Proceedings of the 2010 Winter Simulation Conference
B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, eds.

A REGENERATIVE BOOTSTRAP APPROACH TO ESTIMATING THE INITIAL TRANSIENT

Peter W. Glynn
Xiaowei Zhang

Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, USA

ABSTRACT

We propose a new algorithm for identifying the duration of the initial transient for a regenerative stochastic process. The algorithm involves re-sampling of the simulated cycles, and therefore has a "bootstrap" flavor. The paper includes a derivation of the estimator for the duration of the transient that offers theoretical support for its validity, and provides a preliminary numerical investigation of the estimator's properties.

1 INTRODUCTION

Consider a real-valued stochastic process $Y = (Y(t) : t \geq 0)$ for which there exists a (deterministic) constant α such that

$$\overline{Y}(t) \triangleq t^{-1} \int_0^t Y(s)\,\mathrm{d}s \Rightarrow \alpha$$

as $t \to \infty$, where $\Rightarrow$ denotes weak convergence. The quantity α is then called the *steady-state mean* of Y, and the *steady-state simulation problem* is considered with the efficient Monte Carlo computation of the quantity α.

One fundamental difficulty associated with steady-state simulation is that the estimator $\overline{Y}(t)$ is biased as an estimator of α. In particular, the fact that Y is typically initialized (through a specific choice of the distribution of $Y(0)$) with an initial condition that is atypical of equilibrium behavior implies that Y will generally exhibit an "initial transient" up to the time at which Y equilibrates. Any statistics collected over this initial transient phase will contaminate the estimation of the steady-state quantity α. As a consequence, it is of significant interest to identify the duration of the initial transient phase.

A wide variety of methods for estimating the initial transient have been proposed in the steady-state simulation literature. We differentiate such methods from *low bias estimation methods*, which are intended to produce estimators with better bias characteristics than those associated with the time-average of Y; see, for example, Meketon and Heidelberger (1982), Glynn and Heidelberger (1990), Glynn and Heidelberger (1992), Glynn (1994), Hsieh, Iglehart, and Glynn (2004), and Awad and Glynn (2007). Representative papers that focus on estimating the initial transient are Conway (1963), Fishman (1971), Kelton and Law (1983), Yücesan (1993), White, Cobb, and Spratt (2000); see Robinson (2007) for a recent survey. Most of the existing literature lacks a fully rigorous theoretical support. By contrast, the approach introduced here has a solid theoretical underpinning, as suggested through the derivation provided in Section 2. (A full theoretical exploration of the method's theoretical properties will be provided elsewhere.) A different approach, also with theoretical support, can be found in Awad and Glynn (2010). A discussion of numerical implementation issues, as well as empirical investigation of the method's properties, is the subject of Section 3. Section 4 provides some concluding remarks.

2 OUR PROPOSED ALGORITHM

Throughout this paper, we assume that $Y = (Y(t) : t \geq 0)$ is a (non-delayed) real-valued regenerative process with respect to the regeneration times $0 = T(0) < T(1) < \cdots < T(n-1) < T(n) < T(n+1) < \cdots$. Then, $\tau_n = T(n) - T(n-1)$ is the duration of the n'th cycle. Assume further that Y satisfies

$$\mathrm{E}\int_0^{\tau_1} |Y(s)|\,\mathrm{d}s < \infty$$

and

$$\mathrm{E}\tau_1 < \infty.$$

In this case, Y is a positive recurrent regenerative process and α can be computed as the ratio of expectations given by

$$\alpha = \frac{\mathrm{E}\int_0^{\tau_1} Y(s)\mathrm{d}s < \infty}{\mathrm{E}\tau_1};$$

see, for example, p.170 of Asmussen (2003).

Furthermore, if τ_1 is spread-out (as occurs, for example, if τ_1 has a density), then

$$\mathrm{E}Y(t) \to \alpha$$

as $t \to \infty$; see p.170 of Asmussen (2003). While there are many ways to define the initial transient period, a natural definition (for our purposes) is, for $\varepsilon > 0$, to let $\beta = \beta(\varepsilon)$ be defined by

$$\beta(\varepsilon) = \sup\{t \geq 0 : |\mathrm{E}Y(t) - \alpha| > \varepsilon\};$$

the interval $[0, \beta(\varepsilon)]$ is then the interval over which an ε-transient is present. Our algorithm is intended to estimate $\beta(\varepsilon)$.

Let $\tilde{a}(t) = \mathrm{E}Y(t)$. The function $\tilde{a} = (\tilde{a}(t) : t \geq 0)$ satisfies the renewal equation

$$a = b + F * a,$$

where $b(t) \triangleq \mathrm{E}Y(t)I(\tau_1 > t)$, $F(\mathrm{d}t) \triangleq \mathrm{P}(\tau_1 \in \mathrm{d}t)$, and $*$ denotes the convolution operator. The solution of this renewal equation is then given by

$$\tilde{a} = U * b,$$

where $U = \sum_{n=0}^{\infty} F^{(n)}$ and $F^{(n)}$ is the n-fold convolution of F. Hence, $\beta(\varepsilon)$ can be re-expressed as

$$\beta(\varepsilon) = \sup\{t \geq 0 : |(U * b)(t) - (U * b)(\infty)| > \varepsilon\}.$$

We now consider the estimation of $\beta(\varepsilon)$ based on a single simulation of Y over $[0,t]$. Let

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

be the number of complete regenerative cycles simulated within our single replication of Y. Note that we may estimate b via

$$\hat{b}_t(\cdot) = \begin{cases} \dfrac{1}{N(t)} \displaystyle\sum_{i=0}^{N(t)-1} Y(T(i)+\cdot)I(\tau_{i+1} > \cdot), & N(t) \geq 1 \\ 0, & N(t) = 0 \end{cases}$$

and we can estimate F via

$$\hat{F}_t(\mathrm{d}s) = \begin{cases} \dfrac{1}{N(t)} \displaystyle\sum_{i=0}^{N(t)-1} \delta_{\tau_i}(\mathrm{d}s), & N(t) \geq 1 \\ \delta_1(\mathrm{d}s), & N(t) = 0, \end{cases}$$

where $\delta_z(\cdot)$ is a unit point mass at the point z. Set $\hat{U}_t = \sum_{n=0}^{\infty} \hat{F}_t^{(n)}$, where $\hat{F}_t^{(n)}$ is the n-fold convolution of $\hat{F}_t$. Our estimator for $\beta(\varepsilon)$ is then given by

$$\hat{\beta}_t(\varepsilon) = \sup\{s \geq 0 : |(\hat{U}_t * \hat{b}_t)(s) - (\hat{U}_t * \hat{b}_t)(\infty)| > \varepsilon\}.$$

Note that the renewal theorem, applied to the empirical renewal equation, guarantees that $(\hat{U}_t * \hat{b}_t)(\infty)$ can be easily computed without any need to resort to numerical computation; in particular,

$$(\hat{U}_t * \hat{b}_t)(\infty) = \frac{1}{T(N(t))} \int_0^{T(N(t))} Y(s)\,\mathrm{d}s.$$

We turn next to a discussion of the numerical computation of $\hat{\beta}_t(\varepsilon)$ based on the simulation $(Y(s) : 0 \leq s \leq t)$. One possible approach is to recognize that the Laplace transform of $((\hat{U}_t * \hat{b}_t)(s) : s \geq 0)$ is easy to compute; see, for example, p.181 of Resnick (1992). The challenge then lies in efficiently computing the inverse Laplace transform as a means of recovering $((\hat{U}_t * \hat{b}_t)(s) : s \geq 0)$.

However, given the simulation setting within which we are working, it is particularly natural to consider a sampling-based algorithm for computing $(\hat{U}_t * \hat{b}_t)(s)$. We will re-sample the $N(t)$ completed cycles, and generate a "bootstrap" version of $(Y(s) : s \geq 0)$, thereby motivating our use of the term "regenerative bootstrap estimator" in describing our algorithm. We will now describe our bootstrap algorithm for generating a bootstrap sample $(Y^*(s) : s \geq 0)$ of $(Y(s) : s \geq 0)$.

Algorithm:

i.) **Input:** $((Y(T(j)+r) : 0 \leq r \leq \tau_{j+1}) : 0 \leq j \leq N(t)-1)$
ii.) Initialize $n = 1$, sum $= 0$
iii.) Draw τ_n^* independently from $\hat{F}_t(\cdot)$
iv.) Update $n = n+1$, sum $=$ sum $+ \tau_n^*$
v.) If (sum $< s$), then

return to step iii.)

vi.) **Output:** Set $Y^*(s) = Y(T(k-1) + (s - \tau_1^* - \tau_2^* - \cdots - \tau_{n-1}^*))$ conditional on $\tau_n^* = \tau_k$.

We can therefore compute a bootstrap estimator of $(\hat{U}_t * \hat{b}_t)(s)$ by running the above algorithm m independent and identically distributed (iid) times, thereby yielding $Y_1^*(s), Y_2^*(s), \ldots, Y_m^*(s)$, so that $(\hat{U}_t * \hat{b}_t)(s)$ can be estimated via $m^{-1} \sum_{i=1}^{m} Y_i^*(s)$.

3 NUMERICAL EXPERIMENTS

We apply the proposed regenerative bootstrap algorithm to the number-in-system processes for M/M/1 and M/M/∞ queueing models, respectively. The M/M/1 queue is a model that has structure characteristic of single-server systems, whereas the M/M/∞ queue can be viewed as being representative of many-server models. In each model, we compare $|(U * b)(s) - (U * b)(\infty)|$ and $|(\hat{U}_t * \hat{b}_t)(s) - (\hat{U}_t * \hat{b}_t)(\infty)|$ for different values of s.

Because this algorithm includes a novel bootstrap component, some discussion of the additional computational burden it imposes is warranted. As is typical of single run steady-state simulation procedures, it is important that the simulated time horizon t be large. Of course, this makes $N(t)$ large, so that the population of completed cycles from which to obtain bootstrap samples is large. To determine the duration of the initial transient, we need to compute $(\hat{U}_t * \hat{b}_t)(s)$ at a set of different s values. While the best way to determine this set of s-values requires further research, one obvious heuristic is to select a set of s-values corresponding roughly to times at which the

difference between $(\hat{U}_t * \hat{b}_t)(\cdot)$ is within approximately $j\varepsilon$ of $(\hat{U}_t * \hat{b}_t)(\infty)$. This suggests using a set of the form $s = c(\log(1/\varepsilon) + \log(1/j))$, where c is a constant and j ranges between some integer r and 1. From a practical standpoint, it is rarely essential to estimate the precise time at which the initial transient has completed (e.g. a relative error of 50% may often be acceptable), so that a precise determination of the "correct" s-values at which to compute $(\hat{U}_t * \hat{b}_t)(s)$ will often be unnecessary.

To compute $(\hat{U}_t * \hat{b}_t)(s)$ using our bootstrap algorithm, each of the m bootstrap simulations involves drawing cycle durations from $\hat{F}_t$ until we reach the maximal s-value. Note that these bootstrap simulations will often run *much* faster than did the original single-run simulation, because only the cycle durations are sampled (rather than all the events within each cycle, as occur in the original simulation) and because the bootstrap simulations run only to the maximal s-value. In addition, other variance reduction ideas can be exploited in the setting of these bootstrap simulations (e.g. one can execute step vi.) of the bootstrap algorithm multiple independent times and average the answers, rather than execute step vi.) only once) to further improve the efficiency of the bootstrap simulations.

In this section, we report preliminary numerical results on the use of this regenerative bootstrap algorithm, using only a "plain vanilla" implementation. Rather than use the logarithmic set of s-values discussed above, we use an equally spaced grid of s-values. In addition, we have not explored any use of variance reduction in conjunction with our bootstrap simulations.

3.1 Example 1: M/M/1

Consider a single server queue with Poisson arrivals and independent exponential service times. Let λ be the arrival rate, μ be the service rate, $\rho = \frac{\lambda}{\mu}$ be the traffic intensity, and $Y(t)$ be the number-in-system at time $t \geq 0$. Assume $\lambda < \mu$ and $Y(0) = 0$. Then Y is a non-delayed regenerative process with $T(n)$ being the n'th time that Y visits the origin. The steady-state mean of Y is $\alpha = \frac{\rho}{1-\rho}$. It is known that

$$\mathrm{E}Y(t) = \frac{\rho}{1-\rho} - \frac{2}{\pi}\int_0^\pi \frac{e^{-\gamma(x)\mu t}}{\gamma(x)^2}\sin^2(x)\,\mathrm{d}x$$

where $\gamma(x) = 1 + \rho - 2\sqrt{\rho}\cos x$; see p.27 of Takács (1962). It is reported in Abate and Whitt (1989) that a simple trapezoidal rule works well for numerically evaluating the above trigonometric integral.

For the numerical experiment, we take $\lambda = 0.3$, $\mu = 0.4$, $t = 10^5$. The number of bootstrap iterations are set to be $m = 1000$ and $m = 10000$, respectively. The results of the numerical experiment are shown in Table 1, Table 2, and Figure 1.

Table 1: Selected numerical results for M/M/1 model with $m = 1000$

s	$(U*b)(s)$	$(U*b)(\infty)$	$\lvert(U*b)(s) - (U*b)(\infty)\rvert$	$(\hat{U}_t * \hat{b}_t)(s)$	$(\hat{U}_t * \hat{b}_t)(\infty)$	$\lvert(\hat{U}_t * \hat{b}_t)(s) - (\hat{U}_t * \hat{b}_t)(\infty)\rvert$
50	2.033	3	0.967	2.350	3.057	0.707
100	2.557	3	0.443	2.637	3.057	0.420
150	2.771	3	0.229	2.652	3.057	0.405
200	2.874	3	0.126	2.869	3.057	0.188

Table 2: Selected numerical results for M/M/1 model with $m = 10000$

s	$(U*b)(s)$	$(U*b)(\infty)$	$\lvert(U*b)(s) - (U*b)(\infty)\rvert$	$(\hat{U}_t * \hat{b}_t)(s)$	$(\hat{U}_t * \hat{b}_t)(\infty)$	$\lvert(\hat{U}_t * \hat{b}_t)(s) - (\hat{U}_t * \hat{b}_t)(\infty)\rvert$
50	2.033	3	0.967	2.258	3.057	0.799
100	2.557	3	0.443	2.646	3.057	0.411
150	2.771	3	0.229	2.796	3.057	0.261
200	2.874	3	0.126	2.896	3.057	0.161

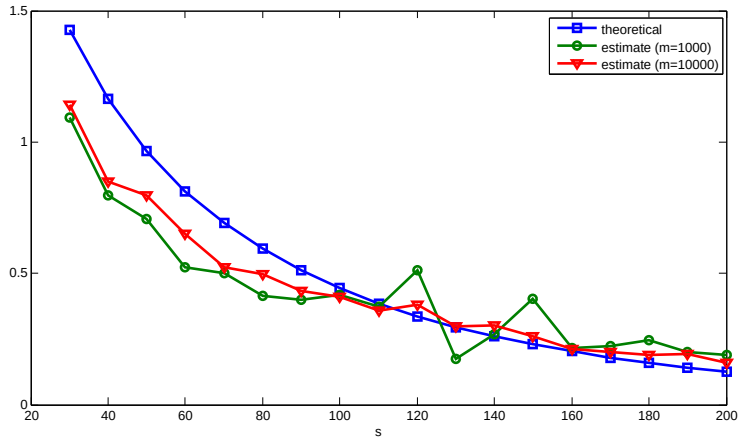

Figure 1: Comparison of $|(U*b)(s)-(U*b)(\infty)|$ (blue squares), $|(\hat{U}_t*\hat{b}_t)(s)-(\hat{U}_t*\hat{b}_t)(\infty)|$ for $m=1000$ (green circles), $|(\hat{U}_t*\hat{b}_t)(s)-(\hat{U}_t*\hat{b}_t)(\infty)|$ for $m=10000$ (red triangles) for M/M/1 model

3.2 Example 2: M/M/∞

Consider a queue with an infinite number of servers. Assume that the arrival process is Poisson and the service times for each server are iid exponential random variables. Let λ be the arrival rate, μ the server rate of one server, and $Y(t)$ be the number-in-system at time $t \geq 0$. Assume $Y(0) = y_0 = \lfloor \frac{\lambda}{\mu} \rfloor$. Then Y is a non-delayed regenerative process with $T(n)$ being the n'th time that Y visits $\lfloor \frac{\lambda}{\mu} \rfloor$. The steady-state mean of Y is $\alpha = \frac{\lambda}{\mu}$. It is well known that

$$\mathrm{E}Y(t) = \frac{\lambda}{\mu}(1-e^{-\mu t}) + y_0 e^{-\mu t};$$

see, for example, Eick, Massey, and Whitt (1993).

For the numerical experiment, we take $\lambda = 3.5$, $\mu = 0.4$, $t = 5000$. The number of bootstrap iterations are set to be $m = 1000$ and $m = 10000$, respectively. The results of the numerical experiment are shown in Table 3, Table 4, and Figure 2.

Table 3: Selected numerical results for M/M/∞ model with $m = 1000$

s	$(U*b)(s)$	$(U*b)(\infty)$	$\lvert(U*b)(s)-(U*b)(\infty)\rvert$	$(\hat{U}_t*\hat{b}_t)(s)$	$(\hat{U}_t*\hat{b}_t)(\infty)$	$\lvert(\hat{U}_t*\hat{b}_t)(s)-(\hat{U}_t*\hat{b}_t)(\infty)\rvert$
1	8.247	8.75	0.503	8.246	8.720	0.474
2	8.413	8.75	0.337	8.347	8.720	0.373
3	8.524	8.75	0.226	8.561	8.720	0.159
4	8.599	8.75	0.151	8.561	8.720	0.159
5	8.648	8.75	0.102	8.608	8.720	0.112

4 CONCLUDING REMARKS

We proposed a sampling-based algorithm for estimating the duration of the initial transient for a regenerative process. Since it can be shown theoretically all "well-posed" steady-state simulation problem are regenerative (see Glynn (1994)), our algorithm therefore has a potentially wide scope of applicability. Moreover, its validity can be supported at a theoretical level and is illustrated by the two numerical experiments we conducted.

Table 4: Selected numerical results for M/M/∞ model with $m = 10000$

s	$(U*b)(s)$	$(U*b)(\infty)$	$\lvert(U*b)(s)-(U*b)(\infty)\rvert$	$(\hat{U}_t*\hat{b}_t)(s)$	$(\hat{U}_t*\hat{b}_t)(\infty)$	$\lvert(\hat{U}_t*\hat{b}_t)(s)-(\hat{U}_t*\hat{b}_t)(\infty)\rvert$
1	8.247	8.75	0.503	8.128	8.720	0.592
2	8.413	8.75	0.337	8.412	8.720	0.309
3	8.524	8.75	0.226	8.498	8.720	0.222
4	8.599	8.75	0.151	8.595	8.720	0.125
5	8.648	8.75	0.102	8.615	8.720	0.104

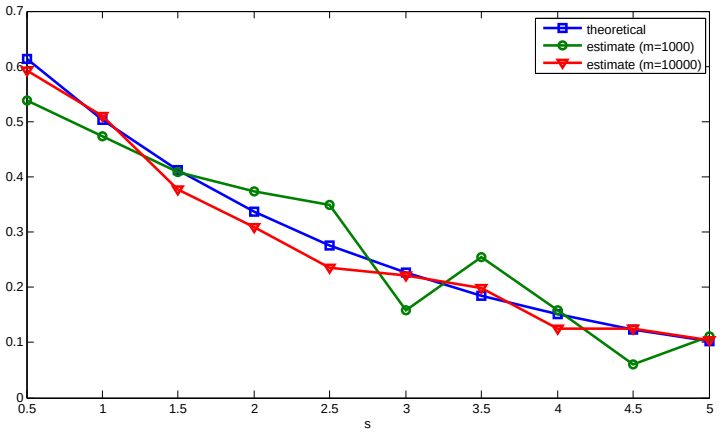

Figure 2: Comparison of $\lvert(U*b)(s)-(U*b)(\infty)\rvert$ (blue squares), $\lvert(\hat{U}_t*\hat{b}_t)(s)-(\hat{U}_t*\hat{b}_t)(\infty)\rvert$ for $m = 1000$ (green circles), $\lvert(\hat{U}_t*\hat{b}_t)(s)-(\hat{U}_t*\hat{b}_t)(\infty)\rvert$ for $m = 10000$ (red triangles) for M/M/∞ model

REFERENCES

Abate, J., and W. Whitt. 1989. Calculating time-dependent performance measures for the M/M/1 queue. *IEEE Trans. Commun.* 37:1102–1104.

Asmussen, S. 2003. *Applied Probability and Queues*. 2 ed. Springer.

Awad, H. P., and P. W. Glynn. 2007. On the theoretical comparison of low-bias steady-state estimators. *ACM Trans. Model. Comput. Simul.* 17 (1).

Awad, H. P., and P. W. Glynn. 2010. A statistically based initial transient deletion rule with rigorous theoretical support. Working paper.

Conway, R. W. 1963. Some tactical problems in digital simulation. *Management Science* 10:47–61.

Eick, S. G., W. A. Massey, and W. Whitt. 1993. The physics of the $M_t/G/\infty$ queue. *Oper. Res.* 41 (4): 731–742.

Fishman, G. 1971. Estimating sample size in computer simulation experiments. *Management Science* 18:21–37.

Glynn, P. W. 1994. Some topics in regenerative steady-state simulation. *Acta Appl. Math.* 34:225–236.

Glynn, P. W., and P. Heidelberger. 1990. Bias properties of budget constrained Monte Carlo simulations. *Oper. Res.* 38:801–814.

Glynn, P. W., and P. Heidelberger. 1992. Jackknifing under a budget constraint. *ORSA J. Comput.* 4:226–234.

Hsieh, M.-H., D. L. Iglehart, and P. W. Glynn. 2004. Empirical performance of bias-reducing estimators for regenerative steady-state simulations. *ACM Trans. Model. Comput. Simul.* 14 (4): 325–343.

Kelton, W. D., and A. M. Law. 1983. A new approach for dealing with the startup problem in discrete event simulation. *Naval Research Logistics Quarterly* 30:641–658.

Meketon, M. S., and P. Heidelberger. 1982. A renewal theoretic approach to bias reduction in regenerative simulations. *Management Science* 28 (2): 173–181.

Resnick, S. I. 1992. *Adventures in Stochastic Processes*. Brikhäuser Boston.

Robinson, S. 2007. A statistical process control approach to selecting a warm-up period for a discrete-event simulation. *European Journal of Operational Research* 176:332–346.

Takács, L. 1962. *Introduction to the Theory of Queues*. Oxford Univ. Press.

White, K., M. Cobb, and S. C. Spratt. 2000. A comparison of five steady-state truncation heuristics for simulation. In *Proceedings of the 2000 Winter Simulation Conference*, ed. J. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 755–760: IEEE Computer Society Press, Loss Alamitos, Calif.

Yücesan, E. 1993. Randomization tests for initialization bias in simulation output. *Naval Research Logistics* 40:643–663.

AUTHOR BIOGRAPHIES

PETER W. GLYNN received his Ph.D. in Operations Research from Stanford University in 1982. He then joined the faculty of the University of Wisconsin at Madison, where he held a joint appointment between the Industrial Engineering Department and Mathematics Research Center, and courtesy appointments in Computer Science and Mathematics. In 1987, he returned to Stanford, where he is now the Thomas Ford Professor of Engineering in the Department of Management Science and Engineering. He also has a courtesy appointment in the Department of Electrical Engineering and serves as Director of the Stanford Institute of Computational and Mathematical Engineering. He is a Fellow of both INFORMS and the Institute of Mathematical Statistics and his research interests include computational probability, simulation, queueing theory, statistical inference for stochastic processes, and stochastic modeling.

XIAOWEI ZHANG is currently a Ph.D. student in the Department of Management Science and Engineering at Stanford University. He graduated with B.Sc. in the Department of Mathematics at Nankai University, China. His research interests include applied probability, stochastic modeling in finance and rare event simulation.