REVIEWS OF STATISTICAL AND ECONOMIC BOOKS

CONTENTS

1.—*A Study in Analysis of Stationary Time Series*. By Herman Wold. Uppsala: Almqvist und Wiksell, 1938. 9″ × 6″. viii + 214 pp. Price kr. 6 (bound kr. 8.)

In a lecture I gave two years ago on the different methods of time series analysis, I divided all the methods applied into two categories, which I called "empirical" and "aprioristic," and remarked that although the empirical approach is much the more common and, at present, the more useful, the future belonged to the aprioristic method.* If we did not apply it at present, it was because we knew too little about the necessary mathematical tools—namely, the branch of the theory of probability dealing with what might be called "random curves" or "random processes." That theory, originated by the Russian mathematicians, Markoff, Khintchine, and Kolmogoroff, was then in an early stage and was to be found only in a rather abstract form in articles in mathematical journals.

This book by Herman Wold fills the need for a publication giving

* J. Neyman: *Lectures and Conferences on Mathematical Statistics*, Washington, D.C., 1937.

the theory in a form adapted to practical applications and it may be warmly recommended to economists and statisticians engaged in the analysis of time series. As stated in the preface, Mr. Wold is a student of H. Cramér, and derived inspiration from the writings of G. U. Yule and A. Khintchine, as well as from the lectures of his immediate teacher. The combination of the names of Yule and Khintchine may seem surprising, for Mr. Yule would not claim to be a pure mathematician, and Mr. Khintchine would never be recognized as a statistician. Yet it appears that they were studying essentially the same things.

The mathematical equivalent of the statistical conception of the time series is the random process. This is defined as follows: Imagine a finite or infinite time interval and denote by t any moment in it. Consider next a random variable $x(t)$ which, in a certain sense, corresponds to the moment t. To each moment t there will correspond some particular random variable $x(t)$. The random variables $x(t')$ and $x(t'')$ corresponding to two different moments t' and t'' need not be independent. We are thus led to consider an infinite system of, in general, mutually dependent variables and their joint probability law. Such a system of variables is just what is called a random process. It may be continuous, if the variable t is allowed to vary continuously, or discrete—if t may have only a denumerable set of equidistant values. An actually observed time series is considered as a set of particular values of the variables $x(t)$.

Mr. Wold limits his considerations to discrete and stationary random processes. Denote by $F(t_1, t_2, \ldots t_n;\ u_1, u_2, \ldots u_n)$ the probability of the simultaneous fulfilment of the n inequalities $x(t_i) \leqslant u_i$ for $i = 1, 2, \ldots, n$, as determined by the joint probability law of the variables $x(t)$ forming a discrete random process. If this function possesses the property that $F(t_1 + k, t_2 + k, \ldots, t_n + k;\ u_1, u_2, \ldots u_n) \equiv F(t_1, t_2, \ldots, t_n;\ u_1, u_2, \ldots, u_n)$, then the random process is called stationary. We may explain this verbally by saying that by a stationary process the probability of various situations at some moments does not depend on the position of those particular moments, but on their mutual distances in time and on what happened before. The process which is not stationary is called evolutive.

The author gives the necessary definitions and a sequence of interesting theorems concerning the discrete stationary processes. Some of these are taken from current literature, but some are new. Various remarkable types of the processes considered are distinguished. The simplest type is, of course, the purely random process by which any variable $x(t')$ is independent of any other $x(t'')$ and all follow the same law of frequency. Next comes the type described as the process of moving averages. If $\eta(t)$ is a random variable belonging to a purely random process, and if for any t the variable $\xi(t)$ is defined by the relation

$$\xi(t) = b_0\eta(t) + b_1\eta(t-1) + \ldots + b_n\eta(t-n),$$

then the process composed of the variables $\xi(t_i)$ is called the process of moving averages. We recognize here a conception previously discussed by Yule, and there are many more similar contacts with the work of Yule in the book.

Having discussed a number of types of random processes, the author furnishes proofs of their various properties, and then gives a theorem of considerable interest, concerning the structure of the most general discrete stationary process. This, in fact, can always be presented as a sum of two components the nature of which is known. When dealing with stationary processes, the author discusses what he calls stochastical difference equations, and proves some of their fundamental properties. All this is done in Chapters II and III. Chapter I is given to a review of some of the current methods of time series analysis. The fourth and last chapter will be the most interesting to a non-mathematical reader, for it contains the applications of the theory to a number of actual time series. The first example is the series of yearly wheat prices in Western Europe, 1518–1869, compiled by Sir William Beveridge. Next come two correlated time series of (i) the level of Lake Vaner and (ii) the rainfall. Finally, there is a detailed discussion of a series representing the cost-of-living index in Sweden. Besides those examples, the discussions of which amount almost to separate studies, the author refers frequently to various time series previously discussed by Mr. Yule, Sir G. Walker and others. The book ends with two appendices on the Cramér-Mises ω^2 test for goodness of fit, and—a very interesting one—on the numerical significance of correlation coefficients.

It seems probable that the publication of Mr. Wold's book, which is, so far as I know, the first of its kind, will considerably influence the development of the time series analysis, but still some of the most important problems involved remain unsolved and, as the author says himself, at present it is difficult to see how to attack them. He describes them as sampling problems, but really they are problems of testing hypotheses and of estimation. Given a time series and a hypothetical probabilistical model, M_0, represented by a random process, shall we agree that the model M_0 does represent the machinery which produced the time series, or shall we reject the model M_0 and look for another? This is the kind of question we must learn how to answer. The general method of attacking it seems to be that common to all problems of testing hypotheses. If we expect that the original model may be wrong, we must be clear about the ways in which it may be wrong. This will involve a definition not of one single random process M_0, but of the whole family of them, that may be denoted by (M). If a time series consists of n figures, then it can be represented by just one point, E, in a space W of n dimensions and, whatever be a region w in that space, each of the models M will define a probability that the point E will fall within w. The process of looking for a criterion appropriate to test the hypothesis that M_0 adequately represents the actual machinery behind the time-series considered, is equivalent to a search of a certain region w_0 in the space W such that, (i) whenever M_0 is true the point E is unlikely to be found in w_0 and (ii) whenever M_0 is not an adequate model of the phenomena considered, then the chance of E being found in w_0 is as great as possible. We see that the questions to solve do involve the problems of sampling. The amount of such problems seems to be even greater than is suggested by the author. But there is something beyond those problems, that of choosing a proper criterion. In view of the author's manifest skill, we may confidently hope that some of these problems will be found solved in future editions of Mr. Wold's very interesting book.

J. N.

2.—*Statistical Methods for Research Workers.* By R. A. Fisher. 7th edition. Edinburgh: Oliver and Boyd, 1938. $8\frac{3}{4}'' \times 5\frac{1}{2}''$. xv + 356 pp. 15*s.*

In the seventh edition of this now classical work the author follows his previous practice of adding new sections on methods that have recently been evolved. In this edition a most valuable note has been added on the discrimination of groups by means of multiple measurements. The statistical technique of the analysis of multiple measurements has recently received considerable attention from a number of quarters, and the utility and wide application of discriminant functions in making possible the exact treatment of previously insoluble problems are now recognized.

The section on orthogonal polynomials has also been expanded so as to give a fuller introduction to their theory, by way of orthogonal comparisons between observations, a revision which is made more valuable by the fact that tables of the polynomical coefficients are now available in *Statistical Tables.*

F. Y.

3.—*Statistical Tables for Biological, Agricultural and Medical Research.* By R. A. Fisher and F. Yates. London: Oliver and Boyd, 1938. $11\frac{1}{4}'' \times 8\frac{1}{2}''$. 90 pp. 12*s.* 6*d.*

There are thirty-four tables in this book, and it is only necessary to enumerate them to give an idea of their value. The first seven are tables of significance points of the normal distribution and associated sampling distributions, t, χ^2, z and the correlation coefficient, and will be familiar to readers of Professor Fisher's *Statistical Methods.* Then come tables for testing significance 2×2 contingency tables, and tables of profits and angular transformations. These are followed by tables of Latin squares and associated quantities; of scores for ranked data; and of orthogonal polynomials. The set concludes with some of the material commonly required in statistical work, logarithms, natural logarithms, squares, reciprocals, sines, tangents, random numbers, and a very useful table of constants, weights and measures which is far superior to the usual perfunctory table with which volumes of this kind generally conclude.

For once a reviewer can use a stock phrase without exaggeration. The book will be indispensable to users of the newer methods in statistics.

M. G. K.