TECHNICAL NOTE

CONFIDENCE INTERVAL ESTIMATION OF THE MEAN OF STATIONARY STOCHASTIC PROCESSES: A COMPARISON OF BATCH MEANS AND WEIGHTED BATCH MEANS APPROACH

S. T. Akhavan-Niaki

Department of Industrial Engineering
Sharif University of Technology
Tehran, Iran

Abstract Suppose that we have one run of n observations of a stochastic process by means of computer simulation and would like to construct a condifence interval for the steady-state mean of the process. Seeking for independent observations, so that the classical statistical methods could be applied, we can divide the n observations into k batches of length m (n= k.m) or alternatively, transform the correlated batch means vector into an independent vector. These methods are known as (ordinary) batch means and weighted batch means, respectively. In this paper, using the probability of coverage and the half length of a confidence interval as criteria for comparison, we empirically show that weighted batch means is superior to ordinary batch means, and that it is less sensitive to batch sizes and total number of observations.

Key Words Simulation, Output Analysis, (Weighted) Batch Means

چکیده فرض کنید که n مشاهده ازیک فرآیند تصادفی از طریق یک بار اجرای کامپیوتری مدل شبیه سازی شده فرآیند دردست است و می خواهیم که یک فاصله اطمینان برای میانگین درحالت پایدار آن بسازیم. اگر در جستجوی مشاهدات مستقل باشیم به نحوی که از روشهای کلاسیک آماری جهت ساخت این فاصله اطمینان استفاده کنیم، می توانیم n مشاهده را به k دسته که هر دسته شامل m مشاهده باشد (n= k.m) تقسیم کنیم و آنگاه یا از روش معمولی میانگینهای دسته ای و یا از روش میانگینهای وزن داده شده دسته ای جهت ساخت فاصله اطمینان فوق الذکر استفاده کنیم. دراین مقاله، با استفاده از معیارهای پوشش و نیم عرض فاصله اطمینان و همچنین با استفاده از فرآیندهای مختلف تصادفی با میانگینهای مشخص، این دو روش با هم مقایسه شده اند و نشان داده شده است که روش میانگینهای وزن داده شده دسته ای نسبت به روش نسبتاً معروف میانگینهای دسته ای به مراتب روش بهتری می باشد.

INTRODUCTION

The determination of confidence intervals on the steady-state mean of a stochastic process arising from simulation experiments has been a problem of long standing interest for computer simulation practitioners and researchers. Six approaches have evolved: independent replications, batching, regeneration, autoregressive representation, spectral analysis, and standardized time series; as discussed in several papers. One of the new procedures is the weighted batch means method [1] and in this paper we only compare the results obtained from the implementation of this method and the well-known ordinary batch means procedure on output sequences produced from different stochastic processes. Both methods are similar in philosophy in that they try to avoid autocorrelation by breaking the data into "independent" segments. Then the sample means and the weighted sample means of the data in each segment are computed and the analysis for i.i.d. observations is applied to construct a confidence interval (c.i.) for the steady-state mean.

BACKGROUND

A few procedures have been developed to construct confidence intervals for the mean response in single steady-state simulation models. Law [2] grouped these procedures in four basic categories: **(i)** those that seek independent observations, e.g. replications and batch means method, **(ii)** those that seek to estimate dependence among the output variables, e.g. the method of spectrum analysis and autoregressive method, **(iii)** those that exploit special structure of the underlying process, e.g. the regenerative method, and **(iv)** those based on standardized time series. Each of the above approaches involves some basic assumptions on the process being simulated, which may not be realized in real-world systems. Some of these methods, such as the independent replications, are wasteful in term of the information obtained from the data. Others, such as spectrum analysis, autoregressive representation, and standardized time series place a heavy requirement on the user to be familiar with sophisticated methods of time series analysis. The regenerative method is simple and easy to understand and to implement, but its applicability to real-world systems is very limited. The remaining approach suggested in the literature, the batch means, is easy to understand and to apply and is based on the availability of i.i.d. observations. However, one key element in the batch means method is the determination of the number of observations per batch, which is highly model dependent. In addition, all of the above procedures require a relatively large number of observations. However, the weighted batch means procedure seems to be not only easy to understand and to use, but also less sensitive to the total number of observations. In addition it seems that the determination of the batch sizes is not a serious problem in this approach.

Since we are to compare the ordinary batch means and the weighted batch means, a very brief description of these approaches are given below.

Ordinary Batch Means Procedure

In this procedure we assume that the simulation output is a convariance stationary process with the mean of μ and lag s covariance $C_s = Cov(X_t, X_{t+s})$. Starting from some initial conditions (or after a warm-up period) a single run of length n is made. This run is then devided into k "bathes" of m consecutive observations each (let n= k.m). Let X(i,j), (i= 1,2,...,m and j= 1,2,..., k) be the ith observation from the ith batch, and define $\overline{X_j}\ (m)$ and $\overline{\overline{X}}\ (k,m)$ to be the sample mean of the m observations in the jth batch and our point estimator of the steady-state mean, respectively. If we choose m sufficiently large, the batch means are i.i.d. normal random variables with the mean of μ. Applying classical statistical methods, a $100(1-\alpha)\%$ c.i. for μ would be:

$$\overline{\overline{X}}\ (k,m) \pm t_{(k-1,1-\alpha/2)}\ \widehat{\sigma}\ (k,m)$$

where $t_{(k-1,\,1-\alpha/2)}$ is the upper $1-\alpha/2$ critical point of the t distribution with k-1 degrees of freedom (d.f.),:

$$\overline{\overline{X}}\ (k,m) = (1/k)\,[\sum_{j=1}^{k} \overline{X_j}\ (m)\,]$$

is a point estimator of μ, and:

$$\widehat{\sigma}^2\,(k,m) = \frac{\sum_{j=1}^{k} \{\overline{X_j}\ (m) - \overline{\overline{X}}\ (k,m)]^2}{k\ (k-1)}$$

is a point estimate of the $Var\ [\overline{\overline{X}}\ (k,m)]$.

There are three potential sources of error when using batch means to construct a c.i. for a steady-state mean: the correlation between the $\overline{X}_j(m)$'s, the fact that the $\overline{X}_j(m)$'s are not identically distributed with mean μ, and the nonnormality of the $\overline{X}_j(m)$.

Weighted Batch Means Procedure

One disadvantage and the most serious source of error of the ordinary batch means method is that the batch means are positively correlated (the case usually encountered in practice), giving a variance estimate that is biased low and a c.i. that is too small such that it does not cover the true mean of the stochastic process. In the weighted batch means method, however, some weight schemes for the batch means are found such that the weighted batch means become i.i.d. random variables. Thus, the most potential source of error involved in the ordinary batch means method is eliminated. This method is specially helpful in situations where observations from simulation outputs are hard or expensive to obtain and only limited data are available.

In this method, given the random vectors $\mathbf{X} = (\overline{X_1}(m), \overline{X_2}(m), \ldots, \overline{X_k}(m))^T$, a 100(1-$\alpha$)% c.i. for the steady-state mean is obtained using:

$$P\left[\sum_{i=1}^{k} W_i \overline{X}_i - \sqrt{\sum_{i=1}^{k} \lambda_i}\; Z_{\alpha/2} \le \mu \sum_{i=1}^{k} W_i \le \sum_{i=1}^{k} W_i \overline{X}_i + \sqrt{\sum_{i=1}^{k} \lambda_i}\; Z_{\alpha/2}\right] = 1-\alpha$$

where $\overline{X}_i$ is the ith batch mean, W_i is equal to the sum of the ith column of the **D** matrix (the transpose of a matrix whose columns are the unit eigenvectors of the variance-covariance matrix of the batch means vector), λ_i's are non-negative eigenvalues of the variance-covariance matrix associated with the random vector containing the batch means, and $Z_{\alpha/2}$ is the upper (1-α/2)% percentile of the standard normal distribution [1].

Possible Sources of Error

When applying the weighted batch means method, there are four possible sources of error, and one has to watch for the following pitfalls:

1. Make sure that the initial transient effects have been removed before collecting the data. Also, in order to estimate the variance-covariance matrix, one has to assume that the output is covariance stationary. This assumption is usually not far from reality after reaching steady-state conditions.

2. The batch size (m) should be large enough, otherwise the sum of the weighted batch means may not be approximately normally distributed.

3. In this method, after replacing the variance-covariance matrix with its estimate, the transformed variables may not be totally independent.

4. The variance-covariance estimator of the batch means vector may not be an unbiased and consistent estimator.

EMPIRICAL COMPARISON

In this section, a brief discussion is given on the results obtained from implementing the weighted batch means method on random outputs obtained from three different stochastic processes: **(1)** M/M/1 queuing systems with different traffic intensities, **(2)** M/M/2 queuing systems with different traffic intensities, and **(3)** An autoregressive processes of order one [AR(1)]. The results are compared with those obtained with the application of the ordinary batch means approach (OBMA). The M/M/1 queuing systems have an arrival rate of λ=1, and service rates of μ= 1.43, 1.25, and 1.11 (i.e. traffic intensities of ρ=0.7, 0.8, and 0.9). In this systems the expected waiting time in the queue, W_q, is known to be equal to the ratio of ρ to $\mu(1-\rho)$. The M/M/2 queuing systems have an arrival rates of λ=10 and service rates of μ= 5.38, 6.25 (i.e. traffic intensities of ρ=0.8 and 0.9). The expected waiting time in the M/M/2 queuing systems is known to be [3]:

$$W_q = \frac{1}{\lambda}\left[\frac{\rho^{c+1}}{(c-1)!\,(c-\rho)^2}\right]P_0$$

where C=2, $\rho = \lambda/2\mu$, and

$$P_0 = \left[\sum_{n=0}^{c-1}\frac{\rho^n}{n!} + \frac{\rho^c}{c!\,(1-\rho/c)}\right]^{-1}$$

The autoregressive model has an autocorrelation $\Phi = 0.7$, a mean $\mu = 0$, and standard normal random terms ε_i's: $X_i = 0.7\,X_{i-1} + \varepsilon_i$ with the initial value of X, X_0, equal to ε_0 [4].

A set of one hundred different output sequences were generated for each of the above processes, each output containing 2560 observations. To evaluate the performance of the weighted batch means method, and to compare the results with the OBMA, confidence intervals were built, in each case, on the C.I. coverage and its half length. Using the central limit theorem and the normal approximation to the binomial distribution, the 100(1-α)% confidence limits on the coverage would be

$$\frac{Y}{R} \pm Z_{\alpha/2}\sqrt{\frac{(Y/R)\,(1-Y/R)}{R}}$$

where R= 100 is the total number of replications of the stochastic processes, Y is the total number of times the c.i. method covered the true mean of the process, and $Z_{\alpha/2}$ is the upper (1-α/2) percentile of the standard normal distribution. Also, using the central limit theorem, with large values of R, the 100(1-α)% confidence limits on the half length of the confidence intervals would be:

$$\overline{H} \pm t_{(R-1,\alpha/2)}\sqrt{(S_H^2)/R}$$

where $\overline{H}$ and S_H^2 are the unbiased estimators of the mean and variance of the half length respectively.

Results Obtained with M/M/1 Queues

Since OBMA is very sensitive to batch size and to the total number of observations in the output process, the c.i. methods were applied with total number of observations n= 160, 320, 640, 1280, and 2560, with the following number of batches for each case: k= 2,4,5,8,10,16,20,32,40, and 64. The 95% confidence limits on the coverage, and half length of the 95% c.i. for the mean waiting time in R=100 replications of an M/M/1 queuing system with λ=1 and μ=1.25 (ρ=0.8) were then calculated. Results show that the weighted batch means method gives much better coverage than the ordinary batch means method. The OBMA simply does not take into account the correlation between the batch means, and hence underestimates the variance of the sample means, and produces narrower confidence intervals. Also for a fixed total number of observations, as the batch size decreases, the OBMA coverage deteriorates. For a batch size of four observations the actual coverage reduces to only 39% when the nominal coverage is 95%. Also, the half length of the confidence intervals obtained with the OBMA gets smaller as the batch size decreases. This is expected because as the number of observations per batch decreases, higher correlations exist between batch means, and the OBMA, which ignores these correlations, underestimates the variance of the batch means and produces smaller intervals than it should. In all applications, the weighted batch means proved to be insensitive to the total number of observations, and consistently produced better results than the ordinary batch means approach. The advantage of the weighted batch means method becomes more significant with short simulations in which the collection of observations is expensive.

Since the weighted batch means approach is needed most with short (expensive) simulation processes, results in the remaining discussion are presented only with a total of 160 observations.

Results Obtained with M/M/2 Queuing Systems

The 95% confidence limits on the coverage of the 95% confidence intervals for W_q of M/M/2 queuing systems presented earlier were calculated. These results are obtained from R= 100 replications, each of which contain 160 observations and show that the OBMA tends to have less coverage with the increase in the number of batches, but the WBMA gives much better coverage than the OBMA. This difference is more recognizable with a higher intensive system like M/M/2 with traffic intensity of 0.9. Also for smaller batch sizes, the gap between the actual coverage associated with the OBMA and the nominal coverage increases. Again, these results show that the weighted batch means method provides more reliable c.i.'s for small batch sizes.

Results Obtained with Autoregressive Processes

The 95% confidence limits on the coverage and half length of the 95% confidence intervals for μ of the autoregressive process presented earlier were calculated at this stage. These results are obtained from R=100 replications, each of which contain 160 observations and show that for large batch sizes, there is no significant difference between the actual and the nominal coverages, but the OBMA yields better confidence intervals because of its narrower half lengths. However, for smaller batch sizes, the gap between the actual coverage associated with the OBMA and the nominal coverage increases. Again, these results show that the weighted batch means method provides more reliable c.i.'s for small batch sizes.

CONCLUSIONS

The weighted batch means method was implemented on output sequences from M/M/1 queuing systems with traffic intensities of 0.7, 0.8, and 0.9, M/M/2 queuing systems with traffic intensities of 0.8 and 0.9, and from an autoregressive model of order one [AR(1)]. The results obtained were compared with those of the ordinary batch means approach for different combinations of run length and batch sizes. The weighted batch means approach consistently produced better coverage than the ordinary batch means method, especially with small batch sizes. In general, the weighted batch means method is not sensitive to the batch sizes and performs well even if the total number of observations is not large. Although the weighted batch means approach is not as simple as the ordinary batch means method, it is still easy to understand and to apply.

In summary, the weighted batch means procedure is easy to understand and to implement, requires a relatively small number of observations, and produces c.i.'s that have good coverage.

REFERENCES

1. S. T. A. Niaki, "A Procedure for Building Confidence Interval on the Mean of Simulation Output Data", *Journal of Engineering, Islamic Republic of Iran*, Vol. 7, No. 2, (1994) 111-118.
2. A. M. Law, "Statistical Analysis of Simulation Output Data". Operations Research, Vol. 31, (1983) 983-1029.
3. H. A. Taha, "Operations Research", 4th Ed., MACMILLAN (1987).
4. G. S. Fishman, "Principles of Discrete Event Simulation", John Wiley & Sons Inc., New York (1978).