Systems Simulation
Output Data Analysis

Fatih Cavdur
fatihcavdur@uludag.edu.tr

May 12, 2014

Introduction

- Output analysis is the examination of data generated by a simulation.
- Its purpose is either to predict the performance of a system or to compare the performance of two or more alternative system designs.
- The need for statistical output analysis is based on the observation that the output data from a simulation exhibits random variability when RN generators are used to produce the values of the input variables.

Introduction

- Some output variables could be autocorrelated, implying the lack of statistical independence, means that classical methods of statistics, which assume independence, are not directly applicable to the analysis of these output data, and hence, must be properly modified.

- In addition to the autocorrelation present in most simulation output data, the specification of the initial conditions of the system at time 0 can pose a problem for the simulation analyst and could influence the output data. For purposes of statistical analysis, the effect of the initial conditions is that the output observations might not be identically distributed and that the initial observations might not be representative of the steady-state behavior of the system.

Types of Simulations with respect to Output Analysis
Terminating Simulations

In the analysis of simulation output data, a distinction is made between terminating or transient simulations and steady-state simulations. A *terminating* simulation is one that runs for some duration of time T_E where E is a specified event (or set of events) that stops the simulation. Such a simulated system "opens" at time 0 under well-specified *initial conditions* and "closes" at the *stopping time* T_E.

Types of Simulations with respect to Output Analysis
Non-Terminating Simulations

A *non-terminating system* is system that runs continuously, or at least over a very long period of time. Examples include assembly lines that shut down infrequently, continuous production systems of many different types, telephone systems, the Internet etc. A simulation of a non-terminating system starts at simulation time 0 under initial conditions defined by the analyst and runs for some analyst-specified period of time T_E. Usually, the analyst wants to study the steady-state or long-run properties of the system -that is, properties that are not influenced by the initial conditions of the model at time 0. A *steady-state simulation* is a simulation whose objective is to study long-run or steady-state behavior of a non-terminating system.

Stochastic Output

Consider one run of a simulation model over a period of time $[0, T_E]$. Since the model is an input-output transformation, and since some of the model input parameters are RVs, it follows that the model output variables are RVs. Example 11.7 and 11.8 are given in the text (DESS) to illustrate the nature of the output data from stochastic simulations and to give a preliminary discussion of several important properties of these data.

Performance Measure Estimations

Consider the estimation of a performance parameter, θ or ϕ of a simulated system. It is desired to have a point estimate and an interval estimate of θ or ϕ. The length of the interval estimate is a measure of the error in the point estimate. The simulation output data are of the form $\{Y_1, \ldots, Y_n\}$ for estimating θ; we refer to such output data as *discrete-time-data*, and the simulation output data are of the form $\{Y(t), 0 \le t \le T_E\}$ for estimating ϕ; we refer to such output data as *continuous-time data*. The parameter θ is an ordinary mean and ϕ is a time-weighted mean. We use two different symbols here simply to provide a distinction between ordinary means and time-weighted means.

The point estimator of θ based on the data $\{Y_1, \ldots, Y_n\}$ is defined by

$$\widehat{\theta} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

where $\widehat{\theta}$ is a sample mean based on a sample of size n. Simulation languages may refer to this as a "discrete-time", "collect" or "observational" statistic. The point estimator $\widehat{\theta}$ is said to be unbiased for θ if its expected value is θ -that is, if $E(\widehat{\theta}) = \theta$. In general, however, $E(\widehat{\theta}) \ne \theta$, and $E(\widehat{\theta}) - \theta$ is called the *bias* in the point estimator $\widehat{\theta}$. It is desirable to have estimators unbiased, or if not possible, have a small bias relative to the magnitude of θ.

The point estimator of ϕ based on the data $\{Y(t), 0 \leq t \leq T_E\}$ is defined by

$$\widehat{\phi} = \frac{1}{T_E} \int_0^{T_E} Y(t)dt$$

and is called a time average of $Y(t)$ over $[0, T_E]$. Simulation languages may refer to this as a "continuous-time", "discrete-change" or "time-persistent" statistic. In general $E(\widehat{\phi}) \neq \phi$. Again, we want to have unbiased or low-biased estimators.

CI Estimation

To understand confidence intervals (CI) fully, it is important to understand the difference between a measure of error and a measure of risk. One way to make the difference clear is to contrast a CI with a prediction interval (PI) (which is another useful output-analysis tool).

Both CIs and PIs are based on the premise that the data being produced by the simulation is represented well by a probability model. Suppose that model is the normal distribution with mean θ and variance σ^2, both unknown. To make the example concrete, let $\bar{Y}_i$ be the average cycle time for parts produced on the ith replication of the simulation. Therefore, θ is the mathematical expectation of $\bar{Y}_i$, and σ represents the day-to-day variation of the average cycle time.

CI Estimation

Suppose we want to estimate θ. The natural estimator for θ is the overall sample mean of R independent replications is given by $\bar{Y}.. = \sum_{i=1}^{R} \bar{Y}_i./R$. Since $\bar{Y}..$ is an estimate of θ, based on a sample, it has an error. A CI is a measure of that error. Let

$$S^2 = \frac{1}{R-1} \sum_{i=1}^{R} (Y_i. - \bar{Y}..)^2$$

be the sample variance across the R replications.

CI Estimation

The usual CI, which assumes the $Y_i.$ are normally distributed, is

$$\bar{Y}.. \pm t_{\alpha/2, R-1} \frac{S}{\sqrt{R}}$$

where $t_{\alpha/2}$ is the quantile of the t distribution with $R-1$ d.o.f. that cuts off $\alpha/2$ of the area of each tail.

PI Estimation

Now suppose that we need to make a promise about what the average cycle time will be on a particular day. A good guess is our estimator $\bar{Y}..$, but it is unlikely to be exactly right. Even θ itself, which is the center of the distribution, is not likely to be the actual average cycle time on any particular day, because the daily average cycle time varies. A PI, on the other hand, is designed to be wide enough to contain the actual average cycle time on any particular day with high probability. A PI is a measure of risk; a CI is a measure of error.

PI Estimation

The normal-theory prediction interval is

$$\bar{Y}.. \pm t_{\alpha/2, R-1} S \sqrt{1 + \frac{1}{R}}$$

The limit of this interval will go to 0 as R increases. In fact, in the limit it becomes $\theta \pm z_{\alpha/2}\sigma$ to reflect the fact that, no matter how much we simulate, our daily average cycle time still varies.
In summary, a PI is a measure of risk and a CI is a measure of error. We can simulate away error by making more replications, but we cannot simulate away risk. We can, however, do a better job of evaluating risk by making more replications. Look at Example 11.9 in the text.

Example

Suppose that overall average cycle time on 120 replications of a manufacturing simulation is 5.80 hours with a sample standard deviation of 1.60 hours. Since $t_{0.025,119} = 1.98$, a 95% CI is

$$\bar{Y}.. \pm t_{\alpha/2,R-1}\frac{S}{\sqrt{R}} \Rightarrow 5.80 \pm 1.98\frac{1.60}{\sqrt{120}} = 5.80 \pm 0.29 \text{ hours}$$

On any particular day, we are 95% confident that the average cycle time for all parts produced on that day will be the PI

$$\bar{Y}.. \pm t_{\alpha/2,R-1}S\sqrt{1+\frac{1}{R}} \Rightarrow 5.80 \pm \sqrt{1+\frac{1}{120}} = 5.80 \pm 3.18 \text{ hours}$$

Output Analysis for Terminating Simulation

Consider a terminating simulation that runs over a simulated time interval $[0, T_E]$ and results in observations $Y_1, \ldots, Y_n$. The sample size, n, may be a fixed number, or it may be a RV. A common goal in simulation is to estimate

$$\theta = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)$$

or, if the data are of the form $\{Y(t), 0 \le t \le T_E\}$,

$$\phi = E\left(\frac{1}{T_E}\int_0^{T_E} Y(t)dt\right)$$

Output Analysis for Terminating Simulation

The method used in each case is the method of independent replications. The simulation is repeated a total of R times, each run using a different RN stream and independently chosen initial conditions. We address this problem.
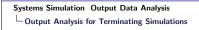
Statistical Background

Perhaps the most confusing aspect of simulation output analysis is distinguishing within-replication data from across-replication data, and understanding the properties and uses of each.

To illustrate the key ideas, think in terms of the simulation of a manufacturing system and two performance measures of that system, the cycle time for parts and the work in process (WIP). In computer applications, these correspond to the response time and the length of the task queue at the CPU etc.

Here is the usual set up for something like cycle time: Let Y_{ij} be the cycle time for the jth part produced in the ith replication. If each replication represents two shifts of production, then the number of parts produced in each replication might differ. Table 1 shows, symbolically, the results of R replications.

Statistical Background

Within-Replication	Across-Replication
$Y_{11}\ Y_{12} \ldots Y_{1n_1}$	$\bar{Y}_1., S_1^2, H_1$
$Y_{11}\ Y_{12} \ldots Y_{1n_2}$	$\bar{Y}_2., S_2^2, H_2$
$\ldots \ldots \ldots \ldots$	$\ldots \ldots \ldots$
$\ldots \ldots \ldots \ldots$	$\ldots \ldots \ldots$
$Y_{R1}\ Y_{R2} \ldots Y_{Rn_R}$	$\bar{Y}_R., S_R^2, H_R$
	$\bar{Y}.., S^2, H$

Table : Within and Across-Replication Cycle Time Data

Statistical Background

The across-replication data are formed by summarizing within-replication data: $\bar{Y}_i.$ is the sample mean of the n_i cycle times from the ith replication, s_i^2 is the sample variance of the same data, and

$$H_i = t_{\alpha/2, n_i - 1} \frac{S_i}{\sqrt{n_i}}$$

is a CI half-width based on this dataset. The average of the daily cycle time averages

$$\bar{Y}.. = \frac{1}{R} \sum_{i=1}^{R} \bar{Y}_i$$

Statistical Background

The sample variance of the daily cycle time averages,

$$S^2 = \frac{1}{R-1} \sum_{i=1}^{R} (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$$

and finally, the CI half-width

$$H = t_{\alpha/2, R-1} \frac{S}{\sqrt{R}}$$

The quantity $S/\sqrt{R}$ is the standard error, which is sometimes interpreted as the average error in $\bar{Y}_{\cdot\cdot}$ as an estimator of θ.

Statistical Background

Within a replication, work in process (WIP) is a continuous-time output, denoted $Y_i(t)$. The stopping time for the ith replication, T_{E_i}, could be a RV, in general; in this example, it is the end of the second shift.

Within-Replication	Across-Replication
$Y_1(t), 0 \le t \le T_{E_1}$	$\bar{Y}_{1\cdot}, S_1^2, H_1$
$Y_2(t), 0 \le t \le T_{E_2}$	$\bar{Y}_{2\cdot}, S_2^2, H_2$
… … … …	… … …
… … … …	… … …
$Y_R(t), 0 \le t \le T_{E_R}$	$\bar{Y}_{R\cdot}, S_R^2, H_R$
	$\bar{Y}_{\cdot\cdot}, S^2, H$

Table : Within and Across-Replication WIP Data

Statistical Background

The within-replication sample mean and variance are defined appropriately for continuous-time data:

$$\bar{Y}_{i\cdot} = \frac{1}{T_{E_i}} \int_0^{T_{E_i}} Y_i(t)\,dt$$

and

$$S_i^2 = \frac{1}{T_{E_i}} \int_0^{T_{E_i}} \left(Y_i(t) - \bar{Y}_{i\cdot} \right)^2 dt$$

A definition for H_i is more problematic, but take it to be

$$H_i = z_{\alpha/2} \frac{S_i}{\sqrt{T_{E_i}}}$$

Example

Consider Example 11.7 in the text (DESS) with $R = 4$ replications given in Table 11.1 in the text (DESS). The four utilization estimates, $\widehat{\rho}_r$, are time averages of the continuous variable Y. The simulation output data of the form

$$Y_r(t) = \begin{cases} 1, & \text{if Able is busy at time } t \\ 0, & \text{otherwise} \end{cases}$$

and $\widehat{\rho}_r = \bar{Y}_{r\cdot}$ as computed by the above time average formula with $T_E = 2$ hours.

Example

If we want a 95% CI for Able's true utilization ρ,

$$\bar{Y}.. = \widehat{\rho} = \frac{0.808 + 0.875 + 0.708 + 0.842}{4} = 0.808$$

and

$$S = \sqrt{\frac{(0.808 - 0.808)^2 + \dots (0.842 - 0.808)^2}{4 - 1}} = 0.072$$

Example

Thus, the standard error of $\widehat{\rho} = 0.808$ is estimated by s.e.$(\widehat{\rho}) = S/\sqrt{4} = 0.036$, and the 95% CI is

$$H = t_{0.025,3}\frac{S}{\sqrt{4}} = (3.18)(0.036) = 0.114$$

and we have 0.808 ± 0.114 or $0.694 \le \rho \le 0.922$ as shown below.

$$0.808 - 0.114 \le \rho \le 0.808 + 0.114 \Rightarrow 0.694 \le \rho \le 0.922$$

Example

Similarly, a 95% CI for mean time in system is obtained as follows:

$$\widehat{w} = \frac{3.74 + \ldots 3.98}{4} = 4.02$$

$$S = \sqrt{\frac{(3.74 - 4.02)^2 + \ldots (3.98 - 4.02)^2}{4 - 1}} = 0.352$$

and

$$H = t_{0.025,3} \frac{S}{\sqrt{4}} = (3.18)(0.176) = 0.560$$

and we have 0.808 ± 0.114 or $0.694 \le \rho \le 0.922$.

Example

A 95% CI for mean time in system is obtained as follows:

$$4.02 - 0.56 \le w \le 4.02 + 0.56 \Rightarrow 3.46 \le w \le 4.58$$

Confidence Intervals with Specified Precision

As mentioned above, the half-length H of a $100(1-\alpha)$% CI for a mean θ, based on the t distribution, is given by

$$H = t_{\alpha/2,R-1}\frac{S}{\sqrt{R}}$$

where S^2 is the sample variance and R is the number of replications. Suppose that an error criterion ϵ is specified; in other words, it is desired to estimate θ by $\bar{Y}..$ to within $\pm\epsilon$ with high probability-say, at least $1-\alpha$. Thus, it is desired that a sufficiently large sample size, R, be taken to satisfy

$$P(|\bar{Y}.. - \theta| < \epsilon) \geq 1 - \alpha$$

Confidence Intervals with Specified Precision

Assume that we initially have R_0 independent replications. We must have $R_0 \geq 2$, with 10 or more being desirable. The R_0 replications will be used to obtain an initial estimate S_0^2 of the population variance σ^2. To meet the half-length criterion, a sample size R must be chosen such that $R \geq R_0$ and

$$H = t_{\alpha/2,R-1}\frac{S_0}{\sqrt{R}} \leq \epsilon$$

Solving for R in above inequality shows that R is the smallest integer satisfying $R \geq R_0$ and

$$R \geq \left(\frac{t_{\alpha/2,R-1}S_0}{\epsilon}\right)^2$$

Confidence Intervals with Specified Precision

Since $t_{\alpha/2,R-1} \geq z_{\alpha/2}$, an estimate for R is given by

$$R \geq \left(\frac{z_{\alpha/2} S_0}{\epsilon} \right)^2 \tag{1}$$

where $z_{\alpha/2}$ is the $100(1-\alpha/2)$ percentage point of the standard normal distribution. Since $t_{\alpha/2,R-1} \approx z_{\alpha/2}$ for large R (say, $R \geq 50$), the second inequality for R is adequate when R is large. A $100(1-\alpha)\%$ CI for θ by

$$\bar{Y}.. - t_{\alpha/2,R-1} \frac{S}{\sqrt{R}} \leq \theta \leq \bar{Y}.. + t_{\alpha/2,R-1} \frac{S}{\sqrt{R}} \tag{2}$$

where $\bar{Y}..$ and S^2 are computed on the basis off all R replications.

Example

Suppose that we want to estimate Able's utilization in Example 11.7 in the text (DESS) to within ± 0.04 with probability 0.95. An initial sample of $R_0 = 4$ is taken, with the results given in Table 11.1 in the text. An initial estimate of the population variance is $S_0^2 = 0.00518$. The error criterion is $\epsilon = 0.04$, and the confidence coefficient is $1 - \alpha = 0.95$. Thus, the final sample size muse be at least as large as

$$R \geq \left(\frac{z_{\alpha/2} S_0}{\epsilon} \right)^2 = \frac{(1.96)^2 (0.00518)}{(0.04)^2} = 12.44 \tag{3}$$

Example

To test the possible candidates $(R = 13, 14, \ldots)$,

$$R = 13 \quad \Rightarrow \quad R \geq \left(\frac{t_{\alpha/2, R-1} S_0}{\epsilon} \right)^2 = 15.39$$

$$R = 14 \quad \Rightarrow \quad R \geq \left(\frac{t_{\alpha/2, R-1} S_0}{\epsilon} \right)^2 = 15.10$$

$$R = 15 \quad \Rightarrow \quad R \geq \left(\frac{t_{\alpha/2, R-1} S_0}{\epsilon} \right)^2 = 14,83$$

Thus, $R = 15$ is the smallest integer satisfying the above inequality, so $R - R_0 = 15 - 4 = 11$ additional replications are needed.

Example

Suppose that we want to estimate the 0.8 quantile of the time to failure (in hours) for the communications system in Example 11.3 and form a 95% CI for it. We have $R = 500$ replications. The point estimator $\widehat{\theta} = 4,644$ hours since the 80% of the data in the histogram is to the left of 4,644 hours. That means, it is the $500 \times 0.8 = 400$th smallest value of the sorted data.

Example

To obtain CI, we compute

$$p_l = p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{R-1}} = 0.8 - 1.96\sqrt{\frac{0.8(0.2)}{499}} = 0.765$$

$$p_u = p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{R-1}} = 0.8 + 1.96\sqrt{\frac{0.8(0.2)}{499}} = 0.835$$

Thus, $\widehat{\theta}_l = 4,173$ (it is the $500(0.765) = 382$nd smallest value, rounding down), and $\widehat{\theta}_u = 5,199$ (it is the $500(0.835) = 418$th smallest value, rounding up)

Estimating Probabilities and Quantiles from Summary Data

Knowing the equation for the CI half-width is important if all the simulation software provides is $\bar{Y}..$ and H and you need to work out the number of replications required to get a pre-specified precision, or if you need to estimate a probability or quantile. Sample standard deviation can be extracted from H by using the formula

$$S = \frac{H\sqrt{R}}{t_{\alpha/2,R-1}}$$

Estimating Probabilities and Quantiles from Summary Data

The more difficult problem is estimating a probability or quantile from summary data. When we only have the sample mean and CI half-width (which gives us the sample standard deviation), we can use a normal-theory approximation for the probabilities or quantities we desire,

$$P\{\bar{Y}_{i\cdot} \le c\} \approx P\left\{Z \le \frac{c - \bar{Y}_{\cdot\cdot}}{S}\right\}$$

and

$$\widehat{\theta} \approx \bar{Y}_{\cdot\cdot} + z_p S \qquad (4)$$

Example

From 25 replications of the manufacturing simulation, a 90% CI for the daily average WIP is 218 ± 32. What is the probability that the daily average WIP is less than 350? What is the 85th percentile of daily average WIP?

$$S = \frac{H\sqrt{R}}{t_{\alpha/2, R-1}} = \frac{32\sqrt{25}}{1.71} = 93$$

Example

Using the normal approximation and the normal table, we get

$$
\begin{aligned}
P\{\bar{Y}_{i\cdot} \le c\} \approx P\left\{Z \le \frac{c - \bar{Y}_{\cdot\cdot}}{S}\right\} &= P\left\{Z \le \frac{350 - 218}{93}\right\} \\
&= P\{Z \le 1.42\} \\
&= 0.92
\end{aligned}
$$

and

$$
\widehat{\theta} \approx \bar{Y}_{\cdot\cdot} + z_{0.85}S = 218 + 1.04(93) = 315 \text{ parts}
$$

Consider a single run of a simulation model whose purpose is to estimate a steady-state or long-run characteristic of the system. Suppose that the single run produces observations $Y_1, Y_2, \ldots$, which, generally, are samples of an autocorrelated time series. The steady-state measure of performance, θ, is defined by

$$
\theta = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} Y_i
$$

with probability 1, where the value of θ is independent of the initial conditions. Similarly, we have for the continuous case,

$$
\phi = \lim_{T_E \to \infty} \frac{1}{T_E} \int_0^{T_E} Y(t)\,dt
$$

with probability 1.

There are several methods of reducing the point-estimator bias caused by using artificial and unrealistic initial conditions in a steady-state simulation. The first method is to initialize the simulation in a state that is more representative of long-run conditions (sometimes called intelligent initialization). A second method is to divide each simulation run into two phases: first, an initialization phase fallowed by a data-collection phase. Look at the examples.

If $\{Y_1, \ldots, Y_n\}$ are not statistically independent, then S^2/n given as above is a biased estimator of the true variance, $V(\widehat{\theta})$. This is almost always the case when $\{Y_1, \ldots, Y_n\}$ is a sequence of output observations from within is a sequence of output observations from within a single replication. If a point estimator for θ is the sample mean $\bar{Y} = \sum_{i=1}^{n} Y_i/n$, a general result from mathematical statistics is that the variance of $\bar{Y}$ is

$$V(\bar{Y}) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \text{cov}(Y_i, Y_j)$$

where $\text{cov}(Y_i, Y_i) = V(Y_i)$.

To construct a CI for θ, an estimate of $V(\bar{Y})$ is required. But obtaining an estimate of the above is pretty much hopeless since each covariance term could be different. Fortunately, systems that have a steady-state will, if simulated long enough to pass the transient phase, produce an output process that is approximately covariance stationary. Intuitively, stationary implies that Y_{i+k} depends on Y_{i+1} in the same manner as Y_k depends on Y_1. In particular, the covariance between two RVs in the time series depends only on the number of observations between them, called the lag.

For a covariance-stationary time series, $\{Y_1, Y_2, \ldots\}$, define the lag-k by

$$\gamma_k = \text{cov}(Y_1, Y_{1+k}) = \text{cov}(Y_i, Y_{i+k})$$

which, by definition of covariance stationarity, is not a function of i. For $k = 0$, γ_0 becomes the population variance σ^2, that is

$$\gamma_0 = \text{cov}(Y_i, Y_{i+0}) = V(Y_i) = \sigma^2$$

The lag-k autocorrelation is the correlation between any two observations k apart. It is defined by

$$\rho_k = \frac{\gamma_k}{\sigma^2}$$

and $-1 \leq \rho_k \leq 1$, $k = 1, 2, \ldots$

If a time series is covariance stationary, then we can simplify the above variance equation as

$$V(\bar{Y}) = \frac{\sigma^2}{n} \left[1 + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n} \right) \rho_k \right]$$

where ρ_k is the lag-k autocorrelation as given above.

Why does autocorrelation make it difficult to estimate $V(\bar{Y})$? Recall that the standard estimate for the variance of a sample mean is S^2/n. By using the above variance expression, we can show that the expected value of the variance estimator S^2/n is

$$E\left(\frac{S^2}{n} \right) = BV(\bar{Y})$$

where

$$B = \frac{n/c - 1}{n - 1}$$

The effect of the autocorrelation on the estimator S^2/n is derived by an examination of the above three equations. There are three possibilities:

Case 1: If the Y_i are independent, then, $\rho_k = 0$ for $k = 1, 2, 3, \ldots$. Therefore, $c = 1 + 2\sum_{k=1} n - 1(1 - k/n)\rho_k = 1$ and equation 11 reduces to familiar σ^2/n. Also note that $B = 1$, so S^2/n is an unbiased estimator of $V(\bar{Y})$. The Y_i will always be independent when they are obtained from different replications.

Case 2: If the autocorrelations ρ_k are primarily positive, then $c = 1 + 2\sum_{k=1} n - 1(1 - k/n)\rho_k > 1$, so that $n/c < n$, and hence $B < 1$. Therefore, S^2/n is biased low as an estimator of $V(\bar{Y})$.

Case 3: If the autocorrelations ρ_k are substantially negative, then $0 \leq c < 1$, it follows that $B > 1$ and S^2/n is biased high for $V(\bar{Y})$.

If initialization bias in the point estimator has been reduced to a negligible level (through some combination of intelligent initialization and deletion), then the method of independent replications can be used to estimate point-estimator variability and to construct a CI. The basic idea is simple: Make R replications, initializing and deleting from each one the same way.

If the analyst decides to delete d observations of the total of n observations in a replication, then the point estimator of θ is $\bar{Y}..(n, d)$, defined as before, that is, the point estimator is the average of the remaining data. The basic raw output data are $\{Y_{rj}, r = 1, \ldots, R; j = 1, \ldots, n\}$. Each Y_{rj} is derived in one of the following ways:

Case 1: Y_{rj} is an individual observation from within replication r

Case 2: Y_{rj} is a batch mean from within replication r of some number of discrete-time observations.

Case 3: Y_{rj} is a batch mean of a continuous-time process over time interval j.

When using the replication method, each replication is regarded as a single sample for estimating θ.

For replication r, we define

$$\bar{Y}_r.(n, d) = \frac{1}{n - d} \sum_{j=d+1}^{n} Y_{rj}$$

as the sample mean of all (non-deleted) observations in replication r. Since all replications use different RN streams and all are initialized at time 0 by the same set of initial conditions (I_0), the replication averages $\bar{Y}_1.(n, d), \ldots, \bar{Y}_R.(n, d)$ are IID RVs, that is, they constitute a random sample from some underlying population having unknown mean

$$\theta_{n,d} = E[\bar{Y}_R.(n, d)]$$

The overall point estimator given above can also given by

$$\bar{Y}..(n, d) = \frac{1}{R} \sum_{r=1}^{R} \bar{Y}_{R}.(n, d)$$

Thus, it follows that

$$E[\bar{Y}..(n, d)] = \theta_{n,d}$$

If d and n are sufficiently large, then $\theta_{n,d} \approx \theta$, and $\bar{Y}..(n, d)$ is an approximately unbiased estimator of θ. The bias in $\bar{Y}..(n, d)$ is $\theta_{n,d} - \theta$.

For convenience, when the value of n and d are understood, abbreviate $\bar{Y}_{r}.(n, d)$ and $\bar{Y}..(n, d)$ by $\bar{Y}_{r}.$ and $\bar{Y}..$, respectively. To estimate the standard error of $\bar{Y}..$, first compute the sample variance

$$S^2 = \frac{1}{R-1} \sum_{r=1}^{R} (\bar{Y}_{r}. - \bar{Y}..)^2 = \frac{1}{R-1} \left(\sum_{r=1}^{R} \bar{Y}_{r.}^2 - R\bar{Y}_{..}^2 \right)$$

The standard error of $\bar{Y}..$ is given by

$$se(\bar{Y}..) = \frac{S}{\sqrt{R}}$$

A $100(1-\alpha)\%$ CI (which is only valid if the bias of $\bar{Y}..$ is approximately zero) is given by

$$\bar{Y}.. - t_{\alpha/2,R-1}\frac{S}{\sqrt{R}} \leq \theta \leq \bar{Y}.. + t_{\alpha/2,R-1}\frac{S}{\sqrt{R}}$$

Example

Consider the $M/G/1$ queuing system of Example 11.8 and 11.14 in the text (DESS). Suppose that the simulation analyst decides to make $R = 10$ replications, each of length $T_E = 15,000$ minutes, each starting at time 0 in the empty and idle state, and each initialized for $T_0 = 2,000$ minutes before data collection begins. The raw data output consists of the batch means; recall that each batch means is simply the average number of customers in queue for a 1,000-minute interval. The first two batch means are deleted ($d = 2$). The purpose of the simulation is to estimate the long-run time-average queue length by a 95% CI.

Example

The replication averages $\bar{Y}_r.(15, 2), r = 1, 2, \ldots, 10$, are shown in Table 11.8 in the text (DESS). The point estimator is computed as

$$\bar{Y}..(15, 2) = 8.43$$

Its standard error is given by

$$\mathrm{se}(\bar{Y}..) = \frac{S}{\sqrt{R}} = \mathrm{se}(\bar{Y}..(15, 2)) = 1.59$$

and thus,

$$8.43 - 2.26(1.59) \leq L_Q \leq 8.43 + 2.26(1.59) \Rightarrow 4.84 \leq L_Q \leq 12.02$$

Suppose that, in Example 11.15, the analyst had decided to delete one batch ($d = 1$) or no batches ($d = 0$). Using the quantities in Table 11.8 in the text again, the 95% CIs are

$$4.66 = 8.21 - 2.26(1.57) \leq L_Q \leq 8.21 + 2.26(1.57) = 11.76$$
$$4.62 = 7.94 - 2.26(1.47) \leq L_Q \leq 7.94 + 2.26(1.47) = 11.26$$

for $d = 1$ and $d = 0$, respectively. Note that, for a fixed total sample size, n, two things happen as fewer data are deleted:

- The CI shifts downward, reflecting the greater downward bias in $\bar{Y}..(15, d)$ as d decreases.
- The standard error of $\bar{Y}..(n, d)$, namely $S/\sqrt{R}$, decreases as d decreases.

Sample Size in Steady-State Simulations

Suppose it is desired to estimate a long-run performance measure, θ, within $\pm\epsilon$, with confidence $100(1-\alpha)\%$. In a steady-state simulation, a specified precision may be achieved either by increasing the number of replications (R) or by increasing the run length (T_E). Let's look at the following examples.

Batch Means

One disadvantage of the replication method is that data must be deleted on each replication and, in one sense, deleted data are wasted. This approach suggests that there might be merit in using an experimental design that is based on a single, long replication. The disadvantage of a single-replication design arises when we try to compute the standard error of the sample mean. Since we only have data from within one replication, the data are dependent and the usual estimate is biased.

Batch Means

One disadvantage of the replication method is that data must be deleted on each replication and, in one sense, deleted data are wasted. This approach suggests that there might be merit in using an experimental design that is based on a single, long replication. The disadvantage of a single-replication design arises when we try to compute the standard error of the sample mean. Since we only have data from within one replication, the data are dependent and the usual estimate is biased.

Batch Means

For a continuous-time process, then we form k batches of size $m = T_E/k$ and compute the batch means as

$$\bar{Y}_j = \frac{1}{m} \int_{(j-1)m}^{jm} Y(t + T_0)dt$$

for $j = 1, 2, \ldots, k$. For a discrete -time process,

$$\bar{Y}_j = \frac{1}{m} \sum_{i=(j-1)m+1}^{jm} Y_{i+d}$$

for $j = 1, 2, \ldots, k$ (if k does not divide $n - d$ evenly, round down to the nearest integer).

For either case, the variance of the sample mean is estimated by

$$\frac{S^2}{k} = \frac{1}{k} \sum_{j=1}^{k} \frac{(\bar{Y}_j - \bar{Y})^2}{k-1} = \frac{\sum_{j=1}^{k} \bar{Y}_j^2 - k\bar{Y}^2}{k(k-1)}$$

where $\bar{Y}$ is the overall sample mean of the data after deletion. Unfortunately, there is no widely accepted and relatively simple method for choosing an acceptable batch size m, but there are some guidelines.

Based on these, we can recommend the following:

- Obtain output data from a single replication and delete as appropriate. Recall the guideline of collecting at least 10 times as much data as is deleted.
- Form up to $k = 400$ batches (but at least 100) with the retained data, and compute the batch means. Estimate the sample lag-1 autocorrelation of the batch means as

$$\widehat{\rho}_1 = \frac{\sum_{j=1}^{k-1} (\bar{Y}_j - \bar{Y})(\bar{Y}_{j+1} - \bar{Y})}{\sum_{j=1}^{k} (\bar{Y}_j - \bar{Y})^2}$$

- Check the correlation to see whether it is sufficiently small.
  - If $\widehat{\rho}_1 \leq 0.2$, then re-batch the data into $30 \leq k \leq 40$ batches, and form a CI using $k - 1$ degrees of freedom for the t distribution and using the variance of the sample mean equation as given above to estimate the variance of $\bar{Y}$.
  - If $\widehat{\rho}_1 > 0.2$, then extend the replication by 50% to 100% and go to Step 2. If it is not possible to extend the replication, then re-batch the data into approximately $k = 10$ batches, and form the CI, using $k - 1$ degrees of freedom for the t distribution and using the variance of the sample mean equation as given above to estimate the variance of $\bar{Y}$.

As an additional check on the CI, examine the batch means for independence, using the following test. Compute the test statistic

$$C = \sqrt{\frac{k^2 - 1}{k - 2} \left[\widehat{\rho}_1 + \frac{(\bar{Y}_1 - \bar{Y})^2 (\bar{Y}_k - \bar{Y})^2}{2 \sum_{j=1}^{k} (\bar{Y}_j - \bar{Y})^2} \right]}$$

If $C < z_\beta$, then accept the independence of the batch means, where β is the Type I error level of the test. Otherwise, extend the replication by 50% to 100% and go to Step 2. If it is not possible to extend the replication, then re-batch the data into approximately $k = 10$ batches, and form the CI, using $k - 1$ degrees of freedom for the t distribution and using the variance of the sample mean equation as given above to estimate the variance of $\bar{Y}$.

Their average is

$$\widehat{\theta}. = \frac{1}{R}\sum_{i=1}^{R}\widehat{\theta}_i$$

It can be used as the point estimator of θ, and an approximate CI is

$$\widehat{\theta}. \pm t_{\alpha/2,R-1}\frac{S}{\sqrt{R}}$$

where S^2 is the usual sample variance of $\widehat{\theta}_1, \ldots, \widehat{\theta}_R$.
If there is a single replication, then the same reasoning applies if
we let $\widehat{\theta}_i$ be the quantile estimate from within the ith batch of
data. This requires sorting the data, or a histogram, within each
batch. If the batches are large enough, then these within-batch
quantile estimates will also be approximately IID.

Summary

- Reading HW: Chapter 11.
- Chapter 11 Exercises.