Monte Carlo Method for Component Sizing

Robert C. Spear*
University of California, Berkeley, Calif.

For preliminary design studies, one often can express some index of component size as a function of various system demands, but there are usually a number of poorly known parameters involved in the descriptions of the demands. In this Note an approach is presented for dealing with such problems. Basically, the parametric uncertainty is described statistically, and a digital computer is used to conduct statistical experiments to determine certain properties of the distribution of a selected component size index. Such Monte Carlo methods[1,2] have been applied to many problems. The particular variant discussed in this Note is distinguished only in the application of statistical methods that are free from assumptions concerning the parent distribution of the random variable under study.

Problem Definition

Let u be a suitable component size index (CSI) which is functionally dependent on a set of parameters $\mathbf{x}$. If $\mathbf{x}$ is known, we assume that $u = f(\mathbf{x})$ can be evaluated; i.e., the functional relationship is specified. Further, it is assumed that if a component with a CSI of u_1 is adequate, then any component with a CSI $> u_1$ will also meet the demand. For example, if u is the maximum flow rate demanded of a hydraulic pump and $\mathbf{x}$ describes the duty cycle of actuators driven by the pump plus their physical characteristics, the functional relationship will involve a set of differential equations so that the notation $u = f(\mathbf{x})$ does not imply only an algebraic relationship.

The supposition is that the parameter set $\mathbf{x}$ is not well known, and the resultant problem is that of choosing a CSI which will deal with the eventual real demand with reasonable probability. Some assumptions on the possible and/or probable values of $\mathbf{x}$ are required. For our hydraulic pump, suppose x_i is the maximum velocity demanded for an actuator. We may induce the control system designer to forecast that $5 \leq x_i \leq 10$ in./sec, for example. For a bounded parameter, it is a simple step to restate the assumptions in the context of a probability density function or "parameter distribution;" i.e., we assume that $5 \leq x_i \leq 10$ with probability 1, and no value in the interval is more likely than another. Then the parameter distribution is the familiar rectangular or uniform distribution. But it must be remembered that the real parameter set is not random and that statistical notions have been introduced only in the process of developing a method to deal with our uncertainty as to the real, constant values of the parameters.

At this point let us assume that for each element of $\mathbf{x}$ there has been defined a probability density function reflecting its possible values and, if sufficient data exist, its probable values. Let us place no restrictions on the parameter distributions except that they lead to a continuous cumulative distribution for the CSI. Hence, if there is reason to believe that x_j has a most probable value, the corresponding parameter distribution might be gaussian (or, alternatively exponential). Further, let us place no restriction on the combination of parameter distributions in any problem. Several or all of the x_j may be assumed jointly distributed if need be. The only difficulties will be the practical problems associated with obtaining random numbers from such distributions.

The combination of the mathematical model and the parameter distributions define an ensemble of systems. The basic assumption implicit in the method is that the model of the real system is a random member of the ensemble. By studying the properties of the ensemble, specifically with respect to the CSI, we intend to reach some conclusions about the probable properties of the real system. Specifically, the model relates the CSI to the parameter set via a function $u = f(\mathbf{x})$. Hence, u becomes a random variable with an unknown cumulative distribution, $F(u)$, and we assume that one of the possible values taken on by u corresponds to the value u^* eventually required of the component. The problem is to choose a value of the CSI, say $\hat{u}_1$, such that $P(u^* < \hat{u}_1)$ is high. (It is possible to treat $u^* > \hat{u}_2$ or $\hat{u}_2 < u^* < \hat{u}_1$ similarly.) Since our knowledge of u^* is assumed to be limited to the fact that it is a possible value taken on by the random variable u, it follows that $P(u^* < \hat{u}_1) = F(\hat{u}_1)$. Hence, a knowledge of $F(u)$ would solve the problem as originally posed.

Suppose the CSI desired is the maximum flow rate of a variable displacement hydraulic pump, e.g., $\hat{u}_1 = 2.5$ gpm. If $F(u)$ were known, $F(2.5)$ could be found. If $F(2.5) = 0.93$, a 2.5 gpm pump should provide sufficient flow capability to deal with 93% of the cases that might occur under the assumed $\mathbf{x}$ distributions. Of course, $F(u)$ is not known and is calculable from the $\mathbf{x}$ distributions only under very specific conditions. The alternative is to conduct experiments on a digital computer to infer from samples values from $F(u)$ a value $\hat{u}_1$ such that $F(\hat{u}_1) = P(u^* < \hat{u}_1)$ is sufficiently large for the purposes at hand.

Calculation of a sample value of the CSI requires a randomly selected set of parameters and computation of u via $u = f(\mathbf{x})$. A general flow chart is shown in Fig. 1. Usually it will be most convenient to supply the program with a set of

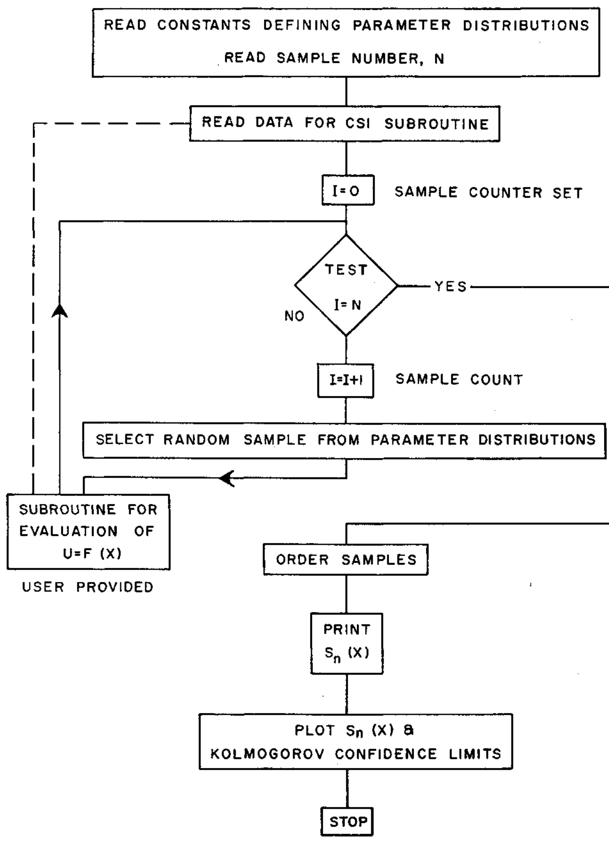

Fig. 1 Sample generating program: flow diagram.

Received March 12, 1970; revision received May 1, 1970. The preparation of this Note was supported by Environmental Health Training Grant No. 8T01 EC00020-04 from the Environmental Control Administration, United States Public Health Service.

* Postdoctoral Scholar, Division of Environmental Health ciences, School of Public Health.

random number generating routines for common distributions and input control parameters specifying the distribution assumed for each parameter, x_i. Most computer centers have such routines in their libraries. The user must provide the $f(\mathbf{x})$ routine, specify the form and statistical parameters of each x_i distribution and, finally, specify the number of samples n required. The number of samples required obviously depends on the problem to be solved.

Several Distribution-Free Methods

In stating that $F(u)$ is unknown, we mean that neither the functional form nor any associated statistical parameter is known; we assume only that $F(u)$ is continuous, i.e., $P(u = \text{const}) = 0$. This is a mild assumption, particularly since it is still possible to achieve practical results. Many statistical results do not rely on assumptions concerning $F(u)$ and are called distribution-free or nonparametric. A number of references are available[3–5] although some require substantial statistical background.

A particularly useful distribution-free result concerns the proportion of a population included between two order statistics from a random sample. Suppose we draw n samples from the population $F(u)$ and arrange them in increasing order of magnitude, $u_1, u_2 \ldots u_n$. If we choose any two of these order statistics, u_r and u_{n-s+1}, it is possible to state, with a given confidence, the proportion of the population included between these values. Specifically, the general result gives the probability α that $100\beta\%$ of the population will be included between u_r and u_{n-s+1}, where $0 < \beta < 1$. The indices are such that u_r is the rth smallest value and u_{n-s+1} the sth largest value counting from the top. The probability α depends only on the number of samples n and the parameter $m = r + s$. That is, given the number of samples and any two indices r and s, it is possible to estimate the proportion of the true population enclosed within these two ordered samples. Suppose we wish to find the number $\hat{u}_1$, such that $P(u < \hat{u}_1) = 0.90$ with 95% confidence. We select $\hat{u}_1 = u_n$ and ask how many samples must be taken such that 90% of the population be less than the largest sample with 95% probability? In this case $r = 0$, $s = 1$, and from Murphy[6] we find $n \simeq 29$. So, using $n = 29$ as an input to the computer program, we find the largest resulting value of the CSI, u_{29}, and it is then possible to state that a component whose CSI $= u_{29}$ will be able to handle 90% of the cases which might arise (under the assumed $\mathbf{x}$ distribution) with a probability (or tolerance level) of 0.95. It is not legitimate, however, to take first a random sample in which $u_n = 1.47$, for example, and then use the tables to predict $P(u^* < 1.47)$, because the statistic is based on the order statistic u_n and not its value. Clearly, this statistic can be used in a variety of ways. A multivariable version of these results is also available.

Cases may exist for which more detailed knowledge of the shape of $F(u)$ would be desirable. For example, if $F(u)$ tails off very slowly there might be a substantial difference in the values of u between the points where $F(u) = 0.98$ and $F(u) = 0.99$ (the 98th and 99th quantiles). If the penalty associated with undersizing the component is not great, then the smaller size will be chosen. But, to gain such information by sampling from $F(u)$, it is necessary to increase the sample size. Fortunately, several results exist which relate the sample size to the accuracy of estimation of $F(u)$ by the sample distribution function $S_n(u)$;

$$S_n(u) = \begin{cases} 0 & u \leq u_1 \\ r/n & u_r < u \leq u_{r+1} \\ 1 & u_n < u \end{cases}$$

where u_j is the jth order statistic. The best known result is due to Kolmogorov who determined the distribution of the statistic

$$D_n = \sup_x |S_n(x) - F(x)|$$

It is possible to invert this result into a confidence statement of the form

$$P\{S_n(x) - d_\alpha \leq F(x) \leq S_n(x) + d_\alpha; \text{all } X\} = 1 - \alpha$$

where d_α is the critical value of D_n at the $1 - \alpha$ confidence level.[7] In practice all that is required is to plot the stairstep function $S_n(x)$ and to construct a band of width $\pm d_\alpha$ about it. Then, $F(x)$ lies entirely within this band with probability $1 - \alpha$. The critical values of D_n are available in many statistical tables and texts.[8,9] Suppose, for example, we wish to determine $F(x)$ such that it lies within $S_n(x \pm 0.1)$ with

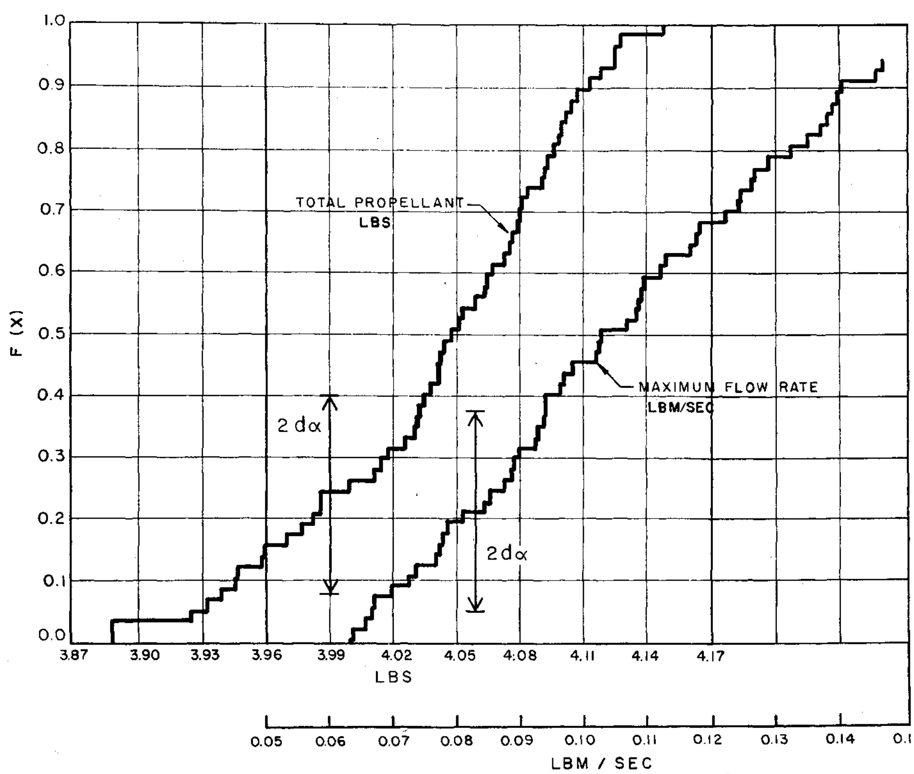

Fig. 2 Total propellant and maximum flow rate sample distribution functions with 90% confidence limits.

95% confidence. From the tables we find $d_{0.05} = 0.1 = 1.36/(n)^{1/2}$, and $n = 185$ samples.

If, as just stated, we are particularly concerned about large or small values of x, then a result due to Rényi[10] is to be preferred to the Kolmogorov statistic. Rényi determined the distribution of a proportional statistic:

$$\sup_x \left| \frac{S_n(x) - F(x)}{F(x)} \right|$$

His result leads to a confidence band of the form:

$$S_n(x)/(1 + y/n^{1/2}) < F(x) < S_n(x)/(1 - y/n^{1/2})$$

However, there are certain subtleties involved with Rényi's statistic, and the interested reader should see Spear[11] or, for detailed theory, Rényi.[10] Both papers contain tables of the asymptotic distribution of the statistic which are valid for $n > 50$. If large x is of particular concern, the variable chosen for study should be $z = 1/x, x \neq 0$.

Example

A problem was to size the propellant supply tank for a gas generation system which is to be used in a positive expulsion system for a liquid rocket engine. The solution rested primarily on an adequate description of the heat transfer taking place from the hot gas to the surrounding components. Seven first-order, nonlinear differential equations were developed to simulate the heat transfer occurring in the system and to predict the gas required for a given expulsion schedule. The principal uncertainties concerned the probable values of three heat-transfer coefficients, h_i. Previous experience had shown that an experimental investigation aimed at estimating these values was very expensive and unwarranted unless the tankage configuration was absolutely fixed. Therefore the design engineers, relying on previous experience and intuition, found it possible to formulate upper and lower bounds for each h_i and agreed to experiment with the Monte Carlo approach.

The simulation equations were set up such that the main engine propellant was expelled in a set time interval at an essentially constant flow rate. These data then specified the pressure/volume requirements of the gas as a function of time, etc. The routine for calculating the total propellant requirement (the model and associated logic) was then incorporated in the random sampling program shown in Fig. 1. The project engineer asked for an estimate of the total propellant volume that would be adequate for 95% of the cases that might arise under the assumptions concerning the h_i with 95% probability. Using the tolerance statistics just discussed with $m = r + s = 0 + 1 = 1$, fifty-eight samples are required to achieve 95% coverage at the 95% tolerance level. Figure 2 shows the sample distribution function resulting from the experiment. The value of u_{58} is 4.15 lb. Hence, by designing the system to carry 4.15 lb of gas generator propellant the probability is 0.95 (with 95% tolerance) that there will be enough propellant to deal with the demand if the assumptions concerning the h_i are valid. Figure 2 also shows the width of the 90% confidence limits, 2 d_α, on the distribution function which result from the Kolmogorov statistic for 58 samples.

The most interesting result was that the total propellant required was rather insensitive to the h_i values. This fact was of much greater interest to the design engineers than the statistical results. However, renewed interest in the statistical method arose when it was decided to consider a size index for the gas generator itself, specifically maximum flow rate, which shows greater variation in Fig. 2. This result indicated that considerable uncertainty existed concerning the dynamic requirements placed on the generator.

In the case of two correlated variables of interest, like maximum flow rate and total propellant, multivariable tolerance level results[6] can be applied, but they are beyond the scope of the present discussions.

Conclusions

An example has demonstrated the simplicity of the application of Monte Carlo methods when the variables of interest can be studied using distribution-free techniques. The three statistical methods discussed herein are but a few of a large number which can be used to advantage in the same context.

References

[1] Hammersley, J. M. and Handscomb, D. C., *Monte Carlo Methods*, Methuen, London, 1964.

[2] Schreider, Y. A., *The Monte Carlo Method*, Pergamon Press, London, 1966.

[3] Siegel, S., *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill, New York, 1956.

[4] Bradley, J. V., *Distribution-Free Statistical Tests*, Prentice-Hall, New York, 1968.

[5] Fraser, D. A. S., *Nonparametric Methods in Statistics*, Wiley, New York, 1959.

[6] Murphy, R. B., "Nonparametric Tolerance Limits," *Annals of Mathematical Statistics*, Vol. 24, 1948, pp. 581–589.

[7] Kendall, M. G. and Stuart, A., *The Advanced Theory of Statistics*, Vol. 2, Charles Griffin, 1961, p. 457.

[8] Hoel, P. G., *Introduction to Mathematical Statistics*, 3rd ed., Wiley, New York, 1962, p. 345.

[9] Owen, D. B., *Handbook of Statistical Tables*, Addison-Wesley, Reading, Mass., 1962, p. 423.

[10] Rényi, A., "On the Theory of Order Statistics," *Acta Mathematica Academiae Scientiarum Hungaricae*, Vol. 4, No. 3, 1953, pp. 191–231.

[11] Spear, R. C., "The Application of Kolmogorov-Rényi Statistics to Problems of Parameter Uncertainty in Systems Design," *International Journal of Control*, Vol. 11, No. 5, 1970.

Comparison of Basic Modes for Imaging the Earth

ALDEN P. COLVOCORESSES*
U.S. Geological Survey, McLean, Va.

IN September 1969, President Nixon in an address to the United Nations General Assembly stated, "I feel it is only right that we should share both the adventures and the benefits of space. As an example of our plans, we have determined to take action with regard to earth resources satellites, as this program proceeds and fulfills its promise." Making this a reality is going to require several complex experiments before an operational system is properly defined.

The Department of the Interior has developed the Earth Resources Observation Satellite (EROS) program so that Earth-sensing systems can be put to practical use. The department further recognizes that surveying the Earth's resources requires remote sensing from spacecraft for the global synoptic approach and from aircraft for localized use. NASA has designated the Earth Resources Technology Satellite (ERTS) as a series of space flights to meet the needs of Interior and other departments, such as Agriculture and Commerce.

The EROS program recognizes that any one satellite system may not fully meet the operational needs and has in fact defined four basic modes for remote sensing of the Earth and its resources, as follows: 1) airborne, generally film return, 2) space, data transmission, global, 3) space, film return, global, and 4) space, data transmission, geosynchronous.

Presented as Paper 70-294 at the AIAA Earth Resources Observations and Information Systems Meeting, Annapolis, Md., March 2–4, 1970; submitted March 6, 1970; revision received May 11, 1970. Publication authorized by the Director, U.S. Geological Survey.

* Cartography Coordinator, EROS Program.