consumption series for New England and, in some cases, separately for each of the six states. Most series are on a yearly basis and end with 1956. Tabular data are provided and procedures and sources are stated. The graphs are effectively drawn and a scanning will give the reader a good picture of the structural changes which have been occurring in New England.

Information Theory and Statistics. *Solomon Kullback.* New York: John Wiley and Sons, Inc.; London: Chapman and Hall, Ltd.; 1959. Pp. xvii, 395. $12.50.

D. V. LINDLEY, *University of Cambridge, England*

THIS book is written around the following fundamental ideas.

(i) Let $f_1(x)$, $f_2(x)$ be two probability densities, then

$$I(1:2) = \int f_1(x) \log [f_1(x)/f_2(x)]dx \tag{1}$$

is a suitable measure of the difference between $f_1(x)$ and $f_2(x)$: Kullback calls it the "mean information for discrimination in favor of $f_1(x)$ against $f_2(x)$ per observation from $f_1(x)$" (p. 5).

(ii) For fixed $f_2(x)$ and fixed $T(x)$, the minimum value of $I(1:2)$, as $f_1(x)$ ranges through the class of distributions for which the expectation of $T(x)$ with respect to $f_1(x)$ is equal to a fixed value θ, is attained when $f_1(x)$ is proportional to

$$e^{\tau T(x)} f_2(x) \tag{2}$$

where τ is a function of θ. (This is a form of the Cramér-Rao inequality.) (p. 38).

(iii) As θ, and therefore τ, varies the family (2) is an exponential family with $T(x)$ as a sufficient statistic, providing an unbiased estimate of θ of minimum variance (p. 44).

(iv) If we wish to test the simple null hypothesis H_2 that the density is $f_2(x)$ against the composite alternative H_1 that the density belongs to the family (2) we may use the statistic $\hat{I}(*:2)$ which is the minimum value of $I(1, 2)$ in (ii) for $\theta = \hat{\theta}(x) = T(x)$, where x is the observed value; rejecting if $\hat{I}(*:2)$ is too large. (Crudely, the idea is to find the member of H_2 which is nearest to—diverges least from—H_2.) (p. 81).

(v) If H_2 is composite the same object may be accomplished by using the minimum of $\hat{I}(*:2)$ as $f_2(x)$ ranges throughout H_2 (p. 85).

We thus have a method of testing hypotheses about exponential families which is applied in Chap. 6 to multinomial populations; in Chap. 7 to Poisson populations; in Chap. 8 to contingency tables, and in Chaps. 9–13 to normal populations, dealing with tests of linear hypotheses (univariate and multivariate), tests of homogeneity and linear discrimination. In all cases the analysis is for samples of fixed size. The earlier chapters develop the fundamental ideas quoted above. The style is severely mathematical and in summarizing the ideas I have ignored the qualifications that are necessary in presenting a rigorous account.

The book proper begins "Consider the *probability spaces* $(\mathfrak{X}, \mathcal{S}, \mu_i)$, $i=1, 2, \cdots$" (p. 3) and continues in this vein. Thus (1) is written with densities defined (by the Radon-Nikodym theorem) with respect to a σ-finite measure $\mu(x)$. This is a perfectly proper and sensible thing to do, but a reviewer is bound to mention it because it affects considerably the class of potential readers. Not only is the mathematics at the beginning abstract but later on it becomes complex: even by p. 27 we have some pretty formidable expressions but by pp. 243–5 one begins to feel that we have moved into a realm where mathematical symbolism is perfection and the use of the

English language almost indecent. Nevertheless, the mathematics seems remarkably free from errors and the style, for someone who can take such heavy doses of symbolism, good. There are lots of problems at the end of each chapter, some of them dealing with practical statistical situations. There are copious references to other authors, numerous examples throughout the text, an excellent bibliography, a glossary and index, and tables of the following functions (i) $\log_e n$ and $n\log_e n$ for $n=1(1)1000$, to ten places of decimals. (ii) $p_1\log(p_1/p_2)+q_1\log(q_1/q_2)$, where $p_i+q_i=1$, for p_1, $p_2=0.01(.01)0.05(.05)0.95(.01)0.99$ to seven places of decimals (iii) the 5% points of Fisher's B^2 (related to non-central χ^2) for $\beta=0(.02)5.00$ and $\nu=1(1)7$.

The merit of this book lies in the unity of the treatment based on the fundamental ideas. Most of the techniques that statisticians use in dealing with samples of fixed size from exponential distributions are derived by use of them. Notable exceptions are certain exact, or small-sample, techniques, such as Fisher's test for the 2×2 contingency table. This is surely valuable for aesthetic and pedagogical reasons, besides the fact that one would expect that such a unified approach to statistics would suggest new developments: there are some indications that this is so in the later chapters on multivariate analysis, which present a compact and coherent account of the subject. One attractive outcome of the information approach is the treatment of contingency tables based essentially on $\sum O\log(O/E)$ instead of the usual χ^2 statistic, $\sum(O-E)^2/E$. The additivity of information, but not of χ^2, is an appealing property. The student (and he will have to be an advanced student, for the basic ideas of probability and statistics are assumed known) will welcome the unity of ideas, but will undoubtedly be worried when reading outside of the text because most writers do not adopt Kullback's approach.

A critical evaluation of the book hinges on a discussion of the statements (i)–(v) above. (ii) and (iii) are statements of mathematical results but they incorporate the concept of unbiasedness, and they restrict the techniques to the exponential family. The latter is not serious—indeed an adequate account of the statistics of this family would be enough for most practical purposes—but the former is. To demand that a statistic has no bias is difficult to justify in any development of the subject, and, with the rising popularity of the use of the likelihood function through the writings of Savage, Barnard and others, is less attractive than it used to be. The use of (1) is the essential theme and is most open to objection. It is a pity that Kullback does not spend more time in the book talking around the notion—in words and not symbols. My own view is that he has been correct in using the logarithmic measure, but that he has used it wrongly. I feel that there is a need for some measure of information in statistical work, quite apart from considerations of utilities and decisions, and that this measure should be additive. Hence the logarithm. But also the information should take account of our prior knowledge. If one has strong prior opinions then a suggested experiment may be expected to be uninformative, but if ones opinions are vague the same experiment may be expected to yield much information. This point arises in Shannon's original work where he points out that information is essentially a statistical idea; that is that a message must be considered in relation to a set of messages: similarly a parameter must be considered in relation to a distribution of parameters. Consequently Kullback's measure (1), which is also used by Jeffreys, seems to me to be wrong in principle, just as Fisher's is. Of course in large samples they are all right (and all equivalent) because the effect of prior knowledge in a large experiment is necessarily slight. All the tests in this book are really large-sample ones, some of them accidentally having small-sample optimum properties: consequently Kullback's ideas do not result in different answers from other peoples.

That criticism of (i) is a personal one. (iv) and (v), however, can be criticized on the grounds that although the technique is appealing, it leads to the same criterion as does the likelihood ratio principle (p. 94). Indeed, Kullback bases the distributional theory of his criterion on the known distributional theory of the likelihood ratio. Perhaps it is only because the ratio has been with us longer, but I would have thought that that principle was intuitively more appealing than Kullback's more complicated one. Also we know that the likelihood ratio test can be absurd in small samples; can the same be said of the information test? Unfortunately there is no discussion of small-sample properties, apart from certain inequalities. In the case of tests of univariate linear hypotheses the property that the F-test has of being uniformly most powerful amongst invariant tests is a stronger reason for its adoption than Kullback's.

To summarize: this is an advanced book on the large-sample theory of tests involving Poisson, multinomial and normal distributions treated mathematically by a new unifying approach which is open to objections but which yields results, seen to be satisfactory by other methods. Incidentally, it is a book only for the expert statistician; others interested in information theory will find almost nothing of value to them.

Alcune memorie matematiche. *F. P. Cantelli.* Pubbl. Facoltà Economia e Commercio, Università di Roma. Milano: A. Giuffré, 1958. Pp. xxx, 448, Ital. Lire 4.000.

LEONARD J. SAVAGE, *University of Chicago*

F. P. CANTELLI's pioneering work in probability theory, financial and actuarial mathematics, and statistics is largely unknown save through indirect references. Many of his papers are amost inaccessible and almost all are written in Italian. This book, dedicated to Cantelli on his 80th birthday, reproduces a selection of 19 of the 90 "principal papers" listed and summarized in the bibliography, which constitutes an important part of the book. Both the reprinted papers and the summaries contribute greatly to the accessibility of Cantelli's work, but it is regrettable that an English or French translation of the summaries was not included, for these would have been of great interest to many who cannot read Italian.

A rough idea of the scope of Cantelli's research is given by listing some of the topics covered by the papers in the bibliography. (The numbers refer to the bibliography, where an asterisk means that the paper is reprinted in the book and "F" means that it is in French.):

The strong law of large numbers (*24, 33, *55, 61F, 80); convergence in probability theory (*23, 41, *58F); foundations of abstract probability theory (*54, 59, 62, 67); if X is normal and Y is normal conditionally on X, when is $X+Y$ normal? (*27, 88); systematic use of random variables in actuarial problems, risk theory, etc. (14, *15, 40, 44, 70, *74, *78); generalizations of the Bienaymé-Tchebycheff inequality (*15, 16, 42); laws of interest, their "divisibility," etc. (10, *19, 34, 73, 81); tables of multi-cause elimination and of "mutuality," social insurance (*19, 37, 38, 66, 79, 87). Many other topics might have been mentioned.

Not only does the book provide access to the works of a great man, it will also cast light on the history of currently widespread ideas, such as the strong law of large numbers, and, even more important, it may attract new attention to certain ideas, such as tables of mutuality, that have perhaps been too lightly passed over.

Formeln und Tabellen der mathematischen Statistik. *Ulrich Graf and Hans-Joachim Henning.* Berlin, Göttingen, Heidelberg: Springer-Verlag, 1958. Pp. vii, 104. DM 12.60.

GOTTFRIED E. NOETHER, *Boston University*

THIS is a new printing (incorporating some minor corrections) of the edition which appeared in 1953. In the preface, the authors state their intentions in approximately the following words: "While the professional mathematical statistician finds