Computing Large Deviations for Continuous State Space Markov Chains

Peter W. Glynn

Stanford University

Joint work with Jose Blanchet and Xiaowei Zhang

RESIM Conference, Rennes (September 24, 2008)

Outline of Talk

1. Rare Event Simulation for Markov-dependent Sums: Review
2. Computation in the Presence of Continuous State Space
3. Computing Exit Time Distributions

Rare Event Simulation for Markov-dependent Sums

- Setting:
  - $X = (X_n : n \geq 0)$ finite state S-valued irreducible Markov chain with unique stationary distribution π
  - $f : S \to \mathbb{R}$ with $\pi f = 0$
  - $S_n = \sum_{i=0}^{n-1} f(X_i)$
- Goal:

  Compute $P(S_n > n\epsilon)$ for n large (with $\epsilon > 0$)

Solution

1. Compute solution to eigenvalue problem

$$K(\theta)h(\theta) = \lambda(\theta)h(\theta)$$

   where $K(\theta, x, y) = \exp(\theta f(x))P(x, y)$

2. Set $\psi(\theta) = \log \lambda(\theta)$. Find the root θ^* of the equation

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\psi(\theta^*) = \epsilon$$

(continued)

3. Simulate X under the transition dynamics associated with the one-step transition matrix

$$\exp(\theta^* f(x) - \psi(\theta^*)) P(x,y) \frac{h(\theta^*, y)}{h(\theta^*, x)}$$

4. Set $W = \mathbb{I}(S_n > n\epsilon) \exp(-\theta^* S_n + n\psi(\theta^*)) \dfrac{h(\theta^*, X_0)}{h(\theta^*, X_n)}$

5. Generate p iid copies $W_1, \ldots, W_p$ of the rv W and return

$$\overline{W}_p = \frac{1}{p} \sum_{i=1}^{p} W_i$$

This algorithm is *logarithmically efficient*, in the sense that the computational effort required to compute $P(S_n > n\epsilon)$ to a given relative precision increases "gracefully" with n (e.g. increases polynomially in n)

Difficulties

- How to compute solutions to the eigenvalue problem when $|S|$ is large?
- How to sample transitions from

$$\exp(\theta^* f(x) - \psi(\theta^*))P(x,y)\frac{h(\theta^*, y)}{h(\theta^*, x)}?$$

A large measure of why we do Monte Carlo is so as to address numerical computation for large state space chains (or to deal with continuous state space chains, without resort to discretization)

One Possible Approach

- Estimate $\psi(\theta)$, θ^*, and the eigenfunction $h(\theta^*)$ numerically (somehow)
- Simulate X under the estimated transition kernel (somehow)

This looks difficult to execute in practice...

Idea

Use *regeneration*.

Specifically, note that the likelihood ratio over a regenerative z-cycle equals

$$\exp(\theta S_{\tau(z)} - \tau(z)\psi(\theta))\frac{h(\theta, X_{\tau(z)})}{h(\theta, X_0)} = \exp(\theta S_{\tau(z)} - \tau(z)\psi(\theta))$$

where $\tau(z) = \inf\{n \geq 1 : X_n = z\}$.

No eigenfunction!

Furthermore, the root $\psi(\theta)$ can be characterized as the root of the equation

$$\mathrm{E}\exp(\theta S_{\tau(z)} - \tau(z)\psi(\theta)) = 1$$

so

$$\widetilde{P}_\theta(\cdot) \triangleq \mathrm{E}\mathbb{I}(\cdot)\exp(\theta S_{\tau(z)} - \tau(z)\psi(\theta))$$

induces a probability measure on the cycle path-space associated with the regenerative process

Let $0 = T(0) < T(1) < T(2) < \ldots$ be the regeneration times at which X returns to z. Put

$$N(n) = \max\{j \geq 0 : T(j) \leq n\}$$

and note that $N(n) + 1 \geq n$ is a stopping time adapted to X. So,

$$P(S_n > n\epsilon) = \widetilde{\mathrm{E}}_\theta \mathbb{I}(S_n > n\epsilon) \exp(-\theta S_{T(N(n)+1)} + T(N(n)+1)\psi(\theta))$$

Finally, the root θ^* can be defined in terms of regenerative quantities:

$$\frac{\mathrm{E}S_{\tau(z)} \exp(\theta^* S_{\tau(z)} - \tau(z)\psi(\theta^*))}{\mathrm{E}\tau(z) \exp(\theta^* S_{\tau(z)} - \tau(z)\psi(\theta^*))} = \epsilon$$

i.e.

$$\frac{\mathrm{E}^*_\theta S_{\tau(z)}}{\mathrm{E}^*_\theta \tau(z)} = \epsilon$$

How do we take advantage of these insights?

1. Generate r iid z-cycles of X (with no importance sampling), yielding copies $(Y_1, \tau_1), \ldots, (Y_r, \tau_r)$ of the rv (Y, τ), where $Y = S_{\tau(z)}$, $\tau = \tau(z)$
2. Compute

$$m_r(\theta, \lambda) = \frac{1}{r} \sum_{i=1}^{r} \exp(\theta Y_i - \lambda \tau_i)$$

3. Compute $\psi_r(\theta)$ as the root of

$$m_r(\theta, \psi_r(\theta)) = 1$$

4. Find the root θ_r^* given by

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\psi_r(\theta_r^*) = \epsilon$$

If we stopped here, we could use:

$$\exp(n(-\theta_r^*\epsilon + \psi_r(\theta_r^*))$$

as a logarithmically accurate estimator of $P(S_n > nx)$, in the sense that

$$\frac{1}{n}\log P(S_n > n\epsilon) - (-\theta_r^*\epsilon + \psi_r(\theta_r^*)) \to 0$$

as r and n tend to ∞

To get an estimator for $P(S_n > n\epsilon)$ itself, we need to do additional *bootstrap-type* sampling:

5. Sample cycle i from collection $(Y_1, \tau_1), \ldots, (Y_r, \tau_r)$ with probability equal to

$$\frac{1}{r} \exp(\theta_r^* Y_i - \psi_r(\theta_r^*)\tau_i)$$

and call it $(\widetilde{Y}_1, \widetilde{\tau}_1)$. Continue sampling such cycles independently until the total length of the sampled cycles $(\widetilde{Y}_1, \widetilde{\tau}_1), \ldots, (\widetilde{Y}_{\widetilde{N}(n)+1}, \widetilde{\tau}_{\widetilde{N}(n)+1})$ exceeds n.

6. Let $\widetilde{S}_n$ be the sum of the first n $f(X_j)$'s associated with the above bootstrapped sequence of cycles, and put

$$W = \mathbb{I}(\widetilde{S}_n > n\epsilon) \exp\left(\sum_{i=1}^{\widetilde{N}(n)+1} (-\theta_r^* \widetilde{Y}_i + \psi_r(\theta_r^*)\widetilde{\tau}_i) \right)$$

7. Repeat steps 5 and 6 m independent times, yielding $W_1, \ldots, W_m$, and returning

$$\overline{W}_m = \frac{1}{m} \sum_{i=1}^{m} W_i$$

Given a computational budget $c \gg n^2$, choose:

$$r \text{ of order } c$$

$$m = \frac{c}{n^2}$$

This algorithm is *logarithmically efficient*, in the sense that the computational effort required to compute $P(S_n > n\epsilon)$ to a given relative precision increases "gracefully" with n (e.g. increases polynomially in n)

Extension to Continuous State Space

Assume the Markov chain $X = (X_n : n \geq 0)$ is a positive Harris recurrent Markov chain with

$$P(x, \mathrm{d}y) \geq p\phi(\mathrm{d}y)$$

for $x \in K$, $y \in S$. Construct regenerations via the "splitting method":

$$P(x, \mathrm{d}y) = p\phi(\mathrm{d}y) + (1-p)Q(x, \mathrm{d}y)$$

i.e. For $x \in K$, chain X regenerates whenever X distributes itself according to ϕ.

Computing Exit Time Distributions

- Setting:
  - $X = (X_n : n \geq 0)$ finite state S-valued irreducible Markov chain
  - A a given subset of S
  - $T = \inf\{n \geq 0 : X_n \in A\}$
- Goal:

  Compute $P(T > n)$ for n large

Here, the eigenvalue problem takes the form:

$$Bh = \lambda h$$

where

$$B = (P(x, y) : x, y \in A^c)$$

and the "ideal zero-variance transition dynamics" are provided by K, where

$$K(x, y) = B(x, y)\frac{h(y)}{\lambda h(x)}$$

The complication here arises as a consequence of the fact that on the time scale of regenerations:

- the first $N(n)$ completed cycles (effectively generated under K) forbid visits to A
- the final cycle in progress at time n forbids visits to A up to time n but permits visits to A after time n

This difficulty can be overcome...

Leads to a logarithmically efficient estimator for $P(T > n)$

Conclusions

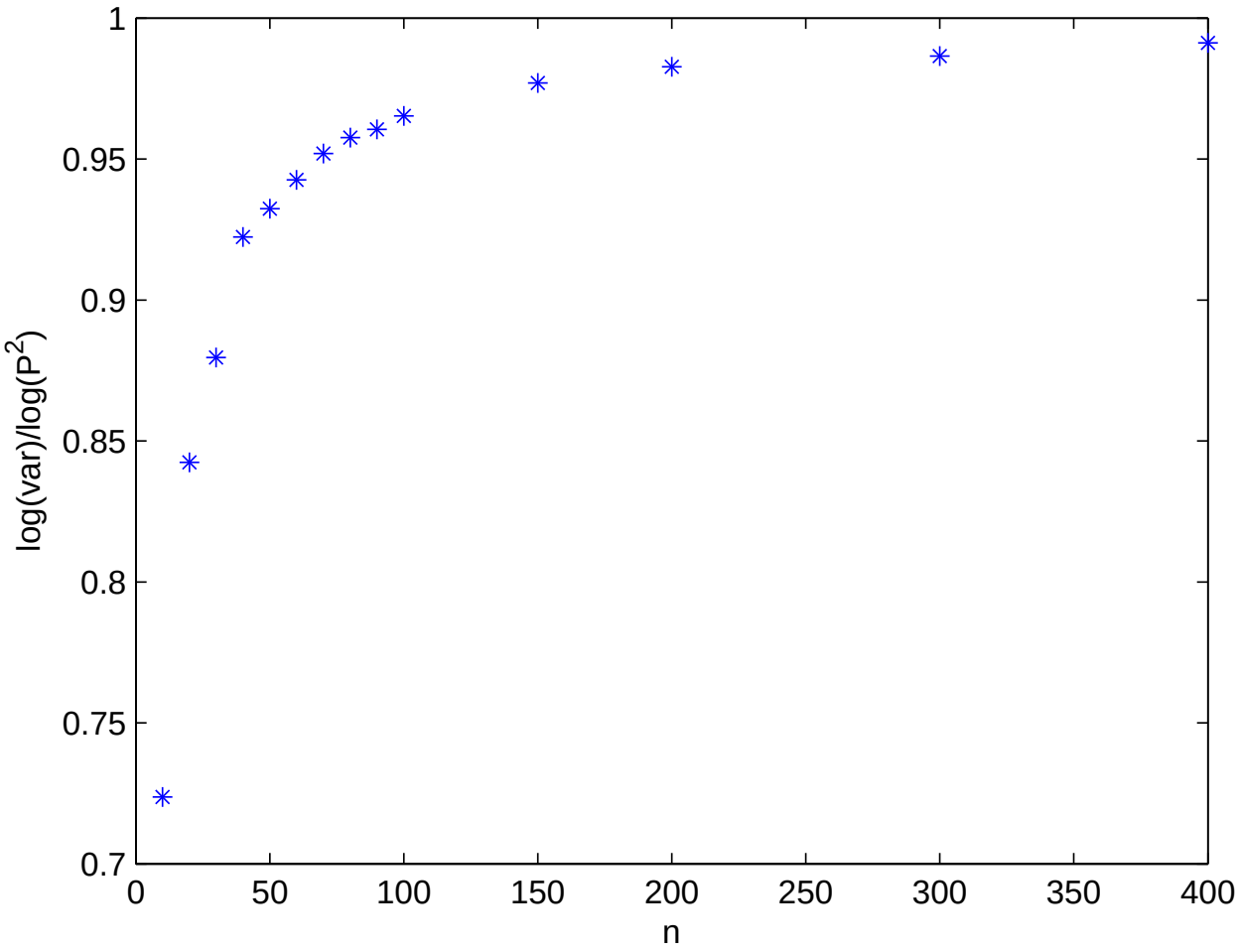
We have shown that regeneration (and associated path-sampling at the cycle level rather than at the level of individual steps) can offer a means of numerically computing large deviations for:

Markov-additive sums
exit time distributions

even in the presence of continuous state space

n	mean	var	log ratio
40	1.51E-15	4.54E-28	0.922361642
60	2.63E-21	1.60E-39	0.942553142
80	8.32E-30	2.04E-56	0.957551067
100	3.72E-37	4.67E-71	0.965306347
150	4.34E-56	6.64E-109	0.976986079
200	1.14E-76	5.50E-150	0.982691846

Table 1: $P(S_n \geq n\epsilon)$, where $S_n = \sum_{i=1}^{n} X_n$, $X_{n+1} = \rho X_n + Z_{n+1}$, $\rho = 0.5$, Z_n's are iid $N(0,1)$ and $\epsilon = 2$.


1
0.95
0.9
0.85
0.8
0.75
0.7
log(var)/log(P^2)
0
50
100
150
200
250
300
350
400
n