NUCLEAR SCIENCE AND ENGINEERING: 22, 384-394 (1965)

Letters to the Editors

Confidence Limits for Monte Carlo Calculations

This note is concerned with calculation of confidence limits based on results of Monte Carlo calculations. While we are especially concerned with neutron and gamma-ray Monte Carlo calculations, the technique is applicable in other contexts.

In a typical neutron or gamma-ray Monte Carlo calculation, one computes random walks for N particles. N is likely to be between 1000 and 100 000. If one is interested in the fraction of neutrons absorbed in a particular region of the physical system, for example, then one obtains from the n'th random walk an estimator, x_n, for this fraction, and computes the mean estimator:

$$\bar{x} = \frac{\sum_1^N x_n}{N} .$$

To give the user an indication of the accuracy of the approximation, the sample variance

$$s^2 = \frac{\sum x_n^2 - N\bar{x}^2}{N-1}$$

may be computed, but lacking knowledge of the probability distribution of the x_n's, one has no basis for setting confidence limits on μ, the true fraction of neutrons, from s^2.

Grouping the particles into, say, G groups with M particles in each ($MG = N$) and writing $\tilde{x}_g$ for the sample mean of the g'th group

$$\tilde{x}_g = \frac{\sum_{m=1}^{M} x_{m,g}}{M} ,$$

one may compute the sample mean $\bar{x}$ and the sample variance $\tilde{s}^2$ of the $\tilde{x}_g$'s. The distribution of the $\tilde{x}_g$'s is, by the Central Limit Theorem, asymptotically (for large M) normal. Assuming normality, confidence limits for μ follow from standard techniques.

There remains the question: Are the groups of particles large enough to make the normality assumption a reasonable one? This question has been neglected in the Monte Carlo programs of which we are aware that calculate confidence intervals in this fashion. Our attention was directed to the question when such a Monte Carlo program gave us a 95% confidence interval of ± 15%, for an estimator which was wrong by 300%.

We propose that a normality test be applied to the group averages ($\tilde{x}_1, \ldots, \tilde{x}_G$).

Such a test has been in use for several months in our DRAM neutron Monte Carlo program. The test used is that of S. Shapiro and M. B. Wilk, described in Ref. 1. When the test is significant at the 10% level (the test will be "significant at the 10% level" only 10% of the time when applied to samples actually drawn from a normal distribution), the program warns the user that the validity of the computed confidence interval is doubtful. When the test is significant at the 1% level, the program warns the user that the reliability of $\bar{x}$, as an estimate of μ, is doubtful. With such small significance levels, the test cannot often detect slight departures from normality but is rarely wrong when it does indicate non-normality. We believe that it is unwise to ignore such a positive indication of non-normality.

The size of N determines the cost of the computation. Within a given cost, one would like to have M as large as possible, in order that the distribution of the $\tilde{x}_g$'s be as nearly normal as possible. On the other hand, if G is too small, the sample variance of the $\tilde{x}_g$'s becomes a poor estimator of their true variance and the normality test

[1] S. SHAPIRO and M. B. WILK, "An Analysis of Variance Test for Normality (Complete Samples)," ATL, G. E. Report No. 64GL140, (Sept. 3, 1964) (Also submitted for publication to *Biometrika*).

loses its power. We have chosen $G = 25$ as a compromise between these considerations.

Many of the Monte Carlo estimators that arise in our work have probability distributions about which little is known. It seems wise, therefore, to use a test that is powerful against a wide variety of non-normal distributions. Shapiro and Wilk have studied the power of their test against a wide variety of alternative distributions that arise in statistical theory (e.g., log normal, chi square with 1, 2, 4, and 10 degrees of freedom, non-central chi square, Cauchy, exponential, rectangular, Poisson, and others). They show the test to be powerful, although of course not equally powerful, against all of them[1]. Our confidence in the test is fundamentally based on their study.

Our experience has added to this confidence. We have been applying the test to Monte Carlo results for over a year. We have not yet had any cases in which a confidence limit that had not been flagged by the normality test turned out to be bad.

For the convenience of those who have not immediate access to the reference, the test of Shapiro and Wilk for a sample of size 25 is:

Let $z_1, z_2, \ldots, z_{25}$ be the 1st, 2nd, . . . , 25th estimators. Arrange the z's in increasing order of magnitude; relabel as $y_1, y_2, \ldots, y_{25}$ where $y_1 \leq y_2 \leq y_3 \leq \ldots \leq y_{25}$. Compute

$$b = \sum_{i=1}^{25} a_i y_i$$

and

$$S^2 = \sum y_i^2 - \frac{\left(\sum y\right)^2}{25}$$

where the $a_i = -a_{26-i}$ are given below.

i	a_i
1	-0.4450
2	-0.3069
3	-0.2543
4	-0.2148
5	-0.1822
6	-0.1539
7	-0.1283
8	-0.1046
9	-0.0823
10	-0.0610
11	-0.0403
12	-0.0200
13	0.0000

The test for departure from normality is based on the statistic

$$w_{25} = b^2/S^2.$$

The 1%, 5% and 10% significance values of w_{25} are 0.888, 0.918 and 0.931, respectively (values smaller than critical are significant).

Our experience with this test has been good, and we recommend the use of this test when confidence intervals are computed by the approach referred to here.

G. L. Burrows
D. B. MacMillan

Knolls Atomic Power Laboratory*
General Electric Company
Schenectady, New York

Received December 3, 1964
Revised February 4, 1965

*Operated for the USAEC by the General Electric Company.

On the Inclusion of Boundary Terms in Time-Dependent Synthesis Techniques

In a recent paper[1] dealing with the application of synthesis techniques to various time-dependent problems, Kaplan, Marlowe, and Bewick presented a variational principle for linear time-dependent group-diffusion theory. The principle, however, is not stationary with respect to arbitrary variations in the functions involved because not all of the end-point (in time) terms in the first variation vanish. The authors postulate that this difficulty can be removed if the variations are limited to functions having the same end points as the selected function. Implicit in such an argument, however, is the assertion that the trial functions which are to be assumed will lead to an approximate solution having the same end-point values as has the exact solution. This is a requirement not easily met.

The difficulty can be avoided by the inclusion of appropriate boundary terms in the functional. For simplicity, consider the one-group-flux and adjoint-flux equations without delayed neutrons:

$$\frac{1}{v}\frac{\partial \phi}{\partial t} = \nabla \cdot D\nabla\phi + (\upsilon F - A)\phi \qquad (1)$$

$$-\frac{1}{v}\frac{\partial \phi^*}{\partial t} = \nabla \cdot D\nabla\phi^* + (\upsilon F - A)^*\phi^* , \qquad (2)$$

[1] S. KAPLAN, O. J. MARLOWE and J. BEWICK, *Nucl. Sci. Eng.*, 18, 163-176 (1964).