LETTERS TO THE EDITORS

The Editors do not hold themselves responsible for opinions expressed by their correspondents. No notice is taken of anonymous communications.

A Suggested Statistical Model of some Time Series which occur in Nature

THE following is an amplification and extension of some work which was described in a general way at a symposium of the Research Section of the Royal Statistical Society on Storage Problems on March 6, 1957.

Annual values of some phenomena such as river discharges, rainfall, temperature, are approximately normally distributed if no account is taken of order of occurrence. Their principal characteristic, however, is the occurrence of periods when, on the whole, values tend to be high and others when they tend to be low, though low values may occur in a high period and vice versa. So far as is known, there is no regularity in the occurrence or the length of these periods, and usually there is no significant correlation over one of them between a year and its successor.

This phenomenon is important in problems of storage. In these a quantity R enters, which is obtained by taking departures of the annual values from their mean for a period of N years, and plotting the continued sums of these departures against the numbers 0 to N. The range of this curve is R and is easily seen to be, in the case of a river, the storage capacity which would have been necessary to enable the mean discharge to flow in every year of the period. In the case of natural phenomena, of which about a hundred have been examined[1], the following equation represents the mean result:

$$R/\sigma = (N/2)^K \qquad (1)$$

where σ is the standard deviation and K is a quantity which has a normal distribution and the mean of which is 0·73.

If, however, the quantities considered are entirely independent events, such, for example, as would arise from tossing a set of, say, twelve coins and recording the differences between the number of heads and number of tails at each throw, R is represented by the following equation:

$$R/\sigma = \sqrt{\tfrac{1}{2}N\pi} = 1{\cdot}25\sqrt{N} \qquad (2)$$

The values of R are the same from the two equations when $N = 22$; but for higher values of N the value of R from equation 1 steadily increases at a greater rate than does R from equation 2. For $N = 100$ the difference is about 25 per cent, and for $N = 1{,}000$, 100 per cent. There is therefore a definite difference between many natural time series and those where the terms are independent of each other.

An apparatus by which either type of series can be imitated is a pack of what I have called 'probability cards'. The first pack contained fifty-two cards numbered ± 1, ± 3, ± 5, ± 7, and the numbers of each kind of card were thirteen of the ones, eight of the threes, four of the fives, and one of the sevens. These numbers are proportional to the ordinates of the normal curve corresponding to departures of 1, 3, 5 and 7 from the mean 0. If cards are cut from this pack, shuffling after each cut, a series of random numbers will be produced which will conform to equation 2, and this has been found to be the case[1].

If, however, the following procedure is adopted, a different result will be obtained. The pack is shuffled and a card is cut, and after its number is noted it is replaced in the pack. Two hands are then dealt, and if, for example, the card cut was $+3$, the highest three positive cards in one hand are transferred to the other, and the highest three negative cards are removed from it. This hand then has a definite positive bias. A joker is now placed in it and it is shuffled and a card cut from it. The number on this card is the first of the series. It is replaced and the hand then reshuffled and another card is cut and recorded. This cutting and shuffling goes on until the joker is cut. Then the joker is removed and all the cards are put together and shuffled, after which the process is repeated. The amount of bias, the number of cuts for which the bias acts, and the actual cards which turn up all depend on random processes.

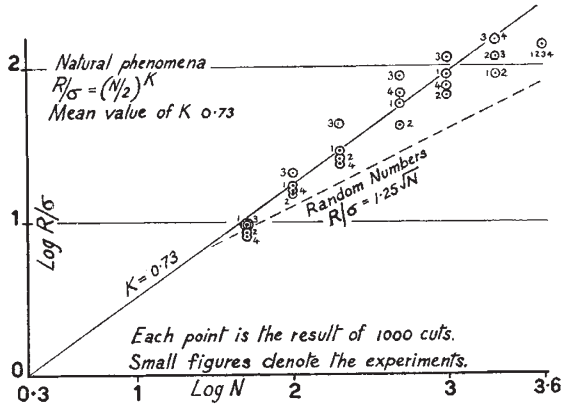

Fig. 1. Cutting probability cards. Pack of fifty-two. Biased hand of twenty-six cards + joker

Four different experiments of 1,000 cuts each were carried out. From each of these R was determined for twenty sections of fifty cuts, ten sections of a hundred, nine overlapping sections of two hundred, etc. The points up to $N = 1{,}000$ are the means of these sets. The four experiments were combined in order of occurrence so as to give values for 2,000 and 4,000 cuts. The results are shown in Fig. 1, where small figures alongside the observed points indicate to which experiment they belong. The mean value of K from all experiments is 0·71.

The card experiments have not been fully analysed as to variation of means and standard deviations, and further experiments are also being made. The result, however, indicates that the process described leads to time series very similar to those which occur in Nature.

H. E. HURST

Sandford on Thames,
Oxford.
May 11.

[1] Hurst, H. E., *Trans. Amer. Soc. Civ. Eng.*, **116**, 770 (1951); *Proc. Inst. Civ. Eng.*, Pt. 1, **5**, 519 (1956).