :

$$p(w) \sim \frac{1}{|EY|} \int_{-\infty}^{\infty} P(Y > s)ds \triangleq v(w)$$

Approximate Q^* by Q :

$$Q(Y_{i+1} \in dy | S_i = z) = P(Y \in dy)v(x-y-z)/w(x-z)$$

where $w(\cdot)$ is the normalization constant

$$w(\cdot) = \int_R P(Y \in dy)v(\cdot - y)$$

Alternatively,

$$Q^*(Y_{i+1} \in dy | S_i = z) = P(Y \in dy | Y + M > x - z)$$

where $M = \max(S_n : n \geq 0)$. Then,

$$Q(Y_{i+1} \in dy | S_i = z) = P(Y \in dy | Y + Z > x - z)$$

where $P(Z > \cdot) = v(\cdot) \approx P(M > \cdot)$.

Use acceptance–rejection to generate such a Y (based on mixture density formed from density of Y).

Theorem (Blanchet & G (2006))

i) If $Y \in S^*$, then the above algorithm enjoys bounded relative variance as $x \to \infty$.

ii) Furthermore, under further restrictions on Y, the expected number of operations, required to simulate under Q is $O(x)$.

Details on Theoretical Analysis

Need to upper bound variance of $I_A L$ under Q. Put

$$s^*(z) = E_z^Q I_A L^2$$

Suppose that Q takes the form

$$Q(y,dz) = P(y,dz)v(z)/w(y)$$

Proposition (Blanchet & G (06)) Suppose there exists a non–negative $h \geq \epsilon$ satisfying

$$w(y) \int P(y,dz)v(z)h(z) \leq h(y)v(y)^2$$

for $y \in A^c$. Then,

$$\frac{s^*(z)}{v^2(z)} \leq \frac{\epsilon^{-1} h(z)}{\inf_{x \in A} v^2(x)}$$

Using a more refined Lyapunov analysis than the above, one can show that bounded relative variance is accomplished when :

- $w(x)/v(x) \to 1$ as $x \to \infty$
- There exists $r \leq 1/2$ and $x_0 < \infty$ such that

  $w(x) \leq v(x) + rP_x(X_1 \in A), \quad x \geq x_0,\ x \in (A \cup D)^c$

Extension to Expectations

$$u^*(x) = E_x \sum_{j=0}^{\tau} f(X_j) \exp\left(\sum_{l=0}^{j-1} h(X_l)\right)$$

u^* satisfies

$$\begin{aligned} u(x) &= f(x) + e^{h(x)} \int_S P(x, dy) u(y) \\ 1 &= \int_S P(x, dy) \frac{e^{h(x)} u^*(y)}{u^*(x) - f(x)} \\ &\triangleq \int_S Q^*(x, dy) \end{aligned}$$

If

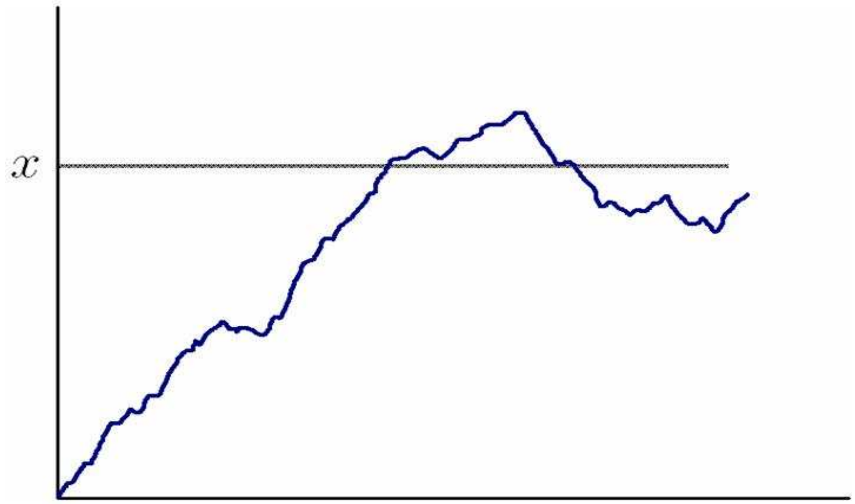
$$W = \sum_{j=0}^{\tau} f(X_j) \exp\left(\sum_{l=0}^{j-1} h(X_l)\right) L_j$$

then

$$W = u^*(X_0) \qquad \text{a.s.}$$

There exists a zero - variance Markovian change - of - measure for such expectations (Awad, Rubinstein, and G (06)).

- Importance algorithm may be non–terminating



- Zero–variance change–of–measure may not be unique

Connection to Dynamic Programming

$$Q(\theta, x, dy) = P(x, dy) l^{-1}(\theta, x, y)$$

Set

$$s(\theta, z) = E_{Q(\theta)}[W^2 | X_0 = z]$$

What is the optimal choice of $\theta(z)$?

Optimal change–of–measure within class $(Q(\theta) : \theta \in \Lambda)$ is determined by HJB equation

$$
\begin{aligned}
s^*(z) \quad = \quad & \inf_{\theta \in \Lambda} \Big[f(z)^2 + 2f(z)e^{h(z)} \int_S P(z, dw)u^*(w) \\
& +e^{2h(z)} \int_S P(z, dw)l(\theta, z, w)s^*(w) \Big]
\end{aligned}
$$

(Dupuis, Wang)

Conclusions:

- In general, a change of measure intended to produce bounded relative variance should exhibit state dependence in the transition structure, even for problems described in terms of random walks
- When analytic approximations to the rare event probabilities exist, these approximations suggest natural state-dependent changes of measure (for both light-tailed and heavy-tailed systems).
- The optimal state-dependent change of measure (within a parametric family of importance kernels) can be computed as the solution to an optimal control problem (Dupuis and Wang)