# A TUTORIAL ON THE REGENERATIVE APPROACH AND DISCRETE-EVENT SIMULATIONS

Michael A. Crane and Austin J. Lemoine
Control Analysis Corporation
800 Welch Road
Palo Alto, California 94304

## ABSTRACT

The purpose of this paper is to provide an introduction to the regenerative method for simulation analysis. The simulations we are concerned with here are simulations of stochastic systems, i.e., systems with random elements. The regenerative approach leads to a statistical methodology for analyzing the output of those simulations which have the property of "starting afresh probabilistically" from time to time. The class of such simulations is very large and very important, including simulations of a broad variety of queues and queueing networks, inventory systems, inspection, maintenance, and repair operations, and numerous other situations.

## I. INTRODUCTION

In simulating systems of a random nature, it is important that a convincing statistical analysis be applied to the output of the simulation. In particular, estimation techniques, e.g., methods of obtaining confidence intervals, are needed which permit the simulator to make valid statistical inferences about the model based on simulation output. Such techniques are also essential so that the simulator may address the important tradeoffs between simulation run length and the level of precision in the estimates.

If a stochastic system is simulated with the goal of estimating some parameter which is indicative of the behavior of the system under "steady-state" conditions, then the simulator faces difficult problems of a "tactical" nature in providing a convincing statistical analysis of the simulation output. These difficult problems include how to start the simulation, when to begin collecting data, and what to do about highly correlated output. If, however, the stochastic simulation has the property of "starting afresh probabilistically" from time to time, then the above problems can be overcome in a very simple way using the regenerative approach. Moreover, in these circumstances the regenerative approach leads to a simple method of obtaining confidence intervals which in turn permit the simulator to make valid statistical inferences about the parameter(s) being estimated based on simulation output.

The research efforts which have led to the regenerative approach have been based at Control Analysis Corporation in Palo Alto, California, and performed under contract to the Office of Naval Research. The study efforts have now reached a certain level of maturity. Many results of practical interest are available, and these are of sufficient scope to justify an informal account of the work done thus far. Such an account is provided in [1] where the basic results of the regenerative approach are presented in a manner which can be easily understood by all potential users. The narrative in [1] is informal but precise, without inundating the reader in theorems, propositions, and formalities, and extensive use is made of examples to motivate and to illustrate fundamental ideas and results. The background required for following [1] is not extensive: a basic introduction to probability and statistics (including the central limit theorem and the notion of a confidence interval).

This conference paper provides a brief account of the regenerative approach and serves as an introduction to the comprehensive tutorial presentation [1]. Section 2 of this paper presents a basic example which serves to illustrate the problems and issues that arise in analyzing the output of stochastic simulations. The traditional "tactical" problems of correlation of simulation output and bias toward intitial conditions are addressed. The example of Section 2 also serves to motivate the regenerative approach as a means of resolving these problems and issues. The fundamental idea of the regenerative method is then spelled out in Section 3. Procedures are given for making valid statistical inferences about model parameters based on simulation output. In particular, a method is given for obtaining a confidence interval for the expected value of an arbitrary function of the steady-state distribution of the process being simulated. The method is based on a random blocking technique which enables the simulator to group the output data into independent and identically distributed blocks.

## 2. BASIC EXAMPLE AND MOTIVATION

In this section we present a simple but important example of a system which occurs frequently in applications: a single-server queue. The point of view will be that of a simulator who is given the task of simulating the system in order to predict how the system will behave. This example will dilineate the difficult issues faced by the simulator in carrying out this task in a satisfactory manner, and will also serve to motivate the basic idea of the regenerative method as a simple approach for resolving these issues.

The standard single-server queue is depicted in Figure 2.1. Customers originate from an input source which generates new customers one at a time. If an arriving customer finds the server free his service commences immediately and he departs from the system when his service requirements have been satisfied. If the arriving customer finds the server occupied, he enters the waiting room and waits his turn to be served. Customers are transferred from the queue into service on a first-come-first-served basis.

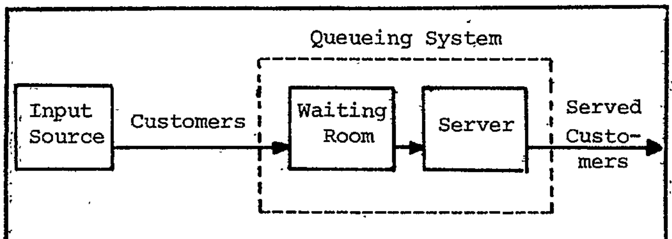


Figure 2.1 - Single-Server Queue

The input source generates customers on a random basis in that the elapsed times between successive customer arrivals are independent random variables having a common distribution. The service requirements of successive customers are independent random variables also having a common distribution. It is assumed that the input source and the service mechanism operate independently of one another.

For the particular queueing system under consideration, suppose that the input source generates customers by sampling from a uniform distribution on the range from 5 to 25. Suppose also that service times are uniformly distributed but on the range from 0 to 20. Assume the measure of performance to be used for the system is $E\{W\}$, the mean value of the waiting time (exclusive of service time) experienced by a customer under steady-state conditions. Since no computationally tractable analytical results are available for computing the exact value of $E\{W\}$ in this particular system, simulation is a natural recourse. Therefore, the task is one of simulating the system and analyzing the output to provide an estimate for the value of $E\{W\}$. Furthermore, it would be desirable to obtain some measure of the "reliability" of the estimate for $E\{W\}$, e.g., a 90% confidence interval for the true value of $E\{W\}$, so that the simulator can determine whether more lengthy runs are required to obtain acceptable precision in the results.

A reasonable way to proceed would seem to be as follows. Let $W_1$ denote the waiting time of the first customer in the simulated system, $W_2$ the waiting time of the second customer, and so on. Then, if the total duration of the simulation run is for N customers, where N might be 1000, for example, then the sample average

$$\frac{W_1 + W_2 + \cdots + W_N}{N} \tag{2.1}$$

is a "consistent estimator" for $E\{W\}$, since it is known that the sample average converges to the true value of $E\{W\}$ with probability one as $N \to \infty$. However, the sample average (2.1) will in general be a "biased estimator" for the true value of $E\{W\}$ due to the initial conditions. For example, since $W_1$ is zero, the next few waiting times will tend to be small. Such bias can be eliminated if the simulator can choose a value for $W_1$ by sampling from the distribution of W itself.[1] Unfortunately, the simulator does not even know the mean of W, let alone its distribution, so that this "solution" is not very practical.

The traditional way of dealing with the difficulty of the initial bias is to run the simulation model for some time without collecting data until it is believed that the simulated queueing system has essentially reached a steady-state condition and then to collect data from that point on.

For example, we might simulate the system for 2000 customers, discard the waiting times of the first 1000 customers, and use the average

$$\frac{W_{1001} + \cdots + W_{2000}}{1000}$$

as an estimate for $E\{W\}$. But, it is by no means clear just how long this "stabilization" period ought to be, so that a great deal of unproductive computer time can be wasted in the process.

There is also another difficulty with using the "seemingly reasonable" estimation procedure which starts with the sample average in (2.1). Based on the output of the simulation experiment, we would like to obtain some measure of the "reliability" of the estimate for $E\{W\}$, indicating the likelihood that similar estimates would result if we were to repeat the simulation. In this spirit, we might wish to construct a 90% confidence interval $\underline{I}$ for the true value of $E\{W\}$, such that in any independent replication of the simulation experiment, the probability would be 0.90 of having the computed interval $\underline{I}$ include the true value for $E\{W\}$. However, in order to construct such confidence intervals using classical statistics, the output data must form a collection of statistically independent and identically distributed samples from some underlying probability distribution. The output data from the queueing simulation is the sequence of waiting times $W_1$,

$W_2, \ldots, W_N$. Note, however that if $W_k$ is large, then the next customer, k + 1, will typically have a large waiting time also; and conversely, if $W_k$ is small, then $W_{k+1}$ will tend to be small. Thus, the samples $W_k$ and $W_{k+1}$ are highly correlated, and this is true whether or not the simulation is begun by sampling from the steady-state distribution of W. Thus, since the waiting times $W_1, W_2, \ldots, W_N$ are not independent, classical statistics appears to be of little use in assessing the "reliability" of (2.1) as an estimate for E{W}.

We see, therefore, that the bias due to the initial conditions and the highly correlated output data pose serious obstacles to using the proposed estimation procedure based on the sample average of (2.1). Since the sample average is such a simple and natural way to estimate E{W}, these "tactical difficulties" raise doubts about whether or not the simulation experiment for the queueing system will lead to meaningful results. We must ask, therefore, if there might be a simple way to overcome these obstacles which does not require the use of sophisticated or cumbersome methods of analysis. Fortunately, the answer to this question is yes, and we can accomplish this by proceeding in a very straightforward manner.

Suppose we begin the simulation by setting $W_1 = 0$, that we then run the simulation for a short while, and that the customer waiting times observed are as in Figure 2.2. We see that customers 1, 4, 5, 10 and 14 are the lucky ones who find the server idle when they arrive and consequently experience no waiting in the queue, while customers 2, 3, 6, 7, 8, 9, 11, 12 and 13 are obliged to wait before being served. Moreover, the server is constantly busy from the time of arrival of customer 1 to the time of departure of customer 3, then constantly idle until customer 4 appears, busy while serving customer 4, idle from the time customer 4 leaves until customer 5 arrives, then constantly busy from the time customer 5 arrives

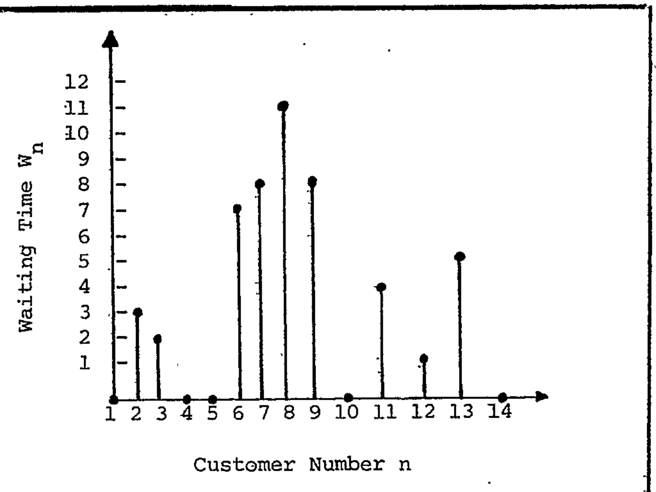


Figure 2.2 - Sample output of queueing simulation

until customer 9 departs, and so on. Now, had we run the simulation beyond customer 14 until we had processed 1000 customers, say, then the simulated queue would undoubtedly exhibit the same pattern of the server being busy, then idle, then busy, then idle, and so on. Suppose we call the operational span of a "busy period" and its ensuing "idle period" a "cycle." Then, our short simulation run has 4 complete cycles during which the following sets of customers are processed by the system: {1, 2, 3}, {4}, {5, 6, 7, 8, 9}, {10, 11, 12, 13}. A fifth cycle begins with the arrival of customer 14. Thus, each new cycle is initiated by a customer who finds the server idle upon arriving.

Note that the system "begins anew" at the dawn of each cycle as it did when customer 1 arrived. That is, at the time when a cycle commences, the future evolution of the simulated queueing system is always independent of its past behavior and always governed by the same probabilistic structure as when customer 1 arrived to find an idle server. The same is also true of the corresponding "real world" queueing system. The start of each cycle is indistinguishable from the arrival of the very first customer to the queue. It seems natural, therefore, to group the simulation output data into blocks, the first block consisting of the waiting times of customers in the first cycle, the second block consisting of the waiting times of customers in the second cycle, and so on. For the short simulation run illustrated above the blocks are $\{W_1, W_2, W_3\}$, $\{W_4\}$, $\{W_5, W_6, W_7, W_8, W_9\}$ and $\{W_{10}, W_{11}, W_{12}, W_{13}\}$. Thus, since each cycle is initiated under the same conditions, and the system "starts afresh" at the times when cycles commence, the blocks of data from successive cycles are statistically independent, and also possess the same distribution. So, for example, if we set $Y_k$ equal to the sum of the waiting times of customers processed in cycle k, and $\alpha_k$ equal to the number of customers processed in cycle k, the pairs $(Y_1, \alpha_1)$, $(Y_2, \alpha_2)$, $(Y_3, \alpha_3)$ and $(Y_4, \alpha_4)$ are independent and identically distributed. (Note, however, that $Y_k$ and $\alpha_k$ are highly correlated.) From Figure 2.1 we have the following:

$$(Y_1, \alpha_1) = (5, 3)$$

$$(Y_2, \alpha_2) = (0, 1)$$

$$(Y_3, \alpha_3) = (34, 5)$$

$$(Y_4, \alpha_4) = (10, 4)$$

Hence, the highly correlated data $\{W_1, W_2, \ldots, W_{13}\}$ has been broken up into statistically independent and identically distributed blocks.

Now, at this point the reader might stop and remark that what we have just observed is all very good but what does it do towards solving the difficult problems we face in analyzing the simulation output, i.e., of obtaining a valid estimate for E{W}. Suppose we resume the simulation run with customer 14 and continue until we have observed n complete cycles of the sort identified above. As before, let $Y_k$ and $\alpha_k$ denote, respectively, the sum of the waiting times of those customers processed in cycle k and the number of customers processed in cycle k, for $k = 1, 2, \ldots, n$. Then, if N is the total number of customers processed over the n cycles, observe that

$$\frac{W_1 + W_2 + \cdots + W_N}{N} = \frac{Y_1 + \cdots + Y_n}{\alpha_1 + \cdots + \alpha_n} \quad (2.2)$$

The right side of (2.2) can be written as

$$\frac{(Y_1 + \cdots + Y_n)/n}{(\alpha_1 + \cdots + \alpha_n)/n} \quad . \quad (2.3)$$

Since each of $\{Y_1, \ldots, Y_n\}$ and $\{\alpha_1, \ldots, \alpha_n\}$ is a collection of independent and identically distributed random variables, we know by the law of large numbers that the numerator in (2.3) converges to the value of $E\{Y_1\}$ as $n\to\infty$ and the denominator converges to the value of $E\{\alpha_1\}$ as $n\to\infty$, both with probability one. Note that $N \geq n$, and that $N\to\infty$ as $n\to\infty$. Moreover, we have already observed that the left side of (2.2) converges to the value of $E\{W\}$ as $N\to\infty$, with probability one. Hence, we have proven that

$$E\{W\} = E\{Y_1\}/E\{\alpha_1\} \quad . \quad (2.4)$$

Therefore, the problem of estimating $E\{W\}$ is the same as estimating the ratio $E\{Y_1\}/E\{\alpha_1\}$. And, since this ratio can be estimated from the independent and identically distributed pairs $(Y_1, \alpha_1), \ldots, (Y_n, \alpha_n)$, classical statistics can be used to make inferences about the true value of $E\{W\}$ based on the simulation output. In particular, we can construct a confidence interval for $E\{Y_1\}/E\{\alpha_1\}$, and a procedure for doing so will be given in Section 3.

Thus, we have arrived at a simple solution of what to do about the highly correlated output data from the simulated queueing system. Moreover, by taking $W_1 = 0$ we begin the simulation by initiating a cycle, and thus no stabilization period whatever is required and every piece of output data generated by the simulation is useful in obtaining a statistically sound estimate of $E\{W\}$.

So, we see that our short simulation run has provided us with insight on how to resolve the hard "tactical" problems we faced at the outset of the proposed simulation experiment; namely, the problem of how to begin the simulation, the problem of when to start collecting data, and the problem of highly correlated output.

At this point it would be well to pause and take stock of what has been done. We began with the goal of estimating the expected steady-state waiting time in a single-server queue by simulating the system and analyzing the output. Before we could even start the simulation experiment, however, we found ourselves facing difficult tactical issues which had to be resolved in order for us to carry out a meaningful statistical analysis of the simulation output. Not to be dissuaded, we then observed that the behavior of the simulated system was characterized by a succession of cycles, that the system regenerated itself probabilistically at the dawn of each cycle, and that if the simulation output were grouped into blocks, with each block consisting of output from a particular cycle, then the blocks were statistically independent, identically distributed, and, in view of (2.4), carried valuable information about the steady-state parameter to be estimated. Thus, the difficult tactical issues were resolved, and we could proceed with our simulation experiment.

Now, if this notion of a stochastic simulation "regenerating itself" probabilistically can be applied to simulations other than queueing, then we have discovered an approach which could be very helpful in analyzing the output of a large class of stochastic simulations. This idea does indeed go well beyond queueing simulations, and the class of simulations having a "regenerative" property is very broad (see [1]). We now go on from here to set down a simple framework in Section 3 for analyzing the output of any stochastic simulation having a "regenerative" property of the sort we observed in the example of this section. Moreover, we will give a simple procedure for constructing a confidence interval (from the simulation output) for a wide variety of steady-state parameters of interest in a "regenerative" simulation.

## 3. THE REGENERATIVE METHOD

The example of Section 2 suggests a unified approach toward analyzing the output of those simulations of stochastic systems which have the property of "regeneration" from time to time. That is, if the simulation output is viewed as a stochastic process, then these "regenerative processes" have the property of always returning to some "regenerative condition" from which the future evolution of the process is always independent of its past behavior and always governed by the same probability law. If the simulation output is then grouped into blocks according to successive returns to the "regenerative condition" then these blocks are statistically independent and identically distributed, and this greatly facilitates statistical analysis of the output by the simulator.

In this section we shall cast all such "regenerative processes" into a common framework and then give a simple technique for obtaining a confidence interval, based on the simulation output, for a variety of steady-state system parameters of practical interest. The method covers any discrete-event simulation that can be modeled as a regenerative process (see [1]). A discrete-event simulation is one in which the state of the system being simulated only changes at a discrete, but possibly random, set of time points. The example considered in Section 2 is of this type.

This section is organized as follows. We first discuss regenerative processes with a discrete time parameter, as in the queueing example. (Regenerative processes with a continuous time parameter are discussed in [1]). After defining regenerative processes, we then give a technique for obtaining confidence intervals for steady-state parameters of such processes.

A sequence $\{X_n, n \geq 1\}$ of random vectors in K dimensions is a regenerative process if there is an increasing sequence $1 \leq \beta_1 < \beta_2 < \ldots$ of random discrete times, called regeneration epochs, such that at each of these epochs the process starts afresh probabilistically according to the same probabilistic structure governing it at epoch $\beta_1$. That is, between any two consecutive regeneration epochs $\beta_j$ and $\beta_{j+1}$, say, the portion $\{X_n, \beta_j \leq n < \beta_{j+1}\}$ of the process is an independent and identically distributed replicate of the portion between any other two consecutive regeneration epochs. However, the portion of the process between epoch 1 and epoch $\beta_1$, while independent of the rest of the process, is allowed to have a different distribution. We will refer to the portion $\{X_n, \beta_j \leq n < \beta_{j+1}\}$ of the process as the jth cycle.

In the queueing example, $X_n = W_n$, and the epochs of regeneration $\{\beta_j, j \geq 1\}$ are the indices of those customers who find the server idle upon their arrival. A typical situation in which the regenerative assumption is satisfied is when $\beta_j$ represents the time of the jth entrance to some fixed state, say $s$. Upon hitting $s$, the simulation can proceed without any knowledge of its past history. Examples of such epochs are the instants when an arriving customer finds all servers idle in a multi-server queueing system and the times when a recurrent irreducible Markov chain hits a fixed state. Not all regenerative behavior, however, is characterized by returns to a fixed state (see [1] ).

Let $\alpha_j = \beta_{j+1} - \beta_j$ for $j \geq 1$. Note that the "sojourn times" $\{\alpha_j, j \geq 1\}$ between consecutive epochs of regeneration are independent and identically distributed. (In the queueing example of Section 2, $\alpha_j$ is the number of customers served in the jth cycle.) We will assume henceforth that $E\{\alpha_1\} < \infty$. This is not a restrictive assumption. For example, it holds in almost any queueing system of practical interest, and it certainly holds for the queueing example of Section 2, as well as for any positive recurrent irreducible Markov chain.

The regenerative property is an extremely powerful tool for obtaining analytical results for the process $X_n, n \geq 1$. Under very mild conditions the process has a limiting or steady-state distribution. These conditions are technical in nature and are discussed in [1]. The main point, however, is that virtually any discrete time parameter regenerative process of practical interest to a simulator has a steady-state distribution in some sense, and most often in the following familiar sense. There is a random K-vector $X$ such that the distribution of $X_n$ converges to the distribution of $X$ as $n \to \infty$, that is, the $\lim_{n\to\infty} P\{X_n \leq x\} = P\{X \leq x\}$ for K-vectors $x$.

Since we now know that regenerative simulations of interest have steady-state distributions, we can turn to the question of estimating characteristics of those steady-state distributions.

Let f be a "nice" function in K dimensions having real values, and suppose the goal of the simulation is to estimate the value of $r \equiv E\{f(X)\}$. (The "nice" functions are the so-called "measurable" functions, and these include virtually all functions of practical interest.) Now, by the appropriate choice of the function f, the simulator can estimate a wide variety of steady-state quantities of interest. To illustrate, suppose first that $X$ is real-valued, so that we replace $X$ by $X$. If f is defined so that $f(x) = x$ for all x, then $r \equiv E\{f(X)\} = E\{X\}$, so that estimating r is equivalent to estimating $E\{X\}$. (This is the function of interest for the queueing example of Section 2.) If $f(x) = x^2$, then $r = E\{X^2\}$; if $f(x) = 1$ for $x \leq a$, where a is fixed, and $f(x)=0$ for $x>a$, then $r = P\{X \leq a\}$; and, if $f(x) = b(x-c)^+$, where b and c are fixed, then $r = b \cdot E\{(X - c)^+\}$.

More generally, if $x_j$ is the jth component of the K-vector $x$ and $X^{(j)}$ is the jth component of $X$, then $f(x) = x_j$ gives $r = E\{X^{(j)}\}$, $f(x) = x_i x_j$ gives $r = E\{X^{(i)} X^{(j)}\}$, and $f(x) = [x_1^2 + \ldots + x_k^2]^{1/2}$ gives r equal to the expected length in K dimensions of the random vector X.

We now observe those properties of the regenerative structure which will be used to obtain a confidence interval for r. Let

$$Y_j = \sum_{i=\beta_j}^{\beta_{j+1}-1} f(X_i) \quad . \tag{3.1}$$

That is, $Y_j$ is the sum of the values of $f(X_i)$ over the jth cycle. (In the queueing example of Section 2, $Y_j$ is the sum of the waiting times of customers in the jth cycle.) Recall that $\alpha_j = \beta_{j+1} - \beta_j$ gives the length of the jth cycle. Then, the fundamental properties of the regenerative process we shall use are given by (3.2) and (3.3).

<u>The sequence</u> $\{(Y_j, \alpha_j), j \geq 1\}$ <u>consists of independent and identically distributed random vectors</u>. (3.2)

If $E\{|f(X)|\} < \infty$ <u>then</u>

$$r \equiv E\{f(X)\} = E\{Y_1\}/E\{\alpha_1\} \,. \tag{3.3}$$

Note that we demonstrated the validity of (3.2) and (3.3) in the queueing example. The same concepts we used in that example can be generalized to establish (3.2) and (3.3) for regenerative processes with a discrete time parameter. The assumption that $E\{|f(X)|\} < \infty$ is not very restrictive, and does not pose an obstacle to application of the regenerative method.

Let us now turn to the main problem of interest here, namely that of estimating the value of $E\{f(X)\}$ based on the simulation output. In view of (3.2) and (3.3) this statistical estimation problem has been reduced to the following:

Given the independent and identically distributed observations

$\{(Y_j, \alpha_j), j \geq 1\}$ , estimate $r \equiv E\{Y_1\}/E\{\alpha_1\}$. (3.4)

Moreover, because we now have independent and identically distributed observations, we can use results from classical statistics to estimate $E\{Y_1\}/E\{\alpha_1\}$ . In particular, we can obtain a confidence interval for this quantity, and a method for doing so will now be given.

In view of (3.4), we thus have the following task: given the independent and identically distributed pairs $(Y_1, \alpha_1)$ , $(Y_2, \alpha_2)$ , . . . $(Y_n, \alpha_n)$ , construct a $100\,(1 - 2\delta)\%$ confidence interval for $E\{Y_1\}/E\{\alpha_1\}$ when n is large. We illustrate one method for obtaining such a confidence interval using the central limit theorem. Further discussion of some alternate confidence intervals is given in [1].

Let $V_j = Y_j - r\alpha_j$. Note that the $V_j$'s are independent and identically distributed and that $E\{V_j\} = E\{Y_j\} - r\,E\{\alpha_j\} = 0$, by virtue of (3.2) and (3.3). Let $\overline{Y}$ , $\overline{\alpha}$, and $\overline{V}$ denote the sample means

$$\overline{Y} = \frac{1}{n} \sum_{j=1}^{n} Y_j \ ,$$

$$\overline{\alpha} = \frac{1}{n} \sum_{j=1}^{n} \alpha_j \ ,$$

and

$$\overline{V} = \frac{1}{n} \sum_{j=1}^{n} V_j \ ,$$

and note that $\overline{V} = \overline{Y} - r\,\overline{\alpha}$ . Putting $\sigma^2 = E\{V_j^2\}$ and assuming $0 < \sigma^2 < \infty$ , the central limit theorem tells us that

$$\lim_{n \to \infty} P \left\{ \frac{n^{1/2}\,\overline{V}}{\sigma} \leq x \right\} = \Phi(x) \qquad (3.5)$$

for each real x, where $\Phi$ is the standard normal distribution function. The assumption that $0 < \sigma^2 < \infty$ is not restrictive for simulation applications. We can rewrite (3.5) as

$$\lim_{n \to \infty} P \left\{ \frac{n^{1/2}\,[\hat{r} - r]}{(\sigma/\overline{\alpha})} \leq x \right\} = \Phi(x) \qquad (3.6)$$

where $\hat{r} = \overline{Y}/\overline{\alpha}$. We cannot, however, produce a confidence interval for r directly from (3.6) since the value of $\sigma$ is unknown. However, we can estimate the value of $\sigma$ as follows. Let $s_{11}$, $s_{22}$ and $s_{12}$ denote, respectively, the sample variance of the $Y_j$'s, the sample variance of the $\alpha_j$'s, and the sample covariance of the $(Y_j, \alpha_j)$'s, i.e.,

$$s_{11} = \frac{1}{n-1} \sum_{j=1}^{n} (Y_j - \overline{Y})^2 = \frac{1}{n-1} \sum_{j=1}^{n} Y_j^2 - \frac{1}{n(n-1)} \left( \sum_{j=1}^{n} Y_j \right)^2$$

$$s_{22} = \frac{1}{n-1} \sum_{j=1}^{n} (\alpha_j - \overline{\alpha})^2 = \frac{1}{n-1} \sum_{j=1}^{n} \alpha_j^2 - \frac{1}{n(n-1)} \left( \sum_{j=1}^{n} \alpha_j \right)^2$$

and

$$s_{12} = \frac{1}{n-1} \sum_{j=1}^{n} (Y_j - \overline{Y})(\alpha_j - \overline{\alpha}) = \frac{1}{n-1} \sum_{j=1}^{n} Y_j\,\alpha_j - \frac{1}{n(n-1)} \left( \sum_{j=1}^{n} Y_j \right) \left( \sum_{j=1}^{n} \alpha_j \right)$$

Now let

$$s^2 = s_{11} - 2\hat{r}s_{12} + \hat{r}^2 s_{22} \ .$$

Then, it can be easily shown that $s^2 \to \sigma^2$ with probability one as $n \to \infty$ . Thus, (3.6) holds with s in place of $\sigma$ , i.e.,

$$\lim_{n \to \infty} P \left\{ \frac{n^{1/2}\,[\hat{r} - r]}{(s/\overline{\alpha})} \leq x \right\} = \Phi(x) \ . \qquad (3.7)$$

Letting $z_\delta^* = \Phi^{-1}(1 - \frac{\delta}{2})$ , i.e., $\Phi(z_\delta^*) = 1 - \frac{\delta}{2}$ ,

then from (3.7) we have

$$P\left\{-z^*_\delta \le \frac{n^{1/2}[\hat{r} - r]}{(s/\bar{\alpha})} \le z^*_\delta\right\} \cong 1 - \delta$$

for large n. This gives the following approximate $100(1 - \delta)\%$ confidence interval for $r = E\{f(X)\}$ :

$$\hat{I} = \left[\hat{r} - \frac{z^*_\delta s}{\bar{\alpha}n^{1/2}}, \hat{r} + \frac{z^*_\delta s}{\bar{\alpha}n^{1/2}}\right] \qquad (3.8)$$

Note that if we let $\hat{J}$ be the width of $\hat{I}$, then

$$\hat{J} \cong \frac{2 z^*_\delta \sigma}{E\{\alpha_1\}n^{1/2}}$$

for large n, with high probability. Thus, in order to reduce the width of the confidence interval $\hat{I}$ by a factor of two (at the same level of confidence), it is necessary to increase the number of cycles simulated by a factor of four.

We now summarize the procedure for obtaining an approximate $100(1 - \delta)\%$ confidence interval for $r = E\{f(X)\}$ :

1. Observe the simulation for n regeneration cycles.
2. Compute $Y_j$ and $\alpha_j$ for each cycle j, where $Y_j$ is the sum of $f(X_i)$ over the jth cycle and $\alpha_j$ is the length of the jth cycle.
3. Compute the sample statistics[1] $\bar{Y}$, $\bar{\alpha}$, $\hat{r}$, $s_{11}$, $s_{12}$, $s_{22}$, and $s^2$.
4. From the confidence interval

$$\hat{r} \pm \frac{z^*_\delta s}{\bar{\alpha}n^{1/2}}$$

where $z^*_\delta = \Phi^{-1}(1 - \frac{\delta}{2})$ and $\Phi$ is the standard normal distribution function.

Note that in the case where the first cycle does not begin immediately at the start of the simulation, the above procedure indicates that the data prior to the first cycle is to be discarded.

[1]In performing these calculations, particularly those of $s_{11}$, $s_{12}$, and $s_{22}$, it is wise to use double-precision arithmetic in order to insure the desired degree of accuracy in computing the sums.

## REFERENCE

[1] CRANE, M. A. and LEMOINE, A.J. (1976). An Introduction to the Regenerative Method for Simulation Analysis. Control Analysis Corporation, Technical Report No. 86-22.