USING PROCESS CENTRAL LIMIT THEOREMS IN ANALYZING SIMULATION OUTPUT

by

Lee Schruben
and
David Goldsman

Cornell University
(Currently on leave to Purdue University)

1. INTRODUCTION

In this article, we focus on two questions of central importance when an estimate is desired for the steady-state mean of a simulation output sequence, Y_i, $i = 1,2,\ldots,n$. This unknown parameter of the system being studied is denoted as,

$$\mu = \lim_{n\to\infty} E[Y_i].$$

The two questions that we address are:

1. Is initialization bias a significant problem?

2. What is the accuracy of our estimate of μ?

We will use a process central limit theorem to attempt to quantify our answers to these two questions. The approach is computationally (but not philosophically) different from other approachs found in the simulation literature [Law and Kelton, 1982].

The usual estimator of μ is the sample mean,

$$\bar{Y}_n = 1/n\sum_{i=1}^{n} Y_i.$$

Most common output analysis techniques attempt to standardize this scalar random variable. By using the Central Limit Theorem, one assumes that this standardized estimator will (for long runs) behave essentially like a standard Normal random variable. Properties of a standard Normal variate can therefore be used to analyze the estimator, $\bar{Y}_n$.

Intuitively, what we will do here is **standardize the entire output series** $\{Y_i\}$. The standardized output series has as its limit (again for long runs) a Brownian bridge stochastic process. Much is known about the behavior of this limiting stochastic process. Properties of this stochastic process are used to analyze the simulation output series. The approach follows the familiar philosophy of classical (non-Bayesian) statistics. The only difference is that a more general Central Limit Theorem is used.

2. STANDARDIZING THE OUTPUT TIME SERIES

Let Y_i, $i=1,2,\ldots,n$, be a scalar time series of system performance measurements obtained from a simulation experiment. We assume, in the absence of initialization bias, that this output series would be stationary and satisfy some mild properties concerning serial dependence [Schruben, 1982a]. The experiment can be a single run of a simulation program or a more complex experiment involving several replications. Y_i might be actual system observations, batch averages of adjacent observations, a sequence of coefficients for a regression model, etc. We will denote the expected value of the i^{th} observation, Y_i, as μ_i. We will say that the output series has no initialization bias if $\mu_i = \mu_j$ for all pairs of observations, Y_i and Y_j.

In using the classical (scalar) Central Limit Theorem, several steps are followed:

1. The data is centered to have a mean equal to zero.

2. The magnitude of the estimator is scaled by dividing by the sample standard deviation.

3. A Central Limit Theorem is used to obtain a probability model for the behavior of the standardized estimator. This model is most commonly a standard scalar Normal random variable.

4. The limiting probability model is used in the analysis of the system being studied. Hypotheses are tested, confidence intervals are constructed, etc. using this asymptotic model.

We will follow these steps, and one additional step, in standardizing the time series, $\{Y_i\}$. Let

$$\bar{Y}_k = 1/k\sum_{i=1}^{k} Y_i$$

Proceedings of the 1982
Winter Simulation Conference
Highland * Chao * Madrigal, Editors

82CH1844-0/82/0000-0599 $00.75 © 1982 IEEE

denote the cumulative average of the first k observations. As in the scalar case, we first perform

> Step 1. Center the output series so that, under the hypothesis that there is no initialization bias, it will have a mean equal to zero.

This is accomplished by considering the transformed series,

$$S_n(k) = \begin{cases} 0 & \text{for } k = 0, \\ \bar{Y}_n - \bar{Y}_k & \text{for } k = 1,2,\ldots,n. \end{cases}$$

Note that, under the hypothesis that all the observations have equal means, $S_n(0) = S_n(n)$ and $E[S_n(k)] = 0$ for all k.

Next, similar to the second step in using the scalar Central Limit Theorem, we perform

> Step 2. Scale the magnitude of the sequence.

Scaling the series magnitude involves dividing $S_n(k)$ by the square root of the variance scaling factor, $n\sigma^2/k$.

Here the constant, σ^2, is given by,

$$\sigma^2 = \sigma_y^2 + 2 \sum_{i=1}^{\infty} \gamma_y(i) = \lim_{n\uparrow\infty} n \operatorname{Var}(\bar{Y}_n)$$

with

$$\sigma_y^2 = \operatorname{Var}(Y_i)$$

and

$$\gamma_y(i) = \operatorname{Cov}(Y_0, Y_i).$$

When the simulation output measurements are serially correlated, σ^2 is in general not equal to $\operatorname{Var}(Y_i)$.

The one additional step required in standardizing the output series is to scale the index to the unit interval. This is done by defining t = k/n as the index for the transformed output process. Therefore, for different run durations (e.g. different n) the index, t, always lies between 0 and 1. This step,

> Step 3. Scale the series index to the unit interval,

results in the standardized time series,

$$T_n(t) = [nt]S_n([nt])/(\sqrt{n}\sigma), \quad 0 \le t \le 1.$$

The original simulation output series, $\{Y_i\}$, can be recovered from the sample mean and the standardized series, $\{T_n(t)\}$.

Next (as in the case of scalar estimators), we perform

> Step 4. apply a Central Limit Theorem to obtain a probability model for the behavior of the sequence, $\{T_n(t)\}$.

In [Schruben, 1982b], a lemma is presented that says that the sequence, $\{T_n(t)\}$ is asymptotically independent of the sample mean, and converges in distribution to a standard Brownian bridge stochastic process. Therefore, we then can perform the final step in the output analysis process (analogous to the fourth step in using scalar estimators).

> Step 5. Use properties of the limiting Brownian bridge process in the analysis of the standardized time series, $\{T_n(t)\}$.

The properties of the standard Brownian Bridge process, denoted here as $\{\beta(t): 0 \le t \le 1\}$, that we will use are,

1. A Wiener Process (i.e. Brownian motion) has independent increments.

2. If t^* is the location of the maximum of $\beta(t)$, and $s^* = \beta(t^*)$, then,

   $$s^{*2}/(t^*(1 - t^*)),$$

   has a χ^2 distribution with 3 degrees of freedom [Schruben, 1982a].

3. The signed area under $\{T_n(k/n)\}$ has an asymptotically Normal distribution with zero mean and variance,

   $$(n^2 - 1)/12.$$

   (see [Schruben, et.al. 1979]).

4. If initialization bias has dissipated by observation $Y_{n\tau}$, for some fraction τ of the run, then the slope of $kS_n(k)$ as a function of $k = nt$ for $t > \frac{n}{\tau}$ is equal to the bias in the sample mean, $\bar{Y}_n$.

3. TESTING THE SIGNIFICANCE OF INITIALIZATION BIAS

Most initialization bias control procedures in the simulation literature are rather elaborate. Therefore, it might be advisable to test the output series for initialization bias before one of the bias control procedures is performed. Many tests for the presence of initialization bias have been developed based on the Brownian bridge model for the standardized output series $\{T_n(t)\}$ [Schruben, et al., 1980, Schruben, 1982a, Heidelberger and Welch, 1982].

A very simple and effective test presented as an afterthought in [Schruben, 1982a] can be generalized as follows,

1. Split the output series into two parts of length n_1, and n_2.

2. Compute,

$$h_1 = s_1^2/(\sigma^2(t_1(1 - t_1))$$

and

$$h_2 = s_2^2/(\sigma^2(t_2(1 - t_2))$$

where t_1 and t_2 are respectively the locations of the (first, if ties occur) maxima of the standardized sequences $\{T_n(t)\}$ and $\{T_n'(t)\}$ corresponding to the first part and the second part of the output series, with $s_1 = T_n(t_1)$ and $s_2 = T_n(t_2)$ for σ set arbitrarily equal to 1 (this constant cancels out of the test ratio in the next step).

3. Test the ratio $h = h_1/h_2$ against an F distribution with degrees of freedom equal to 3 and 3.

4. The hypothesis that the observation means in the two parts of the output series are the same is rejected at the α level if the upper tail of the F distribution above h is greater than α.

This test is asymptotically valid (for large n_1 and n_2) and is theoretically based on properties 1 and 2 of $\beta(t)$. The choice of n_1 and n_2 can be varied as long as they are fairly large. The experimenter is reminded to pay attention to error control if several tests are run on the same output series. The above procedure has been used successfully by students in their first simulation course both at Cornell University and at Syracuse University [Sargent, 1982]. Note that the test does not work well in the presence of a large initial transient that persists throughout the entire run. This case, however, can usually be detected by visual inspection of a plots of $\{Y_i\}$ and $\{T_n(t)\}$ (see [Schruben, 1982a] for a discussion of the behavior of $\{T_n(t)\}$ in the presence of initialization bias).

4. TRUNCATING THE OUTPUT SERIES

Property 4 of the standardized output series can be used to truncate a simulation output series that has suspected initialization bias. A plot of $kS_n(k)$ as a function of $k = nt$ is first printed. An estimate of the slope near $k = n$ is then subtracted from the sample mean for the entire output series. This correction for initialization bias is algebraically equivalent to truncating the output series at the first place (smallest value of k) a line fixed at (n,0) intersects a plot of $kS_n(k)$ (if that value of k is an integer). This is a slight generalization of truncation since this value of k need not be integer (i.e. effectively a fractional part of an observation will be truncated). The authors have found that a string held at (n,0) on a plot of $kS_n(k)$ versus $k = nt$ to estimate the slope near $k = n$ is a quick method of truncating a simulation output series.

5. CONFIDENCE INTERVALS FOR μ

Properties 2 and 3 of $\beta(t)$ have been used to obtain confidence intervals for μ [Schruben, 1982b]. These confidence intervals can be used with the output of a single run and perform as well or better than the common techniques of batched means and replication [Law and Kelton, 1982].

The confidence interval estimator for a single run based on property 2 of $\beta(t)$ is as follows. Let

$$M = (\sum_{i=1}^{n}(\bar{Y}_n - Y_j))^2/(nt^*(1 - t^*))$$

Then an asymptotically valid 100(1-α)% confidence interval for μ is given by,

$$\mu \in \bar{Y}_n \pm t_3(\alpha)\,\mathrm{sqrt}(M/n)$$

Here $t_3(\alpha)$ is the 100(1-α/2) quantile of a t distribution with 3 degree of freedom.

The confidence interval estimator for a single run based on property 3 of $\beta(t)$ is as follows. Let

$$A' = (\sum_{i=1}^{n}\sum_{j=1}^{i}(\bar{Y}_n - Y_j))^2$$

and

$$A = (12/(n^3 - n)A'$$

Then an asymptotically valid 100(1-α)% confidence interval for μ is given by,

$$\mu \in \bar{Y}_n \pm t_1(\alpha)\,\mathrm{sqrt}(A/n)$$

Here $t_1(\alpha)$ is the 100(1-α/2) quantile of a t distribution with 1 degree of freedom. To obtain additional degrees of freedom with either estimator, replications can be run or the output batched in the usual manner. For large batch sizes or run durations these two confidence intervals have theoretical properties that are superior to the methods of batching and replication. Most notably, the confidence interval based on property 2 has a interval half-width variance that is at most only 1/3 of that for the methods of replication and batching [Goldsman and Schruben, 1982]. This indicated that sequential confidence interval methods based on the estimator using property 2 might not suffer from the tendency of ''picking losers'' that methods currently in the literature have.

6. CLOSING REMARKS

The purpose of this article is to present a new approach to the analysis of simulation output. The authors hope that practitioners will try the techniques already developed and that researchers will be able to improve on these methods. There are potentially many applications of the standardized time series, $\{T_n(t)\}$, presented in this article. Other applications are currently under study by the authors.

The authors are grateful to Professors D. Heath, A. Law, R. Sargent, W. Vervaat, and J. R. Wilson who have aided and encouraged this line of research.

References

Goldsman, D. and Schruben, L. (1982), ''Comparison of Some Confidence Intervals for Simulation Output'', Tech. Rept. 544, School of O.R.I.E., Cornell University, Ithaca, NY, 14853

Heidelberger, P., and Welch, P. D. (1982), ''Simulation Run Length Control in the Presence of an Initial Transient'', Report RC 9393, Computer Science, I.B.M. Watson Research Center, Yorktown Heights, NY, 10598.

Law, A.M., and Kelton, W.D. (1982), Simulation Modeling and Analysis, McGraw-Hill, New York, NY,

Sargent, R. G., personal communication

Schruben, L., H. Singh, and L. Tierney, (1980), ''Optimal Tests for Initialization Bias in Simulation Output'', Tech. Rept. 471, School of O.R.I.E., Cornell University, Ithaca, NY, 14853

Schruben, L. (1982a), ''Detecting Initialization Bias in Simulation Output'', Operations Reserach, Vol. 30, No. 3, pp. 569-590.

Schruben, L. (1982b), ''Confidence Interval Estimation from Standardized Simulation Output'', Tech. Rept. 518, School of O.R.I.E., Cornell University, Ithaca, NY, 14853