A Partition Test for Pseudo-Random Numbers

By J. C. Butcher.

A frequently used test on random digit sequences consists in forming from them sets of n integers; $\nu_1, \nu_2, \cdots, \nu_n$ say, which, on the hypothesis of randomness, are independent and have equal probabilities, $1/x$, of being each of the integers $1, 2, \cdots, x$ (x usually being chosen as a power of 2). From any equalities that may exist between $\nu_1, \nu_2, \cdots, \nu_n$ a partition of n is defined and the actual test is on the frequency of occurrence of the different partitions of n.

For example, a popular form of the test known as the "poker test" distinguishes the different partitions of $n = 5$ known by the descriptive names "all different," "one pair," "two pairs," "three of a kind," "full house" (three of one kind, two of another), "four of a kind" and "five of a kind."

It is obvious that much computing time must be absorbed in distinguishing between the various possibilities, and that if n is much greater than 5, the number of possible partitions becomes very large. For this reason it may be advantageous to group the partitions together in some way that lowers the number of cases and also simplifies the programming necessary to distinguish these cases.

A convenient way of doing this will now be described. It has the added advantage that the calculation of the expected probabilities is extremely simple. Let us distinguish n classes of partitions $C_1, C_2, \cdots, C_n$ where C_r includes all partitions defined from exactly r different integers occurring amongst $\nu_1, \nu_2, \cdots, \nu_n$. In the example $n = 5$, C_1 would include the single partition "five of a kind," C_2 would include the two partitions "full house" and "four of a kind," C_3 would include "three of a kind" and "two pairs," while C_4 and C_5 would each contain the single partitions "one pair" and "all different," respectively.

If the n integers $\nu_1, \nu_2, \cdots, \nu_n$ are generated one after the other, and if we define $S_i (i = 1, 2, \cdots, n)$ as the set of integers $1, 2, \cdots, x$, other than those identical with ν_j for some $j < i$, then the value of r that characterizes the class C_r is given by

$$r = \sum_{i=1}^{n} \epsilon_i ,$$

where ϵ_i is equal to 1 or 0 according as ν_i is or is not a member of S_i.

When x is no more than the computer word length, then S_i is conveniently represented by the binary digits of a single word, and this word is quickly computed from the word representing S_{i-1} if, for example, a logical conjunction instruction is available.

To find the expected frequency of occurrence of the class C_r for $r = 1, 2, \cdots, n$ we must find the probability p_r that the class C_r will arise in a single trial. It is easy to see that

$$p_1 = \frac{1}{x^{n-1}} a_1 ,$$

Received June 30, 1960.

$$p_2 = \frac{(x-1)}{x^{n-1}} a_2 ,$$

$$\vdots$$

$$p_n = \frac{(x-1)(x-2)\cdots(x-n+1)}{x^{n-1}} a_n ,$$

where $a_1 , a_2 , \cdots , a_n$ are dependent on n, but not on x. Since $p_1 + p_2 + \cdots + p_n$ is identically equal to unity, we find

$$(1) \qquad x^n = a_1 x + a_2 x(x-1) + \cdots + a_n x(x-1)\cdots(x-n+1)$$

for all values of x. The coefficients $a_1 , a_2 , \cdots , a_n$ are thus identical with Stirling's numbers of the second kind of order n [1], and they can be evaluated in turn by substituting $x = 1, 2, \cdots , n$ in (1) and using as a check the value $a_n = 1$.

Department of Applied Mathematics
University of Sydney

1. C. JORDAN, *The Calculus of Finite Differences*, Chelsea Publishing Co., New York, 1947, p. 170.