ESTIMATING RELIABILITY IN SIMULATION EXPERIMENTS

George S. Fishman
The RAND Corporation
Santa Monica, California

In all simulations containing random phenomena, the investigator wishes to estimate the reliability of his results. This study describes several approaches and suggests an easily implemented two-stage procedure.

DEFINITIONS

In most simulations the basic elements of collected information are interdependent. These elements, known as observations, are quantitatively summarized to yield the experimental result of interest. To analyze experimental results properly, we require a probability model that explicitly recognizes the interdependence among observations. To a great extent, covariance stationary stochastic processes are useful for this purpose.

Let Y_t be the t^{th} observation collected in an experiment. The sequence of observations $\{Y_t;\ t=1,2,\ldots,\infty\}$ defines a covariance stationary stochastic process if

(1) $$\mu = E(Y_t),$$

(2) $$R_\tau = E[(Y_t-\mu)(Y_{t+\tau}-\mu)] < \infty,$$

$$\tau = 0,\pm 1,\pm 2,\ldots,\pm\infty.$$

This means that the process $\{Y_t\}$ has an index invariant mean μ and an autocovariance function R that is a function only of the difference between indices t and $t+\tau$. This model is commonly used to study a series of events, especially in time series analysis.

The random variable Y_t may assume one of two meanings. It may measure system response over the interval $[t-1,t)$, or it may denote some characteristic of the t^{th} event in a series, for example, the waiting time of the t^{th} job to receive service. In the first case the index t refers to time. In the second it simply denots order, which may or may not agree with chronological ordering. We use the first definition, but our analysis applies equally to either definition.

We impose the condition that

(3) $$\lim_{\tau\to\infty} R_\tau = 0,$$

which implies that the influence of the past wears off as time elapses. This accords with our intuition and simplifies certain limiting procedures regarding the variance of the sample mean.

Suppose we collect N observations on $\{Y_t\}$ without error in a simulation experiment. The sample mean is then

(4) $$y_N = N^{-1}\sum_{t=1}^{N} Y_t$$

with variance

(5) $$\mathrm{var}(y_N) = N^{-1}\sum_{\tau=-N+1}^{N-1}(1-|\tau|/N)R_\tau .$$

Notice that y_N is a sample time integrated average over $[0,N)$. As N becomes large, we have

(6a) $$\lim_{N\to\infty} N \operatorname{var}(y_N) = m,$$

(6b) $$m = \sum_{\tau=-\infty}^{\infty} R_\tau < \infty,$$

so that

(6c) $$\operatorname{var}(y_N) \sim m/N.$$

The truth of (6a) is a consequence of (3). In addition the statistic $N^{\frac{1}{2}}(y_N-\mu)/m^{\frac{1}{2}}$ asymptotically has the unit normal distribution. These asymptotic properties are especially useful in determining sample size. Hereafter we assume N sufficiently large so that the asymptotic properties hold.

RELIABILITY

The sample mean y_N is a random variable which we take as an estimate of the population mean μ. By increasing N we theoretically improve our estimate of μ. Two questions now arise. The first is how reliable an estimate of μ we require. Since y_N is asymptotically normal we may write

(7a) $$\operatorname{Prob}(|y_N-\mu|<\delta) \sim 1-\alpha,$$

(7b) $$\delta^2 = Q^2 \operatorname{var}(y_N),$$

where

(7c) $$(2\pi)^{-\frac{1}{2}} \int_{-\infty}^{Q} e^{-z^2/2} dz = 1-\alpha/2.$$

For N observations collected in one replication we have

(8a) $$\delta^2 \sim Q^2 m/N.$$

The quantity Q is the $\alpha/2$ probability point of the unit normal distribution.

The second question is how large N should be so that y_N satisfies (7a). From (8a) we have

(8b) $$N \sim (Q/\delta)^2 m.$$

We defer our discussion of the choice of δ until Sec. 10 of [3] and concentrate here on estimating m and then N.

If $\{Y_t\}$ were a sequence of uncorrelated random variables, then

(9a) $$R_\tau = 0, \tau \neq 0,$$

so that

(9b) $$m = R_0.$$

An unbiased estimator of m would be

(9c) $$\hat{m} = \hat{R}_0 = (N-1)^{-1} \sum_{t=1}^{N} (Y_t - y_N)^2,$$

a simple straightforward result. The process $\{Y_t\}$ is generally autocorrelated, and hence (9a), which applies only for an uncorrelated sequence, is inappropriate. The quantity m may be estimated by one of several methods. They are described in Secs. 4 and 5 of [3].

INDEPENDENT REPLICATIONS

The most obvious estimating approach is to replicate the experiment. Let $Y_{j,t}$ be an observation at time t on the j^{th} replication.

For M independent replications each with S observations, we have the sample means

(10a) $$y_{j,S} = S^{-1}\sum_{t=1}^{S} Y_{j,t}, \quad j=1,2,\ldots,M,$$

(10b) $$y_S = M^{-1}\sum_{j=1}^{M} y_{j,S} = (MS)^{-1}\sum_{j=1}^{M}\sum_{t=1}^{S} Y_{j,t},$$

with variance

(10c) $$\mathrm{var}(y_S) = \mathrm{var}(y_{j,S})/M \sim m/(MS).$$

To estimate $\mathrm{var}(y_S)$ we use

(10d) $$\widehat{\mathrm{var}}(y_S) = [M(M-1)]^{-1}\sum_{j=1}^{M}(y_{j,S}-y_S)^2$$

so that

(10e) $$\hat{m} = MS\,\widehat{\mathrm{var}}(y_S).$$

To meet the reliability criterion (7a), each replication must have

(11a) $$N^* \sim m(Q/\delta)^2/M$$

observations. We estimate N^* from

(11b) $$\hat{N}^* = \max[S, \hat{m}(Q/\delta)^2/M].$$

If $\hat{N}^*-S > 0$, we run each replication for $\hat{N}^*-S$ additional observations. If $\hat{N}^*-S = 0$, we have enough observations and we use y_S as the estimate of μ, its estimated variance being $\hat{m}/(MS)$.

As an alternative, we may also run one replication, say the M^{th} one, for $N'-S$ additional observations, where $N'-S$ is the additional number of observations required to obtain the specified reliability. The sample mean is then

(12a) $$y_{S,N'} = M^{-1}\left(\sum_{j=1}^{M-1} y_{j,S} + N'^{-1}\sum_{t=1}^{N'} Y_{M,t}\right)$$

with variance

(12b) $$\mathrm{var}(y_{S,N'}) = m[(M-1)/S + 1/N']/M^2 .$$

To satisfy (7a), we require

(12c) $$N' \sim mS/[S(\delta M/Q)^2-m(M-1)],$$

which we estimate from

(12d) $$\hat{N}' = S\max\{1, \hat{m}/[S(\delta M/Q)^2-\hat{m}(M-1)]\}.$$

If $\hat{N}'-S > 0$, collect $\hat{N}'-S$ more observations on the M^{th} replication. Otherwise use y_S as the estimate of μ, its estimated variance being $\hat{m}/(MS)$.

It is helpful to compare the two ways of meeting the specified reliability when each requires more observations. We have

(13a) $$N'-S-(N^*-S)M = N' + (M-1)S-MN^*$$

$$= (M-1)[m(Q/\delta)^2-MS]^2/[SM^2-m(M-1)\cdot(Q/\delta)^2] > 0,$$

since

(13b) $$S(\delta M/Q)^2-m(M-1) > 0.$$

This implies that if we wish to minimize the number of observations, we should collect $\hat{N}^*$ observations on each of the M replications. From a practical viewpoint, it is more desirable simply to collect $\hat{N}'$ observations on the M^{th} replication, since we then need no longer account for the remaining (M-1) replications. This practical consideration suggests that the experimenter evaluate (13) and decide whether or not the operating advantage of using one replication overcomes the penalty of the additional computer time required.

AN ALTERNATIVE ESTIMATOR OF m

Using independent replications to determine sample size carries with it the necessity of handling many replications if $var(y_S)$ is to be estimated reliably. We suggest an alternative approach given in [2]. Here we describe its major points.

Suppose we initially collect S observations on an experiment. We form the sample autocovariance function

(14a) $$\hat{R}_\tau = S^{-1} \sum_{t=1}^{S-\tau} (Y_t - y_S)(Y_{t+\tau} - y_S)$$

$$\tau = 0,1,\ldots,P < S - 1.$$

(14b) $$y_S = S^{-1} \sum_{t=1}^{S} Y_t.$$

A rough estimate of m is

(15a) $$\tilde{m} = \hat{R}_0 + 2 \sum_{\tau=1}^{P} (1-\tau/P) \hat{R}_\tau \qquad P \leq S.$$

For large S we have

(15b) $$E(\hat{R}_\tau) \sim R_\tau - m/S,$$

so that for sufficiently large P

(15c) $$E(\tilde{m}) \sim m(1-P/S).$$

To remove the bias we modify our estimate of m:

(16a) $$\hat{m} = \tilde{m}/(1-P/S).$$

If $\{X_t\}$ is a normal process, then

(16b) $$var(\hat{m}) \sim \Psi m^2$$

(16c) $$\Psi = 4P/(3S).$$

The experimenter must specify the design parameter P. To understand its significance we note that as P increases,

(17) $$\lim_{P\to\infty} \sum_{\tau=-P}^{P} (1-|\tau|/P) R_\tau = m.$$

In the mathematical sense, this is the correct limit. In (15a) we use $\hat{R}_\tau$ instead of R_τ; therefore as P increases, the estimate $\hat{m}$ becomes less statistically reliable. Now we desire both good resolution (mathematical convergence) and good reliability (convergence in probability). One alternative is to make S very large, but this defeats our purpose. A compromise is therefore necessary that accomodates the conflicting objectives.

Let

(18a) $$\phi = [var(\hat{m})]^{\frac{1}{2}}/E(\hat{m}) \sim \Psi^{\frac{1}{2}}.$$

If

(18b) $$P = 3S/4,$$

then sampling fluctuations would be of the same order as m, thus yielding a poor estimate. If

(18c) $P=S/4$,

then $\phi \sim 0.57$. As a rule of thumb, we suggest that P not exceed S/4 and generally be much less. One may compute

$$\text{(18d)} \qquad \hat{m}_i = [\hat{R}_0 + 2 \sum_{\tau=1}^{P_i - 1} (1-\tau/P_i)\hat{R}_\tau]/(1-P_i/S),$$

$$\text{(18e)} \qquad P_i = i\Delta P \qquad i = 1,2,\ldots[S/4]/\Delta P,$$

where ΔP is some specified increment, plot the $\hat{m}$'s, and subjectively determine the value to which $\hat{m}$ seems to be converging. Alternatively, we may compute $\hat{m}$ for P=S/16, S/8 and S/4 and observe where convergence occurs. In some cases convergence may not occur. Then it is necessary to increase S before one can reasonably estimate m.

With this estimate of m, one determines the required sample size from

$$\text{(18f)} \qquad \hat{N} = (Q/\delta)^2 \hat{m}$$

and collects $\hat{N}$-S more observations, provided that $\hat{N}-S > 0$. As a final check we suggest estimating m with all $\hat{N}$ observations.

REFERENCES

1. Draper, N. R., and H. Smith, Applied Regression Analysis, John Wiley and Sons, New York, 1966.

2. Fishman, G. S., "Problems in the Statistical Analysis and the Length of Sample Records," Comm. ACM, Vol. 10, No. 2, February 1967.

3. Fishman, G. S., Digital Computer Simulation: Input-Output Analysis, RM-5540-PR, The RAND Corporation, February 1968.