*Proceedings of the 2009 Winter Simulation Conference*
*M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, eds.*

# NEW ESTIMATORS FOR PARALLEL STEADY-STATE SIMULATIONS

Ming-hua Hsieh

Department of Management Information Systems
National Chengchi University
Taipei, 11605, TAIWAN

Peter W. Glynn

Department of Management Science and Engineering
Stanford University
Stanford, C.A. 94305, U.S.A.

## ABSTRACT

When estimating steady-state parameters in parallel discrete event simulation, initial transient is an important issue to consider. To mitigate the impact of initial condition on the quality of the estimator, we consider a class of estimators obtained by putting different weights on the sampling average across replications at selected time points. The weights are chosen to maximize their Gaussian likelihood. Then we apply model selection criterion due to Akaike and Schwarz to select two of them as our proposed estimators. In terms of relative root MSE, the proposed estimators compared favorably to the standard time average estimator in a typical test problem with significant initial transient.

## 1 INTRODUCTION

Let $Y = (Y(t) : t \geq 0)$ be a real-valued stochastic process representing the output of a discrete event simulation. Suppose that $Y$ satisfies a law of large number (LLN) of the form

$$\alpha(t) \equiv \frac{1}{t}\int_0^t Y(s)ds \Rightarrow \alpha \tag{1}$$

as $t \to \infty$, for some constant $\alpha$, where $\Rightarrow$ denotes convergence in distribution. As suggested by (1), the time-average $\alpha(t)$ is an obvious point estimator of $\alpha$. Suppose that $P$ processors are available for simulation, and each processor simulates the process $Y$ independently up to (deterministic) time $T$. Let $\alpha_p(T)$ denote the time-average $\alpha(T)$ generated by processor $p$. Set

$$\alpha(T,P) = \frac{1}{P}\sum_{p=1}^{P}\alpha_p(T). \tag{2}$$

Then one might expect $\alpha(T,P)$ is an asymptotically unbiased estimate for $\alpha$. We call estimator (2) the standard estimator. The standard estimator will converge to a wrong value if simulated time $T$ and the number of processors $P$ are not chosen appropriately (Glynn and Heidelberger 1991b, Glynn and Heidelberger 1992a, and Glynn and Heidelberger 1992b.) Such statistical problems basically arise because any bias effects on a single replication are magnified on multiple replications. This type of problems also arise in transient simulation context; see (Heidelberger 1988) and (Glynn and Heidelberger 1991a), for example.

The process $Y$ is typically initialized via a distribution for $Y(0)$ that is not characteristic of the steady-state behavior. As a consequence, $\alpha(T)$ is biased as an estimator of $\alpha$. In other words, $E\alpha(T) \neq \alpha$. For the same reason, $E\alpha(T,P) \neq \alpha$. The bias in $\alpha(T)$ as an estimator of $\alpha$ is known as the "initial bias". The initial bias problem, in the single processor context, can be mitigated in two different ways. One approach is to delete that initial segment of the simulation that is "contaminated" by initial bias. Such an initial bias deletion approach has been studied by many authors; see, for example, Glynn (1995), Schruben (1982), Schruben, Singh, and Tierney (1983), and White (1997). An alternative is to consider an estimator, based on simulating $Y$ over $[0,T]$, that attempts to compensate for the bias present in $\alpha(T)$. We refer to such estimators as "bias reducing" estimators. The bias reducing estimators usually need to make use of independent identically distributed quantities. Exploiting the regenerative structure of the process $Y$ is a possible approach; see Hsieh, Iglehart, and Glynn (2004) for a survey.

In the parallel processors (multiple replications) context, Glynn and Heidelberger (1992a) and Glynn and Heidelberger (1992b) have studied the initial bias deletion approach. Both theoretical and empirical results in their study show that the standard estimator (2) and its variant with initial bias deletion are not statistical efficient and ratio estimators are more appropriate.

We present a new method for mitigating the effect of this "initial bias" when the number $P$ of processors is large. We first take advantage of the fact that when we average the simulation output across the independent replications, we end up with an approximately Gaussian process. We then introduce a family of Gaussian models, indexed by the duration of the initial transient, that is intended to "explain" the Gaussian data that has been collected via averaging. Because these models involve differing numbers of estimated parameters, we penalize the highly parameterized models via model selection criterion due to Akaike and Schwarz. This yields new steady-state estimators that are intended to reduce the impact of the initial transient.

This paper is organized as follows. In Section 2, we describe the proposed estimators and discuss their theoretical properties. In Section 3, we discuss some of our computational experience with the procedures introduced in Section 2. Finally, Section 4 offers some concluding remarks.

## 2 THE PROPOSED ESTIMATORS

Suppose that $Y$ is the simulation output that is derived from the simulation of a stochastic system from processor $p$. Assume the simulated time on each of the $P$ processors is $T$. Let $0 = t_0 < t_1 < \cdots < t_m = T$ be a selection of time points. We define

$$\bar{Y}(t_j) = \frac{1}{P}\sum_{p=1}^{P} Y(t_j), \quad j = 0,1,\ldots,m \tag{3}$$

When $P$ is large, we can take advantage of the central limit theorem (CLT) to develop a large-sample approximation to the distribution of $\bar{Y}(t_j)$. Specifically, if $\bar{Y} \triangleq (\bar{Y}(t_0),\ldots,\bar{Y}(t_m))^T$ and $\mu^* = (\mu^*(0),\ldots,\mu^*(m))^T$ with $\mu^*(j) = EY(t_j)$, then

$$\sqrt{P}(\bar{Y} - \mu^*) \Rightarrow N(0,C)$$

as $P \to \infty$, where $N(0,C)$ is a multivariate normal random vector (of dimension $n+1$) with mean 0 and covariance matrix $C \triangleq (C(k,l) : 0 \le k \le n, 0 \le l \le n)$, with entries $C(k,l) \triangleq \mathrm{Cov}(Y(t_k),Y(t_l))$.

It follows that when $p$ is large,

$$\bar{Y} \overset{D}{\approx} N(\mu^*, \frac{1}{P}C).$$

Our interest is in using the data $\bar{Y}$ to estimate $\alpha$.

Thus, if the initial transient finishes roughly at time $l-1$, the mean vector $\mu^*$ should be such that $\mu^*$ is approximately of the form $\mu^* = (\mu^*(0),\mu^*(1),\ldots,\mu^*(l-1),\mu^*(\infty),\mu^*(\infty),\ldots,\mu^*(\infty))^T$ (i.e. $\mu^*(j) = \mu^*(\infty)$ for $j \ge l$). So, the approximate likelihood for $\bar{Y}$ when the model is such that the system is in equilibrium from time $l$ onwards is

$$(2\pi)^{-(n+1)/2}|\det\ C|^{-1/2} \cdot \exp\left(-\frac{1}{2}(\bar{Y} - \mu_{[l]})^T \left(\frac{C}{P}\right)^{-1} (\bar{Y} - \mu_{[l]})\right), \tag{4}$$

where $\mu_{[l]} = (\mu(0),\mu(1),\ldots,\mu(l-1),\mu(\infty),\ldots,\mu(\infty))$. The maximum likelihood estimator for $\mu_{[l]}$ is the minimizer of the quadratic programming problem

$$\min_{\mu_{[l]}} \quad (\bar{Y} - \mu_{[l]})^T \left(\frac{C}{P}\right)^{-1} (\bar{Y} - \mu_{[l]}) \tag{5}$$

Since $C$ is unknown, we replace (5) with

$$\min_{\mu_{[l]}} \quad (\bar{Y}-\mu_{[l]})^T \left(\frac{\hat{C}}{P}\right)^{-1} (\bar{Y}-\mu_{[l]}) \tag{6}$$

where $\hat{C}$ is the sample covariance matrix given by

$$\hat{C} = \frac{1}{P-1}\sum_{i=1}^{P}(Y_i-\bar{Y})(Y_i-\bar{Y})^T,$$

with $Y_i \stackrel{\triangle}{=} (Y_i(t_0),\ldots,Y_i(t_m))^T$. Let $\hat{\mu}_{[l]}$ be the minimizer of (6).

The model that describes an initial transient that finishes at time $l-1$ (i.e. the model associated with estimating $\mu^*$ via $\hat{\mu}_{[l]}$) is nested within the model that describes an initial transient finishing at time $l$, in the sense that the range of $\mu_{[l]}$ is contained within that of $\mu_{[l+1]}$. As a consequence, we get a better fit as we increase $l$. In particular, the quality of the fit, as measured by the quadratic form

$$\min_{\mu_{[l]}} \quad (\bar{Y}-\hat{\mu}_{[l]})^T \left(\frac{\hat{C}}{P}\right)^{-1} (\bar{Y}-\hat{\mu}_{[l]}) \tag{7}$$

improves as $l$ increases. In fact, when $l=n$, $\hat{\mu}_{[n]}=\bar{Y}$, the residual sum of squares (7) is zero, and $\hat{\mu}_{[n]}(\infty)=\bar{Y}(n)$.

However, this discussion does not take into account the model complexity. Specifically, the vector $\mu_{[l]}$ associated with $l$-th model contains $l+1$ free parameters that the data (as represented by $\bar{Y}$) must attempt to accurately estimate. It is now well understood within the statistical literature that one must "penalize" models with a large complexity (i.e. a large number of parameters). The two most widely used such criterion are the Akaike information criterion (AIC) and the Bayesian information criterion (BIC); see (Akaike 1974) and (Schwarz 1978) for details. In our current setting, the AIC criterion demands that one maximize

$$-\frac{1}{2}(\bar{Y}-\hat{\mu}_{[l]})^T \left(\frac{\hat{C}}{P}\right)^{-1} (\bar{Y}-\hat{\mu}_{[l]}) - (l+1), \tag{8}$$

over $l \in \{0,1,\ldots,n\}$; the inclusion of the term $l+1$ penalizes models with a large number of estimated parameters. If $L_A$ is the maximizing value of $l$ in (8), then the AIC-based estimator for $EY(\infty)$ is $\hat{\mu}_{[L_A]}(\infty)$.

The closely related BIC criterion penalizes more complex models (as measured through the number of estimated parameters) more heavily than does the AIC criterion. Specifically, the BIC criterion requires that one maximize

$$-\frac{1}{2}(\bar{Y}-\hat{\mu}_{[l]})^T \left(\frac{\hat{C}}{P}\right)^{-1} (\bar{Y}-\hat{\mu}_{[l]}) - \frac{(l+1)}{2}\log P, \tag{9}$$

over $l \in \{0,1,\ldots,n\}$ in this context. Letting $L_B$ be the maximizing value of $l$ in (9), the BIC-based estimator for $EY(\infty)$ is then $\hat{\mu}_{[L_B]}(\infty)$.

## 3 EMPIRICAL RESULTS

In this section, we present computational results which illustrate the properties of the AIC and BIC estimators. In particular, we expect MSE of the proposed AIC- and BIC-based estimators are smaller than that of standard estimators (2). We start with a description of the test problem and associated parameters. The problem had been used in (Steiger et al. 2005) to test ASAP3 and other batch means procedures. This test problem has significant initial bias.

1. *First-Order Autoregressive* (AR(1)) *Process.* This test problem is to estimate the steady-state mean of a first-order autoregressive process $X=(X_n : n \geq 0)$ with initial condition $X_0 = 0$. The process is defined as

$$X_n = (1-\rho)\mu + \rho X_{n-1} + \varepsilon_n, \quad n \geq 1$$

Table 1: The estimated relative RMSE of each estimator, estimated $EL_A$ and $EL_B$ for the First-Order Autoregressive Process. Relative RMSE, $EL_A$ and $EL_B$ are estimated by 100 independent copies.

| Estimators | $\alpha(T,P)$ | AIC | BIC | $EL_A$ | $EL_B$ |
|---|---|---|---|---|---|
| $P = 512$; $T = 1000$ | | | | | |
| $m = 5$ | 0.1984 | 0.0070 | 0.0087 | 4.77 | 4.41 |
| $m = 10$ | | 0.0084 | 0.0109 | 9.15 | 8.33 |
| $P = 2048$; $T = 1000$ | | | | | |
| $m = 5$ | 0.1975 | 0.0065 | 0.0067 | 5 | 4.94 |
| $m = 10$ | | 0.0069 | 0.0081 | 9.78 | 9.18 |
| $P = 8192$; $T = 1000$ | | | | | |
| $m = 5$ | 0.1974 | 0.0067 | 0.0067 | 5 | 5 |
| $m = 10$ | | 0.0068 | 0.0070 | 10 | 9.86 |
| $P = 512$; $T = 2000$ | | | | | |
| $m = 5$ | 0.0994 | 0.0030 | 0.0030 | 3.34 | 2.82 |
| $m = 10$ | | 0.0024 | 0.0026 | 5.97 | 4.82 |
| $P = 2048$; $T = 2000$ | | | | | |
| $m = 5$ | 0.0995 | 0.0015 | 0.0015 | 3.50 | 3.05 |
| $m = 10$ | | 0.0012 | 0.0016 | 6.57 | 5.31 |
| $P = 8192$; $T = 2000$ | | | | | |
| $m = 5$ | 0.0995 | 0.0008 | 0.0009 | 3.80 | 3.08 |
| $m = 10$ | | 0.0007 | 0.0008 | 7.14 | 6.02 |
| $P = 512$; $T = 4000$ | | | | | |
| $m = 5$ | 0.0496 | 0.0020 | 0.0017 | 2.33 | 1.90 |
| $m = 10$ | | 0.0019 | 0.0018 | 3.87 | 2.87 |
| $P = 2048$; $T = 4000$ | | | | | |
| $m = 5$ | 0.0496 | 0.0011 | 0.0009 | 2.42 | 2.03 |
| $m = 10$ | | 0.0009 | 0.0008 | 3.89 | 3.01 |
| $P = 8192$; $T = 4000$ | | | | | |
| $m = 5$ | 0.0497 | 0.0006 | 0.0005 | 2.55 | 2.01 |
| $m = 10$ | | 0.0005 | 0.0005 | 4.37 | 3.25 |

where $\mu = 100$, $\rho = 0.995$ and $\varepsilon_n$'s are independent standard normal random variables. Therefore, the steady-state mean of the first-order autoregressive process is 100.

The parameters of the proposed estimators include the number of processors $P$, the simulated time $T$ and the time points $t_1,\ldots,t_m$. We selected 18 combinations for the test problem by varying the value of each parameter. In particular, they include the following combinations:

1. $P = 512, 2048$, or 8192;
2. $T = 1000, 2000$ or 4000;
3. Three different selections of the time points $t_1,\ldots,t_m$. The time points were selected by first fixing $m$ and then the time points are equally spaced within the interval $(0,T)$. The values of $m$ are 5 or 10.

Table 1 show the detail information of these parameter combinations.

We use relative root mean squared error (RMSE) as a performance measure for the estimators. In particular, for estimator $\hat{\alpha}$, the relative RMSE is defined as

$$\text{relRMSE} = \frac{\sqrt{E[(\hat{\alpha}-\alpha)^2]}}{\alpha} \tag{10}$$

We performed $N = 100$ independent parallel simulation runs for each combination of $T$, $P$, and $m$, and based on each parallel simulation run (include $P$ replications), we compute standard estimator $\alpha(T,P)$ and the proposed estimators $\alpha_i(T,P)$, $i = 1,2,3,4$. That is, for each estimator $\hat{\alpha}$, we have $N$ independent copies. Denote these $N$ independent copies of $\hat{\alpha}$ as $\hat{\alpha}^{(k)}$, $k = 1,\ldots,N$. Then the relative RMSE of $\hat{\alpha}$ can be estimated by

$$\frac{1}{\alpha N}\sqrt{\sum_{k=1}^{N}(\hat{\alpha}^{(k)}-\alpha)^2}$$

The empirical results are shown in Table 1.

From the experiment results, the AIC- and BIC-based estimators, in terms of relative root MSE, compared favorably to the standard time average estimator. Their performance are about the same, but AIC-based estimator seems a little bit better than BIC-based estimator. If the simulated time is not long enough ($T = 1000$) to let the process $Y$ be in the steady state, both method will select $L_A$ and $L_B$ close to $m$ which is an indication of the simulated time is too short. When the simulated time are long enough ($T = 2000$ and $T = 4000$) to let the process $Y$ be in the steady state, AIC-based estimator select $EL_A = 5.97 to 7.14$ (when $T = 2000$) and $EL_A = 3.87$ to $4.37$ (when $T = 4000$) predicts initial transient terminate around $T = 1200$ to $1600$; BIC-based estimator select $EL_B = 4.82$ to $6.02$ (when $T = 2000$) and $EL_B = 2.87 to 3.25$ (when $T = 4000$) predicts initial transient terminate around $T = 1000$ to $1300$. Also, $EL_A$ increases when $P$ increases. This is actually a theoretical property of AIC-based estimator, which we will discuss in the expansion version of this paper.

## 4 CONCLUSIONS AND FUTURE WORKS

In this paper, we proposed AIC and BIC-based estimators for parallel steady state simulation. Based on the numerical experiments we conducted, they seem promising. We will develop the confidence interval construction procedures for these estimators and test their coverage probability.

## ACKNOWLEDGMENTS

This research was partially supported by the National Science Council of Taiwan under contract number NSC97-2410-H-004-041-MY2.

## REFERENCES

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19 (6): 716–723.

Glynn, P. W. 1995. Some new results on the initial transient problem. In *Proceedings of the 1995 Winter Simulation Conference*, eds. C. Alexopoulos, and K. Kang, 165–170. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Glynn, P. W., and P. Heidelberger. 1991a. Analysis of initial transient deletion for replicated steady-state simulations. *Operations Research Letters* 10:437–443.

Glynn, P. W., and P. Heidelberger. 1991b. Analysis of parallel, replicated simulations under a completion time constraint. *ACM Trans. on Modeling and Computer Simulation* 1:3–23.

Glynn, P. W., and P. Heidelberger. 1992a. Analysis of initial transient deletion for parallel steady-state simulations. *SIAM Journal Scientific Stat. Computing* 13:909–922.

Glynn, P. W., and P. Heidelberger. 1992b. Experiments with initial transient deletion for parallel, replicated steady-state simulations. *Management Science* 38:400–418.

Heidelberger, P. 1988. Discrete event simulations and parallel processing: Statistical properties. *SIAM J. Scientific and Statist. Computing* 9:1114–1132.

Hsieh, M.-H., D. L. Iglehart, and P. W. Glynn. 2004. Empirical performance of bias-reducing estimators for regenerative steady-state simulations. *ACM Trans. on Modeling and Computer Simulation* 14:325–343.

Schruben, L. W. 1982. Detecting initialization bias in simulation output. *Oper. Res.* 30 (3): 151–153.

Schruben, L. W., H. Singh, and L. Tierney. 1983. Optimal tests for initialization bais in simulation output. *Oper. Res.* 31 (6): 1167–1178.

Schwarz, G. 1978. Estimating the dimension of a model. *The Annals of Statistics* 6 (2): 461–464.

Steiger, N., E. Lada, J. Wilson, J. Joines, C. Alexopoulos, and D. Goldsman. 2005. ASAP 3: a batch means procedure for steady-state simulation analysis. *ACM Transactions on Modeling and Computer Simulation* 15 (1): 39–73.

White, K. P. 1997. An effective truncation heuristic for bias reduction in simulation output. *Simulation* 69 (6): 323–334.

## AUTHOR BIOGRAPHIES

**MING-HUA HSIEH** is an Associate Professor of the Department of Management Information Systems at National Chengchi University, Taiwan. From 1997 to 1999, he was a software designer at Hewlett Packard company, California. He is a member of INFORMS. His research interests include simulation methodology and financial engineering. He has served as a Proceedings Co-editor for WSC 2007. His e-mail address is <mhsieh@nccu.edu.tw>.

**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 INFORMS and a Fellow of the Institute of Mathematical Statistics and his research interests in computational probability, simulation, queueing theory, statistical inference for stochastic processes, and stochastic modeling. His e-mail address is <glynn@stanford.edu>.