The Annals of Mathematical Statistics
1972, Vol. 43, No. 2, 697-700

BOOK REVIEW

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.

BECHHOFER, R. E., KIEFER, J. AND SOBEL, M. *Sequential Identification and Ranking Procedures* (*with special reference to Koopman-Darmois populations*). The University of Chicago Press, Chicago and London, 1968. xvii+420 pp. $17.50.

Review by SHANTI S. GUPTA
Purdue University

In many practical situations the experimenter is confronted with the problem of choosing the best (or t best) from a group of k populations (categories, processes, etc.). The classical tests of homogeneity, though long used in these situations, did not adequately answer the real questions. A more meaningful formulation was provided in response to this need in the form of what is now commonly known as multiple decision or ranking and selection procedures. Since the early investigations of Bechhofer [*Ann. Math. Statist.* **25** (1954), 16–39], who considered the normal means problem under a formulation now commonly referred to as the indifference zone approach, many authors have contributed to various aspects of the problem with different modifications in the goal. Most of the early endeavors were devoted to single stage procedures. Some investigations were made relating to two-stage procedures. The initial efforts in the area of sequential procedures by two of the present authors date back to the mid-fifties. With challenging questions arising one after another and the enthusiasm of three very competent mathematical statisticians to match them, it is no surprise that what was intended to be a short paper grew to become the present monograph. It is true that during the years of preparation of this monograph some papers have been published in the area of sequential ranking procedures, but there has been practically no overlap and the importance of the contribution made by this work has in no way been diminished.

The basic problem investigated by the authors can be briefly described as follows. Suppose there are k populations with parameters $\theta_1, \theta_2, \cdots, \theta_k$. For an identification problem, it is assumed that the θ_i are known but not the true pairing of the populations and the ordered θ_i which are denoted by $\theta^0_{[1]} \leqq \theta^0_{[2]} \leqq \cdots \leqq \theta^0_{[k]}$. Assuming that $\theta^0_{[k-1]} < \theta^0_{[k]}$, a possible identification goal would be "to identify the population associated with $\theta^0_{[k]}$." For a ranking problem, it is assumed that the θ_i are unknown and that the experimenter has no prior knowledge relevant to the true pairing of the populations and the ordered θ_i which are now denoted by $\theta_{[1]} \leqq \theta_{[2]} \leqq \cdots \leqq \theta_{[k]}$. Two goals are considered: Goal I is to select an unordered set of t populations associated with the set $\{\theta_{[k-t+1]}, \theta_{[k-t+2]}, \cdots, \theta_{[k]}\}$ and Goal II is to select a population associated with $\theta_{[k-t+1]}$, a population associated with $\theta_{[k-t+2]}, \cdots$, and a population associated with $\theta_{[k]}$. The statement of a ranking problem is not complete without specifying certain constants and an associated probability requirement. To be precise, it is required that the probability of a correct selection (i.e., a selection of t populations satisfying the goal) is at least a preassigned number P^* $(\binom{k}{t}^{-1} < P^* < 1)$ whenever $\varphi_{k-t+1,k-t} \geqq \varphi^*{}_{k-t+1,k-t}$ in the case of Goal I and $\varphi_{k-i+1,k-i} \geqq \varphi^*{}_{k-i+1,k-i}$ $(i = 1, \cdots, t)$ in the case of Goal II where $\varphi_{i,j}$ denotes the distance (suitably defined) between the populations associated with $\theta_{[i]}$ and $\theta_{[j]}$ $(i < j)$ and the starred quantities are positive constants specified in advance.

The first chapter of the monograph gives a somewhat informal description of the subject and the approach adopted in the presentation of it. The remaining seventeen chapters have been organized into three parts. Part I (Chapters 2–9) embodies an analytical treatment of the general theory whereas Part II (Chapters 10–16), written primarily for the benefit of the practitioners, contains applications of the general theory to ranking of Koopman-Darmois (D-K) parameters with special reference to the normal means. Part III describes some Monte Carlo sampling results for the normal means problem with a common known variance.

Part I discusses a sequential procedure (Chapter 3) which is a generalization of Wald's Sequential Probability Ratio Test (WSPRT) to multiple-decision problems and which coincides with WSPRT for the case of $k = 2$ populations. For sequential procedures of a certain type useful upper bounds are obtained for the expected loss and a method is given for obtaining Monte Carlo sampling estimates of the expected loss. For sequential stopping rules of a fairly general form (including the rules discussed later in detail) the certainty of termination and the finiteness of moments of n (the number of stages to terminate) are established. A general lower bound for the ASN, namely, the expected value of n, has been obtained by using a result of Hoeffding. Chapter 3 also presents a good discussion on the algebraic structure of identification problems. Though the identification problems are of interest in themselves, the emphasis is on the insight they provide into the solution of a class of ranking problems. The authors have obtained a sufficient condition (called the rankability condition) for $k \geqq 2$ and, a necessary and sufficient condition for $k = 2$ such that the solution of an identification problem leads to a solution of the corresponding ranking problem.

Two basic sequential procedures are defined; $\mathscr{P}_B$, an identification procedure (Chapter 4) and $\mathscr{P}_B{}^*$, a ranking procedure (Chapter 5). These two procedures are very fundamental to the subsequent developments in the monograph. A number of specific identification problems and the corresponding ranking problems are discussed. A substantial part of the discussion is devoted to some special cases of K-D family of populations, namely, (A) normal populations with unknown means, (B) normal populations with unknown variances, (C) exponential populations with known location parameter and unknown scale parameter, (D) binomial, (E) Poisson and (F) negative binomial distributions.

The main analytical results concerning $\mathscr{P}_B$ and $\mathscr{P}_B{}^*$ are given in Chapter 6. These results relate to K-D populations and include (i) lower bounds for the probability of a correct selection (PCS) for $\mathscr{P}_B{}^*$ under Goals I and II, (ii) lower bounds for the ASN under Goal I when $t = 1$, (iii) the behavior of the PCS- and ASN-functions as the specified δ^*'s (which are φ^*'s for a particular choice of distance function) tend to zero and (iv) the behavior of the ASN as P^* approaches one. These results provide the necessary insight into the "large sample" properties of $\mathscr{P}_B{}^*$ and lead, in the case of ranking the normal means, to a study of the asymptotic efficiency as $P^* \to 1$ of $\mathscr{P}_B{}^*$ relative to the best single-stage procedure guaranteeing the same probability requirement. For $k > 2$, it is shown that $\mathscr{P}_B$ and $\mathscr{P}_B{}^*$ are not in general optimal for a fixed P^*, but are asymptotically optimal as $P^* \to 1$. In the K-D cases (B) through (F) a single-stage procedure guaranteeing the probability requirement does not exist. It is shown that two-stage procedures can be constructed for ranking normal variances or exponential scale parameters but that L-stage procedures ($2 \leqq L < \infty$) cannot be constructed for ranking Bernoulli, Poisson or negative binomial parameters.

Other important results of Part I include (i) exact results for the PCS- and ASN-functions associated with $\mathscr{P}_B$ in the case of certain K-D populations and the Weiner Process when $k = 2$, and in the case of the Poisson processes when $k \geqq 2$ (Chapter 7) and (ii) analytic approximations and bounds for these functions under Goal I when $k = 2$ and $t = 1$ (Chapter 8).

The first part ends with a discussion of several modifications that can be adopted in the basic procedure. A simplification of $\mathscr{P}_B{}^*$ is described. This is achieved at the cost of increasing the ASN which can be justified in situations where observations are relatively inexpensive and complex computations are relatively costly. Another modification is truncation. In the case of the normal means, this helps partially overcome the problem of a large value of the ASN under the EM (equal means) configuration (i.e., $\theta_{[1]} = \cdots = \theta_{[k]}$), especially when P^* is close to unity. It is shown that $\mathscr{P}_B{}^*$, truncated at $-4 \log(1 - P^*)$ stages, asymptotically minimizes the ASN both for the LF (least favorable) and the EM configurations. Some applications of the technique of simultaneous tests for the construction of modified procedures are also described. The basic procedures discussed in the present work are of the non-eliminating type in which observations are taken from each population until the terminal decision is made. In this context the authors have discussed some procedures with elimination by Stein [*Ann. Math. Statist.* **19** (1948) 429] and Paulson [*Ann. Math. Statist.* **35** (1964) 174–180]. However, the main emphasis in the monograph is on untruncated procedures without elimination since the authors feel that sampling from all populations at each stage is often reasonable in practice and makes it easy to obtain important quantitative properties.

As remarked earlier, Part II is intended to be a somewhat technical handbook for the benefit of the practitioner, who is sure to profit by the self-contained presentation of the material. The authors have also presented in this part some material complementing that of Part I. The main discussions of this part include the application of the sequential ranking approach under Goal I with $t = 1$ to the several special cases of K-D family mentioned earlier (Chapter 12) and a brief consideration of the problem under Goal II in the special case of $k = 3$, $t = 2$ (Chapter 13). The basic ranking procedure for Goal I and the implications of several theorems providing lower bounds and approximations for the PCS, the ASN and the relative efficiency are discussed (Chapter 14). Some special analytical results for the PCS- and ASN-functions under Goal I are obtained (Chapter 15) in the special case of $k = 3$, $t = 2$. With the objective of stimulating further theoretical research, the authors have provided a list of unsolved problems (Chapter 16) with critical comments on them.

Part III describes some Monte Carlo sampling results for the problem of ranking the means of normal populations with a common known variance. A fairly updated bibliography is given at the end which is followed by a glossary of symbols and a subject index.

The monograph is not of the easy-to-read type, partly due to the complicated notation that is concomitant with the nature of the subject matter. Though the authors feel that the monograph appears at a stage of development earlier than is customary in the literature of a subject, they have done an essential and timely job in bringing together a large body of results obtained well over a decade. There is no doubt in the mind of this reviewer that this monograph will serve as a useful reference to the research worker in this area and as a handy manual for the applied statistician and the experimenter who should find increasing use for these procedures.