PARAMETRIC EXPRESSION FOR MSER WITH GEOMETRICALLY DECAYING BIAS

K. Preston White, Jr.
William W. Franklin

University of Virginia
Department of Systems and Information Engineering
P.O. Box 400747
Charlottesville, VA, 22904-4747, USA

ABSTRACT

The MSER algorithm for determining the warm-up period in steady-state simulations was introduced in 1990. Over the past two decades, empirical evaluations by different research teams have demonstrated the effectiveness of the procedure. These extensive empirical results illustrate the relative advantage of MSER over other approaches under alternative bias scenarios. In this paper, we develop parametric expressions to explore how MSER should behave in the case of simulation output with geometrically decaying bias and white noise. We derive a closed-form expression for the expected optimal truncation point for this scenario. This permits computation of the threshold bias-to-noise ratio for bias detection in terms of the rate of decay of the initial bias and the strength of autocorrelation in the output sequence.

1   INTRODUCTION

Initialization bias arises when the initial conditions for a steady-state simulation are chosen as a matter of convenience. If these conditions are unrepresentative of steady-state operation (have small stationary probabilities and values that are far from the steady-state mean), bias is introduced in the estimation of output statistics. The most common most means for reducing this bias is simply to delete the offending observations from the beginning of the trial and to compute statistics using only the remaining observations. This technique is generally called *truncation* or *deletion*.

The question remains as to how to determine the appropriate number of observations to truncate. Inadequate truncation fails to remove the bias sufficiently; over-truncation comes with a loss of precision. One method for selecting the truncation point is the MSER-5 algorithm (Spratt 1998), a modification of the original MSER heuristic described by White (1997). (See, also, White *et al.* 2000 for a more readily available discussion of MSER-5.) Though MSER-5 has been shown to work exceptionally well, requiring little computational bandwidth or experimenter intervention (Hoad, Robinson, and Davies 2009), prior analyses have been largely empirical in nature.

In this paper we introduce a parametric approach to analyzing the expected behavior of the MSER algorithm and apply this approach to an output model with geometrically decaying bias and constant-parameter AR(1) white noise. With a parametric expression for the expected behavior, we can begin to drive to more fundamental understandings of MSER and can explore boundary conditions. Such explorations can also yield broader insights into simulation output analysis.

2   THE MSER CRITERION

The MSER statistic is

$$MSER(n,d) = \frac{1}{(n-d)^2} \sum_{t=d+1}^{n} (y_t - \bar{y}(n,d)))^2$$

where n is the simulation run length, d is an arbitrary truncation point, and

$$\bar{y}(n,d) = \sum_{d+1}^{n} y_t$$

is the mean of the output series $\{y_t; t=1,2,\ldots,n\}$ truncated at d. The MSER criterion specifies the optimal truncation point as

$$d^* = \arg \min_{d \ll n}[MSER(n,d)]$$

3   A SURROGATE MODEL

To explore the behavior of MSER, consider a general surrogate model for a simulation output variable

$$y_t = \mu + b_t + \varepsilon_t$$

The output has three components:

(i)   the steady-state mean of the output, the constant term μ,
(ii) a bias term, which vanishes in the limit, i.e., $b_t \to 0$ as $t \to \infty$, and
(iii) a zero-mean, stationary process ε_t, which introduces white noise into the output observations.

For this surrogate model, the expectation of the MSER statistic is

$$E[MSER(n,d)] = \frac{1}{(n-d)^2} \sum_{t=d+1}^{n} (b_t - \bar{b}(n,d)))^2 + E\left[\frac{1}{(n-d)^2} \sum_{t=d+1}^{n} (\varepsilon_t - \bar{\varepsilon}(n,d)))^2 \right]$$

Consider a special case of the surrogate model suggested by Nelson (2009):

(i) a geometrically decaying bias term

$$b_t = br^t$$

where $1 > r \geq 0$ is the rate of bias decrease and $b > 0$ is a scaling constant. If the run length n is sufficiently large, $b_n \approx 0$ and the bias component of the MSER expectation is

$$\frac{1}{(n-d)^2} \sum_{t=d+1}^{n} (b_t - \bar{b}(n,d)))^2 \approx \frac{b^2 r^{2(d+1)}}{(n-d)^2}$$

(ii) a constant-parameter, AR(1) noise process

$$\varepsilon_t = \phi\varepsilon_{t-1} + \eta_t$$

with positive parameter $1 \geq \varphi > 0$ and η_t a zero-mean, stationary process with variance σ^2 and initial condition $\eta_0 = 0$. The variance of this noise process is

$$Var[\varepsilon_t] = \frac{\sigma^2}{1 - \varphi^2}$$

and the autocorrelation function is

$$Corr[\varepsilon_t, \varepsilon_{t+h}] = \varphi^h$$

For n sufficiently large, the correlation component of the MSER expectation is

$$E\left[\frac{1}{(n-d)^2} \sum_{t=d+1}^{n} (\varepsilon_t - \bar{\varepsilon}(n,d)))^2\right] \approx \frac{1}{(n-d)}\left[\frac{\sigma^2}{1-\phi^2}\right].$$

For this special case, the optimal truncation point corresponding to the expectation is

$$d^* = \arg\min_{d} \left[\sigma^2 E\left[\frac{\beta^2 r^{2(d+1)}}{(n(r)-d)^2} + \frac{1}{(1-\varphi^2)(n(r)-d)}\right]\right] \tag{1}$$

where $\beta = b/\sigma$ is the peak bias-to-noise ratio. This normalizing constant allows us to express an arbitrary truncation point in terms of the ratio of the magnitude of the initial bias to the variability inherent in the noise process. For the run length to be "sufficiently large", define an effectively zero bias as, say, $b_n = \gamma \ll 1$. Then, a run length of

$$n(r) = \frac{\log(\gamma)}{\log(r)}$$

is sufficient to allow a decay of $100(1-\gamma)\%$ the bias term at the termination of a run.

4 ANALYSIS

4.1 The optimal truncation point

To determine the relationship between the optimal truncation point, d^*, and the parameters of the surrogate model (β, r, φ) we solve for the zeros of the derivative of *E[MSER]*. Differentiating,

$$\frac{\partial}{\partial d} E[MSER(n,d)] = \left[\frac{2\beta^2 r^{2d+2}((n(r)-d)\log(r)+1)}{(n(r)-d)^3} + \frac{1-r^2}{(1-\varphi^2)(n(r)-d)^2}\right]$$

Therefore, $d = d^*$ (integer) yields the approximate zeros of the expectation when

$$2\beta^2 r^{2(d^*+1)}[(n(r)-d^*)\log(r)+1](1-\varphi^2) + (1-r^2)(n(r)-d^*) \approx 0$$

For $n >> d$

$$r^{2(d*+1)} \approx \frac{r^2 - 1}{2\beta^2 \left[\log(r) + \dfrac{1}{n(r)} \right](1 - \varphi^2)}$$

and the optimal truncation point is

$$d* \approx \frac{1}{2\ln(r)} \ln\left(\frac{r^2 - 1}{2\beta^2 \left[\log(r) + \dfrac{1}{n(r)} \right](1 - \varphi^2)} \right) - 1$$

Note that for *n* sufficiently large, since $1 > r \geq 0$, the term

$$2\beta^2 r^{2(d*+1)} \left[\log(r) + \frac{1}{n(r)} \right](1 - \varphi^2) < 0$$

and is monotone increasing in *d*, approaching zero in the limit. Therefore there will be *at most* one change in sign in the derivative (negative to positive) as *d* increases. This change indicates that the zero has been "trapped" between succeeding integer values *d*.

4.2   Interpretation

Consider first the combinations of (β, r, φ) which imply no truncation. The detection threshold values of β correspond to $d* = d + \delta < 1$, where δ is vanishingly small. Hence

$$2\beta^2 r^4 (\log(r) + 1/n)(1 - \varphi^2) + (1 - r^2) \geq 0$$

and

$$\beta \leq \frac{1}{r^2} \sqrt{\frac{r^2 - 1}{2(\log(r) + 1/n)(1 - \varphi^2)}}$$

The threshold values for β are plotted in Figure 1(a) as a function of *r* for selected values of φ. These curves define the initial signal-to-noise ratio β below which bias cannot be detected. In effect, *the bias gets lost in the noise*. As a check, the treshold values also were determined empirically. These values are shown in Figure 1(b). While the empirical curves are modestly steeper, the shape of the curves agree.

As shown, the threshold increases with increasing autocorrelation—greater variance in the noise process makes the bias increasingly difficult to identify. The threshold also increases as the rate of bias decay approaches zero. As $r \to 0$, the bias decays to zero almost instantaneously, contributing little to the sample mean; as $r \to 1$, the bias persists and $E[MSER]$ adjusts by requiring a lower initial bias in order to allow later samples to overwhelm the effects of that initial bias. These results are consistent with the general observations of White and Robinson (2010). These observations also explain why MSER-5 provides improvement over MSER, smoothing the output time-series and thereby increasing the bias-to-noise ratio.
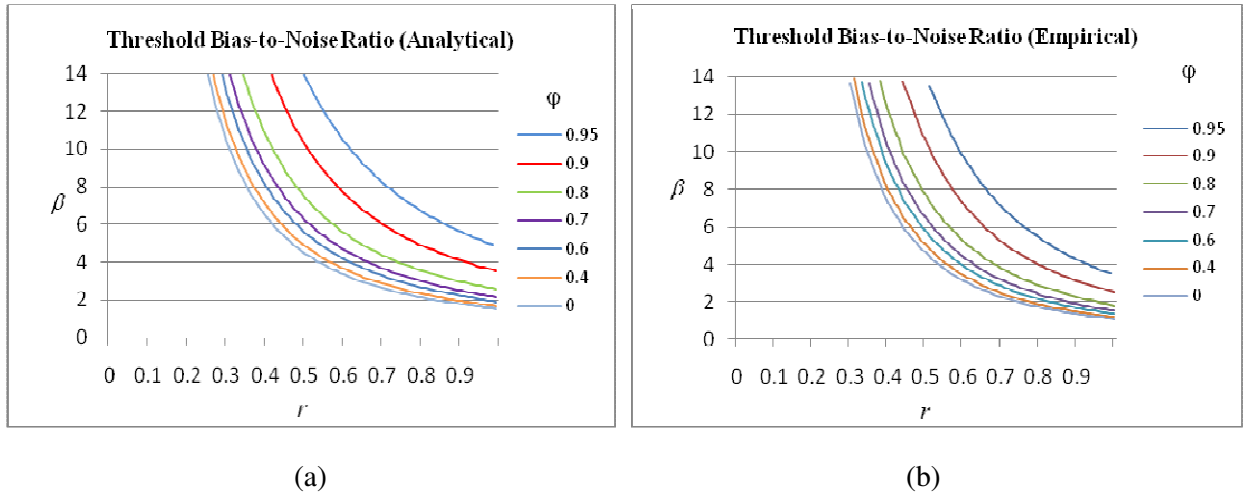
(a)   (b)

Figure 1: Threshold values of β, r, and φ that give $E[d^*] \approx 0$.

Next consider the general case where $t \leq d^* < t+1$, and t observations are truncated. In this instance, the maximum threshold is $\beta_t = \beta r^t$ and

$$\beta_t \leq \frac{1}{r^2} \sqrt{\frac{r^2 - 1}{2(\log(r) + 1/n)(1 - \varphi^2)}} \tag{2}$$

Therefore, d^* is the smallest integer t such that Inequality (2) is true.

4.3   An Example

Consider the special case of the surrogate model with parameters $\mu=100$, $r=0.9$, $b=10$, $\varphi=0.9$, $\sigma=10$, and $\gamma=0.001$. For these parameter values, $\beta=1$, $Var(\varepsilon_t)=1000$, and $n(r)=65.56$. β is well below the threshold value of 1.80 and therefore $d^*=0$. An instance of the output is shown in Figure 2(a).

Consider also an instance of the output shown in Figure 2(b), where the initial bias is increased ten-fold ($b=100$) and all of the other parameters are unchanged. $\beta=10$ is well above the threshold and the truncation point is $d^*=13$.

The "trunc mean" line in Figures 2(a) and 2(b) is the computation of the expected sample mean if we truncate prior to time index t. In Figure 2(a), we notice that this line doesn't move noticeably from the mean of $\mu = 100$, which agrees with our expectation that $E[d^*] = 0$. In the example of Figure 2(b), where we expect d^* to be equal to 13, we indeed see that the truncated mean values approach the mean of $\mu = 100$ between $t = 10$ and $t = 20$.
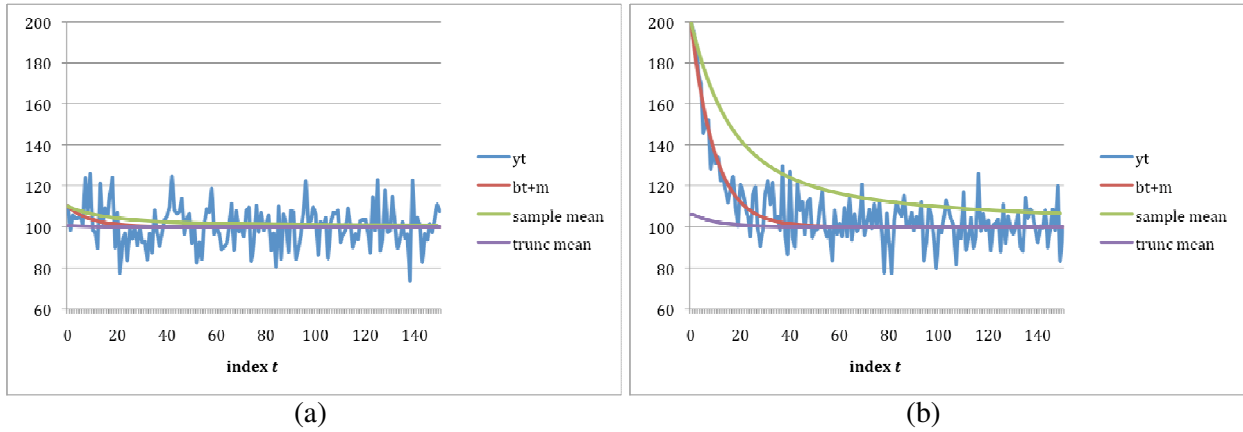
(a)

(b)

Figure 2: Example of initial bias below (a) and above (b) detection threshold

4.4 Using larger *n*

In practice we don't know *n(r)*. Implementations of MSER include a rule that forces more data to be collected if the selected *d** falls in the second half of the collected sample. Practical guidance also suggests that we should use *n* much larger than 65, regardless of the mean estimation method used, simply to improve our confidence in the estimate about the mean. So, what happens when we use *n* much larger? It turns out that the estimate of *d** generated in the case we've been examining is highly resilient to increases in the value of *n* chosen. In Figure 3 we repeat the bias-to-noise plot of Figure 1, but replace the computed *n(r)* value with a fixed value of *n* = 1000. The effect is roughly to shift the curves for each value of φ left by *r* = -0.03 and slightly increase the second partial derivative with respect to *r*.
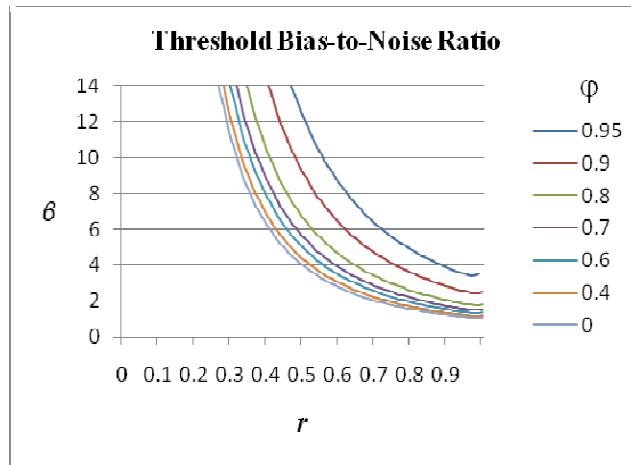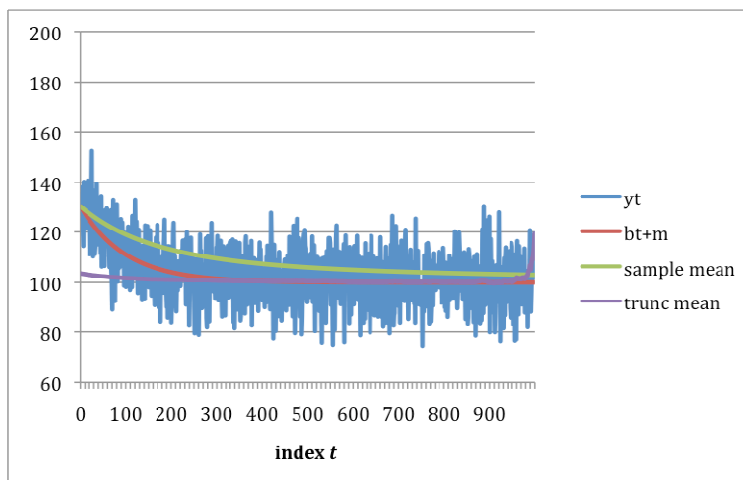

Figure 3: Threshold values of β, *r*, and φ that give $E[d^*] = 0$, where *n* = 1000

Of special interest is a slight non-monotonicity introduced at values of *r* approaching 1. At these extreme values of *r*, the initial bias damps out so slowly that a larger bias signal, relative to the variance of the noise, is required to create a condition where MSER concludes that truncation is required. An example where *r* = 0.99 is shown in Figure 4. The other values are $\mu = 100$, *b* = 10, $\varphi = 0.9$, $\sigma = 10$, and, thus, $\beta = 1$.

Figure 4: Example case where *r* = 0.99

5   CONCLUSIONS

In this paper, we introduced a technique for analyzing the expected behavior of the MSER bias truncation algorithm. This technique was used to analyze the expected behavior for the particular case where the bias takes the form of a geometrically decaying bias signal. We found that for a given noise correlation value and bias decay parameter, we can find the threshold bias-to-noise ratio required for MSER to return any specified truncation value. For bias magnitudes below this threshold, the bias cannot be distinguished from the noise. We noted that higher correlation values lead to higher values of the threshold. Likewise, cases where the bias decays more rapidly, the number of biased samples is small enough that a larger bias relative to the noise variance is required for MSER to recommend truncation. These results are consistent with the observation of White and Robinson (2010) that initialization bias increases not only with the magnitude of the bias, but also with the autocorrelation of the output. These observations also explain why MSER-5 provides improvement over MSER, smoothing the output time-series and thereby increasing the bias-to-noise ratio.

The techniques used in this paper can be used for other types of bias descriptions to examine the relationship between the expected results of MSER and the parameters of the bias description.

ACKNOWLEDGEMENT

REFERENCES

Hamilton, J. D. 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press

Hoad, K., S. Robinson, S., and R. Davies. 2009. Automating warm-up length estimation. *Journal of the Operational Research Society*, doi:10.1057/jors.2009.121.

Nelson, B. L. 2009. An Analysis of MSER. Unpublished manuscript.

Spratt, S. C. 1998. Heuristics for the startup problem. M.S. Thesis, Department of Systems Engineering, University of Virginia.

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

White, K. P, Jr., M. J. Cobb, and S. C. Spratt. 2000. A comparison of five steady-state truncation heuristics for simulation. In *Proceedings of the 2000 Winter Simulation Conference*, eds. J. A. Joines, R. R.

Barton, K. Kang, and P. A. Fishwick, 755–760. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

White, K. P., Jr., and S. Robinson, S. 2010. The initial transient problem (again), or why MSER works. *Journal of Simulation*, in press.

AUTHOR BIOGRAPHIES

K. PRESTON WHITE, JR., is Professor of Systems Engineering at the University of Virginia. He is a past member of the WSC Board of Directors and General Chairman for WSC2011. He received the B.S.E., M.S., and Ph.D. degrees from Duke University. He has held faculty appointments at Polytechnic University and Carnegie-Mellon University and served as Distinguished Visiting Professor at Newport News Shipbuilding and at SEMATECH. He is a member of INFORMS and a senior member of IEEE and IIE. He sits on the Advisory Board of VMASC. His email address is <kpwhite@virginia.edu>.

WILLIAM W. FRANKLIN holds a Ph.D. in systems engineering from the University of Virginia and continues to collaborate with UVA faculty in research. He is also Director of Data Management at a large financial services firm. Dr. Franklin received a B.S. and an M.S. in electrical engineering from the University of Texas-Austin. In addition to his work in information technology in the financial services industry, he has been a statistical analyst and worked in the defense industry, where he specialized in signal processing. His email address is <wwf6r@viginia.edu>.