Journal of Statistical Planning and Inference 85 (2000) 127–135

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Simulating the maximum of a random walk

Katherine Bennett Ensor[a, 1], Peter W. Glynn[b, *, 2]

[a] Department of Statistics, Rice University, Houston, TX 77251-1892, USA

[b] Terman Engineering Center, Department of Engineering-Economic Systems and Operations Research, Stanford University, Stanford, CA 94305-4023, USA

Abstract

In this paper, we show how to exactly sample from the distribution of the maximum of a random walk with negative drift. We also explore related variance reduction methods.

MSC: primary 60J15; 60K25; 60-04; secondary 60J10; 90B22; 62L10

Keywords: Random walk; Single-server queue; Simulation; Stationary distribution

1. Introduction

Let $X = (X_n\colon n \geqslant 1)$ be a sequence of real-valued independent identically distributed (i.i.d.) random variables. Put $S_0 = 0$, $S_n = X_1 + \cdots + X_n$, so that $S = (S_n\colon n \geqslant 0)$ is the associated random walk. Our concern, in this paper, is with the development of an exact sampling technique for generating i.i.d. replicates of the random variable M, where

$$M = \max\{S_n\colon n \geqslant 0\}.$$

Obviously, if $EX_1 > 0$ or if $EX_1 = 0$ with $\operatorname{var} X_1 > 0$, then $M = +\infty$ a.s. Hence, our focus here is on the case in which S has negative drift (or, equivalently, $EX_1 < 0$).

The random variable M plays a key role in the Wiener–Hopf theory for random walk (see, for example, Chapter 8 of Chung, 1974). In addition, the study of M arises naturally in several important applications. For example, the tail probability $\alpha = P(M > x)$ arises in the risk setting, in which α can be interpreted as the probability that an insurance company with initial capitalization x will eventually go bankrupt

* Corresponding author. Tel.: +1-650-725-0554; Fax: +1-650-723-4107.

E-mail address: glynn@leland.stanford.edu (P.W. Glynn)

[1] This author acknowledges support from the National Science Foundation under Grant No. DMS-9704648.

[2] The research of this author was supported by the Army Research Office under Contract No. DAAL03-91-G-0319 and the National Science Foundation under Grant No. DMS-9704732.

PII: S0378-3758(99)00075-0

(see, for example, Asmussen, 1987). The probability $\alpha = P(M > x)$ also is of interest in the consideration of "one-sided" sequential probability ratio tests (see Siegmund, 1985, Chapter 8 for details).

Furthermore, in queueing theory, the random variable M has a distribution identical to that of the random variable $W(\infty)$, where $W(\infty)$ is the steady-state waiting time associated with the single-server G/G/1 queue. In particular, if W_n is the waiting time (exclusive of service) of the nth customer in the single-server first-come first-serve queue in which the inter-arrival times are i.i.d., as are the service times, then (provided the mean inter-arrival time is greater than the mean service time) $W = (W_n : n \geqslant 0)$ is a Markov chain for which $W_n \Rightarrow W(\infty)$ as $n \to \infty$, where $W(\infty) \stackrel{\mathscr{D}}{=} M$ ($\stackrel{\mathscr{D}}{=}$ denotes equality in distribution). Here, the random variable X_i is formed by subtracting the $(i+1)$st inter-arrival time from the ith service time. (See Asmussen, 1987 for details.)

One approach to generating M is, of course, to approximate M by $M_n = \max\{S_k : 0 \leqslant k \leqslant n\}$ with n large. However, this approximation introduces bias into any estimate corresponding to expectations associated with M. Perhaps even more serious is that generating the random variable M_n requires simulating the random walk for n time units and the resulting estimator is therefore computationally expensive for n large.

An alternative is to simulate the Markov chain W and to average over the chain's trajectory in order to estimate the expectation of any given functional of M. While this approach is widely used in practice, it suffers from the fact that the estimator produced tends to be biased, due to the inability to generate a stationary version of W. In addition, the values over which the estimator is averaged are highly correlated due to the Markov dependence inherent in W.

Our main contribution in this paper is Theorem 1, in which we describe how to generate the random variable M in finite time, or equivalently how to generate a stationary version of the Markov chain W. Thus, our contribution can be viewed as being in the same spirit as the recent activity in which various researchers have produced algorithms intended to simulate stationary versions of certain Markov chains that arise in applied probability, statistical physics, and Bayesian statistics; see Asmussen et al. (1992), and Propp and Wilson (1996) for representative examples. Theorem 1 also may be of interest in its own right as a potential theoretical tool for studying M, as it provides a "coupling" between M and a certain closely related exponential random variable. In addition, the ability to generate i.i.d. replicates of M (or equivalently $W(\infty)$) allows one to apply standard nonparametric methodology to the estimation of various functionals corresponding to M. For example, one can now potentially estimate quantiles of $W(\infty)$ by appealing to confidence intervals based on order statistics (see, for example, p. 103 of Serfling, 1980); such intervals remove the need to estimate certain statistically challenging parameters (e.g. the density of $W(\infty)$).

This paper is organized as follows. Section 2 establishes our main result, namely Theorem 1, in which we provide a means of simulating i.i.d. replicates of M. In Section 3, we provide some refinements that are pertinent to obtaining variance reductions in computing expectations associated with M; these variance reduction methods are closely related to ones introduced by Asmussen (1990). In Section 4, we apply our estimators to the examples considered in Asmussen (1990). Finally, in Section 5, we compare the efficiency of the various estimators.

2. The basic idea

Set $\psi(\theta) = \log E \exp(\theta X_1)$ and assume that

A. (i) There exists $\theta^* > 0$ such that $\psi(\theta^*) = 0$;
(ii) There exists $\varepsilon > 0$ such that $\psi(\theta^* + \varepsilon) < \infty$.

By the convexity of $\psi(\cdot)$, such a root θ^* will necessarily be unique. While A fails to hold for random variables X_1 for which $P(X_1 > x)$ decreases as a power of x as $x \to \infty$, it typically is in force for distributions in which the right tail goes to zero at least exponentially fast (e.g. bounded random variables and normal random variables).

Put

$$F^*(\mathrm{d}x) = \exp(\theta^* x) P(X_1 \in \mathrm{d}x)$$

for $x \in \mathbb{R}$, and note that F^* is a probability, due to the definition of θ^*. Let P^* be the probability measure under which $X = (X_n\colon n \geqslant 1)$ evolves as an i.i.d. sequence with increment distribution given by $P^*(X_i \in \cdot) = F^*(\cdot)$, we further suppose that our probability space is sufficiently rich so as to support an additional random variable U which is independent of X under P^* and has the uniform distribution. Let $E^*(\cdot)$ be the expectation operator corresponding to P^*.

It is easily verified that under P^*, $(S_n\colon n \geqslant 0)$ is a random walk having positive drift. In fact, $E^* X_1 = \psi'(\theta^*)$ (and the convexity of $\psi(\cdot)$, together with the observation that $\psi(0) = \psi(\theta^*) = 0$, ensures that $\psi'(\theta^*) > 0$). As a consequence, $N(t) < \infty$ P^* a.s., where $N(t) = \min\{n \geqslant 1 : S_n > t\} - 1$. (We define $N(t)$ in this way to preserve the notational similarity with renewal counting processes in which the X_i's are non-negative.) Furthermore, it is well known (see, for example, Chapter 12 of Asmussen, 1987) that

$$P(M > x) = E^* \exp(-\theta^* S_{N(x)+1}). \tag{2.1}$$

Noting that $0 \leqslant \exp(-\theta^* S_{N(x)+1}) \leqslant 1$, we can rewrite (2.1) as

$$\begin{aligned} P(M > x) &= E^*[P^*(U \leqslant \exp(-\theta^* S_{N(x)+1}) | S_n : n \geqslant 0)] \\ &= P^*(U \leqslant \exp(-\theta^* S_{N(x)+1})) \\ &= P^*\left(S_{N(x)+1} \leqslant -\frac{1}{\theta^*} \log U\right). \end{aligned} \tag{2.2}$$

Put $Z = -(1/\theta^*)\log(U)$ and note that Z is exponentially distributed with parameter θ^*. To proceed further, we define the (strict) ascending ladder height epochs of the random walk $(S_n\colon n \geqslant 0)$ under P^* via the formula

$$\tau(n+1) = \inf\{m > \tau(n)\colon S_m > S_{\tau(n)}\},$$

subject to the initial condition $\tau(0) = 0$. For $x \geqslant 0$, let

$$\tilde{N}(x) = \max\{n \geqslant 0\colon S_{\tau(n)} \leqslant x\}$$

be the number of (strict) ascending ladder heights in $(0, x]$, and set

$$\Gamma = S_{\tau(\tilde{N}(Z))}.$$

Observe that because $S_{N(x)+1}$ must coincide with a (strict) ascending ladder height, $S_{N(x)+1} = S_{\tau(\tilde{N}(x)+1)}$ so that

$$\begin{aligned} P^*(S_{N(x)+1} \leqslant Z) &= P^*(S_{\tau(\tilde{N}(x)+1)} \leqslant Z) \\ &= P^*(S_{\tau(\tilde{N}(Z))} > x) \\ &= P^*(\Gamma > x). \end{aligned}$$

Relation (2.2) therefore yields the following result.

Theorem 1. *Under* A, $P(M \in \cdot) = P^*(\Gamma \in \cdot)$.

Consequently, M has the same distribution as the last strict ascending ladder height of $(S_n\colon n \geqslant 0)$ under P^* prior to the random walk's passing above the level Γ. Therefore, under assumption A, we may simulate i.i.d. replicates of M associated with distribution P, by equivalently simulating i.i.d. replicates of Γ under P^*. It follows that the algorithm below produces a replicate of M (associated with distribution P):

1. Generate an exponential random variable Z having parameter θ^*.
2. Independently simulate the random walk $(S_n : n \geqslant 0)$, using increment distribution F^*, until $N(Z) + 1$.
3. Set $M = S_{\tau(\tilde{N}(Z))}$.

In terms of computational efficiency, it should be noted that the random walk needs to be simulated to time $N(Z) + 1$. By Wald's identity,

$$E^* S_{N(x)+1} = E^*(N(Z) + 1) E^* X_1.$$

If $P(X_i \leqslant a) = 1$ for some $a < \infty$ (so that the positive increments of the random walk are bounded above by a under both P and P^*), then

$$|S_{N(x)+1} - Z| \leqslant a,$$

so that

$$\frac{1}{\psi'(\theta^*)\theta^*} \leqslant E^*(N(Z) + 1) \leqslant \frac{1}{\psi'(\theta^*)} \left(\frac{1}{\theta^*} + a \right). \tag{2.3}$$

The bounds (2.3) suggest that a rough measure of the number of X_i's needed to generate M is something on the order of $1/(\psi'(\theta^*)\theta^*)$.

3. A refinement

As indicated in the Introduction, the ability to generate M in finite time guarantees that one can simulate time-stationary versions of the waiting time sequence $(W_n : n \geqslant 0)$ corresponding to the G/G/1 queue.

However, suppose that the task at hand is instead to compute, for a given functional $f : [0, \infty) \to \mathbb{R}$, the expectation

$$\alpha = E\, f(M).$$

Noting that

$$0 \leqslant Z - S_{\tau(\tilde{N}(Z))} \leqslant X_{N(Z)+1},$$

it is evident that $S_{\tau(\tilde{N}(Z))}$ is strongly correlated with Z. This suggests using Z as a control variate, much as in the spirit of Asmussen (1990). Assuming that $E^* f(Z)$ can be computed (either analytically or numerically), it is clearly the case that

$$\chi(\lambda) = f(S_{\tau(\tilde{N}(Z))}) - \lambda(f(Z) - E^*\, f(Z))$$

has mean α under P^*. (The quantity $f(Z) - E^*\, f(Z)$ is known as a "control variate" in the simulation literature, and λ is the associated "control coefficient"; see Bratley et al. (1987) for details.) The optimal control coefficient λ^* which minimizes the variance of $\chi(\cdot)$ is given by

$$\lambda^* = \frac{\mathrm{cov}^*(f(S_{\tau(\tilde{N}(Z))}), f(Z))}{\mathrm{var}^*\, f(Z)}$$

(where the notation $\mathrm{cov}^*(\cdot)$, $\mathrm{var}^*(\cdot)$ reflects the fact that the covariance and variance must be computed under P^*). Since λ^* is typically unknown, it can be estimated by substituting the sample covariance and sample variance of the $(f(\Gamma_i), f(Z_i))$'s produced by simulating replicates of (Γ, Z), thereby producing an estimator λ_n that is consistent for λ^*. The corresponding point estimate for α, based on a sample of size n, is

$$\hat{\alpha}_2 = \frac{1}{n} \sum_{i=1}^{n} \chi_i(\lambda_n). \tag{3.1}$$

Note that estimating λ^* in this way introduces bias into the estimator $\hat{\alpha}_2$ of α. If one wishes to avoid the bias and/or the additional computation associated with calculating λ_n, a reasonable choice of λ is to use $\lambda = 1$.

With this choice of λ, the estimator for $\alpha = E\, f(M)$ takes the form

$$\hat{\alpha}_1 = E^*\, f(Z) + \frac{1}{n} \sum_{i=1}^{n} (f(\Gamma_i) - f(Z_i)). \tag{3.2}$$

Another class of estimators for α can be derived by noting that

$$\begin{aligned} E\, f(M) &= E^*\, f(S_{\tau(\tilde{N}(Z))}) \\ &= E^*[E^*[f(S_{\tau(\tilde{N}(Z))}) | S_n\colon n \geqslant 0]] \end{aligned}$$

$$= E^*\left[\int_0^\infty f(S_{\tau(\tilde{N}(t))})\theta^* e^{-\theta^* t}\, dt\right]$$
$$= \theta^* E^*\left[\int_0^\infty f(S_{\tau(\tilde{N}(t))})P^*(Z > t)\, dt\right]$$
$$= \theta^* E^*\left[E^*\left[\int_0^\infty f(S_{\tau(\tilde{N}(t))})I(Z > t)\, dt \mid S_n: n \geqslant 0\right]\right]$$
$$= \theta^* E^*\left[E^*\left[\int_0^Z f(S_{\tau(\tilde{N}(t))})\, dt \mid S_n: n \geqslant 0\right]\right]$$
$$= \theta^* E^* \int_0^Z f(S_{\tau(\tilde{N}(t))})\, dt.$$

Consequently, $\alpha = E f(M)$ can be estimated by simulating i.i.d. replicates of the random variable

$$\theta^* \int_0^Z f(S_{N(t)})\, dt \tag{3.3}$$

under P^*. A natural control variate to use in conjunction with (3.3) is

$$\theta^* \int_0^Z f(t)\, dt - E^* f(Z). \tag{3.4}$$

This random variable has mean zero and can be used analogously to how $(f(Z) - E^* f(Z))$ was used earlier to "control" $f(\Gamma)$. Denote the resulting estimators by $\widetilde{\alpha}_0$, $\widetilde{\alpha}_1$ and $\widetilde{\alpha}_2$, corresponding to no control variate (control coefficient is zero), control coefficient of one, and control coefficient estimated from the observed series.

Estimator (3.3) is closely related to one introduced by Asmussen (1990, Eq. (4.1)). However, his approach requires that f be differentiable and works with f' directly.

4. Examples

As an illustration of the proposed estimators, we consider estimation of $\alpha = E(M)$ (so that $f(t) = t$) for the M/M/1, M/D/1 and D/M/1 queues considered in Asmussen (1990). The parameters for each queue are given in Table 1. The rates are given for the exponential random variables, whereas the mean is given for the deterministic components. Also, provided in Table 1 are the parameters under P^*. The

Table 1
Examples considered

Queue	Parameters under P			Parameters under P^*		
	T	U	μ	θ^*	T	U
M/M/1	0.9	1	9	0.1	1	0.9
M/D/1	0.9	1	4.5	0.207147	1.107147	1
D/M/1	1/0.9	1	4.179	0.193100	1/0.09	0.806900

Table 2
Point estimates (PI) and standard errors (SE) for M/M/1, M/D/1, and D/M/1 queues

n		$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\tilde{\alpha}_0$	$\tilde{\alpha}_1$	$\tilde{\alpha}_2$	$\bar{k}$
				M/M/1 queue with $\mu = 9$				
100	PE	8.671	9.113	9.114	7.757	9.136	9.055	95.3
	SE	0.910	0.080	0.079	1.444	0.098	0.041	
1000	PE	9.064	9.001	9.002	8.884	8.993	8.988	101.1
	SE	0.309	0.030	0.030	0.616	0.037	0.021	
9000	PE	8.847	9.010	9.008	8.674	9.023	9.008	99.1
	SE	0.103	0.010	0.010	0.218	0.013	0.007	
				M/D/1 queue with $\mu = 4.5$				
100	PE	4.760	4.541	4.542	4.861	4.484	4.493	61.2
	SE	0.498	0.022	0.022	1.285	0.036	0.016	
1000	PE	4.456	4.502	4.502	4.559	4.502	4.504	52.8
	SE	0.156	0.008	0.008	0.351	0.011	0.005	
9000	PE	4.525	4.497	4.497	4.604	4.500	4.500	53.8
	SE	0.052	0.002	0.002	0.118	0.004	0.002	
				D/M/1 queue with $\mu = 4.179$				
100	PE	3.897	4.258	4.235	3.565	4.216	4.135	43.5
	SE	0.468	0.102	0.101	0.764	0.127	0.075	
1000	PE	4.512	4.227	4.237	4.967	4.116	4.193	51.4
	SE	0.177	0.032	0.032	0.392	0.047	0.028	
9000	PE	4.168	4.173	4.172	4.129	4.178	4.173	50.0
	SE	0.053	0.011	0.011	0.105	0.014	0.009	

random walk is simulated based on the increment distribution $X_1 = U - T$ defined under P^*, where U and T are independent random variables representing the service and inter-arrival times, respectively.

The proposed point estimators are based on n i.i.d. replicates of $S_{\tau(\tilde{N}(Z))}$ or (3.3) (refer to these observations as $y_1,\dots,y_n$) and the respective controls (referred to as $q_1,\dots,q_n$); and thus for large n, derive their distributional properties from the central limit theorem. An estimate of the standard error is given by s^2/n where: $s^2 = s_y^2$, the sample variance of $y_1,\dots,y_n$ if no control variable is used (i.e., estimates $\hat{\alpha}_0$ and $\tilde{\alpha}_0$); $s^2 = s_d^2$, the sample variance of $d_i = y_i - q_i$ if the control variable is used with control coefficient set to one ($\hat{\alpha}_1$ and $\tilde{\alpha}_1$), and $s^2 = (1-r^2)s_y^2$, with r denoting the sample correlation coefficient between $y_1,\dots,y_n$ and $q_1,\dots,q_n$, if the control coefficient λ^* is estimated ($\hat{\alpha}_2$ and $\tilde{\alpha}_2$). Application of these six estimators to the three examples given in Table 1 is summarized in Table 2. As a measure of the computation time required to simulate a replicate of M for each example, the mean number of increments in the n simulated random walks ($\bar{k}$) is reported in Table 2.

From Asmussen (1990), application of the Minh–Sorli estimator (see Asmussen, 1990, pp. 1855 and 1890) based on 9000 cycles to the M/M/1 queue yielded an estimate of 8.993 with a standard error of approximately 0.0165 compared to our observed standard error for $\tilde{\alpha}_2$ of 0.007 based on 9000 replications. For the D/M/1 queue, Asmussen reports an estimate of 4.593 with a standard error of approximately 0.0015 when estimating μ using the Minh–Sorli method. Again, this standard error is comparable to our observed standard error for $\tilde{\alpha}_2$ of 0.009 based on 9000 replications.

5. Efficiency comparison

Given the various estimators that we have introduced here, it is of some interest to compare them from an efficiency standpoint. An appropriate "figure of merit" to use in making such comparisons is the product of the mean time required to generate each observation with the variance per observation (Glynn and Whitt, 1992). We will compare this "figure of merit" in a certain asymptotic regime.

Specifically, our interest lies in the behavior of the estimators when the drift of the random walk under P is small and negative, so that the random walk is close to null recurrent; this corresponds, in the queueing setting, to the queue being in "heavy traffic". To be precise, let V be a spread-out random variable having zero mean, and suppose that its moment generating function converges in a neighborhood of the origin. We further require $\mathrm{var}(V) > 0$. Set $\psi_V(\theta) = \log E \exp(\theta V)$. Then, for $\varepsilon > 0$ and small,

$$P_\varepsilon(\mathrm{d}x) = \exp(-\varepsilon x - \psi_V(-\varepsilon))P(V \in \mathrm{d}x)$$

is the distribution of a random variable having mean $\psi_V'(-\varepsilon) < 0$. Suppose that

$$P(X_1 \in \cdot) = P_\varepsilon(\cdot)$$

for ε small and positive. Our interest is in considering the efficiency issue as $\varepsilon \downarrow 0$ for the choice $f(x) = x$ (in which case we are computing EM).

Firstly, the computer time required to generate M via our algorithm of Section 2 is of order ε^{-2} (see Asmussen, 1990, Theorem 2). In addition, it is a standard result that εM converges in distribution to an exponential random variable as $\varepsilon \downarrow 0$. This suggests that $\mathrm{var}(M)$ is of order ε^{-2} (see Asmussen, 1990 for complete details). Consequently, the estimator, $\hat{\alpha}_0$, based on averaging copies of M (and using our algorithm to generate the M_i's) has a figure of merit of order ε^{-4}. This matches the figure of merit associated with estimating EM by simulating the Markov chain $(W_n : n \geqslant 0)$; see Asmussen (1992). Turning now to the estimator $\hat{\alpha}_1$ given by (3.2), $\mathrm{var}^*(\Gamma - Z)$ is $O(1)$ as $\varepsilon \downarrow 0$ (see the proof of Theorem 1 of Asmussen, 1990); the figure of merit for this estimator is of order ε^{-2} as $\varepsilon \downarrow 0$. Finally, the estimator based on averaging replicates of

$$E^* f(Z) + \theta^* \int_0^Z [f(S_{N(t)}) - f(t)]\, \mathrm{d}t$$

has a variance that is $O(1)$ as $\varepsilon \downarrow 0$. (Again, see Theorem 1 of Asmussen (1990) for a similar proof.) Hence, the figure of merit for the estimator ($\tilde{\alpha}_1$) constructed by combining (3.3) and (3.4), is of order ε^{-2} as $\varepsilon \downarrow 0$. However, it can be shown that if one estimates the optimal control coefficient in combining (3.3) with the control (3.4) (estimator $\tilde{\alpha}_2$) that the variance is then of order $O(\varepsilon)$, so that the figure of merit is of order ε^{-1} as $\varepsilon \downarrow 0$. Summarizing the above discussion:

the figure of merit defined above, as $\varepsilon \downarrow 0$, it is of the order of ε^{-4} for estimator $\hat{\alpha}_0$, ε^{-2} for estimators $\hat{\alpha}_1$ and $\tilde{\alpha}_1$ and ε^{-1} for estimator $\tilde{\alpha}_2$.

Thus the performance of $\tilde{\alpha}_2$ matches the performance of any of the estimators considered in Asmussen (1990), including the "heavy traffic" estimator of Minh and Sorli (1983).

References

Asmussen, S., 1987. Applied Probability and Queues. Wiley, New York.

Asmussen, S., 1990. Exponential families and regression in the Monte Carlo study of queues and random walks. Ann. Statist. 18, 1851–1867.

Asmussen, S., 1992. Queueing simulation in heavy traffic. Math. Oper. Res. 17, 84–111.

Asmussen, S., Glynn, P.W., Thorisson, H., 1992. Stationary detection in the initial transient problem. ACM Trans. Modelling Comput. Simulation 2, 130–157.

Bratley, P., Fox, B.L., Schrage, L., 1987. A Guide to Simulation. Springer, New York.

Chung, K.L., 1974. A Course in Probability Theory. Academic Press, New York.

Glynn, P.W., Whitt, W., 1992. The asymptotic efficiency of simulation estimators. Oper. Res. 40, 505–520.

Minh, D.L., Sorli, R.M., 1983. Simulating the GI/G/1 queue in heavy traffic. Oper. Res. 31, 966–971.

Propp, J.G., Wilson, D.B., 1996. Exact sampling with coupled Markov chains and applications to statistical mechanics. Random Struct. Algorithms 9, 223–252.

Serfling, R.J., 1980. Approximation Theorems in Statistics. Wiley, New York.

Siegmund, D., 1985. Sequential Analysis. Springer, New York.