Proceedings of the 1985 Winter Simulation Conference
D. Gantz, G. Blais, S. Solomon (eds.)

OVERVIEW OF STANDARDIZED TIME SERIES

Lee Schruben
School of Operations Research and Industrial Engineering
Cornell University
Ithaca, NY 14853

This paper presents the basic concepts and motivation for standardized time series analysis. Various applications to the analysis of simulation output are mentioned.

BACKGROUND:

Standardized time series can be motivated by first considering the role of standardization in scalar statistical analysis. Standardization has long been one of the central concepts in objective statistical analysis. The basic idea is to transform a statistic in such a manner that (under postulated conditions) its probabilistic behavior can be accurately approximated with a known probability law. This probability law is then used as a model for making inferences about the population under study. We will simply do an analogous transformation of an entire time series rather than just a scalar statistic.

The utility of standardization comes from the fact that the same probability model can be used to analyze data from a variety of sources. A statistician using the same mathematical methods, if applied correctly, can draw valid inferences using data from fields as diverse as economics, psychology, manufacturing, military systems, and medicine. Whatever the source of the data standardized statistics will be familiar to the analyst with a knowledge of the appropriate methodology. The technique of standardization thus serves as a mathematical surrogate for experience with the system under study. While standardization is not a substitute for such experience, it certainly augments it. The technique of standardizing statistics makes it possible to use the same statistical tables and computer packages in many situations and to a large extent is why statistics is so broadly applicable.

STANDARDIZING A SCALAR STATISTIC:

We break the process of standardizing a scalar statistic into elementary steps and carefully examine each step. This serves as our guide in how to standardize an entire time series.

The steps of standardizing a scalar statistic can be illustrated with the familiar example of the t-statistic. Suppose one has observations, Y_1, Y_2, ..., Y_n, that are independent and identically distributed and wishes to make inferences about the unknown mean μ of the population from which the data was taken. The statistic used is the average of the data, $\bar{\bar{Y}}_n$. A major problem is that the variance, σ^2, of the population is also unknown.

The following steps are followed.

(STEP 1) CENTER THE STATISTIC: We consider the random variable $\bar{\bar{Y}}_n - \mu$ which has an expected value of zero.

(STEP 2) SCALE THE STATISTIC MAGNITUDE: Standardized statistics are expressed in a common unit of measurement called a standard deviation. Here we scale the magnitude of our statistic by dividing by $\sigma/\sqrt{n}$. Our statistic is now

$$Z_n = (\bar{\bar{Y}}_n - \mu)/(\sigma/\sqrt{n})$$

Steps 1 and 2 result in the standardized statistic, Z_n. All sample means standardized in this manner will have the same first two moments; they all have a zero mean and unit standard deviation.

The problem we are faced with now is that the scaling parameter, σ, is not known to us. At this point we have at least two choices; we can consistently estimate σ or we can form a statistic where it cancels out. (In the reference by Glynn and Iglehart several interesting aspects of this choice in a simulation context are discussed.) The second choice, cancelling σ out of a ratio statistic, is usually followed and we make this choice here.

(STEP 3) CANCEL THE SCALE PARAMETER: Here the data is aggregated or batched into b exclusive adjacent groups of size m (assume $b = [n/m]$). The average of each group (often called a "batched mean") is computed. We denote these b batched means as, $\bar{Y}_{i,m}$, $i = 1,2,\ldots,b$. The sample variance of these batched means is computed as

$$S^2 = (n-1)^{-1} \sum_{i=1}^{b} (\bar{Y}_{i,m} - \bar{\bar{Y}}_n)^2.$$

We next form the random ratio,

$$T_{b-1} = ((\bar{\bar{Y}}_n - \mu)/(\sigma/\sqrt{n}))/\sqrt{(((b-1)S^2)/(\sigma^2(b-1)))}$$
$$= (\bar{\bar{Y}}_n - \mu)/(S/\sqrt{n}).$$

The important features of this ratio are that it not depend on σ and that its limiting distributions are known.

(STEP 4) APPLY LIMIT THEOREMS: We know that as $m \to \infty$ the distribution function of the random variable $(b-1)S^2/\sigma^2$ converges to that of a χ^2 random variable with b-1 degrees of freedom. Also as $n \to \infty$ $\bar{\bar{Y}}_n$ will converge to the constant μ and the distribution function of Z_n will converge to that of a standard Normal random variable. The distribution function of the random ratio, T_{b-1}, (being a continuous mapping) will thus converge to that of a t random variable with b-1 degrees of freedom. Recall that a t random variable is the ratio of a standard normal random variable divided the root an independent χ^2 random variable scaled by its degrees of freedom.

(STEP 5) USE THE LIMITING PROBABILITY MODEL FOR INFERENCE: The distribution of statistic, T_{b-1}, is approximated by the t distribution with b-1 degrees of freedom. This probability model is then used for hypothesis tests, confidence intervals on μ, comparison of two or more populations, etc. using widely available statistical tables or computer packages.

STANDARDIZING A TIME SERIES:

We will apply the concept of standardization to an entire time series of m observations generated by a run of a simulation program. Denote this output series as $Y_1, Y_2, \ldots, Y_m$. It might represent one of b "batches" from a single long run or one of b independent replications. Following the notation in the previous example n = mb will denote the total number of observations. Unlike the scalar example, it is possible that b = 1 making n = m. That is we will transform the original observations into a standardized sequence of observations. The hypotheses analogous to assuming that the data is identically distributed is to assume that the output is a sequence of observations of a stationary stochastic process. We will see in the next section how this can be tested leading to tests for the presence of initialization bias in the output. We no longer need to assume that the data is independent; however, we will assume that there is some minimal amount of randomness in the process. The mathematical assumptions needed are given in [Schruben, 1983] where it is argued that simulations on a computer will meet the imposed restrictions for applicability.

The sequence of statistics we will standardize will be the cumulative means up to and including the k^{th} observations, given by,

$$\bar{Y}_{i,k} = (1/k) \sum_{i=1}^{k} Y_i$$

We will follow steps in standardizing the sequence of cumulative means analogous that are followed in scalar standardization.

(STEP 1) CENTER THE SERIES: We will consider the zero mean sequence given by

$$S_m(k) = \bar{Y}_{i,m} - \bar{Y}_{i,k}$$

(STEP 2) SCALE THE SERIES MAGNITUDE: For dependent sequences the scaling constant σ^2 generalizes to,

$$\sigma^2 = \lim_{m\to\infty} m \operatorname{Var}(\bar{Y}_{i,m})$$

which is equal to the previously defined population variance in the special case of independent identically distributed data. To scale the standardized sequence to a unit standard deviation we divide by $\sqrt{m}(\sigma)/k$. Of course the scaling constant is again unknown. As expected, this unknown population parameter will cancel out of our statistics as before.

Now there is one step required that was not necessary in the scalar standardization case. Different time series can be of different length so we must also scale the index of the series. Thus we have the additional step

(STEP 2') SCALE THE SERIES INDEX: We will define the continuous index, $t = k/m$. Our previous index is thus given by $k = [mt]$. We also add the starting point $S_0 = 0$ so that $0 \leq t \leq 1$. The result is that all standardized time series have indices on the unit interval.

We now have what we will call a standardized time series given by

$$T_m(t) = ([mt])S_m([mt])/(\sqrt{m}(\sigma)).$$

(STEP 3) CANCEL THE SCALE PARAMETER: There are several functions that might be considered for the denominator of of a ratio that cancels σ. We will consider here only one such function, the sum (or limiting area under the function $T_m(t)$.

$$A = \sum_{k=1}^{m} T(k/m).$$

(STEP 4) APPLY LIMIT THEOREMS: It is shown in [Schruben, 1983] that the standardized series, $T_m(t)$, will converge in probability distribution to that of a Brownian Bridge stochastic process. Thus the Brownian Bridge process plays the role in time series standardization that the normal random variable played in the scalar standardization. An the important feature of the standardized series, $T_m(t)$, is that it is constructed to be asymptotically independent of the sample mean, $\bar{Y}_{i,m}$.

There are several functions of $T_m(t)$ that will also be asymptotically χ^2 distributed. The area, A, will have a limiting normal distribution with zero mean and variance $V = 12\sigma^2/(m^3 - m)$. Therefore, A^2/V will have a limiting χ^2 distribution with one degree of freedom.

Now consider where each of b independent replications (or b batches of data) are standardized in the manner above. We can then add the resulting χ^2 random variables, A^2/V, for each replication or batch to obtain a χ^2 random variable with b degrees of freedom. Also each of the replication of batch means can be treated as a set of scalar random variables and standardized giving another χ^2 random variable $(b-1)S^2/\sigma^2$ (given above). Due to the independence of $T_m(t)$ and the $\bar{Y}_{i,m}$'s, these two χ^2 random variables can be added giving a χ^2 random variable with 2b-1 degrees of freedom. This can be considered as a "pooled" estimator of σ^2 which we will denote as Q^2.

(STEP 5) USE THE LIMITING PROBABILITY MODEL FOR INFERENCE: Exactly like for the scalar case, the standardized (scalar) sample mean of all of the data can be divided by the square root of Q^2 over 2b-1 to form a ratio (independent of the scale parameter σ). For large values of m the distribution of this ratio can be accurately modeled as having a t distribution with 2b-1 degrees of freedom. The same types of inferences can be made for the dependent time simulation output series as were applicable in the independent data case. The resulting "t variate" is given by,

$$T_{b-1} = (\bar{\bar{Y}}_n - \mu)/(Q/\sqrt{n}).$$

Theoretical properties of confidence intervals formed using standardized time series are presented in [Goldsman and Schruben, 1984]. This paper compares the standardized time series approach to conventional methods.

APPLICATIONS OF STANDARDIZED TIME SERIES:

Standardized time series has been implemented in several simulation analysis packages. Most notably at IBM [Heidelberger and Welch, 1983], at Bell Labs [Nozari, 1985], and at G.E. [Schruben, 1986]. These packages typically control initialization bias (see also [Schruben, Singh and Tierney, 1983] and [Schruben, 1982]. and run duration as well as produce confidence intervals. Other applications of standardized time series have been to selection and ranking problems [Goldsman, 1983] and simulation model validation [Chen and Sargent, 1984].

CLOSING REMARKS

1) In the above derivations Q^2 (or S^2) must be asymptotically independent of $\bar{\bar{Y}}_n$ and have a limiting χ^2 distribution with known degrees of freedom. This is valid in the scalar case due to the asypmtotic normality of the batched means, $\bar{Y}_{i,m}$. For standardizing time series this is valid as shown in the lemma in [Schruben, 1983]. Both these results require that the batch size m become large as the sample size, n, increases. The common method to allow the batch size to grow as the sample size increases is to fix the number of batches, b. Fixing b at say 10 or 20 seems to be a reasonable thing to do in most applications as long as the sample size is large (see [Schmeiser, 1983]). However, as pointed out in [Glynn and Iglehart, 1985], fixing b leads to statistics that lack the asymptotic efficiency of estimating σ directly rather than cancelling it in a ratio statistic. The χ^2 quantity Q^2 can of course be used as a basis for consistently estimating σ. The theoretical objection to having a fixed number of batches may be essentially overcome by allowing both m and b to grow as n increases. Say, as n gets very large, one

might eventually set b approximately equal to n^{δ} where δ is arbitrarily close to but less than 1. The value of allowing the number of batches to increase must be weighed against the validity of having a relatively small number of batches. When the sample size is finite, the statistical validity of the approach is improved by having the batch size be as large as possible.

2) The limiting probability model of a t random variable has as its highest (and perhaps only) virtue the fact that it is widely tabled and has been studied extensively. There certainly exist other limiting models that might be used but none have been developed to the extent of the t model. The point is that there are alternative ways of performing each step in any standardization procedure. In this paper we wish to emphasize the importance of the concept of standardization and not the mechanics of any particular application. The specific examples of standardization presented were chosen as illustrations because they appear to work well in practice.

References:

Chen, B., and R. G. Sargent, 1984, "Confidence Interval Estimation for the Difference Between Two Stochastic Processes Using Standardized Time Series" Working paper 84-004, Dept. of IE and OR, Syracuse University.

Glynn, P., and D. Iglehart, 1985, "The Theory of Standardized Time Series" (forthcoming).

Goldsman, D., 1983 "Ranking and Selection in Simulation", Proceedings of the 1983 Winter Simulation Conference, pp. 387-393.

Goldsman, D., and L. Schruben, 1984, "Asymptotic Properties of Some Confidence Interval Estimators for Simulation Output", Management Science, Vol. 30, pp. 1217-1225.

Heidelberger, P., and P. Welch, 1983, "Simulation Run Length Control in the Presence of an Initial Transient", Operations Research, Vol. 31, pp. 1109-1144.

Nozari, A., 1985, "Confidence Intervals Based on Steady-State Continuous-Time Statistics" Bell Labs Tech. Rpt. (forthcoming)

Schmeiser, B. W., 1983, "Batch Size Effects in the Analysis of Simulation Output", Operations Research Vol. 30, pp. 565-568.

Schruben, L., 1982, "Detecting Initialization Bias in Simulation Output", Operations Research Vol. 30, pp. 569-590.

Schruben, L., 1983, "Confidence Interval Estimation Using Standardized Time Series", Operations Research, Vol. 31, pp. 1090-1107.

Schruben, L., H. Singh, and L. Tierney, 1983, "Optimal Tests for Initialization Bias in Simulation Output", Operations Research, Vol. 31, pp. 1167-1177.

Schruben, L., 1986, "A Procedure for Simulation Run Control and Analysis" (forthcoming).

LEE SCHRUBEN

Prof. Schruben received his undergraduate degree in engineering from Cornell University and a Masters degree from the University of North Carolina. His Ph.D. is from Yale University. He has been on the Cornell faculty in Operations Research and Industrial Engineering for the past nine years. Prior to coming to Cornell he was the Associate Director of the Health Systems Research Division of the Medical School at the University of Florida. His research interests are in the statistical design and analysis of large scale simulation experiments. His consulting activities have been primarily in the area of manufacturing systems simulation.