Proceedings of the 2015 Winter Simulation Conference
L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds.

OBM CONFIDENCE INTERVALS: SOMETHING FOR NOTHING?

Yingchieh Yeh

Institute of Industrial Management
National Central University
Taoyuan, TAIWAN

Bruce Schmeiser

School of Industrial Engineering
Purdue University
West Lafayette, IN 47907, USA

ABSTRACT

Since the 1950s, nonoverlapping batch means (NBM) has been a basis for confidence-interval procedures (CIPs) for the mean of a steady-state time series. In 1985, overlapping batch means (OBM) was introduced as an alternative to NBM for estimating the standard error of the sample mean. Despite OBM's inherent efficiency, because the OBM statistic does not approach normality via the chi-squared distribution, no OBM CIP was introduced. We define two fixed-sample-size OBM CIPs. OBM1 is based on the result that asymptotically OBM has half again as many degrees of freedom as NBM. OBM2 does the same, but increases degrees of freedom. We argue that OBM's sampling distribution has skewness and kurtosis closer to normal than the chi-squared distribution. We show experimentally that for AR(1) processes the OBM CIPs perform better than NBM CIPs in terms of classic criteria and the VAMP1RE criterion. Finally, we introduce the concept of VAMP1RE-optimal batch sizes.

1 INTRODUCTION

Output analysis addresses the problem of determining and reporting the precision of point estimates of performance measures that arise when probability models are analyzed with Monte Carlo simulation. Since Mechanic and and McKay (1966a, 1966b) through Alexopoulos et al. (2014), confidence intervals have been the paradigm of simulation output analysis. From the beginning through Law and Kelton (1984), the focus was on fixed sample sizes and the performance measure was the mean. More recently, the focus has been on sequential procedures and other performance measures. We return to fixed sample sizes and the mean to discuss the use of overlapping batch means (OBM) as the basis for creating confidence intervals.

In Section 2, we review simulation output analysis, in Section 3 we state two closely related OBM confidence-interval procedures (CIPs), and in Section 4 we provide Monte Carlo simulation results of AR(1) data processes to show that OBM CIPs compare favorably to the NBM CIPs when using the same batch size. We conclude with the concept the VAMP1RE-optimal batch size, an idea that arose during our NBM/OBM Monte Carlo experiments.

2 SIMULATION OUTPUT ANALYSIS: REVIEW

A classic use of stochastic simulation is to estimate the mean μ of a time series $(Y_1, Y_2, \ldots, Y_n)$ from a steady-state stochastic process. After possibly deleting some initial data to avoid initialization bias, the point estimator of μ is the sample mean $\overline{Y} = n^{-1}\sum_{i=1}^{n} Y_i$. Asymptotically (as n goes to infinity) $\overline{Y}$ is normally distributed with mean μ and variance inversely proportional to n. More precisely, there is a process parameter τ that is the limiting value of $n\text{var}(\overline{Y})$. The limiting value is $\tau = \gamma_0\sigma^2$; here σ^2 is the variance of any one observation Y_i and $\gamma_0 = 1 + 2\sum_{h=1}^{\infty}\rho_h$ is the asymptotic sum of all lag-h autocorrelations.

Because the value of τ is seldom known, using only the data $(Y_1, Y_2, \ldots, Y_n)$ to estimate τ is at the heart of simulation output analysis for μ. For iid data that are normally distributed, the autocorrelations are zero, $\gamma_0 = 1$, and $S^2 = (n-1)^{-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$ estimates σ^2. Deviations from the assumptions opens the research world of simulation output analysis.

2.1 Reporting Precision

Having (somehow) an estimate, $\hat{\tau}$, of τ, one has an estimate of the standard error of $\overline{Y}$. In particular, $\hat{\text{se}}(\overline{Y}) = \sqrt{\hat{\tau}/n}$. This value can be used in its own right to report the precision of an observed $\overline{Y}$. Estimating the standard error, rather than developing confidence interval procedures (CIPs), has been followed in the Purdue dissertations Song (1988), Pedrosa (1994), and Yeh (2002); Song and Schmeiser (2009, 2011) discuss using the estimated standard error to determine the number of digits to report as well as how to format them.

The more-traditional approach to simulation output analysis is to calculate a confidence interval for the performance measure of interest, in our case the mean μ. With few exceptions, the confidence interval for the mean is

$$\overline{Y} \pm t_{\nu;\alpha}\sqrt{\hat{\tau}/n}, \tag{1}$$

where the interval is designed to *cover* the mean μ with nominal probability $1-\alpha$. For independent and identically distributed (iid) data, the coefficient $t_{\nu;\alpha}$ provides good (that is, actual matches nominal) coverage when it is the α quantile of the Student T distribution with ν degrees of freedom. Because data are often not independent and often not normally distributed, many CIPs for the mean have been published. They usually differ in their choice of the estimator $\hat{\tau}$ and the corresponding choice of the value of ν.

2.2 Comparing CIPs

Traditionally multiple CIPs are compared by considering multiple data processes and multiple nominal probabilities α. In addition to reporting the relationship between actual and nominal coverage probabilities, the expected interval half widths and standard deviations of the half widths are compared to those of other CIPs.

Schmeiser and Yeh (2002) and Yeh and Schmeiser (2015) advocate a single criterion to compare CIPs. The criterion, which we now refer to as VAMP1RE, is $\text{E}[(\Delta - \Psi)^2]$, where Ψ is Schruben's coverage value for the CIP of interest and Δ is the coverage value of an ideal CIP. Here our ideal CIP uses $\hat{\tau} = \tau$ and sets ν to infinity. The VAMP1RE criterion combines a CIPs lack of validity (distance between actual and nominal coverage probabilities) and inability to mimic (the ideal CIP); small VAMP1RE values are good. In our Monte Carlo experiments below, we report both traditional measures and VAMP1RE values.

2.3 Estimating τ

We review two methods for estimating τ: NBM and OBM. They are the foundation upon which the NBM and OBM CIPs are built.

The first method, NBM, partitions the data into $k = \lfloor n/m \rfloor$ batches, where m is the *batch size* and

$$\overline{Y}_j = m^{-1}\sum_{i=(j-1)m+1}^{jm} Y_i \tag{2}$$

is the jth batch mean. Typically n is much larger than m, so almost all of the n observations lie in one of the k batches;we ignore the remaining few. The NBM estimator of τ is

$$\hat{\tau}_{\text{nbm}} = d_{\text{nbm}}^{-1}\sum_{j=1}^{k}(\overline{Y}_j - \overline{Y})^2 \tag{3}$$

where $d_{\text{nbm}} = k(k-1)$. The NBM estimator assumes that when m grows large the batch means are iid and normally distributed. Then, after scaling, the distribution of $\hat{\tau}_{\text{nbm}}$ is chi-squared with $\nu_0 = (n/m) - 1$

degrees of freedom. The NBM estimator replaces the original n observations with the k batch means, estimates the variance of the sample mean as the variance of the batch means divided by k, as is appropriate if the batch means are independent.

The second method, OBM, is analogous, but it uses all $n-m+1$ possible batches of m observations, so now let

$$\overline{Y}_j = m^{-1} \sum_{i=j}^{j+m-1} Y_i \tag{4}$$

denote the jth batch mean. The OBM estimator of τ is then

$$\hat{\tau}_{\text{obm}} = d_{\text{obm}}^{-1} \sum_{j=1}^{n-m+1} (\overline{Y}_j - \overline{Y})^2 \tag{5}$$

where $d_{\text{obm}} = (n-m)(n-m+1)/m$. In Meketon and Schmeiser (1985), where OBM was introduced (but with a slightly different denominator), confidence interval procedures are not discussed.

2.4 Thoughts on OBM CIPs

The reason that Meketon and Schmeiser (1985), and researchers since, have not considered CIPs based on the OBM estimator of the sample mean is that (for $m > 1$) the overlapping batches are highly dependent; therefore, the OBM estimator goes to normality on a path that does not follow the chi-squared distribution with increasing degrees of freedom. The question that we address is what happens if an OBM-based CIP assumes, falsely, that the chi-squared distribution is appropriate.

We begin with a quick analysis of the relationship between the NBM and OBM estimators. For now, ignore end effects. The key is that, for batch size m, the OBM estimator $\hat{\tau}_{\text{obm}}$ is the simple average of m identically distributed NBM estimators, $\hat{\tau}_{\text{nbm}}$, with the ith NBM estimator beginning with observation i. Because the overlapping batch means are dependent, the m NMB estimators are dependent with positive correlation. Based on this relationship, what can we say about the sampling distribution of the OBM estimator, $\hat{\tau}_{\text{obm}}$?

First, the expected values of $\hat{\tau}_{\text{obm}}$ and $\hat{\tau}_{\text{nbm}}$ are the same, since dependence does not affect the expected values of sums; that the m NBM estimators are dependent is irrelevant.

Second, because of the central limit theorem, the shape of the sampling distribution of $\hat{\tau}_{\text{obm}}$ is closer to normal than the chi-squared sampling distribution of each of the m NBM estimators $\hat{\tau}_{\text{nbm}}$. As a consequence, and as we show in Section 6, (for the same batch size) the skewness and kurtosis of the OBM sampling distribution is closer to the normal distribution than to the chi-squared distribution; that is, the higher-order moments act as if the degrees of freedom where larger.

What remains is the variance of the OBM estimator, $\hat{\tau}_{\text{obm}}$. The variance does depend upon the correlations between the m NBM estimators, which is why asymptotically the variance of $\hat{\tau}_{\text{obm}}$ is two thirds that of $\hat{\tau}_{\text{nbm}}$, rather than the ratio being $1/m$ (as would happen if the NBM estimators were independent). For the chi-squared distribution, variance and degrees of freedom are inversely related. If the variance of $\hat{\tau}_{\text{obm}}$ were known, then assuming (again, falsely) the chi-squared distribution impies an number of degrees of freedom to use. With Monte Carlo experiments, we estimated the OBM variance, and therefore the implied OBM degrees of freedom, for iid normal data across many batch sizes. Similar in spirit to Schmeiser (1982), we extrapolate the iid results to auto-dependent data.

3 THREE BATCH-MEAN CIPS

We state three CIPs, one the classic NBM CIP and two new OBM CIPS. All three CIPs use the confidence-interval form of Equation (1). They differ in their method of estimating τ and in their choice of degrees of freedom.

The NBM CIP uses $\hat{\tau}_{\text{nbm}}$ and ν_0 degrees of freedom. In practice, the choice of the number of batches, k, or equivalently the batch size m, is central. There is always a tradeoff between choosing large values of m, where the resulting independence and normality provides good coverage probabilities, and small values of m, where the resulting large degrees of freedom provide short expected interval widths (high apparent precision) having small standard deviations (stable intervals). Rather than focusing on the batch size m, Schmeiser (1982) argued that avoiding the extremes by choosing k (the number of batches) greater than ten and no greater than thirty is a reasonable rule of thumb.

The two OBM CIPS use the OBM estimate of the variance of the sample mean of Equation (5). They differ only in their degrees of freedom for given run length n and batch size m.

The first CIP, which we refer to as OBM1, uses the asymptotic degrees of freedom, $\nu_1 = 1.5((n/m)-1)$ from Meketon and Schmeiser (1985). Our intention was to have only one OBM CIP, but after seeing the performance of OBM1, we created a modification, which we refer to as OBM2. The degrees of freedom for OBM2 are

$$\nu_2 = -0.4\nu_1^2 + 1.3\nu_1 + 2 \tag{6}$$

for $0 < \nu_1 \leq 2$ and otherwise

$$\nu_2 = \nu_1 + 4/\nu_1^2. \tag{7}$$

OBM2 has more degrees of freedom, especially when n/m is small. When the OBM1 degrees of freedom are $\nu_1 = 0$, the OBM2 degrees of freedom are $\nu_2 = 2$; when the OBM1 degrees of freedom are $\nu_1 = 2$, the OBM2 degrees of freedom are $\nu_2 = 3$. Asymptotically (as n/m grows large) the degrees of freedom are equal, $\nu_1 = \nu_2$. Because they use the same estimator of τ, by construction OBM1 has higher coverage probability and larger interval widths than OBM2.

The formula for ν_2 is our empirical approximation, chosen to balance simplicity and accuracy. But if accuracy, then what is the target? We considered using the degrees of freedom that matched actual and nominal coverage probabilities. But that choice would require a different formula for every nominal probability α. What we did was to appeal to the chi-squared distribution, despite that distribution not being applicable to OBM (or to NBM when m is too small). Recall that if $\nu\hat{\tau}/\tau$ has the chi-squared distribution with ν degrees of freedom, then its mean is ν and its variance is 2ν. But $\text{var}[\nu\hat{\tau}/\tau] = 2\nu$ implies that $\nu\ \text{var}(\hat{\tau}) = 2\tau^2$. Both the value of ν and the form of $\hat{\tau}$ depend on the estimation method, while the right-hand side is a function of only the data process. If both NBM and OBM2 satisfied the chi-squared conditions, then $\nu_{\text{nbm}}\text{var}(\hat{\tau}_{\text{nbm}}) = \nu_{\text{obm}}\text{var}(\hat{\tau}_{\text{obm}})$ would be true. The implication would be that

$$\nu_{\text{obm}} = \nu_{\text{nbm}} \left[\frac{\text{var}(\hat{\tau}_{\text{nbm}})}{\text{var}(\hat{\tau}_{\text{obm}})} \right].$$

We simulated iid normal data (with some large value of n) and various values of m. For each $\nu_{\text{nbm}} = (n/m) - 1$ value, we estimated the ratio $\text{var}(S^2_{\text{nbm}})/\text{var}(S^2_{\text{obm}})$ to obtain the target OBM degrees of freedom. OBM2 is the CIP that uses $\nu_2 = \nu_{\text{obm}}$ as approximated by Equations (6) and (7).

The OBM1 and OBM2 degrees of freedom formulas seem inappropriate for small values of m. In the extreme, when $m = 1$, the NBM and OBM estimators of the variance of the sample mean are identical; there is no overlapping. When n is reasonably large, approximately $\nu_1 = \nu_2$, but their value is half again larger than the value of ν_0. But large degrees of freedom all behave like infinite degrees of freedom, so the mismatch has negligible effect.

4 CIP COMPARISONS

We compare the NBM, OBM1, and OBM2 CIPs with a Monte Carlo experiment that considers only AR(1) data processes, for which $Y_i = \phi(Y_{i-1} - \mu) + \varepsilon_i$ and ε_i values are iid normal. The processes are at steady state. Our results do not depend on the values of the mean or variance. The first process has $\phi = 0$, which provides iid Y_i values. The second has $\phi = 0.9$, corresponding to $\gamma_0 = (1+\phi)/(1-\phi) = 19$, which implies that 19 autocorrelated observations contains the same information as one independent observation. Although this is a modest amount of autocorrelation, it allows us to investigate departure from independence. Our Monte Carlo results are precise to within one unit of the least-significant reported digit; the irony that we don't use confidence intervals is not lost on us, but they would have negligible width.

For a large value of n and various batch sizes m, we compare coverage probabilities in Section 4.1, half width moments in Section 4.2, and the VAMP1RE criterion values in Section 4.4. Not surprisingly, we conclude that OBM2 is performs best in this small experiment.

4.1 Coverage Probabilities

First consider iid normal data. For batch sizes $m = 1,2,5,10,20,50,100$, NBM's actual and nominal coverage probabilities match because the chi-squared distribution holds. Obtained with Monte Carlo simulation with $n = 200$, Table 1 shows OBM1 and OBM2 actual coverage probabilities for nominal coverage probabilities 0.5, 0.9, 0.95, and 0.99 for batch sizes ranging from $m = 1$ to $m = 199$. Other than for $m = 1$, where NBM and OBM are the sample estimator, the chi-squared distribution does not hold. Nevertheless, both OBM1 and OBM2 provide at least the nominal coverage probability. We created OBM2 because the OBM1 coverage probabilities were substantially greater than the nominal values for larger sample sizes (corresponding to fewer degrees of freedom). The bottom row, for $m = 199$ where only two batchs are overlapped, is interesting because the actual coverage probabilities for OBM2 are essentially perfect.

Table 1: For nominal coverage probabilities 0.5, 0.9, 0.95, and 0.99, actual coverage probabilities for OBM1 and OBM2 from iid normal data. Monte Carlo results for $n = 200$, with negligible sampling error (affecting only the last reported digit). For NBM, actual equal nominal; not applicable for $m > 100$.

$1-\alpha$	0.5		0.9		0.95		0.99	
m	OBM1	OBM2	OBM1	OBM2	OBM1	OBM2	OBM1	OBM2
1	0.500	0.500	0.900	0.900	0.950	0.950	0.990	0.990
2	0.500	0.500	0.900	0.900	0.950	0.950	0.990	0.990
5	0.500	0.500	0.900	0.900	0.950	0.950	0.990	0.990
10	0.500	0.500	0.900	0.900	0.950	0.950	0.990	0.990
20	0.501	0.501	0.901	0.901	0.951	0.950	0.991	0.991
50	0.505	0.503	0.910	0.910	0.963	0.960	0.996	0.995
100	0.562	0.508	0.983	0.919	0.998	0.969	1.000	0.998
150	0.785	0.511	1.000	0.927	1.000	0.975	1.000	0.999
199	NA	0.501	NA	0.901	NA	0.950	NA	0.990

Table 2 is analogous Table 1 except that the data process is now AR(1; $\phi = 0.9$). Because the NBM batches are not independent, the chi-squared distribution no longer holds, so NBM results are included. The Monte Carlo results are now based on run length $n = 2000$. As m increases to 1000, which corresponds to two batches and one degree of freedom, NBM's coverage is essentially perfect. Both OBM1 and OBM2 have results similar to NBM for small batch sizes, but as m increases OBM1's coverage is substantially too high; OBM2's coverage is also too high, but the excess is not as great.

In summary, ignoring some exceptions in the third digit, coverage probabilities favor OBM1, followed by OBM2, followed by NBM. By construction, because $\nu_1 \leq \nu 2$, OBM1 has coverage probability no less than OBM2.

4.2 Expected Interval Half Widths

First consider iid normal data. For NBM, OBM1, and OBM2, Table 3 shows expected interval half widths for batch sizes $m = 1,2,5,10,20,50,100,150,199$ with $n = 200$. For small batch sizes m, all interval half widths are similar. OBM1 has intervals shorter than NBM everywhere except for $m = 100$ (that is, two

Table 2: Actual coverage probabilities for NBM, OBM1, and OBM2 for AR(1; $\phi = 0.9$) data.

$1-\alpha$	0.5			0.9			0.95			0.99		
m	NBM	OBM1	OBM2	NBM	OBM1	OBM2	NBM	OBM1	OBM2	NBM	OBM1	OBM2
10	0.324	0.324	0.324	0.693	0.692	0.692	0.777	0.777	0.777	0.891	0.891	0.891
20	0.395	0.395	0.395	0.793	0.792	0.792	0.868	0.867	0.867	0.954	0.952	0.952
50	0.458	0.458	0.458	0.864	0.863	0.863	0.925	0.923	0.923	0.981	0.980	0.980
100	0.481	0.48	0.48	0.885	0.883	0.883	0.941	0.938	0.938	0.987	0.986	0.986
200	0.492	0.49	0.49	0.894	0.893	0.892	0.946	0.945	0.945	0.989	0.989	0.989
500	0.498	0.501	0.499	0.899	0.91	0.906	0.95	0.960	0.957	0.990	0.996	0.995
500	0.498	0.501	0.499	0.899	0.91	0.906	0.95	0.960	0.957	0.990	0.996	0.995
1000	0.502	0.557	0.504	0.901	0.980	0.915	0.951	0.998	0.965	0.990	0.999	0.998
1500	NA	0.774	0.502	NA	1.000	0.918	NA	1.000	0.969	NA	1.000	0.999
1990	NA	NA	0.296	NA	NA	0.742	NA	NA	0.856	NA	NA	0.974

NBM batches). By construction, OBM2 has intervals shorter than OBM1, because the degrees of freedom are larger. For large batch sizes, OBM2 has substantially shorter intervals.

Table 3: Expected half widths for NBM, OBM1, and OBM2. Monte Carlo results for iid normal data with $n = 200$.

$1-\alpha$	0.5			0.9			0.95			0.99		
m	NBM	OBM1	OBM2	NBM	OBM1	OBM2	NBM	OBM1	OBM2	NBM	OBM1	OBM2
1	0.675	0.675	0.675	1.65	1.65	1.65	1.97	1.97	1.97	2.60	2.59	2.60
2	0.675	0.675	0.675	1.66	1.65	1.65	1.98	1.97	1.97	2.62	2.60	2.61
5	0.676	0.676	0.676	1.67	1.66	1.66	2.01	1.99	1.99	2.69	2.65	2.65
10	0.679	0.677	0.677	1.71	1.69	1.69	2.07	2.03	2.03	2.82	2.74	2.74
20	0.684	0.681	0.681	1.78	1.74	1.73	2.20	2.11	2.11	3.16	2.94	2.93
50	0.705	0.700	0.697	2.17	1.97	1.94	2.93	2.54	2.49	5.38	4.08	3.95
100	0.798	0.813	0.713	5.04	3.45	2.19	10.14	5.60	2.97	50.8	16.6	5.45
150	NA	1.43	0.723	NA	37.91	2.36	NA	151.7	3.31	NA	3783	6.63
199	NA	NA	0.724	NA	NA	2.59	NA	NA	3.81	NA	NA	8.80

Now consider AR(1; $\phi = 0.9$) data. For NBM, OBM1, and OBM2, Table 4 shows expected interval half widths for batch sizes $m = 10, 20, 50, 100, 200, 500, 1000, 1990$ with $n = 2000$. The relative results are similar to those of iid data. For small batch sizes m, all interval half widths are similar. OBM1 has intervals no wider than NBM intervals everywhere except for $m = 1000$ (that is, two NBM batches). By construction, OBM2 has intervals no wider than OBM1, because the degrees of freedom are larger. For large batch sizes, OBM2 has substantially shorter intervals.

Table 4: Expected half widths for NBM, OBM1, and OBM2. Monte Carlo results for AR(1; $\phi = 0.9$) data with $n = 2000$.

$1-\alpha$	0.5			0.9			0.95			0.99		
m	NBM	OBM1	OBM2	NBM	OBM1	OBM2	NBM	OBM1	OBM2	NBM	OBM1	OBM2
10	0.417	0.417	0.417	1.02	1.02	1.02	1.22	1.21	1.22	1.60	1.60	1.60
20	0.515	0.515	0.515	1.26	1.26	1.26	1.51	1.50	1.51	2.00	1.99	1.99
50	0.609	0.608	0.608	1.51	1.50	1.50	1.81	1.80	1.80	2.42	2.39	2.39
100	0.646	0.643	0.643	1.62	1.60	1.60	1.97	1.93	1.93	2.69	2.60	2.60
200	0.668	0.663	0.663	1.74	1.69	1.69	2.15	2.06	2.05	3.09	2.86	2.85
500	0.669	0.691	0.688	2.15	1.95	1.92	2.91	2.51	2.46	5.34	4.03	3.90
1000	0.800	0.803	0.704	5.05	3.41	2.17	10.16	5.54	2.93	50.9	16.4	5.38
1500	NA	1.401	0.709	NA	37.1	2.31	NA	148	3.23	NA	3700	6.48
1990	NA	NA	0.388	NA	NA	1.39	NA	NA	2.04	NA	NA	4.71

Expected interval half width favors OBM2, followed by OBM1, followed by NBM. By construction, because $v_1 \leq v2$, OBM2 intervals are no wider than the corresponding OBM1 interval.

4.3 Half-Width Coefficients of Variation

A measure of the stability of a CIP's intervals is the coefficient of variation (CoV), the ratio of the half widths' standard deviation and their expected value. A zero CoV indicates that intervals all have the same half width; large values indicate that the interval half widths differ greatly data set to data set, in which case the confidence interval is not a good measure of the precision of the point estimator. For the CIPs considered here, the Student T multiplier cancels in the definition of CoV. Therefore CoV is a property of only the estimator of the variance of the sample mean. The result is that (1) when the mean and standard deviation exist, OBM1 and OBM2 have the same CoV values and (2) CoV values do not depend on the nominal coverage values.

Table 5: Coefficients of variation for NBM, OBM1, and OBM2 half widths. Monte Carlo results for iid normal data with $n = 200$.

m	NBM	OBM1	OBM2
1	0.050	0.050	0.050
2	0.071	0.062	0.062
5	0.114	0.093	0.093
10	0.163	0.133	0.133
20	0.239	0.191	0.191
50	0.422	0.311	0.311
100	0.755	0.394	0.394
150	NA	0.426	0.426
199	NA	NA	0.523

First consider iid normal data. For NBM, OBM1, and OBM2, Table 5 shows CoV values for batch sizes $m = 1, 2, 5, 10, 20, 50, 100, 150, 199$ with $n = 200$. For $m = 1$, all CoV values are identical because the intervals are identical; otherwise OBM CoV values are smaller than the NBM values. The OBM1 and OBM2 CoV values are identical, except for $m = 199$, where the OBM1 CoV is undefined.

Now consider AR(1; $\phi = 0.9$) data. For NBM, OBM1, and OBM2, Table 6 shows half-width CoV values for batch sizes $m = 10, 20, 50, 100, 200, 500, 1000, 1990$ with $n = 2000$. The relative results, and inferences, are the same as for iid data.

Table 6: Coefficients of variation for NBM, OBM1, and OBM2 half widths. Monte Carlo results for AR(1; $\phi = 0.9$) data with $n = 2000$.

m	NBM	OBM1	OBM2
10	0.064	0.064	0.064
20	0.078	0.076	0.076
50	0.115	0.106	0.106
100	0.164	0.144	0.144
200	0.239	0.201	0.201
500	0.422	0.321	0.321
1000	0.756	0.403	0.403
1500	NA	0.440	0.440
1990	NA	NA	0.534

In summary, the OBM estimator (of the variance of the sample mean) yields confidence intervals that are substantially more stable than NBM.

4.4 The VAMP1RE Criterion

Again, first consider iid normal data. For NBM, OBM1, and OBM2, Table 7 shows VAMP1RE criterion values for various batch sizes m with run length $n = 2000$.

Table 7: The VAMP1RE criterion for NBM, OBM1, and OBM2. Monte Carlo results for iid normal data with $n = 2000$.

m	NBM	OBM1	OBM2
1	0.000031	0.000031	0.000031
10	0.00031	0.00021	0.00021
20	0.00062	0.00041	0.00041
50	0.0016	0.0010	0.0010
100	0.0032	0.0021	0.0021
200	0.0069	0.0043	0.0043
500	0.021	0.010	0.010
1000	0.054	0.018	0.015
1500	NA	NA	0.017
1999	NA	NA	0.019

Now consider AR(1; $\phi = 0.9$) data. Table 8 is analogous to Table 7. Again, for small batch sizes, the three CIPs are essentially the same, so the VAMP1RE values are equal. Also, the OBM1 and OBM2 VAMP1RE values is similar for all batch sizes.

Table 8: The VAMP1RE criterion for NBM, OBM1, and OBM2. Monte Carlo results for AR(1, $\phi = 0.9$) data with $n = 2000$.

m	NBM	OBM1	OBM2
1	0.164	0.164	0.164
10	0.027	0.027	0.027
20	0.0094	0.0095	0.0095
50	0.0029	0.0027	0.0027
100	0.0035	0.0028	0.0028
200	0.0070	0.0048	0.0048
500	0.021	0.011	0.011
1000	0.054	0.018	0.016
1500	NA	NA	0.019
1999	NA	NA	0.051

In summary, consistent with the traditional measures, all three CIPs have the same VAMP1RE criterion values for small batch sizes m. Over the entire range of batch sizes, OBM1 and OBM2 have similar VAMP1RE values, except that OBM2 is defined over a wider range of batch sizes. For larger batch sizes, OBM1 and OBM2 have smaller VAMP1RE criterion values.

5 VAMP1RE-OPTIMAL BATCH SIZES

Tables 7 and 8 illustrate the concept of VAMP1RE-optimal batch sizes, the batch size that minimizes the VAMP1RE-criterion value. For iid data, not surprisingly, the VAMP1RE-optimal batch size is $m = 1$ for all three CIPs. For AR(1; $\phi = 0.9$) data, the VAMP1RE-optimal batch size for NBM is at about $m^* = 60$ whereas the OBM1 and OBM2 optimal batch sizes are a bit larger, because their VAMP1RE values grow slower for larger batch sizes.

For when the purpose is to estimate the variance of the sample mean, rather than to create a confidence interval, Song and Schmeiser (1995) discuss the concept of MSE-optimal batch sizes. Pedrosa (1994) discussed "1-2-1 OBM", a method to estimate the MSE-optimal batch size in practice. Yeh (2002) improves storage efficiency. An open research topic, then, is to find a way to determine VAMP1RE-optimal batch size in practice (rather than as we have done as researchers, knowing the data process).

6 SAMPLING DISTRIBUTIONS

OBM1 and OBM2 perform well despite using the Student T coefficient while not satisfying the independence condition for the chi-squared distribution to hold. Both use the same estimator of the variance of the sample mean, so we compare NBM and OBM.

Consider iid normal data. Table 9 shows mean, variance, skewness, and kurtosis for the NBM and OBM estimators of the variance of the sample mean. Both NBM and OBM are unbiased estimators of the variance of the sample mean, so all entries are 1. For the variances, small batch sizes yield similar results, but as batch sizes increase the asymptotic $2/3$ relationship appears. Skewness and kurtosis values are similar for all batch sizes other than the very large.

Table 9: Moments of the sampling distribution of the NBM and OBM estimators of the variance of the sample mean. IID data with $n = 2000$.

	Mean		Variance		Skewness		Kurtosis	
m	NBM	OBM	NBM	OBM	NBM	OBM	NBM	OBM
1	1	1	0.010	0.010	0.065	0.065	3.01	3.01
10	1	1	0.010	0.007	0.20	0.20	3.06	3.06
20	1	1	0.020	0.014	0.28	0.29	3.11	3.11
50	1	1	0.051	0.034	0.45	0.46	3.30	3.33
100	1	1	0.105	0.071	0.65	0.66	3.62	3.69
200	1	1	0.223	0.150	0.94	0.98	4.31	4.52
500	1	1	0.671	0.421	1.62	1.72	6.94	7.73
1000	1	1	2.045	0.675	2.81	1.95	14.8	8.75
1500	NA	1	NA	0.830	NA	2.27	NA	11.14
1999	NA	1	NA	0.943	NA	2.44	NA	12.34

Consider AR(1; $\phi = 0.9$) data. Table 10 is analogous to Table 9. Now, until batch sizes grow, NBM and OBM are biased estimators of the variance of the sample mean, but their biases are equal. For the variances, again small batch sizes yield similar results, but as batch sizes increase the asymptotic $2/3$ relationship appears. NBM and OBM skewness, as well as kurtosis, values are similar for all batch sizes other than the very large. Maybe surprisingly, introducing autocorrelation in the data causes little change in the the skewness and kurtosis values.

Table 10: Moments of the sampling distribution of the NBM and OBM estimators of the variance of the sample mean. AR(1; $\phi = 0.9$) data with $n = 2000$.

	Mean		Variance		Skewness		Kurtosis	
m	NBM	OBM	NBM	OBM	NBM	OBM	NBM	OBM
1	0.052	0.052	0.000026	0.000026	0.29	0.29	3.14	3.14
10	0.38	0.38	0.0024	0.0024	0.31	0.32	3.16	3.16
20	0.58	0.58	0.0082	0.0079	0.35	0.36	3.19	3.21
50	0.81	0.81	0.035	0.030	0.47	0.49	3.33	3.38
100	0.91	0.91	0.087	0.068	0.65	0.68	3.63	3.73
200	0.95	0.95	0.202	0.149	0.94	0.99	4.32	4.54
500	0.99	0.98	0.646	0.424	1.63	1.72	6.94	7.75
1000	1.00	0.99	2.007	0.669	2.81	1.87	14.8	8.10
1500	NA	0.97	NA	0.803	NA	2.26	NA	11.1
1999	NA	0.29	NA	0.095	NA	2.40	NA	12.1

In summary, the NBM and OBM sampling distributions are similar, except for the variances, which are equal for small batch sizes and have the $2/3$ ratio for large batch sizes. For iid data, the NBM moments are those of the chi-squared distribution. Therefore, although the OBM estimator goes to normality in a path other than the chi-squared distribution, in the sense of the first four moments the OBM is more favorable than the chi-squared distribution.

7 SUMMARY

We show that OBM, despite a false appeal to the chi-squared distribution, can be the foundation for fixed-sample-size CIPs. We define two OBM CIPs, OBM1 and OBM2, and compare them to the classic NBM CIP. OBM1, which uses the asymptotic degrees of freedom for all batch sizes, has the disadvantage of excess coverage probability; nevertheless, its interval half widths are never much larger than those of NBM. OBM2, which uses more degrees of freedom than OBM1 for small batch sizes, has less (but still some) excess coverage probability; its interval expected half widths are shorter than those of OBM1. Interval stability as measured by coefficient of variation, which depends only upon the NBM and OBM estimators of the variance of the sample mean, favors OBM.

In a single measure, the VAMP1RE criterion summarizes the previous measures; for large batch sizes, the OBM1 and OBM2 VAMP1RE criterion values are smaller than those of NBM. Finally, we introduce the concept of the VAMP1RE-optimal batch size, which is applicable to all CIPs based on batching ideas.

Although our Monte Carlo results are from only AR(1) processes, the conclusions are likely to generalize. The reason is that (ignoring end effects) the OBM estimator is the simple average of m identically distributed, but dependent, NMB estimators. The central limit theorem says that the OBM sampling distribution will be closer to normal than NBM's chi-squared distribution. In that sense, allowing the OBM CIPs to assume the chi-squared distribution is conservative. The result is that OBM CIPs may dominate NBM CIPs in general, the "something for nothing?" in the title.

ACKNOWLEDGMENTS

We thank two referees for comments that improved presentation.

REFERENCES

Alexopoulos, C., D. Goldsman, A. Mokashi, R, Nie, Q. Sun, K-W. Tien, and J.R. Wilson. "Sequest: A Sequential Procedure for Estimating Steady-State Quantiles". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S.Y. Diallo, I.O. Ryzhov, L. Yilmaz, S. Buckley, and J.A. Miller, 662–673. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Law, A.M., and W.D. Kelton (1984). "Confidence Intervals of Steady-State Simulations, I: A Survey of Fixed Sample Size Procedures." *Operations Research* 32: 1221–1239.

Mechanic, H., and W. McKay (1966a). "Confidence Intervals for Averages of Dependent Data in Simulations, I." Technical Report ASDD 17-201, Advanced Systems Development Division, IBM Corporation, Yorktown Heights, New York.

Mechanic, H., and W. McKay (1966b). "Confidence Intervals for Averages of Dependent Data in Simulations, II." Technical Report ASDD 17-202, Advanced Systems Development Division, IBM Corporation, Yorktown Heights, New York.

Meketon, M.S., and B.W. Schmeiser. "Overlapping Batch Means: Something for Nothing?" In *Proceedings of the 1985 Winter Simulation Conference*, edited by S. Sheppard, U. Pooch, and D. Pegden, 227–230. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Schmeiser, B.W., and M.D. Scott. "SERVO: Simulation Experiments with Random-Vector Output." In *Proceedings of the 1991 Winter Simulation Conference*, edited by B.L. Nelson, W.D. Kelton, and G.M. Clark, 928–936. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., Schmeiser, B.W. 1982. "Batch Size Effects in the Analysis of Simulation Output." *Operations Research* 30: 556–568.

Schmeiser, B.W. "Simulation Output Analysis: A Tutorial Based on One Research Thread." In *Proceedings of the 2004 Winter Simulation Conference*, edited by R.G. Ingalls, M.D. Rossetti, J.S. Smith, and B.A. Peters, 162–170. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Schmeiser, B.W., and Y. Yeh. "On Choosing a Single Criterion for Confidence-Interval Procedures." In *Proceedings of the 2002 Winter Simulation Conference*, edited by E. Yucesan, C.-H. Chen, J.L. Snowdon, and J.M. Charnes, 345–352. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Pedrosa, A. 1994. *Automatic Batching in Simulation Output Analysis*. Ph.D. thesis, Department of Industrial Engineering, Purdue University, West Lafayette, Indiana.

Song, W-M. T. 1988. *On Quadratic-Form Variance Estimators of the Sample Mean in the Analysis of Simulation Output*. Ph.D. thesis, Department of Industrial Engineering, Purdue University, West Lafayette, Indiana.

Song, W-M. T., and B.W. Schmeiser. 1995. "Optimal Mean-Squared-Error Batch Sizes." *Management Science* 41: 110–123.

Song, W.-M. T., and B. W. Schmeiser. 2009. "Omitting Meaningless Digits in Point Estimates: The Probability Guarantee of Leading-Digit Rules." *Operations Research* 57: 109–117.

Song, W-M. T., and B. W. Schmeiser. 2011. "Displaying Statistical Point Estimates Using the Leading-Digit Rule." *IIE Transactions* 43: 851-862.

Yeh, Y. 2002. *Steady-State Simulation Output Analysis: MSE-Optimal Dynamic Batch Means with Parsimonious Storage*. Ph.D. thesis, Department of Industrial Engineering, Purdue University, West Lafayette, Indiana.

Yeh , Y., and B.W. Schmeiser. 2003. "On the MSE Robustness of Batching Estimators." *Operations Research Letters* 32: 293-298.

Yeh, Y., and B.W. Schmeiser. 2015. "VAMP1RE: A Single Criterion for Rating and Ranking Confidence-Interval Procedures." *IIE Transactions*, forthcoming.

AUTHOR BIOGRAPHIES

YINGCHIEH YEH is an assistant professor of Institute of Industrial Management at National Central University, Taiwan. He received his Ph.D. from the School of Industrial Engineering at Purdue University. His primary research interests include simulation output analysis, applied probability and statistics, and applied operations research. His email address is yeh@mgt.ncu.edu.tw.

BRUCE SCHMEISER is a Professor Emeritus in the School of Industrial Engineering at Purdue University. His research interests center on developing methods for better simulation experiments. He is a fellow of INFORMS and IIE, as well as recipient of the Informs Simulation Society's 2014 Lifetime Professional Achievement Award. A long-time participant in the WSC, he served as the 1983 Program Chair and the 1988–1990 President of the Board of Directors. His e-mail address is bruceschmeiser@gmail.com.