Importance sampling algorithms for the fork-join queue

Ad Ridder
Department Econometrics and Operations Research
Vrije University Amsterdam, Netherlands
email `aridder@feweb.vu.nl`

February 20, 2009

Abstract

In this paper we consider a rare-event problem in the fork-join queue for which we develop an efficient importance sampling algorithm.

Keywords: Rare Events, Importance Sampling, Fork-Join Queue, Universal Distributions.

1 Introduction

We consider the discrete-time Markov chain on the two-dimensional positive quadrant of integers that results by embedding a two-dimensional fork-join queue [5]. This is queueing system with a single Poisson arrival process with rate λ, and where any arriving job splits itself in two subjobs each joining a single server queue. These two queues act as independent $M/M/1$ queues with service rates μ_1, μ_2, respectively. For stability we demand $\lambda < \min(\mu_1, \mu_2)$. The folklore application of this queueing model is a system with two bathrooms, one for men and one for women, with arrivals of heterosexual couples only. The original motivation was to study a machine with parallel coupled processors, and an inventory control problem of data base systems.

The associated discrete-time Markov chain that results by embedding at jump times, is denoted by $(S(k))_{k=0}^{\infty}$ and has state space $\mathbb{Z}_+^2$, representing the backlogs at the two queues (including the servers). We are interested in transient probabilities of large backlogs of at least one of the queues:

$$\gamma_n(x, y, T) = \mathbb{P}(S_1(nT) \geq ny_1 \text{ or } S_2(nT) \geq ny_2 | S(0) = nx), \tag{1}$$

for fixed scaled initial state $x = (x_1, x_2) \in \mathbb{R}_+^2$, fixed scaled threshold $y = (y_1, y_2) \in \mathbb{R}_+^2$, fixed scaled horizon $T > 0$, and parameter $n \to \infty$. The set of interest is scaled by n and then called the rarity set:

$$D = \{\eta \in \mathbb{R}_+^2 : \eta_1 \geq y_1 \text{ or } \eta_2 \geq y_2\}.$$

The difficulty here is twofold: (i) the rarity set is not convex which may cause troubles for developing an efficient importance sampling scheme [4]; and (ii) the set D cannot decomposed in two disjoint sets such that the separate probabilities are estimated by efficient importance sampling estimators [6].

In this paper we investigate the method of universal simulation distributions for finding an efficient importance sampling estimator. This method has been introduced originally in [9], and further developed in [1], see also [2, Chapter 10]. The method is based on large deviations for sequences of random variables, however, we need to adapt it to process level because we deal with sample path large deviations. In Section 2 we shall briefly review the method of universal simulation distributions, and the sample path large deviations for the folk-join queue. In Section 3 we give our algorithm and we conclude with a few numerical results in Section 4.

2 Preliminaries

Universal simulation distributions.

Suppose that we can write the target of our problem to be

$$\gamma_n(x, y, T) = \mathbb{P}\left(f_n(Y_n)/n \in E\right),$$

where Y_n is a random variable on some appropriate space $\mathcal{S}_n$, and f_n a function $\mathcal{S}_n \to \mathbb{R}^d$. Let P_n be the probability measure on $\mathcal{S}_n$ induced by Y_n. Under some regularity conditions the Gärtner-Ellis theorem holds [2, Section 3.2],

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(f_n(Y_n)/n \in E\right) = -I(E).$$

The large deviations rate function $I(v), v \in \mathbb{R}^d$, is the convex conjugate of the asymptotic log moment generating function of $(f_n(Y_n))$, with optimising argument θ_v. Suppose that there is also a sequence $(Z_n)_n$ of $\mathcal{S}_n$-valued random variables, with induced probability measure Q_n, such that $P_n \ll Q_n$. The importance sampling estimator of the target probability γ_n is defined by

$$\widehat{\gamma}_n = \frac{dP_n}{dQ_n}(Z_n)\, 1\{f_n(Y_n)/n \in E\}. \tag{2}$$

The universal simulation distribution method says that if there are $m < \infty$ points $v_1, \ldots, v_m \in \mathbb{R}^d$, such that

$$\text{(i) } I(v_i) \geq I(E) \quad (i = 1, \ldots, m); \quad \text{(ii) } E \subset \cup_{i=1}^m \mathcal{H}(v_i), \tag{3}$$

where $\mathcal{H}(v)$ is the half-space $\{w \in \mathbb{R}^d : \langle \theta_v, w - v\rangle \geq 0\}$, then for any probability vector $\pi = (\pi_i)$ with positive elements, the change of measure

$$dQ_n(s) = \left(\sum_{i=1}^m \pi_i \exp\left(\langle \theta_{v_i}, f_n(s)\rangle - \psi_n(\theta_{v_i})\right)\right) dP_n(s)$$

gives an asymptotically optimal importance sampling estimator (2), see [1, 9].

The random walk associated with the fork-join queue.

The discrete-time Markov chain $(S(k))_{k=0}^\infty$ representing the backlogs at the queues at their jump times is a face-homogeneous random walk on the positive quadrant $\mathbb{Z}_+^2$, which means the following [8]: for any $s \in \mathbb{R}_+^2$, let $\Lambda(s)$ be the set of indices i for which $s_i > 0$. For any subset $\Lambda \subset \{1, 2\}$, define face F_Λ by

$$F_\Lambda = \{s \in \mathbb{R}_+^2 : \Lambda(s) = \Lambda\}.$$

Notice that $F_\emptyset = \{0\}$. The transition probabilities $p_{ss'}$ of the chain are the same for all s in the same face, and depend only the jump $s' - s$:

$$p_{ss'} = p_{\Lambda(s)}(s - s').$$

Thus, for our two-dimensional fork-join queue, there are four random variables X_Λ that represent the jumps.

Sample path large deviations.

Since the chain is ergodic, a sample path large deviations hold for continuous piecewise linear paths [3, 8]. Firstly, we define the scaled continuous-time processes $(S^{[n]}(t))_{0\le t\le T}$, $n = 1, 2, \ldots$, by $S^{[n]}(t) = S(nt)/n$ for $t = 0, 1/n, 2/n \ldots, T$ and linear interpolation in the other points. Secondly, consider the affine function $\phi(t) = x + vt$ for $t \in [0, T]$ with $x \in \mathbb{R}^d_+$ and $v \in \mathbb{R}^d$, such that $\phi(t) \in F_\Lambda$ for all $0 < t < T$. We call the gradient v the constant speed of the path. Then for the scaled processes $S^{[n]}$, starting in $S^{[n]}(0) = x$, it holds that there exists a locate rate $\ell_\Lambda(v)$ such that

$$\lim_{\epsilon\downarrow 0} \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\sup_{0\le t\le T} \left|S^{[n]}(t) - \phi(t)\right| < \epsilon \right) = -T\ell_\Lambda(v). \tag{4}$$

For continuous piecewise linear paths ϕ, such that each piece lies entirely in some face, we may apply (4) to each piece and add these 'costs'.

The main issue remains to identify the local rate functions $\ell_\Lambda(v)$. A general method has been developed in [7] for face-homogeneous random walks which can be applied to our two-dimensional fork-join queue. Clearly $\ell_\emptyset = 0$ because the process is ergodic. All the other local rate functions are convex conjugates of (adapted) log moment generating functions:

$$\ell_\Lambda(v) = \sup_{\theta\in\mathbb{R}} \big(\langle\theta, v\rangle - \psi(\theta)\big). \tag{5}$$

For the interior face $F_{\{1,2\}}$ it is just the function associated with the jump variable $X_{\{1,2\}}$, for the boudary faces $F_{\{1\}}$ and $F_{\{2\}}$ we have to consider both boundary and interior. The optimisation program (5) is solved numerically and the optimiser denoted by θ_v.

3 Importance sampling algorithm

We assume that the random walk starts off at state 0. Then we rewrite the target probability (1) as

$$\gamma_n(0, y, T) = \mathbb{P}\left(S(nT)/n \in D | S(0) = 0\right) = \mathbb{P}\left(S^{[n]} \in E\right),$$

where E is an appropriate set of absolute continuous paths $\phi : [0, T] \to \mathbb{R}^2_+$ with specifically $\phi(0) = 0$ and $\phi(T) \in D$. Let $\tilde{E} \subset E$ be the subset of piecewise linear paths of the following form.

> $\phi = \phi_{\tau,v}$: it stays in 0 until time τ $(0 \leq \tau < T)$ and then it goes straight at constant speed v to the point $\eta = (T-\tau)v \in D$.

The theory of sample path large deviations [3, 7, 10] says that

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(S^{[n]} \in E\right) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(S^{[n]} \in \tilde{E}\right) = -I(\tilde{E}).$$

So, we are in the business of the universal simulation distributions by finding $m < \infty$ pairs $\left(\tau, v^{(i)}\right)$ $(i = 1, \ldots, m)$ (same τ!) such that (see (3)),

$$\text{(i) } I(\phi^{(i)}) \geq I(\tilde{E}) \quad (i = 1, \ldots, m); \quad \text{(ii) } V_\tau \subset \cup_{i=1}^m \mathcal{H}(v^{(i)}),$$

where

(a.) $\phi^{(i)} = \phi_{\tau,v^{(i)}} \in \tilde{E}$ with cost $I(\phi^{(i)}) = \tau \ell_{\Lambda(v^{(i)})}(v^{(i)})$;

(b.) $V_\tau = \{v \in \mathbb{R}^2_+ : (T-\tau)v \in D\}$, set of speed vectors to the rarity set;

(c.) $\mathcal{H}(v)$ is the half space $\{w \in \mathbb{R}^d : \langle \theta_v, w - v\rangle \geq 0\}$, where θ_v is the optimiser for the local rate function $\ell_{\Lambda(v)}(v)$, see (5).

Hence, the remarkable observation is that we apply the method to drift vectors of affine paths of the (limiting scaled) random walk. From (b) and (c) we deduce that two pairs suffice, one on each of the two boundaries:

$$\begin{cases} v_1^{(1)} = y_1/(T-\tau), \quad 0 \leq v_2^{(1)} < y_2/(T-\tau); \\ 0 \leq v_1^{(2)} < y_1/(T-\tau), \quad v_2^{(2)} = y_2/(T-\tau); \\ (\theta_{v^{(1)}})_2 = 0, \quad (\theta_{v^{(2)}})_1 = 0. \end{cases} \tag{6}$$

However, in most examples there are no two such pairs, but there are two pairs with different zero-sojourn times $\tau^{(i)}$ that satisfy (6). For each of the drift vectors $v^{(i)}$ we determine the associated jump probabilities $q_\Lambda^{(i)}$ of the jump variables X_Λ by an exponential change of measure given by the shift parameters $\theta_{v^{(i)}}$.

4 Example

Let $\lambda = 1, \mu_1 = 1.5, \mu_2 = 2, x = (0.0), y = (1, 1.2), T = 10$. The optimal path $\phi^* = \arg\inf_{\phi \in E} I(\phi)$, has a zero-sojourn time of $\tau = 2$ time units and then runs at a constant speed $v = (1/8, 0)$ along face $F_{\{1\}}$ to state $(1, 0)$.

The naive importance sampling algorithm is based on simulating the random walk with the original jump probabilities p_Λ until time $n\tau$, and next with the exponentially twisted jump probabilities obtained by the parameter θ_v until time nT. It can be shown that the associated estimator is not efficient. Its behaviour is illustrated by plotting the estimated target probability, and the estimated variance of the estimator as functions of the sample size k. The estimates are irregular overestimates of the true value, and the variances get a shock upward whenever the rare event is hit by a 'wrong' path.

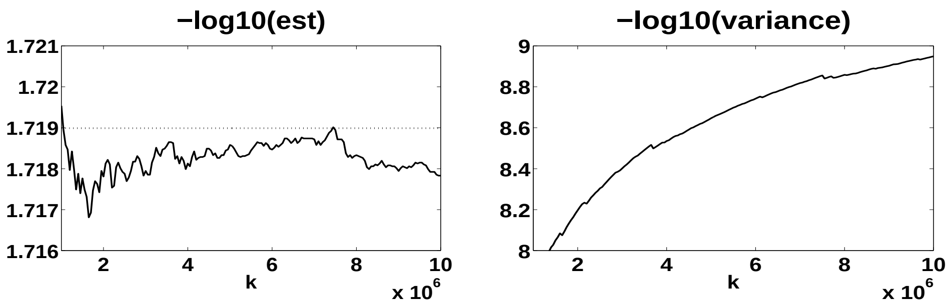

Figure 1. *Estimates from the naive importance sampling algorithm for sample sizes* $k = 10^5$–10^7. *Scaling* $n = 10$. *The dotted line is the exact probability.*

We find numerically a solution of the system (6) with different zero-sojourn times. The first pair is $\left(\tau^{(1)}, v^{(1)}\right) = (2, (1/8, 0))$, which gives the same path ϕ^* as above, the other pair is $\left(\tau^{(2)}, v^{(2)}\right) = (4.6, (1/9, 2/9))$, which gives a path running in the interior face $F_{\{1,2\}}$. In the simulations we used mixing probabilties $(0.8, 0.2)$.

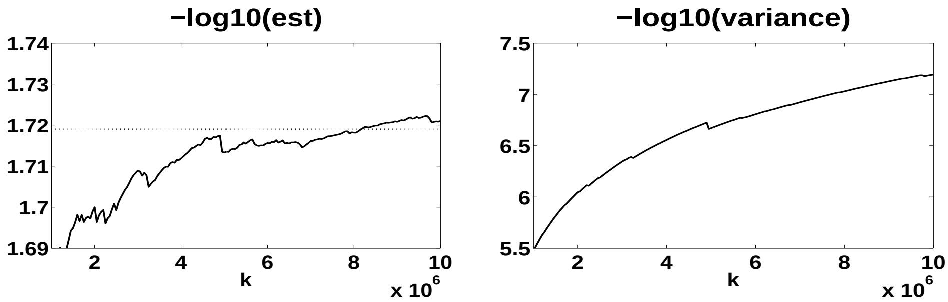

Figure 2. *Estimates from the importance sampling algorithm with universal simulation distribution for sample sizes* $k = 10^5$–10^7*. Scaling* $n = 10$*. The dotted line is the exact probability.*

We experimented with various scalings $n = 10$–200 with sample size $k = 50000$ and collected the relative half width of the 95% confidence interval RHW (efficient estimators have RHW that grows at most polynomially), and the ratio $\mathrm{RAT} = \log \mathbb{E}[(\widehat{\gamma_n})^2]/\log \mathbb{E}[\widehat{\gamma_n}]$ (efficient estimators have RAT that converge to 2).

The conclusion is that the mixed importance sampling estimator improves the naive one, and is asymptotically optimal, although its variance is still irregular. Further investigations are needed, also for other starting points, and other algorithms, for instance with mixing probailities that depend on state and time [4].

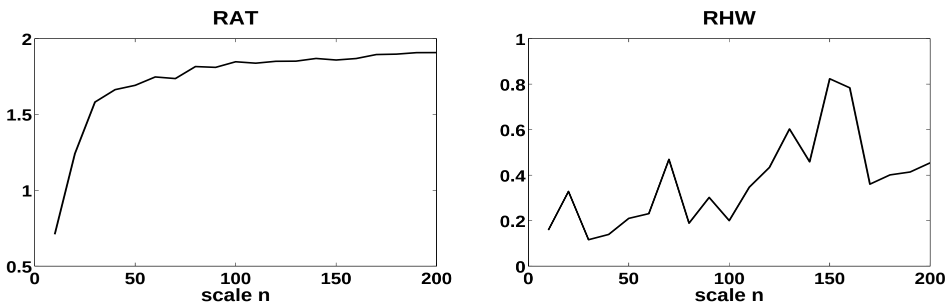

Figure 3. *Performance of the importance sampling estimator with universal simulation distribution for scalings* $n = 10$–200.

References

[1] Bucklew, J.A., Nitinarawat, S., Wierer, J., 2004. Universal simulation distributions, *IEEE Transactions on Information Theory* 50, pp. 2674-2685.

[2] Bucklew, J.A., 2004. *Introduction to Rare Event Simulation.* Springer, New York.

[3] Dupuis, P., and Ellis, R., 1995. The large deviation principle for a general class of queueing systems, *Transactions of the American Mathematical Society* 347, pp. 2689-2751.

[4] Dupuis, P., and Wang, H., 2007. Subsolutions of an Isaacs equation and efficient schemes for importance sampling. *Mathematics of Operations Research* 32, pp. 723-757.

[5] Flatto, L., and Hahn, S., 1984. Two parallel queues created by arrivals with two demands, *SIAM Journal on Applied Mathematics* 44, pp. 1041-1053.

[6] Glassermann, P., and Wang, Y., 1997. Counterexamples in importance sampling for large deviations probabilities, *Annals of Applied Probability* 7, pp. 731-746.

[7] Ignatiouk-Robert, I., 2001. Sample path large deviations and convergence parameters, *Annals of Applied Probability* 11, pp. 1292-1329.

[8] Ignatiouk-Robert, I., 2005. Large deviations for processes with discontinous statistics, *Annals of Probability* 33, pp. 1479-1508.

[9] Sadowsky, J.S., and Bucklew, J.A., 1990. On large deviations theory and asymptotically efficient Monte Carlo estimation. *IEEE Transactions on Information Theory* 36, pp. 579-588.

[10] Shwartz, A., and Weiss, A., 1995. *Large Deviations for Performance Analysis*, Chapman & Hall.