Proceedings of the 2011 Winter Simulation Conference
S. Jain, R.R. Creasey, J. Himmelspach, K.P. White, and M. Fu, eds.

INTERVAL ESTIMATION USING REPLICATION/DELETION AND MSER TRUNCATION

Paul J. Sánchez

Operations Research Department
Naval Postgraduate School
1411 Cunningham Road
Monterey, CA 93943, USA

K. Preston White, Jr.

Department of Systems Engineering
University of Virginia
PO Box 400747
Charlottesville, VA 22976, USA

ABSTRACT

This paper addresses the construction of a consistent interval estimator for the steady-state mean within a replication/deletion framework for output analysis when MSER truncation is applied. Because the MSER truncation point is a random variable, the truncated output sequences for each replication typically are unequal in length. A weighting scheme is applied to the replication means to correct for unequal sample sizes, as is standard in ANOVA. A numerical example is provided to illustrate the procedure and consequences.

1 INTRODUCTION

Over a half-century ago, Conway recognized the initial transient as the first among three principle tactical problems in steady-state simulation (Conway 1959; Conway, *et al*. 1963; Goldsman, *et al*. 2010). An arbitrary selection of initial conditions for simulation runs introduces bias in the estimation of output statistics, such as the stationary mean. The most common approach for mitigating such bias is to truncate or delete some number of observations from the beginning of the output sequence and to compute statistics using only the remaining observations.

Alternate criteria for determining a good truncation point—one that adequately removes the bias without undue loss of precision—have been the subject of continuing invention beginning with Conway himself. Recently, however, a consensus has emerged among researchers that MSER has all of the properties most desired in a truncation criterion. It is effective and efficient at mitigating bias, robust across alternate forms of biasing functions, computationally trivial, easily understood, and does not require experimenter intervention to establish parameters.

MSER was initially developed by McClarnon (1990), White and Minnox (1994), and White (1997) and was applied and extended by Rossetti *et al.* (1995), Spratt(1998), Cobb (2000), White, *et al* (2000), and Franklin (2009). Mahajan and Ingalls (2004) determined three truncation criteria adequate, with MSER-5 recommended for its efficiency and robustness. Oh and Park (2006) compared their EVR method "with the method MSER-m known as the most sensitive rule in detecting bias and most consistent rule in mitigating its effects." MSER was shown to outperform EVR in almost all experiments. Sandikci and Sabuncuogy (2006) automated MSER-5 as their means for studying transients. Bertoli, Casale, and Serazzi (2007, 2009) selected MSER-5 as the initialization approach for their Java Modeling Tools package and included a usage wizard. The criterion gained additional traction with an exhaustive empirical evaluation by Hoad, *et al* (2008, 2011), who chose MSER-5 as the most suitable for automation over a wide range of published approaches to the transient problem, including heuristics, graphical procedures, initialization bias tests, statistical methods, and hybrid approaches.

White and Franklin (2010) confirm the empirical findings of White and Robinson (2010) regarding the relationship between the MSER truncation point and the degree of mean bias and autocorrelation in an output sequence. They introduce a parametric approach to analyzing the expected behavior of MSER and apply this approach to an output model with geometrically decaying bias and constant-parameter AR(1) white noise. Franklin *et al* (2009) explore the intuition that MSER minimizes the mean squared error (MSE) of the mean estimator. This empirical result is confirmed by Pasupathy and Schmeiser (2010). They reason that MSE is the most appropriate criterion for evaluating alternate truncation criteria, show that the MSER statistic is asymptotically proportional to the MSE, and conclude that the MSER statistic is a solid foundation for initial-transient algorithms. Pasupathy and Schmeiser also suggest two new algorithms using the MSER statistic and compare these to the original MSER algorithm using empirical results for M/M/1 and AR(1) data processes. Mokashi *et al* (2010) compared their N-Skart method with MSER-5 and achieved only modest improvements with considerably greater computational effort. Most recently, Hoad and Robinson (2011) consider the practical implementation of MSER-5.

2 TRUNCATION AND THE MSER CRITERION

Denote the output of a single replication of a simulation as the time series [y_i: i=1,2,…,n]. Truncation divides this into two subseries [(y_i: i=1,2,…,d),(y_i: i=d+1,2,…,n)], where d is the truncation point. For an output that is tallied, under truncation the estimator for the mean output is the sample mean of the second (reserved) subseries

$$\bar{Y}(n,d \mid y_0) = \frac{1}{n-d}\sum_{i=d+1}^{n} Y_i .$$

The MSER criterion for the optimal truncation point is

$$d^* = \underset{n>>d\geq 0}{\arg\min}\left[MSER(n,d \mid y_0)\right]$$

where the MSER statistic is the square of the estimated standard error of the mean $MSER(n,d \mid y_0) = \widetilde{SE}^2_{\bar{Y}}(n,d \mid y_0) = S^2_{\bar{Y}}(n,d \mid y_0)/n$ obtained using the large-sample variance

$$S^2_{\bar{Y}}(n,d \mid y_0) = \frac{1}{(n-d)^2}\sum_{\substack{i=d+1 \\ n>>d\geq 0}}^{n}\left(Y_i - \bar{Y}(n,d \mid y_0)\right)^2 .$$

Note that for correlated data the sample variance is a biased estimator. For a covariance-stationary process the *actual* squared standard error is

$$SE^2_{\bar{Y}}(\bar{x}) = \frac{\sigma^2}{n}\left[1 + 2\sum_{i=1}^{n-1}\left(1-\frac{i}{n}\right)\rho_i\right]$$

where σ^2 is the lag-zero autocovariance and ρ_k is the lag k –lag autocorrelation. Stationarity requires that the bracketed term is finite as $n\to\infty$. In practical terms this means that for sufficiently large n the bracketed term becomes *de facto* a constant. Therefore we can consider $SE^2_{\bar{Y}} = c\sigma^2/n$ and estimate it as $\overline{SE}^2_{\bar{Y}} = cS^2_{\bar{Y}}/n$. Since the MSER truncation point is based on *minimizing* $SE^2_{\bar{Y}}$ (rather than *estimating* it), we can effectively ignore the constant and determine a suitable MSER truncation point based on just the familiar variance estimator. This explains why Franklin and White (2008) found that, as expected, the Phillips-Perron variance estimator performed no better in practice than the naïve variance estimator.

3 APPLYING MSER WITHIN A REPLICATION/DELETION FRAMEWORK

Denote the output of m independent replications of a simulation as the set of m time series [y_{ij}: i=1,2,…,n_i; j=1,…,m]. Without loss of generality, consider that each of these series has the same initial condition y_0 and the same run length $n_j=n\ \forall j$. As before, truncation divides each time series into two subseries [(y_{ij}: i=1,2,…,d_j),(y_i: $i=d_j$+1,2,…,n; j=1,…,m], where d_j is the truncation point for the j^{th} series. For an output that is tallied, under replication/deletion we obtain a random sample of m values for the mean, each estimated from one of the reserved subseries as the corresponding sample mean

$$\bar{Y}_j(n,d_j \mid y_0)=\frac{1}{n-d_j}\sum_{i=d_j+1}^{n}Y_{ij};\ j=1,\ldots,m$$

Note, however, that the MSER truncation point for the j^{th} replication is the integer random variable D_j*. This means that attempts to create interval estimators using run-based replication (Conway's second principle problem) are biased if constructed from independent point estimates based on different sample sizes. Two solutions present themselves immediately: (1) find and apply the maximum truncation amount to all runs to reduce them to a common size; or (2) use weighted estimators for both the mean and variance. In this paper we will investigate the second option, which is standard in ANOVA and preserves as much usable data as possible.

Denote the total number of observations reserved across all runs as $N=\sum_{j=1}^{m}\left(n-d_j^*\right)$. We will regard the $\bar{Y}_j$'s as having a common expected value and underlying variance, but different standard errors because of the different sample sizes. We adjust for this using a weighted average

$$\bar{\bar{Y}}=\sum_{j=1}^{m}w_i\bar{Y}_j$$

If $\sum_{j=1}^{m}w_j=1$, i.e., we have a convex combination, then $\bar{\bar{Y}}$ is an unbiased estimator of the common mean. A well-known result is that variance of $\bar{\bar{Y}}$ is minimized when $w_i=(n-d_j^*)/N$. Finally, in order to obtain an interval estimator we need an estimator for the variance of $\bar{\bar{Y}}$

$$Var(\bar{\bar{Y}})=\frac{1}{1-\sum_{j=1}^{m}w_j^2}\sum_{j=1}^{m}w_j\left(\bar{Y}_j-\bar{\bar{Y}}\right)^2$$

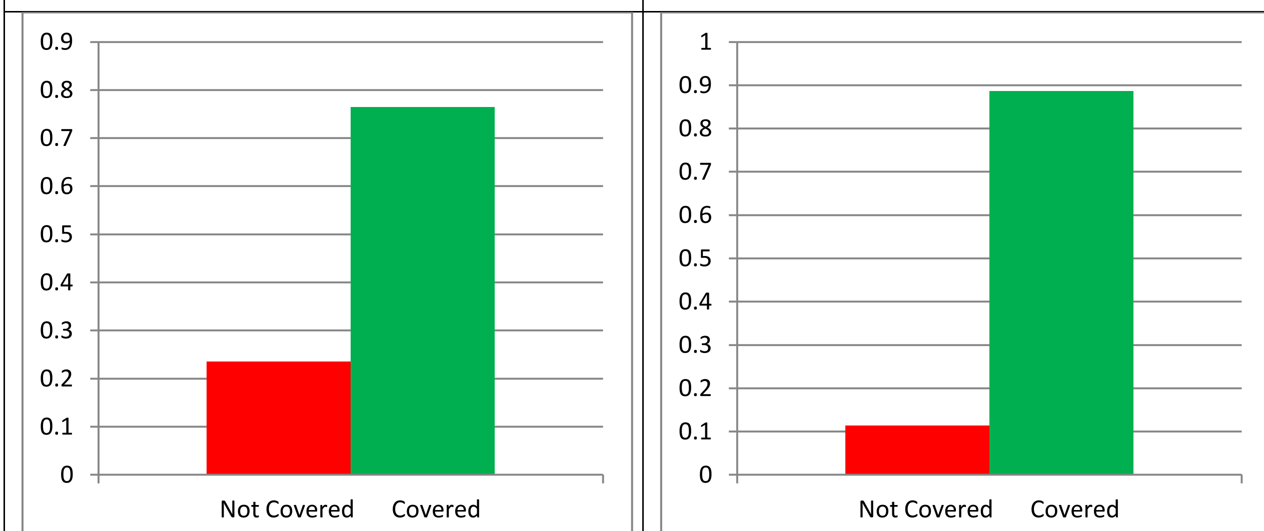
which is an unbiased estimator with $m\left(1-\sum_{j=1}^{m}w_j^2\right)$ degrees of freedom. Note that if the weights are all equal ($w_j=1/m\ \forall j$) this reduces to the familiar sample variance formula with m-1 degrees of freedom. Note also that for unequal weighting the degrees of freedom in general will not be integer and the corresponding t-value will need to be generated with software.

4 AN EXAMPLE

We ran a simulation of an M/M/1 queueing system at traffic intensity 0.95 (arrival rate = 19/time unit, service rate = 20/time unit) for n=10,000 and n=100,000 observations. We used delay in queue as our performance measure, which is suitable for tally statistics as described in Section 3. The output from each of 10 runs was truncated based on MSER-5 and the 10 resulting sample means were pooled using the weighting process described above to form a 90% confidence interval. This process was repeated 1,000 times to create 1,000 confidence intervals. Nominally we would expect 90% of such intervals to cover the true answer (μ= 0.95) obtained from theory. We created an indicator variable for each confidence interval to record whether it covered or did not cover the true mean. The results were analyzed using JMP-9 and are summarized in Figure 1.

Using a run length of 10,000 we obtained empirical coverage of 79% for a nominal 90% CI. When run lengths of 100,000 were used, the empirical coverage improved to 86.5%. Since all CIs were based on 10 runs, the degrees of freedom would be 9 if all run lengths were equal. Actual degrees of freedom varied, ranging from 9 down to 8.68 in our 2000 sets of experiments. Using a conservative 8 degrees of freedom had virtually no impact on the coverage, improving it only from 86.5% to 86.9% when run lengths of 100,000 were used. Upon inspection, failure to achieve nominal coverage seems to be because the confidence intervals are centered at the MSER estimate of the mean, which is known to be biased. As we observed, the bias has a greater impact when shorter runs are used.

Run length n = 10,000	Run length = 100,000



Mean	0.79	Mean	0.865
Std Dev	0.407512	Std Dev	0.3418946
Std Err Mean	0.0128867	Std Err Mean	0.0108117
Upper 95% Mean	0.815288	Upper 95% Mean	0.8862162
Lower 95% Mean	0.764712	Lower 95% Mean	0.8437838
Replications m	1000	Replications m	1000

Figure 1: Statistics of the indicator variable recording coverage of the 90% confidence interval on the mean delay in queue for a simulated M/M/1 queue.

5 CONCLUSION

In this paper we provide a brief overview of the MSER truncation criterion, a technical note on the construction of interval estimates using replication/deletion and MSER truncation, a weighting scheme that can be applied to this end, and an example of such an application. Empirical results from the example suggest that the weighting of unequal samples modestly underestimates the width of confidence intervals on the mean, with underestimation decreasing as a function of increasing run lengths.

ACKNOWLEDGMENTS

This work was completed while the second author was on leave as Visiting Professor in the Operations Research Department at the Naval Postgraduate School.

REFERENCES

Bertoli, M., G. Casale, and G. Serazzi. 2007. "The JMT Simulator for Performance Evaluation of Non-Product-Form Queueing Networks." *Proceedings of the 40th Annual Simulation Symposium*, 3-10.

Bertoli, M., G. Casale, and G. Serazzi. 2009. "JMT: Performance Engineering Tools for System Modeling." *ACM Metrics Performance Evaluation Review*, **36**(4)10-15.

Cobb, M. 2000. *Controlling Initialization Bias and Run Length in Steady-State Simulation*. M.S.thesis, School of Engineering and Applied Science, University of Virginia, Charlottesville, Virginia. Available from University of Viginia Library, lib-lend@virginia.edu, 434–982–3094.

Conway, R. W. 1963. "Some tactical problems in digital simulation." *Management Science* 10(1):47-61.

Conway, R. W., B. M. Johnson, and W. L. Maxwell. 1959. "Some problems of digital systems simulation." *Management Science* 6(1): 92-110.

Franklin, W. W. 2009. *Validation of the Theoretical Basis for the Mean Squared Error Reduction (MSER) Algorithm Used For Reducing Initialization Bias in Discrete Event Simulation Experiments*. Ph. D. dissertation, School of Engineering and Applied Science, University of Virginia, Charlottesville, Virginia. Available from University of Viginia Library, lib-lend@virginia.edu, 434–982–3094.

Franklin, W.W, and K.P. White, Jr. 2008. "Stationarity tests and MSER-5: Exploring the intuition behindmean-squared-error reduction in detecting and correcting initialization bias." In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. J. Mason, R. Hill, L. Moench, and O. Rose, 541-546. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Franklin, W. W., K. P. White, Jr., and K. A. Hoad. 2009. Comparing Warm-up Methods using Mean-Squared Error, Working paper, Department of Systems and Information, University of Virginia, Charlottesville, VA. Available from the authors.

Goldsman, D., R. E. Nance, and J. R. Wilson. 2010. "A Brief history of simulation revisited." In *Proceedings* of the *2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yucesan, 184-197. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Hoad, K. A., and S. Robinson. 2011. "Practical implementation of MSER-5." In *Proceedings* of the *2011 Winter Simulation Conference,* edited by B. Johansson, S. Jain, S. Jain, R.R. Creasey, J. Himmelspach, K.P. White, and M. Fu. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Hoad, K., S. Robinson, and R. Davies. 2008. "Automating warm-up length estimation." In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. J. Mason, R. R. Hill, L. Monch, O. Rose, T. Jefferson, and J. W. Fowler, 532–540. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Hoad K., S. Robinson, and R. Davies. 2011. "AutoSimOA: A framework for automated analysis of simulation output." *Journal of Simulation* 5:9-24.

Mahajan, P. S., and R.G. Ingalls. 2004. "Evaluation of methods used to detect warm-up period in steady state simulation." In *Proceedings of the 2004 Winter Simulation Conference*, edited by R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, 663-671. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Mokashi, A. C., J. J. Tejada, S. Yousefi, T. Xu, J. R. Wilson, A. Tafazzoli, and N. M. Steiger. 2010. "Performance comparison of MSER-5 and N-skart on the simulation start-up problem." In *Proceedings* of the *2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yucesan, 184-197. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

McClarnon, M. A. 1990. *Detection of steady state in discrete event dynamic systems: An analysis of heuristics*. M.S.thesis, School of Engineering and Applied Science, University of Virginia, Charlottesville, Virginia. Available from University of Vigínia Library, lib-lend@virginia.edu, 434–982–3094.

Oh, H-S, and K-J Park. 2006. "An Effective Heuristic to Detect Warm-Up Period in Simulation Output." In *Proceedings of the KISE Spring Conference*, 182-188.

Pasupathy, R. and B. Schmeiser. 2010. "The initial transient in steady-state point estimation: contexts, a bibliography, the MSE criterion, and the MSER statistic.". In *Proceedings* of the *2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yucesan, 184-197. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Rossetti, M. D., P. J. Delaney, and K. P. White, Jr. 1995. "Generalizing the half-width minimization heuristic for mitigating initialization bias." In *Proceedings of International Conference on Systems, Man, and Cybernetics*, 212–216. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Sandikci, B., and I. Sabuncuoglu. 2006. "Analysis of the behavior of the transient period in non-terminating simulations." *European Journal of Operational Research* 173: 252-267.

White, K. P., Jr. 1995. "A simple rule for mitigating initialization bias in simulation output: Comparative results." *In Proceedings of the International Conference on Systems, Man, and Cybernetics*, 206–211. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

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 *Proceeding of the 2000 Winter Simulation Conference*, edited by J. A. Joines, R. R. Barton, K. Kang, and P. A. Fischwick, 755–760. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

White, K. P., and M. A. Minnox. 1994. "Minimizing initialization bias in simulation output using a simple heuristic." In *Proceedings of International Conference on Systems, Man, and Cybernetics*, 215–220. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

White, K. P., Jr., and S. Robinson, 2010. "The initial transient problem (again), or why MSER works." *Journal of Simulation*, 4(3)268-272.

AUTHOR BIOGRAPHIES

PAUL J. SÁNCHEZ is a faculty member in the Operations Research Department at the Naval Postgraduate School. His research focuses on the intersection between computer modeling and statistics. In his spare time he enjoys reading science fiction and mysteries. His email address is pjsanche@nps.edu.

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 currently on leave as visiting professor at the Naval Postgraduate School. 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>.