BOOK REVIEWS

Correspondence concerning reviews should be addressed to the Book Review Editor, Professor Jack Kiefer, Department of Mathematics, Cornell University, Ithaca, New York 14850.

RICHARD VON MISES (edited and complemented by HILDA GEIRINGER), *Mathematical Theory of Probability and Statistics*. Academic Press, New York and London, 1964. $22.00, £7/17/0. xiv + 694 pp.

Review by D. V. LINDLEY

University College of Wales

This book is based on the author's earlier work on probability [5], on the notes of lectures that he gave at the universities of Harvard, Rome and Zurich up to 1952, and on material in his papers and notebooks. These have been put together, edited and complemented by Hilda Geiringer. As a result we have a reasonably complete account of von Mises's own work in probability and statistics and of his attitude toward the work of his contemporaries. As such it deserves the greatest attention from other workers in these two fields. The book has been written as an advanced textbook in the hope that both students and research workers will find it useful. Von Mises made many important contributions covering a great range of topics and it is scarcely within the competence of a single reviewer to do justice to everything in this large book that is so rich in stimulating ideas. I therefore propose reviewing the material a chapter or so at a time, the comments reflecting my own interests and therefore my own limited competence to judge.

Chapter 1 (49 pages) contains the material for which von Mises is best known; the basic ideas of a collective for a discrete label space. In any mathematical treatment of the physical world there is considerable arbitrariness about the starting point. Von Mises felt that the usual approaches to probability did not have enough contact with the physical reality. He preferred to start somewhat earlier in the chain and consider an axiomatic system that dealt directly with the indefinite repetitions upon which probability theory is usually held to be based, rather than to take the abstractions from these repetitions (the frequency limits) and use these to suggest, for example, the measure theory axioms. Of course, when he first put forward his ideas in 1919, the currently accepted measure theory approach did not exist. Consequently von Mises is concerned with infinite sequences. In the simplest case of a label space containing only two elements, 0 and 1, the sequence will consist of 0's and 1's and possesses the basic probability property that the proportion of 1's in the first n terms of the sequence tends to a limit as $n \to \infty$. But the sequence must have another property, it must in some sense be random and the idea of randomness has to be incorporated into the axioms. It is the difficulty in doing this that is the stumbling block of von Mises's approach.

His idea is to consider a selector function, alternatively called a place selection, which selects some members of the sequence so forming a new sequence. For suitable place selections the resulting new sequence should also have the limiting frequency property and the same limit. Thus if the sequence is one of heads (0) and tails (1) resulting from the toss of a coin, the place selection which chooses alternate members of the sequence will, with probability one, result in a sequence having the same limiting frequency of heads as the original. This is an expression of the impossibility of a successful gambling system with a random sequence. On the other hand if the original sequence is a strict alternation of 0's and 1's (limit $\frac{1}{2}$) the same place selection will produce a sequence entirely of 0's (limit 1). The original sequence is not random.

This leads to the idea of a collective defined as follows (p. 12):

"Let $S = \{a_i\}$ be a discrete label set and K a sequence of elements of S. Let G be a system of denumerably many place selections. We assume that

(1) For every label a_i, the limiting frequency p_i exists in K;

(2) $\sum p_i = 1$, the sum extending over all elements of S;

(3) Any place selection Γ belonging to G applied to K produces an infinite subsequence of K in which again, for every a_i, the limiting frequency exists and is equal to p_i.

Then K, or, more explicitly, $K(G, S)$, is called a collective (with respect to G and S) and $p(a_i) \equiv p_i$ is the probability of encountering the label a_i in K."

I quote the definition because it seems to me to be quite unacceptable. Notice that the number of place selections must not be too great (G is usually denumerable) because otherwise collectives would not exist. Notice also that collectives are defined relative to G. As explained above the object of place selection is to remove from consideration regular sequences like $010101 \cdots$, which do not have the randomization property, from consideration. But the definition does not achieve the removal: an example will illustrate. Let G be the class of place selections Γ_r ($r = 0, 1, 2 \cdots$) which select only those places of the form $k(2r + 1)$ for $k = 1, 2, \cdots$. For any Γ_r, the result of applying it to the sequence which is alternately 0 or 1 is to produce the same sequence and therefore the same limiting property. Therefore if $K = (010101 \cdots)$, $K(G, [0, 1])$ is a collective. This is in conflict with von Mises's ideas. (See, for example, the discussion of Bernoulli's theorem on p. 175).

How does this come about? I think it arises from his admission that the collective, K, can only be defined with respect to the system, G, of place selections; and then to omit reference to G. For in the subsequent discussion he talks of K, K^1, etc. and hardly ever $K(G)$, $K^1(G^1)$. Let me give an example. After the definition, several operations on collectives are defined. One of these is the operation of partitioning, in which we take a subset of S and consider only those members of K which belong to the subset. This enables conditional probability with respect to the subset to be defined. In the definition of partition (p. 24) no reference is made to G. To see that G is relevant consider $K = (012012012 \cdots)$. This is a collective relative to the system of place selections Γ_r^* ($r = 0, 1, 2 \cdots$) which selects only those places of the form $k(3r + 1)$ for $k = 1, 2, \cdots$. Now consider a partition of K with respect to the subset $(0, 1)$ of $S = (0, 1, 2)$. The result is the sequence 010101 $\cdots$ which is not a collective with respect to $G^* = \{\Gamma_r^*\}$ though it is with respect to $G = \{\Gamma_r\}$. There is no discussion anywhere that I could see of the need to consider G in any of the operations on collectives.

It may be objected that the place selections in the examples just cited violate the definition of a place selection. It is an essential part of the definition that the decision whether or not to select the nth member of the sequence depends *only* on n and the previous members of the sequence, and *not* on the nth or subsequent members. The above selection rules depend only on n and therefore, I submit, do satisfy the definition. (Von Mises gives an example of a place selection that depends only on n on p. 9).

Criticisms of von Mises's formulation have been made before. For example, Church [1] says of essentially the same definition as appears in the book, "However, this definition . . . while clear as to general intent, is too inexact in form to serve satisfactorily as a basis of a mathematical theory." Church then goes on to formulate a definition which considers a fixed class G of place selections, namely a class consisting of all effectively calculable functions: a notion due to him and explored by Turing (references are in the paper just referred to). If this restricted, but sufficiently broad, class of place selections were used the counter-examples cited above would be invalid. It appears to your reviewer, though the logical ideas involved here are outside his special field of knowledge, that Church's idea avoids many, if not all, of the difficulties, and it is surprising that von Mises makes no mention of it. (There is no reference to Church in the book: I am indebted to G. A. Barnard for drawing it to my attention). It would be necessary to show that the operations, such as partition, generate collectives with respect to the same class of calculable functions, but this is probably not difficult. I am encouraged in my belief that Church's formulation is satisfactory by a lecture of Kolmogorov's in Belgrade in September 1965 in which computable functions were apparently similarly used. Professor L. J. Savage draws my attention to a most comprehensive and valuable discussion of von Mises's ideas by Jean Ville, [6].

The position, as I see it, of von Mises's work on the foundations of probability is that it is an avenue which appears to lead to a dead end. Such avenues have to be explored, for how else do we know they are dead? To build up the notion of probability from sequences of repetitions has not succeeded. The successful avenue appears to be that based on measure theory. There is some discussion of this by von Mises in the second appendix to Chapter 1. The answer to his argument seems to have been given by Kolmogorov in 1933. (The easiest reference in English is Kolmogorov [3].) Essentially the answer is that probability is distinguishable from the general run of measure theory in the emphasis the former places on independence, a concept which is basic to probability and replaces the notion of a place selection.

Chapter 2 (62 pages) deals with the extension of the ideas of the first chapter to label spaces other than discrete ones. The material is interesting because not all events that have a probability in the usual measure-theoretic approach possess one in von Mises's. For example, one can entertain the idea that in a sequence an event will only occur a finite number of times. It needs care in formulating this as a frequency limit and hence as a frequency probability. Von Mises classifies events into those that have, for him, a probability, and others which have only a measure. I do not feel myself competent to judge the material here: the upshot is a situation which is more complicated than in the usual approach.

With the axiomatics out of the way, Chapter 3 (29 pages) is concerned with the basic properties of distributions on the lines of any text on probability theory. These are extended in Chapter 4 (40 pages) and the topics covered include Bernoulli's theorem, theory of runs, applications to genetics and a brief section on Markov chains. (Much of his work on stochastic processes is not included in this volume). The discussion of Bernoulli's theorem (the weak law of large numbers for binomial sequences) is especially interesting because if not carefully stated it would appear to be without content in a theory based so strongly on a frequency limit. The discussion ends with a false statement: "If one uses a probability concept that does not relate to the frequency of an event, neither the Bernoulli theorem, nor any of its generalizations, lead to any statement concerning the frequencies in a sequence of trials" (pp. 176–7). This ignores the fact that de Finetti, using a non-frequency type of probability based on degree of belief, has derived frequency limits. Incidentally, it is somewhat surprising to see von Mises make so little reference to de Finetti's work. There are two mentions of him (pp. 73 and 99) and both refer back to Appendix One for details. But Appendix One contains no reference to de Finetti or his work. The most accessible reference for English readers is to be found in Kyburg and Smokler [4], though the most important of de Finetti's papers is much earlier [2]. It seems to me that the elusive concept of randomness that von Mises tried to formalize in the notion of a collective is satisfactorily captured by de Finetti in his notion of exchangeability.

Chapters 5 and 6 (a total of 105 pages) deal with technical problems in probability theory and in particular with the central limit theorem. I found the treatment here complicated and longwinded. It would seem to be based on the 1931 material, though there are many additions, and little advantage has been taken of developments by other workers since then. A probable reason for many of the difficulties is the added complication introduced by von Mises's restricted idea of probability and the consequent inability to use modern measure theory ideas. It is also not clear what position the concept of random variable holds for von Mises. The only reference in the index is to p. 52, but there is no definition there and I was not able to find one despite hints that it has occurred earlier. Without the usual probability σ-algebra and the mappings onto the reals, it is clear that he is severely handicapped. Confirmation of this is obtained by the fact that most of the limiting results are expressed in terms of distributions (either densities or distribution functions) and not in terms of random variables. As a simple and immediately comprehensible example of the treatment, reference can be made to the two-page proof (pp. 228–230) of the formulae for the mean and variance of a sum: this is expressed in terms of convolutions rather than by means of random variables.

This is perhaps a convenient point to comment on the style. Von Mises is quite capable of stating a result incorrectly. A simple but striking example is the italicized theorem on p. 279 ("a fundamental theorem of the probability calculus"). As stated it is false but afterwards (without italics) he goes on to say "the conditions . . . have all been listed in the foregoing derivation." It is quite hard to organize all the conditions together, though with them I have no doubts that the result is correct. (There is another statement of the same theorem, under unnecessarily severe conditions, on p. 282). Even the definition of a collective that I quoted at length above is not complete: for a few lines later a restriction, which never seems to be used, is placed on G. (The counterexamples cited above satisfy the restriction). Surely, in an exposition which purports to be rigorous the conditions should be explicitly stated in one place. It may be just a personal impression, but I never feel 'safe' with von Mises as I do with the better writers on pure mathematics. Always there is present the feeling that the proof is lacking in some detail, or is indeed incorrect. This arises from the defect of style. Actual errors are rare and misprints are commendably absent.

In Chapter 7 (39 pages) the study of statistics is begun. The first topic discussed is the problem of inference with a binomial distribution. The outlook is unashamedly Bayesian, though it is important to notice that the prior distribution is firmly based on frequency ideas: the urn or die being used is extracted from a collection of urns or dice. Obviously this makes the prior distribution less appealing than it would be if the prior probability were separated from the idea of repetition. One of von Mises's main tasks is to show that the final, posterior inference is unaffected, in the limit as the sample size increases, by the prior distribution; and there are several limit theorems of varying degrees of generality that express ideas of this sort either in the form of a weak law or in the form of a limiting posterior distribution. For example the first and simplest of them (curiously called Bayes's theorem, p. 340) states that "If the observation of an n-times repeated alternative shows a relative frequency r of 'success', then, if n is sufficiently large, the chance that the probability of success lies between $r - \epsilon$ and $r + \epsilon$ is arbitrarily close to one, no matter how small the ϵ". (The probability is the posterior probability and refers to the true success ratio.) The conditions for this to be true are, as explained above, omitted from the statement but included in the proof. Other theorems of this form are discussed but in almost all of them it is important to notice that the observation (in the example, r, the observed frequency) is held fixed. (See, for example, problem 9 on p. 348). Now this is not what happens in practice, where the parameter is fixed and the sample size tends to infinity, the observation fluctuating. One wants to be able to say that with probability one (according to the prior) the posterior probability will, in the limit, concentrate about the true value. This is a harder idea to formulate than von Mises's, and the wanted result much harder to prove. Personally it came as a revelation to me that his inference limit theorems were based on fixed r: in the original German I had failed to appreciate the point. It should be remembered that these results are prior to 1931 so that he has appreciably anticipated modern ideas. Even if not quite what one wants, they are important limit theorems.

Chapter 8 (63 pages) is concerned with rather more orthodox statistical ideas: variances of sample moments, χ^2-, t- and F-distributions and multivariate normal distributions. However, Section B on moments and inequalities is novel. This is a book on statistical theory and practical methods play a minor role, but one can question the few pieces of data that he uses. Thus problem 1 (p. 371) asks the reader to compute the mean and dispersion of the 15 figures which are the number of annual suicides in New York State in the years 1927–1941. This is not a good example since factors like the total population of the State in the periods are relevant. Again, on p. 433, as an illustration of a test for dispersion he uses the frequency of the letter a in Caesar's *Gallic Wars* where the successive letters are not independent. (On the other hand, to anticipate, the Bayesian treatment of the water-supply problem on p. 498 is excellent). There is one extraordinary statement: in discussing whether the sample sum of squares about the sample mean should be divided by n or $n - 1$ he sticks to n. "We see no reason to modify our definition". Had he never heard of Fisher's k-statistics or degrees of freedom? He smartly switches to $n - 1$ when it comes to the F-test (p. 446).

Analysis of Statistical Data is the title of Chapter 9 (63 pages). The bulk of the chapter is devoted to the χ^2-, ω^2- and Kolmogorov-Smirnov tests, but it begins with Lexis theory and the t- and F-tests. Von Mises's attitude to significance tests is disappointingly unclear. I think he has the attitude that many Bayesians have: namely that they are obviously very useful things to have around the place so we had better include them, but they don't fit into the Bayesian framework very well, so keep quiet about their basis. (This does not apply to the formal Neyman-Pearson theory: see below). Thus in the Lexis theory (and even with Student's t) he suggests statistics that might be useful, calculates their means and variances and compares the difference between observed and expected relative to the standard deviation. (He is most vague about the form of this comparison: e.g. p. 433). He never explains how this is good Bayesian statistics—if it is—and posterior or prior distributions never get a mention in this context. There are suggestions (on p. 495 and elsewhere) that the inference problem, where Bayes's formula is used, is distinct from the hypothesis testing problem: but I wish the distinction had been made sharper. Incidentally, the usual error about tail areas is made: "If the 'observed result' has a 'small' probability under the assumed hypothesis, we reject the hypothesis." (p. 441). It is the probability of the 'observed result and more extreme ones' that is used.

There is a delightful, brief passage of 12 lines on pp. 445–6. It begins, "Unfortunately, the Student test is often applied in an illegitimate way. In discussing the just-quoted experiments, one author concludes . . .", and ends, "the reasoning and its result are equally absurd." The anonymous author is Fisher, the reasoning is the fiducial argument. Only the truly great can call the great absurd!

The mathematical treatment of certain statistical problems is not without interest. Both in the discussion of χ^2, when parameters have to be estimated, and of the asymptotic behaviour of the maximum likelihood estimate, difficult situations to discuss with proper mathematical rigour, he uses proofs devised by van der Waerden. Although, in some places, the material seems outdated, in others it is modern. For example, the results of Chernoff and Lehmann on the asymptotic behaviour of the χ^2-statistic, and those of Robbins on empirical Bayes procedures, are mentioned.

Chapter 10 (72 pages) marks a return to the inference problem with a prior distribution. The three main topics discussed are the Neyman-Pearson theory of testing hypotheses, confidence limits and maximum likelihood estimation. Von Mises's treatment is always interesting and novel, but nowhere more so than in this chapter: at almost every point he makes some remark which illuminates the subject from an unusual and often rewarding angle. The chapter begins with some interesting inequalities for the posterior distribution in terms of the posterior distribution with a uniform prior. These do not have the objections I levelled against his earlier limit theorems. They express what Savage has much more recently called 'the principle of precise measurement'. The approach on p. 503 is close to the modern use of conjugate distributions. (This is attributed to H. A. Thomas, Jr., but the date is unclear). The Neyman-Pearson theory is presented with one eye on the prior distribution: thus the probabilities of errors (of either kind) are integrated with respect to the prior to give proper unconditional error rates. Similarly the treatment of confidence intervals expresses the idea that the statement made is one which is valid for any prior distribution. It is interesting to note that von Mises, even if he had been aware of it, would not have accepted some modern Bayesian criticism of confidence intervals or significance tests. I refer to the criticism that such ideas require the experiment under discussion to be imbedded in a sequence of similar experiments—in order that the appropriate repetitions be generated—and that this imbedding is unnatural and arbitrary, and also violates the likelihood principle. He could not agree since in order to obtain his frequency prior he has to do a similar imbedding, though in a different sequence. It is not therefore surprising to see these classical ideas in a Bayesian book by him.

Chapter 11 (49 pages) is rather dull after the excitement of the previous chapter. It deals with regression and correlation. Separated from least-squares theory and the analysis of variance, it seems sterile. It is most surprising that neither of these last two important topics find mention in a book of this size.

The final Chapter 12 (54 pages) is devoted to von Mises's own speciality, the theory of statistical functions. A statistical function is essentially a function of random variables. But with von Mises's reluctance to use random variables he prefers to refer to functions of distribution functions (either sample or population). Thus the sample mean is a statistical function either as a function of random variables or as a function of the sample distribution function. (Lest other readers be confused like I was, let me point out that the statements on p. 616 and elsewhere that a statistical function is a function of the probabilities $p_1, p_2, \cdots, p_k$ is not strictly correct: it also involves the values $x_1, x_2, \cdots, x_k$—as in the mean $\sum x_i p_i$—with which the probabilities are associated). Definitions of continuity and differentiability of statistical functions are given: these are used in a form of Taylor's theorem to obtain two types of laws of large numbers. The distinction between the two types is not at all well indicated by the notation though is explained in the text (p. 631 and elsewhere). They both refer to probability statements about the difference between the statistical function of the sample distribution function, $f\{S_n(x)\}$, and the same statistical function of the population distribution function, $f\{V(x)\}$. In the first type the former is the random variable—and the results are generalizations of the central limit theorem. In the second type the latter is the random variable—and the results are generalizations of what von Mises calls Bayes's theorem (see above). Those of the second type are statements of posterior probability. He says that without the concept of a collective the two types are not clearly distinguished (p. 631) but this is not so: they are distinguished by what is the random variable, $S_n(x)$ or $V(x)$. Most statisticians would currently prefer the approach of Mann and Wald, Chernoff and Pratt using generalizations of the o, O notations to o_p, O_p, "orders in probability", which is discussed in the book. Nevertheless von Mises's approach is interesting, earlier and important. In reading this book it is as well to bear in mind the dates of the original work: in this chapter the material dates from 1935.

The book concludes with a list of 35 selected reference books, a set of 11 statistical tables and an inadequate index.

In summary, this book seems to me to be most interesting, especially the parts on statistics which are at their best novel and important, and always attractively unusual. The parts on probability are less successful: the foundations seem dubious and the calculus is outdated. Defects of style make it seem to me to be unsuitable as a textbook: others may disagree. We should all be most grateful to Hilda Geiringer for having assembled this important material of a very great man in an attractive way, and in English.

REFERENCES

[1] Church, A. (1940). On the concept of a random sequence. *Bull. Amer. Math. Soc.* **46**, 130–135.

[2] Finetti, Bruno de (1937). La prévision: ses Cois Cogiques, ses sources subjectives. *Ann. de l'Institut Henri Poincare* **7**, 1–68.

[3] Kolmogorov, A. N. (1956). *Foundations of the Theory of Probability*. Chelsea Publ. Co., New York. (2nd English edition of a translation by Nathan Morrison from the original German)

[4] Kyburg, Henry E. Jr. and Smokler, Howard E. (1964). *Studies in Subjective Probability*. John Wiley and Sons Inc., New York.

[5] Mises, Richard von (1931) *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*. Leipzig and Vienna.

[6] Ville, Jean (1939). *Étude critique de la Nonon de collectif*. Gauthier-Villars, Paris. A part of the "Borel series", *Monographies des Probabilités*.