Proceedings of the 1990 Winter Simulation Conference
Osman Balci, Randall P. Sadowski, Richard E. Nance (eds.)

USE OF COMMON RANDOM NUMBERS IN COMPARING ALTERNATIVES

Gordon M. Clark

Department of Industrial and Systems Engineering
The Ohio State University
Columbus, Ohio 43210

ABSTRACT

A fundamental use for system simulation is to compare a number of different approaches to achieve a stated objective. The analyst often uses simulation to aid in evaluating multiple system designs and/or operation procedures. We say that simulation is used to compare alternatives. This paper assumes that the simulation output is stochastic and has a distribution that is unknown to the analyst. This paper presents methods for making comparisons by either ranking alternatives or selecting the best alternative with respect to a single output performance measure. To increase the effectiveness of the comparisons, these methods are able to use the variance reduction technique of common random numbers. The finite horizon simulation case is considered in this paper. Examples illustrate the application and utility of the methods.

1. INTRODUCTION

The analyst uses simulation to compare system performance in different situations or with different system designs. For example, an industrial engineer may use a simulation of a manufacturing shop to investigate various methods for reducing lead time. We define lead time as the time to produce an assembly subsequent to releasing materials to production. The emphasis is on comparing the simulation output obtained from multiple sets of inputs. We are assuming that the analyst will rarely use a simulation to estimate performance for a single set of inputs. We will call each set of inputs defined for comparison purposes an *alternative*.

Output analysis is complicated by the stochastic nature of output performance measures. This paper makes the following assumptions concerning the probability distribution of output performance measures:

- The form of the true underlying distribution is unknown.
- The variance can and will change with the alternative.

The above assumptions are common occurrences in system simulation. We use the term *replication* to specify repetition of a simulation with fixed inputs but different outputs due to different random numbers. In a stochastic environment, the analyst can increase his/her confidence in the validity of the comparison made by increasing the number of replications. An important question is how many replications are required to reduce the risk of making inaccurate comparisons. This paper assumes that the experimental objectives require simulations with finite horizons and do not require estimation of performance measures in steady state.

Variance reduction techniques offer the potential for reducing the number of replications required for an experimental objective. Brately, Fox, and Schrage [1987] give a comprehensive description of variance reduction in simulation.The ability to use the same random numbers for each alternative offers the potential for sharpening the comparisons. This is an example of a variance reduction technique and is called *common random numbers*. It is easier to apply than many of the other variance reduction methods. Nelson [1987] reviews the variance reduction method of common random numbers as well as two other commonly used methods. All procedures discussed in this paper can make use of common random numbers.

This paper considers comparisons in two forms:

1. The analyst wants to simultaneously compare all combinations of the alternatives, e.g., generate a ranking with respect to a single output measure, or
2. The analyst wants to select the best alternative based on an output measure.

The paper presents a procedure for each form in the above order. Examples illustrate the application of each procedure.

1.1 Comparison Problem

The simulation experiments must compare γ alternatives. Denote alternative j by Π_j. The value of the performance measure observed for Π_j on replication i is $X_i^{(j)}$. That is, $X_i^{(j)}$ is the simulation output performance statistic on replication i for Π_j. The mean value or expected value of $X_i^{(j)}$ is θ_j. Our objective is to compare the individual alternatives based on their respective values of θ_j. Throughout this paper, we assume that alternatives giving lower values of θ_j are preferred. For example, we will prefer alternatives giving lower values of mean waiting time. Since we do not know the true values of θ_j, we must estimate them from the simulation output statistics. The unbiased point estimator for θ_j is the sample average of the output statistics collected for Π_j. That is,

$$\bar{X}_n^{(j)} = \frac{1}{n}\sum_{i=1}^{n} X_i^{(j)},$$

based upon n replications of Π_j. Our inferences concerning the relative values of θ_j will never invert the ordering of $\bar{X}_n^{(j)}$. Important questions which we will address are:

- How many replications of each alternative must we observe in order to achieve a desired probability of selecting the best alternative?
- Are the observed differences in $\bar{X}_n^{(j)}$ sufficiently large to give us confidence in a ranking of the values of θ_j ?

1.2 Common Random Numbers

The objective is to use the same random numbers for each alternative to obtain a large positive correlation among their performance measure values on each replication. This reduces the variance of the difference between their respective performance measure values. For alternatives 1 and 2,

$$V(X_i^{(1)} - X_i^{(2)}) = V(X_i^{(1)}) + V(X_i^{(2)}) - 2\rho(X_i^{(1)}, X_i^{(2)})\sqrt{V(X_i^{(1)})}\sqrt{V(X_i^{(2)})}, \quad (1)$$

where

$V(X)$ = Variance of the random variable X

$\rho(X, Y)$ = Correlation between the random variables X and Y

For example, if we want to estimate the difference in mean waiting times for jobs entering a queue when two alternatives have unequal mean service times, we could use the same random number(s) for the m^{th} interarrival time (as well as for all other interarrival times) for each alternative. This may increase the correlation between $X_i^{(1)}$ and $X_i^{(2)}$, where these quantities are mean waiting times. The problem of coordinating random numbers so that identical random numbers determine analogous events in each alternative is called the *synchronization* problem. One can obtain dramatic differences in the effectiveness of common random numbers depending on how one performs synchronization. See Bratley, Fox, and Schrage [1987] for suggestions on how to obtain synchronization.

2. RANKING

This section presents an approach that can be used to simultaneously compare γ alternatives and generate a ranking. Kleijnen [1975] outlines a multiple comparison procedure that meets our requirements with respect to unknown variances, unequal variances, and correlated observations. Kleijnen outlines use of the Bonferroni Inequality to satisfy these requirements. However, the procedure leads to inconsistencies when attempting a ranking. We suggest a set of rules in this section for resolving these inconsistencies.

2.1 Bonferroni Inequality

Consider the single alternative case with two performance measures, e.g., tardiness and throughput. Assume we calculated an average tardiness value from a set of n replications and an average throughput from another independent set of n' replications. If we estimate a 95 % confidence interval for tardiness and a 95 % confidence interval for throughput, then the probability that one or both of these confidence intervals do not cover the true value of their respective means is $1 - .95^2$ or .0975. In this case when the confidence intervals are independent, we would need $\sqrt{.95} \times 100\% = 97.47\%$ confidence intervals for each performance measure to give us 95 % confidence that both intervals cover their respective means.

In actual practice, we use data from the same replications to calculate confidence intervals on both performance measures. This practice means that the confidence intervals are correlated. Kleijnen [1975, 1987] describes how one can use the Bonferroni inequality to calculate a conservative joint confidence interval when the statistics used in constructing the confidence intervals are correlated. Assume that we want simultaneous confidence intervals for two unknown mean performance measure values, e.g., μ_1 and μ_2. Let $L_k(\alpha_k)$ be the lower $(1-\alpha_k)\times 100\%$ confidence boundary for μ_k, and let $U_k(\alpha_k)$ be its corresponding upper boundary. If S_k implies the statement that $L_k(\alpha_k) \le \mu_k \le U_k(\alpha_k)$ and $P_{1-\alpha}$ is the probability that both S_1 and S_2 are true, then by the Bonferroni Inequality $P_{1-\alpha} \ge 1 - \alpha_1 - \alpha_2$ when $\alpha = \alpha_1 + \alpha_2$. More generally, when we want to make ψ correlated confidence statements, i.e., $S_u, u = 1, \ldots, \psi$, that are simultaneously true with probability $1 - \alpha$,

$$P_{1-\alpha} \ge 1 - \sum_{u=1}^{\psi} \alpha_u , \quad (2)$$

where

$$\alpha = \sum_{u=1}^{\psi} \alpha_u .$$

Usually, we let $\alpha_u = \alpha/\psi$.

Comparison	Statistic
θ_1 - θ_2	$X_i^{(1)}$ - $X_i^{(2)}$
θ_1 - θ_3	$X_i^{(1)}$ - $X_i^{(3)}$
θ_1 - θ_4	$X_i^{(1)}$ - $X_i^{(4)}$
θ_2 - θ_3	$X_i^{(2)}$ - $X_i^{(3)}$
θ_2 - θ_4	$X_i^{(2)}$ - $X_i^{(4)}$
θ_3 - θ_4	$X_i^{(3)}$ - $X_i^{(4)}$

Table 1. Pairwise Comparisons with 4 Alternatives

2.2 Ranking Procedure

To enhance the power of the comparisons, we will simulate γ different alternatives using common random numbers. To rank the alternatives, we will in essence be making

$$\psi = \binom{\gamma}{2}$$

pairwise comparisons. For example, with four alternatives we would make six pairwise comparisons as specified in Table 1.

The statistics $X_i^{(j)}$ - $X_i^{(k)}$ and $X_i^{(j)}$ - $X_i^{(u)}$ are correlated because of the common random numbers and the common quantity $X_i^{(j)}$. Regardless of the correlation, we estimate ψ simultaneous confidence intervals that jointly hold with probability of at least $1 - \alpha$ by using the Bonferroni Inequality. This may allow us to rank the alternatives if we allow three possibilities:

1. Regard $\theta_j = \theta_k$ when the confidence interval for θ_j - θ_k includes zero.

2. Regard $\theta_j > \theta_k$ when the lower bound on the confidence interval for θ_j - θ_k is positive.

3. Delete equality relations when they introduce contradictions. For example, delete all comparisons with Π_2 when the above rules give $\theta_1 = \theta_2$, $\theta_2 = \theta_3$, and $\theta_1 > \theta_3$.

In practice, the third possibility can usually be resolved by adding replications to reduce the confidence interval widths.

We will illustrate the ranking procedure using results from a simulation of a bus system where the primary performance measure is total passenger system time in minutes. The objective is to rank the following four alternatives with 95 % confidence:

1. One bus of capacity 10,
2. Two buses of capacity 10,
3. One bus of capacity 20, and
4. Two Buses of capacity 20.

The initial simulation experiment consisted of 100 independent replications for each alternative, and the same random numbers were used for each alternative.

The usual way to estimate θ_j - θ_k, for two different alternatives, Π_j and Π_k, when using common random numbers is to form pairs for each replication, i.e.,

$$D_i^{(jk)} = X_i^{(j)} - X_i^{(k)} \tag{3}$$

and calculate:

$$\bar{D}_n^{(jk)} = \sum_{i=1}^{n} \frac{D_i^{(jk)}}{n} \text{ and} \tag{4}$$

$$s_n^2(D_i^{(jk)}) = \sum_{i=1}^{n} \frac{(D_i^{(jk)} - \bar{D}_n^{(jk)})^2}{(n-1)},$$

the unbiased estimator of $V(D_i^{(jk)})$. (5)

Parameter	$\bar{X}_{100}^{(j)} - \bar{X}_{100}^{(k)}$	$L_u(\alpha_u)$	$U_u(\alpha_u)$
θ_1 - θ_2	1.97	1.73	2.21
θ_1 - θ_3	.17	-.03	.37
θ_1 - θ_4	3.77	3.58	3.96
θ_2 - θ_3	-1.80	-2.01	-1.59
θ_2 - θ_4	1.80	1.60	2.00
θ_3 - θ_4	3.60	3.45	3.75

Table 2. Bus System Comparison Statistics

Parameter	$\bar{X}_{10}^{(j)} - \bar{X}_{10}^{(k)}$	$L_u(\alpha_u)$	$U_u(\alpha_u)$
θ_1 - θ_2	5	-1	11
θ_1 - θ_3	10	4	16
θ_2 - θ_3	5	-1	11

Table 3. Hypothetical Comparison Statistics

Then calculate confidence intervals for each θ_j - θ_k using:

$t_{d,\beta}$ = Quantile of the t distribution with d degrees of freedom which is exceeded with probability β

$H_u(\alpha_u)$ = $t_{n-1,1-\alpha_u/2} s_n(D_i^{(jk)})/\sqrt{n}$, half width of the $(1-\alpha_u) \times 100\%$ confidence interval

$L_u(\alpha_u)$ = $\bar{D}_n^{(jk)} - H_u(\alpha_u)$, lower $(1-\alpha_u) \times 100\%$ confidence boundary for θ_j - θ_k when $X_i^{(j)}$ - $X_i^{(k)}$ has a normal distribution

$U_u(\alpha_u)$ = $\bar{D}_n^{(jk)} + H_u(\alpha_u)$, upper confidence boundary analogous to $L_u(\alpha_u)$

Nelson [1985] has shown that calculation of the confidence interval in this way is computationally equivalent to using (1) and replacing values in (1) with their respective estimates. In this case, $\alpha_u = .05/6 = .004167$ since $\psi = 6$.

Table 2 displays the results based on 100 independent replications. The confidence intervals appearing in Table 2 indicate that:

- $\theta_4 < \theta_2 < \theta_1$
- $\theta_4 < \theta_2 < \theta_3$

The hypothetical results shown in Table 3 illustrate the necessity for deleting an alternative from the ranking. Given the results shown in the table, we would delete Π_2 from the ranking and estimate that θ_3 ¡ θ_1. Presented with these results, we may want to simulate addition replications to reduce the confidence interval widths and include Π_2 in the ranking.

3. SELECTION OF THE BEST

We assume that less is better so that we want to identify the alternative having the smallest mean performance measure value. We attempt to do this by running a number of replications on each alternative and declaring the alternative with the smallest average performance measure value as the "best." The challenge is to determine the number of replications which give a specified probability of making a correct selection. . When other competing alternatives have means almost as small as the best alternative, the required number of replications can be quite large. That is why we resort to the indifference zone approach. The length of the indifference zone is δ which is specified by the analyst before conducting the simulation experiments. Any alternative with a mean within δ of the truly best alternative is a satisfactory choice or a correct selection. Let $P(CS)$ represent the probability of making a correct selection. The object of the procedure is to make $P(CS) \geq P^*$, where the analysts specifies a value for $P(CS)$ prior to the simulation experiments.

When we assume normally distributed outputs, unknown variances, and unequal variances, we consider two possible procedures for implementing the indifference zone approach. Dudewicz and Dalal [1975] present an exact procedure for these assumptions when the outputs for each alternative are independent. We should use common random numbers so the specified number of observations by the Dudewicz and Dalal procedure are probably in excess of the minimum requirement. Clark and Yang [1986] present a procedure, described in the following section, that does consider the correlations induced by common random numbers, but they use the Bonferroni Inequality and another bounding assumption which renders their procedure conservative. Clark and Yang's procedure is preferred when the correlations are large and the number of alternatives is not large.

3.1 Bonferroni Selection Procedure

Since the alternative variances are unknown, the selection procedure requires an initial sample to estimate:

m_j = Number of replications required of Π_j to achieve a $P(CS) \geq P^*$

The Bonferroni selection procedure is:

1. Simulate n replications for each alternative.
2. Determine the ψ paired comparisons.
3. For $u = 1, \ldots, \psi$, perform the following calculations:
   - Let (j, k) be the two alternatives in comparison u.

- Calculate $\bar{D}_n^{(jk)}$ using equation 4.
- Calculate $s_n^2(D_i^{(jk)})$ using equation 5.

4. Determine

$$h = t_{n-1,\beta},$$

where

$$\beta = \frac{1 - P^*}{\gamma - 1}$$

5. For $j = 1, \ldots, \gamma$, perform the following calculations:

- Let

$$v_j^2 = \max_{w \neq j}(s_n^2(D_i^{(jw)}))$$

- Let

$$m_j = \max(n, [v_j^2/(\delta/h)^2]),$$

where [x] is the smallest integer greater than or equal to x.

- Simulate $m_j - n$ additional replications for Π_j.
- Calculate $\bar{X}_{m_j}^{(j)}$.

6. Select the alternative with maximum $\bar{X}_{m_j}^{(j)}$.

Clark and Yang [1990] have determined sufficient conditions for increasing correlation among alternatives reducing the expected sample size in the above procedure.

3.2 Application

This section summarizes an application of the Bonferroni selection procedure to identify a preferred sequencing rule in scheduling a cell having three machines. Identify the machines as machines 1, 2, and 3. The primary performance measure is mean job tardiness where:

$$\begin{aligned} T &= \text{Job tardiness} \\ T &= \max(0, C - D) \\ C &= \text{Job completion time} \\ D &= \text{Job due time} \end{aligned}$$

Jobs arrive to the cell from upstream sources to machines 1 and 2, and after processing by these machines all jobs go to machine 3 for testing. Thus, an arriving job must go through two operations before departing the cell. The first operation is at machine 1 or 2 depending on where it arrives. The second operation is at machine 3 for testing. The arrival processes for machines 1 and 2 are approximated as independent Poisson processes, and each process has a mean time between arrivals of 33.33 minutes. The production planning department assigns to each job upon arrival a standard operation time and a due time for completion of each operation. The cell may use these times in sequencing work to reduce job tardiness. Let S_1 be the standard time assigned to a job for its first operation, and S_2 be the assigned standard time for its second operation. For analysis purposes, one can represent the standard times as having log-normal distributions with the following parameters:

- Mean for S_1 is 30 minutes.
- Standard deviation for S_1 is 5 minutes.
- Mean for S_2 is 15 minutes.
- Standard deviation for S_2 is 2.5 minutes.

The actual operation times at each machine have log-normal distributions where the means are equal to S_1 or S_2 depending on which operation is to be performed, and the standard deviations are 10 % of S_1 or S_2 depending on the particular operation involved. Given the cell arrival time of A, the due times are:

- $A + 2S_1$ for the first operation and
- $A + 2(S_1 + S_2)$ for the second operation.

The experimental objective is to compare three queue sequencing rules with respect to their effect on average job tardiness for the next shift. The three rules are:

1. Shortest Standard Time (SST)
2. Earliest Job Due Time (EDT)
3. Earliest Job Operation Due Time (EOT)

The job due time is the time the job is due for completion by the cell. The job operation due time is the time the current operation for the job is due to be completed. The probability must be at least .9 of making a correct selection, and the decision maker is indifferent when mean tardiness differs by one minute or less.

3.3 Approaching Normality

The first step in implementing the selection procedure is to investigate whether the average shift tardiness values are approximately normal. Nelson [1982] describes the use of probability plotting to identify a distribution for representing a particular process. A normal probability plot will give a straight line if the process appears normal; however, in this case, the probability plot of the average shift tardiness values clearly indicated that they were not normally distributed. The usual approach to approximate a normal distribution is to batch the data. For example, one can form batches of size 5 replications by taking the average of the first 5 replications and letting it be the first batch, the second 5 observations would become the second batch, and this process would be continued until all replications are grouped into batches. Under weak assumptions as the batch size becomes large, the central limit theorem shows that the values of the batch averages approach a normal distribution. The output analysis proceeds by setting the values of $X_i^{(j)}$ to the batch averages. However, large batches means fewer batches which will increase the width of the confidence interval. Schmeiser [1982] shows that the effect on confidence interval width is negligible even when the data is truly normal so long as the number of batches is not smaller than 30. For the simulations discussed in this paper, we are never sure the output data is normal. The conclusion is that we should always use batches and limit the number of batches to 30 even with very large samples. The probability plot for a batch size of 5 appeared to be approximately normal.

3.4 Effectiveness of Common Random Numbers

One selects a synchronization method in order to implement common random numbers. A common method is to assign a random number stream for each sampling step in the simulation. The first approach in this case was to use the following streams:

1. Individual streams for interarrival times at machines 1 and 2
2. Individual streams for generating standard times at each machine when a job arrives at machines 1 and 2.
3. Individual streams for sampling actual operation times at each machine.

This method of synchronization yielded very poor results as evidenced by the sample correlation matrix. The change that dramatically improved the correlation among alternatives was to sample for the actual operation times as the jobs entered the system rather than as they started their operations. Table 4 gives the sample correlation matrix.

	SST	EDT	EOT
SST	1.0	.9919	.9915
EDT	.9919	1.0	.9995
EOT	.9915	.9995	1.0

Table 4. Sample Correlation Matrix

3.5 Results

The initial sample consisted of 10 batches of size 5. The results appear in Table 5, and they indicate that the shortest standard time (SST) rule is preferred. To show the advantage of the Bonferroni procedure, the table estimates the sample size required by the Dudewicz and Dalal procedure which ignores the correlation due to common random numbers. The Dudewicz and Dalal procedure would require 21 times as many replications as the Bonferroni procedure.

Alternative	Average Tardiness	Bonferroni Batches	Dudewicz and Dalal Batches
SST	76.96	19	497
EDT	83.80	17	596
EOT	84.11	19	613

Table 5. Tardiness Experiment Results

4. CONCLUSIONS

The primary motivation for the analyst to use proper output analysis procedures when using stochastic simulations is to reduce the risk in making incorrect conclusions. The results clearly show the utility of the procedures presented in this paper to make comparisons using a stochastic simulation. They are easily implemented, and the results indicate that proper experimental analysis may significantly reduce the simulation effort required.

REFERENCES

Bratley,P.,B.L.Fox, and L.Schrage(1987), *A Guide to Simulation*,Second Edition,Springer-Verlag,New York.

Clark,G.M. and W.Yang(1986), "A Bonferroni Selection Procedure When Using Common Random Numbers with Unknown Variances," In *Proceedings of the 1986 Winter Simulation Conference*, J.R. Wilson, J.O. Henriksen, and S.D. Roberts, Eds. IEEE, Piscataway, NJ, 313-315.

Clark,G.M. and W.Yang(1990), "Output Analysis when Using Common Random Numbers to Compare Alternatives," Working Paper, Department of Industrial and Systems Engineering, The Ohio State University, Columbus, OH.

Dudewicz,E.J. and S.Dalal(1975), "Allocation of Observations in Ranking and Selection with Unequal Variances," *Sankhya 37*, 28-78.

Kleijnen,J.P.C.(1975), *Statistical Techniques in Simulation, Part II*, Marcel Dekker, New York.

Kleijnen,J.P.C.(1987),*Statistical Tools for Simulation Practitioners*, Marcel Dekker, New York.

Nelson,B.L.(1985), "The Sample Variance of Antithetic Variates," *TIMS College on Simulation Newsletter 9*, 2, 7.

Nelson,B.L.(1987), "Variance Reduction for Simulation Practitioners," In*Proceedings of the 1987 Winter Simulation Conference*, A.Thesen, H.Grant, W.D.Kelton,Eds. IEEE, Piscataway,NJ, 43-51.

Nelson,W.(1982), *Applied Life Data Analysis*, Wilely, New York.

Schmeiser,B.(1982), "Batch Size Effects in the Analysis of Simulation Output," *Operations Research 30*, 556-568.