The Annals of Mathematical Statistics
1968, Vol. 39, No. 6, 2163–2167

BOOK REVIEWS

Correspondence concerning reviews should be addressed to the Book Review Editor, Professor James F. Hannan, Department of Statistics, Michigan State University, East Lansing, Michigan 48823

THOMAS S. FERGUSON, *Mathematical Statistics, A Decision Theoretic Approach.* Academic Press, New York, 1967. xi + 396 pp. $14.50.

Review by ROBERT A. WIJSMAN

Columbia University and University of Illinois

This is an excellent textbook that contains a large chunk of modern mathematical statistics in spite of its moderate size. It is intended for first-year graduate students, but will undoubtedly benefit a much wider audience. In fact, I think almost everybody can learn something new from it. For graduate students, mastering the material in the book and working all the exercises will probably go a long way toward preparation for the Ph.D. preliminary examination.

The author states in the Preface that as a rule he has included only topics that could be justified from a decision-theoretic viewpoint (the only exception being confidence sets). This certainly leaves out large chunks of specialized topics such as nonparametric, large sample, multivariate and analysis of variance (some basics of linear hypothesis is included, though). Some special tools, such as the Cramér-Rao inequality in estimation, are not included either. But the amount of material that is covered is impressive enough, including a whole chapter on testing hypotheses. Here is a list of chapter titles: Chapter 1: Game theory and decision theory; Chapter 2: The main theorems of decision theory; Chapter 3: Distributions and sufficient statistics; Chapter 4: Invariant statistical decision problems; Chapter 5: Testing hypotheses; Chapter 6: Multiple decision problems; Chapter 7: Sequential decision problems. No special chapter on estimation is included since many of the applications concern estimation. I think it is fair to say that what is covered in the book forms the basic core of statistical inference, and is the conceptually most difficult part of it.

For a text at the graduate level this book presents one unusual feature: it does not strive for complete mathematical rigor. The student is not expected to know measure theory and Lebesgue integration, and thus in certain theorems precise conditions and rigorous proofs are out of the question. This has both advantages and disadvantages. An advantage is that the essential features are not obscured by bothersome measure-theoretic difficulties. A drawback is that the student is left with many gaps to be filled in later. However, the author supplies a generous amount of references, giving the student ample opportunity to supplement the material in the book. Another slight drawback is that sometimes the precise conditions under which a theorem is valid may be lost out of sight. Two examples follow. In Section 1.8 on Bayes estimators the author very wisely states rules rather than theorems because conditions for the validity of the method are not given. Some of the properties to be implied by these conditions are listed, but one important one is left out: that $L(\theta, d(x))$ be measurable in (θ,x). As a second example consider Theorem 3.4.1 which states that given a sufficient statistic T and any decision rule, there is a decision rule dependent only on T with the same risk function. A student reading this theorem is unlikely to realize that the restrictive definition of sufficient statistic used in Chapter 3 places some restriction on the sample space. (If a general definition of sufficiency is used, a restriction has to be placed on the action space; see Bahadur, 1954.) Thus, this theorem may further spread the apparently rather widely held belief that the completeness theorem concerning sufficiency is valid without any conditions. However, even after all this has been said, I would still defend the author's choice of sacrificing rigor as a wise one, both from a pedagogical point of view, and from a practical one: to keep the book down to a reasonable size.

The organization of the material is excellent and the presentation very lucid. The arguments used in the proofs are often very concise, though, and occasionally the reader has to do some hard thinking to supply the missing steps. This is all the better, of course. Concerning the background expected of the student, although no measure theory is required, the student is expected to know what a sigma-field and what a null set is. It is likely that a student will get more out of the book if he is also familiar with measure and integration.

To the best of my knowledge, the only book to which Ferguson's is close in contents is the one by Blackwell and Girshick. Therefore, a comparison is in order. Compared to Ferguson's book, Blackwell and Girshick's has more on games, has more on sequential (e.g. examples of trichotomies), has a chapter on comparison of experiments and presents a mathematically rigorous treatment (except for their existence part in the proof of the optimum property of SPRT's). Compared to their book, Ferguson has more on invariance, has a chapter on testing, and does not restrict himself to discrete distributions as do Blackwell and Girshick. Also, in Chapter 6, a section is included on slippage problems. Blackwell and Girshick's book is somewhat more advanced but also a lot harder to read. Of course, Ferguson had the great advantage of having had Blackwell and Girshick's book as a spring board, and of having had the benefit of statistical work that appeared after the publication of Blackwell and Girshick's book. For instance, Stein's inadmissibility result for the usual estimator of the multivariate normal mean, and Bahadur's elegant proof of Wald's fundamental identity.

While in a textbook at this level one does not expect original contributions by the author, there remains on the part of the author the responsibility of selection, reworking and presentation. In this respect Ferguson has succeded very well indeed. Certain advanced topics are exposed now for the first time and in a simple way, e.g. Bahadur's notion of transitivity (in Section 7.3). Another topic that is notoriously hard to teach, the characterization of Bayes sequential rules, is also presented very clearly (Sections 7.2 and 7.5).

There are many other bonuses. The observation (Section 7.4) that invariant sequential rules can be derived in the same manner as Bayes sequential rules, provided the group is transitive over the parameter space, was new to me. Wald's example of sequential estimation of the mean of a uniform distribution is completely worked out here as a best invariant procedure. I was also unfamiliar with Kemperman's example (Section 7.7) of a family of continuous distributions for which the error probabilities and expected sample sizes of a sequential probability ratio test can be evaluated exactly. (Incidentally, there seems to be an error in formula (7.90), where $(1 - \theta)^2$ in the numerator should be $1 - \theta^2$.) The nonparametric problem of invariant estimation of a distribution function (Section 4.8) is another of the less standard topics. The pretty minimax and complete class theorems (Theorems 2.9.2 and 2.10.3) are no doubt suggested by LeCam's 1955 paper, but I believe not entirely a special case of the latter.

The author distinguishes carefully between two ways of randomization. If D is a class of nonrandomized decision rules, D^* (called class of randomized decision rules) consists of all probability distributions over D. More general is the class $\mathfrak{D}$ (called class of behavioral decision rules) with members δ, where for each x in the sample space X, $\delta(x)$ is a probability distribution over the action space $\mathcal{A}$. The author gives an excellent discussion of the distinction between D^* and $\mathfrak{D}$, and an indication of conditions when they are the same (with a reference to Wald and Wolfowitz). The question of $D^* = \mathfrak{D}$ is taken up again in Section 4.2, when D consists of invariant nonrandomized rules and it turns out that then it is the rule rather than the exception to have $\mathfrak{D}$ larger than D^* (this was new to me). The author derives a necessary condition for $D^* = \mathfrak{D}$, which under some additional regularity is also sufficient. It may be worth-while to restate this condition here in a different language. Let G_x be the isotropy group of G at a point $x \varepsilon X$ i.e. the subgroup of G that leaves x fixed. Similarly, G_a at a point $a \varepsilon \mathcal{A}$. Let $A_x = \{a \varepsilon \mathcal{A}: G_x \subset G_a\}$. Then in order for $\delta \varepsilon \mathfrak{D}$ to be equivalent to a member of D^* it is necessary that $\delta(x)$ puts all its probability on A_x. This is automatically true in either of two cases: G_x trivial for each $x \varepsilon X$ (the author calls this orbits of multiplicity one) and $G_a = G$ for each $a \varepsilon \mathcal{A}$ (i.e. the action of G on $\mathcal{A}$ is trivial).

With so much to praise it is only fair to offer a few criticisms. Chapter 5, Testing hypotheses, is perhaps least justified from the decision-theoretic viewpoint. At least, in my admittedly very subjective opinion, much of what goes on in hypotheses testing gets along quite well without decision theory. Still, there are connections, one of the simplest and most obvious being the fact that in testing a simple hypothesis against a simple alternative the Neyman-Pearson tests are Bayes procedures. Surprisingly, however, Ferguson makes hardly any mention of this. Another topic in testing hypotheses that has received considerable attention in recent years, and is not mentioned in the book, is the question of admissibility and minimaxity of certain popular tests, the main tool being to exhibit these tests as Bayes tests. The inclusion of this topic would have been most appropriate in the spirit of the book, and would have contributed to give the student some idea of how much Bayes decision rules are used as a tool in modern statistical research. Perhaps a simple example of this kind could be incorporated in a possible future edition of the book?

Section 1.4 ("Utility and subjective probability") I liked least. It is written least clearly and comes at a very early stage in the book. For any uninitiated student who thinks this section is vital to the understanding of the rest of the book, it must form a formidable barrier. Even if he can follow the mathematics he is likely to be at a loss to what is going on since in Sections 1.1–1.3 he has been told that the loss function L is numerical, so why now suddently bother with weird nonnumerical L? The situation becomes even more involved in the second half of section 1.4, where "horse lottery" is introduced as opposed to "roulette lottery" (the terminology is Anscombe and Aumann's). At this point I would advise any student who is persistent enough to pursue the matter to consult the original paper by Anscombe and Aumann. At a first reading of the book Section 1.4 could very well be skipped, and I think that the section should be labeled as such. (In fact, the author admits in the Preface that he tends to skip Section 1.4 in his lectures.)

The purpose of Section 1.4 is, I suppose, to show that if certain postulates on a preference pattern are accepted, then even if the payoffs are not numerical, the player acts as if they are, and preference is determined solely on the basis of expected utility. The theory of utility is extremely clever, but also deceptive because the postulates seem so reasonable (even after the author gives a critical discussion of the postulates they seem reasonable) but the consequence is so strong, and, in fact, not acceptable to everyone. Instead of emphasizing the role of utility, the author could simply have stated that the justification for evaluating the consequence of a decision function d as $EL(\theta, d(X))$ is mainly because of mathematical convenience, and partly because it seems quite reasonable in most cases.

There are a few more smaller things to which I would like to draw attention. On p. 156 it seems to me that together with Kiefer and Kudo should be mentioned Wesler (1959). I realize that the Hunt-Stein theorem is out of bounds since it would need measure theory, but should it not at least be mentioned?

The name "invariant decision rule" (Section 4.2) is customary but not very good. Strictly speaking, an invariant function f is constant on orbits: $f(gx) = f(x)$. This is true of invariant tests but not usually in the case of so-called invariant estimators. This causes confusion because when one uses the principle of invariance to derive decision rules, one also often utilizes auxiliary functions that are truly invariant. It would therefore be better to observe the distinction. Now a so-called invariant decision function (nonrandomized) d is defined as $d(gx) = g[d(x)]$, i.e. d commutes with each $g \varepsilon G$. Transformation group mathematicians call such functions *equivariant*, and I propose that the nomenclature equivariant decision rules be used. Thus, in the translation parameter problem, having observed $X = (X_1, \cdots, X_n)$, we may write $d(X) = X_1 + b(X)$, where b is invariant and d is equivariant. Of course, in the case of hypotheses testing, where the group acts trivially on the action space, equivariant and invariant are the same.

Theorem 6.3.1 needs no proof. At least I think the result follows immediately from Theorem 2.3.2. Why Section 7.5 is done only for testing a simple hypothesis against a simple alternative is not clear to me, since most of it carries over, practically unchanged, to arbitrary Θ. The statment of Theorem 7.6.2 (the optimum property of SPRT's) is slightly incomplete since it does not exclude the possibility that another test has smaller error probabilities and the same expected sample sizes, which would make the SPRT inadmissible. That this actually does not happen can be stated by saying that if there is at least one strict inequality among the error probabilities, there is at least one among the expected sample sizes.

The book contains a large number of exercises; in fact, almost every section is followed by a list of problems. I enjoyed working a good number of them and encountered several that I found quite challenging. Exercise 7.7.9 (assuming that the common value of w_{01} and w_{10} is intended to be 1) I could only do by solving a transcendental equation numerically. If this is indeed the intention of the author, the students using his book would undoubtedly appreciate that hint. There is a relatively small number of minor misprints in the text, most of which the reader will have no difficulty spotting. One answer to an exercise seems to be in error: the formula for $E(N \mid H_1)$ in Exercise 7.7.15, which can be written in the form $f(j, k)/g(j, k)$. According to my computations the result is $[f(j, k) - f(-j, -k)]/[g(j, k) - g(-j, -k)]$. That the formula given in the book cannot be right can be checked by taking $j = k = 1$, in which case $N \equiv 1$ but the formula gives 5/3.

In conclusion, I expect Ferguson's book to be popular among all students (in the wide sense) of statistics for many years to come.