Partial Prior Information and Decisionmaking

JOHN M. POTTER, STUDENT, IEEE, AND BRIAN D. O. ANDERSON, FELLOW, IEEE

Abstract—**In a decisionmaking situation if prior information is vague, indecision may result. A model for partial prior information is established and from this arises the notion of a set of feasible decisions. The existence of more than one feasible decision represents the existence of indecision. In many practical situations the partial prior information may be given by linear inequalities on the prior probabilities in which case there is a computationally attractive method for determining the set of feasible decisions.**

I. INTRODUCTION

IN THE STANDARD statistical decisionmaking procedures, certain assumptions are made on the availability of prior information. Thus the Bayesian decisionmaking procedures [1]–[3] assume the existence *and* precise knowledge of the prior probabilities of certain events and use these probabilities in making the decisions; on the other hand, some other statistical inference procedures such as maximum likelihood or the Neyman–Pearson approach [1], [3], [4], make no use at all of any prior information available and base their inferences solely on experimental results. A degree of undesirability is inherent in both approaches as was implied by Pearson in [5] who said, "We [Neyman and Pearson] were certainly aware that inferences must make use of prior information... it was so rarely possible to give sure numerical values to these entities that our line of approach must proceed otherwise." This paper is aimed at a compromise between maximum likelihood and similar approaches on the one hand and Bayesian approaches on the other. Prior information is available and is made use of, but unlike the Bayesian approach, precise knowledge of the prior probabilities is not necessary.

As in [6], the partial prior information to be considered in this paper is taken as the knowledge of a set of prior probabilities $\mathcal{P}$, it being assumed that the true prior probability is a member of $\mathcal{P}$. With a little thought it is not difficult to think of a whole gamut of realistic examples where the actual prior information available is equivalent to the specification of $\mathcal{P}$, so this does appear to be a sensible model for the prior information. (In contrast, other attempts at defining a model for partial prior information [7], [8], do not seem to be so widely applicable). Often the prior information arises from qualified assignments or comparisons of prior probabilities—such a situation tends to lead to $\mathcal{P}$ being defined by linear inequalities, and, as will be shown in Section III, this leads to the use of linear programs in the decisionmaking process formulated in this paper.

In Section II the framework for the standard Bayesian decisionmaking procedure is reviewed and the decision criteria written in a novel manner appropriate for use in later sections. The first innovation of this paper is made in Section III where it is shown how the Bayesian procedure may be generalized to allow for partial prior information. The generalized procedure is computationally effective, at least in many situations of interest. Briefly the idea is to make a decision using standard criteria for each prior probability in $\mathcal{P}$ and to call the combination of these decisions the set of "feasible" decisions. With partial prior information it may not be possible to choose one decision as the best one; rather, all that may be possible is to state that any of a number of feasible decisions could be the best one. On reflection it can be seen that this does indeed model real world decisionmaking processes: if information is vague, it is probable that some indecision will result, with the degree of vagueness affecting the degree of indecision. An hypothesis testing example is developed in Section IV, serving to illustrate the concepts of partial prior information and sets of feasible decisions as they could arise in a practical situation. Section V looks in detail at minimum probability of error (MPE) (or maximum *a posteriori* (MAP)) hypothesis testing and shows that for particular types of prior information, the calculations involved in finding the set of feasible decisions are very simple. To remove indecision which arises with a number of feasible decisions, it is reasonable to suggest taking further observations in the hope that they will provide more information—this idea is investigated in Section VI.

Other workers have dealt with the idea of partial prior information [6]–[8]; an excellent summary of their suggested decisionmaking procedures is available in [1]. Furthermore there have been many papers written on the assignment of prior probabilities according to various criteria; the approaches encountered include minimax [1], [2], [6], maximum entropy [9], [10], and minimax average uncertainty [11], [12]. Two criticisms may be leveled at these approaches: one is that some of the approaches do not always agree with a common-sense evaluation of a real decisionmaking process; the other is that the calculations involved are usually rather difficult. Another ap-

Manuscript received August 15, 1979; revised November 16, 1979. This work was supported by the Australian Research Grants Committee.

The authors are with the Department of Electrical Engineering, University of Newcastle, Newcastle, New South Wales, 2308, Australia.

0018-9472/80/0300-0125$00.75

proach to decisionmaking under uncertainty uses the concept of fuzziness [13], [14]; it would be interesting to consider the approach of the present paper, where uncertain prior information is modeled as the specification of a set of prior probabilities, when the costs are considered to be fuzzy—it has been suggested in [15] that this sort of model, with probabilistic prior information and fuzzy cost information, is more realistic than other existing decisionmaking models.

Most of the results of this paper have been recorded in [16].

II. Review of the Bayesian Decision Rule

This section is devoted to introducing notation, reviewing the standard Bayesian decision rule [1]–[3], and reformulating this rule in a way which emphasizes the linear dependence of the rule on the prior probabilities.

The basic task is to make a decision d_k out of a set M of possible decisions, $M=\{d_1,d_2,\cdots,d_m\}$. This decision is to be based on an observation x which depends on a set N of parameters, $N=\{\theta_1,\theta_2,\cdots,\theta_n\}$; the dependence of the observation on the parameters is characterized by the known conditional probabilities $p_j(x)$ = probability of observing x when θ_j is the true parameter, $j=1,\cdots,n$. (If the set of all possible observations is a continuum, we consider the $p_j(x)$ to be probability densities). Each of the parameters in N occur with a probability P_j, $j=1,\cdots,n$; the prior probability $\boldsymbol{P}$ is written as $\boldsymbol{P}=(P_1,P_2,\cdots,P_n)^T$. If $\boldsymbol{P}$, $p_j(\cdot)$, $j=1,\cdots,n$, and x are known, the posterior probability $Q(x)=(Q_1(x),\cdots,Q_n(x))^T$ may be evaluated; it is given by

$$Q_j(x)=\frac{p_j(x)P_j}{\sum_{l=1}^{n}p_l(x)P_l}$$

and is the conditional probability of θ_j being the true parameter, given x.

In order to make a meaningful decision, there must be some criterion to judge whether a decision is "good" or not. Within a Bayesian framework, this is accomplished by assigning a matrix $C=[c_{ij}]$ of costs; here c_{ij} is the cost of choosing decision $d_i\in M$ when $\theta_j\in N$ is the true parameter. The ith row of C is the vector of costs associated with decision d_i, written as $\boldsymbol{c}_i^T$. Now the Bayesian decision rule is simply to choose that decision which minimizes the expected value of the cost, given the observation. So the rule is to choose k such that

$$\sum_j c_{kj}Q_j(x)\leqslant\sum_j c_{ij}Q_j(x),\qquad\text{for all } i$$

i.e.,

$$\boldsymbol{c}_k^TQ(x)\leqslant\boldsymbol{c}_i^TQ(x),\qquad\text{for all } i. \tag{1}$$

Note that the decision rule does not always choose a decision uniquely since there may be equality of expected cost for different decisions; this slight difficulty is automatically catered for in the procedure to be developed in the next section. The most common Bayesian decisionmaking procedure occurs when $M=N$ and $c_{ii}=0$, $c_{ij}=1$, $i\neq j$. This is an hypothesis testing case where the decision is to select *a posteriori* that hypothesis with the smallest probability of error—hence it is called the minimum probability of error (MPE) hypothesis testing case; this will be looked at further in Section V.

To emphasize the linear dependence of the Bayesian decision rule (1) on the prior probability $\boldsymbol{P}$, it is reformulated as

$$D_k(x)\boldsymbol{P}\geqslant\boldsymbol{0} \tag{2}$$

where the $m\times n$ matrix $D_k(x)$ is defined by

$$D_k(x)=\left(C-\mathbf{1}\boldsymbol{c}_k^T\right)\operatorname{diag}\{p_j(x)\} \tag{3}$$

(where $\mathbf{1}$ is an m-vector with all elements equal to 1).

III. Decisions with Partial Prior Information

As explained in Section I, with only partial prior information available it is conceivable that even after the observation x is made, there may not be enough information to justify choosing a single decision; all that may be possible is for some decisions to be ruled out with the remainder left as "feasible" decisions. In a Bayesian decisionmaking framework, this notion may be quantified. To achieve this the partial prior information must be specified as a set $\mathcal{P}$ of feasible prior probabilities. Given $\mathcal{P}$ and an observation x, it follows that the set $K=K(x,\mathcal{P})$ of feasible decisions may be found (in theory). Each prior probability $\boldsymbol{P}$ in $\mathcal{P}$ provides a feasible decision, and the totality of these feasible decisions $K(x,\mathcal{P})$ is therefore given by

$$K=\bigcup_{\boldsymbol{P}\in\mathcal{P}}\{d_k\colon D_k(x)\boldsymbol{P}\geqslant\boldsymbol{0}\}. \tag{4}$$

A straightforward consequence of (4) is the fact that $\mathcal{P}_1\subset\mathcal{P}_2$ implies $K(x,\mathcal{P}_1)\subset K(x,\mathcal{P}_2)$; this therefore models the intuitive notion that vaguer information results in greater indecision.

In practice it is not clear that K may actually be evaluated since the union is taken over $\mathcal{P}$ which may be a continuum. However, in many practical situations, $\mathcal{P}$ could well be specified by linear inequalities: such common statements as θ_1 is ten times more likely than θ_2, θ_1 is less likely to occur than θ_2 and θ_3, and θ_1 is odds-on *can all be written as linear inequalities on the prior probabilities.* So suppose the partial information is given as

$$\mathcal{P}=\{\boldsymbol{P}\colon A\boldsymbol{P}\geqslant\boldsymbol{0},\ \mathbf{1}^T\boldsymbol{P}=1,\boldsymbol{P}\geqslant\boldsymbol{0}\} \tag{5}$$

(linear inequalities of the form $B\boldsymbol{P}\geqslant\boldsymbol{b}$ are easily reduced to the form $A\boldsymbol{P}\geqslant\boldsymbol{0}$, using the normalizing relationship $\mathbf{1}^T\boldsymbol{P}=1$). Then the feasible decisions are given by (s.t. means subject to)

$$\begin{aligned}&d_k\in K(x,\mathcal{P})\\&\quad\Leftrightarrow\text{ there exists }\boldsymbol{P}\in\mathcal{P}\quad\text{s.t.}\quad D_k(x)\boldsymbol{P}\geqslant\boldsymbol{0}\\&\quad\Leftrightarrow\text{ there exists }\boldsymbol{P}\geqslant\boldsymbol{0}\quad\text{s.t.}\quad A\boldsymbol{P}\geqslant\boldsymbol{0}\\&\qquad\qquad\qquad\qquad\quad\ \mathbf{1}^T\boldsymbol{P}=1\\&\qquad\qquad\qquad\qquad\quad\ D_k(x)\boldsymbol{P}\geqslant\boldsymbol{0}.\end{aligned} \tag{6}$$

Clearly the test in (6) is, for each k, simply a test for a feasible solution to a set of linear inequalities which can be done with standard linear programming techniques.

In summary, a theoretical expression for the set of feasible decisions K may be found for arbitrary partial prior information $\mathcal{P}$. In practice $\mathcal{P}$ could well be given by linear inequalities, in which case K may be found by looking exhaustively at all $d_k \in M$ and solving a linear program for each k. It has been shown elsewhere [16] that the amount of computation required to evaluate K can be substantially reduced by taking advantage of the manner in which the set of inequalities $D_k(x)\boldsymbol{P} \geq \mathbf{0}$ partitions the prior probability simplex[1] and the fact that the other inequalities in (6) are independent of k; this point will be illustrated in the next section. In Section V it is shown that K can be found very efficiently for MPE testing with particular types of partial prior information. A subsequent paper will provide other particular cases where K can be found easily.

IV. An Example

Consider the following situation which could arise in a medical diagnosis problem. A patient exhibits certain symptoms which suggest to a diagnostician that the patient has probably contracted Disease 1. However the symptoms are also consistent with two rarer diseases, Disease 2 and Disease 3. The diagnostician is prepared to say that the chances of it being Disease 3 are at least one in ten thousand and that Disease 2 is more likely than Disease 3. In an attempt to differentiate between the diseases, he runs a white blood cell count x (number of cells/mm^3), where he knows that x is a Gaussian random variable with mean μ_j and variance σ_j^2 depending on the particular disease as specified in Table I.

This problem can be formulated in the framework of Section III. We have the parameter set $N = \{$Diseases 1, 2, and 3$\}$. Now $N = M$, the set of decisions, since the diagnostician's task is to decide which disease the patient actually has. The partial prior information which the diagnostician has, can be written as

$$\mathcal{P} = \{P\colon P_1 \geq 0.5, P_3 \geq 0.0001, P_2 \geq P_3, P_1 + P_2 + P_3 = 1\}.$$

(See Fig. 1. Note that the constraint $P_3 \geq 0.0001$ is indistinguishable in the figure from $P_3 \geq 0$). The decision criterion to be used is to choose the most probable disease given the white cell count x, e.g., Disease 1 is chosen if $Q_1(x) \geq Q_2(x)$ and $Q_1(x) \geq Q_3(x)$—this is just MPE testing. If we define the kth decision region $\mathcal{D}_k(x)$ as $\{\boldsymbol{P}\colon D_k(x)\boldsymbol{P} \geq \mathbf{0}, \mathbf{1}^T\boldsymbol{P} = 1, \boldsymbol{P} \geq \mathbf{0}\}$, $k = 1,2,3$, then given an observation x, the *a priori* probability simplex is partitioned by the three decision regions with the general layout depicted in Fig. 2. Suppose that the patient has 5000 white cells per mm^3 of blood ($x = 5000$), whence $p_1(x) = 5.40 \times 10^{-5}$, $p_2(x) = 9.13 \times 10^{-5}$, $p_3(x) = 0.23 \times 10^{-5}$. Now, with partial prior information $\mathcal{P}$, it is clear from (6) that a decision d_k is feasible if and only if $\mathcal{P}$ and $\mathcal{D}_k(x)$ have nonempty intersection. From Fig. 3, which superimposes $\mathcal{P}$ and the partitioning into decision regions associated with $x = 5000$, it is easily seen that $K(x, \mathcal{P}) = \{$Diseases 1 and 2$\}$. A list of feasible decisions corresponding to any white cell count is given in Table II; this table is not difficult to check if Theorem 4 (Section VI) is used.

This example illustrates how the framework of Section III is to be used in practice; it also serves to emphasize what is really a "feasible decision." When it is said that

[1] Redundancies in the constraints $D_k(x)\boldsymbol{P} \geq 0$ (if there are any) may be removed by adopting the same approach as that taken in an appendix of [17].

TABLE I
Means and Variances for White Cell Counts

Disease j	μ_j (cells/mm^3)	σ_j (cells/mm^3)
1	3,000	1,000
2	7,500	2,000
3	16,000	4,000

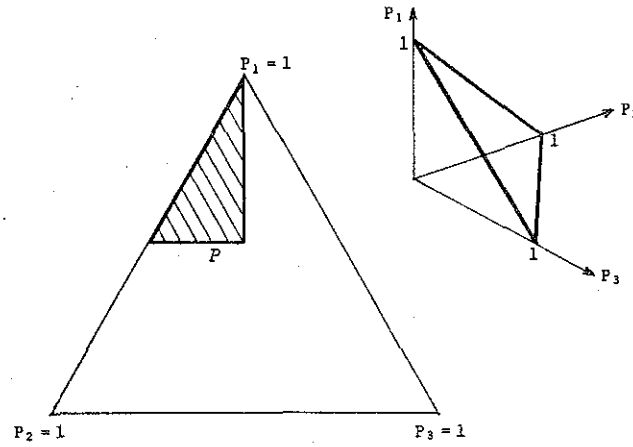

Fig. 1. Illustration of $\mathcal{P}$ in the 3-D probability simplex.

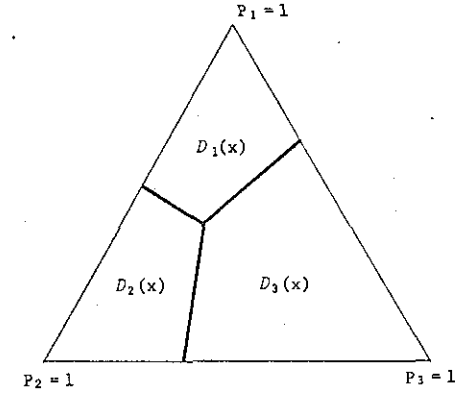

Fig. 2. Partition of probability simplex into decision regions.

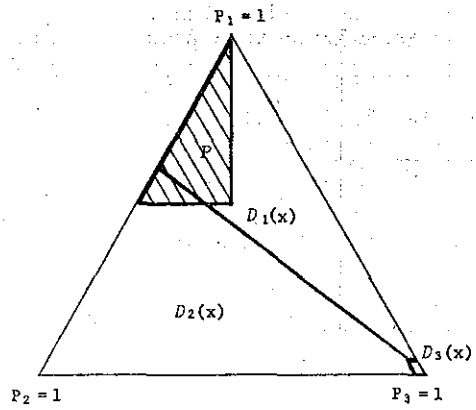

Fig. 3. Decision regions when $x=5000$ cells/mm^3, showing that $K(x,\mathcal{P})=$ {Diseases 1 and 2}.

TABLE II
SET OF FEASIBLE DECISIONS FOR DIFFERENT WHITE CELL COUNTS

Count x (cells/mm^3)	Feasible decisions K(x, P)
x < 4,500	{Disease 1}
4,500 ≤ x ≤ 7,293	{Diseases 1 and 2}
7,293 < x < 10,333	{Disease 2}
10,333 ≤ x ≤ 15,755	{Diseases 2 and 3}
15,755 < x	{Disease 3}

Disease 3 is not feasible, it does not mean that it is not possible—merely that it was less probable (*a posteriori*) than at least one of Disease 1 or 2, no matter what the prior probability was in $\mathcal{P}$. As far as the diagnostician is concerned the set of feasible decisions will, in practice, merely serve to guide the direction of future diagnostic tests (e.g., by "eliminating" Disease 3 from contention).

An understanding of the graphical approach to finding $K(x,\mathcal{P})$, introduced in this example (Fig. 3), can provide valuable insight into the idea of partial prior information and the consequent set of feasible decisions. For example, it is clear from Fig. 3 that $K(x,\mathcal{P})$ will be unchanged for relatively large changes of $\mathcal{P}$, for the given x; i.e., the figure gives an idea of the sensitivity of the decisions to the prior assignment of $\mathcal{P}$.

V. MPE Hypothesis Testing

It was noted in Section II that the most common application for the decisionmaking procedure which was outlined there arises when $M=N$ and $c_{ii}=0$, $c_{ij}=1$, $i\neq j$, which results in minimum probability of error (MPE) hypothesis testing. In this section it will be shown that for particular forms of partial prior information which may reasonably be expected to occur in many practical situations, the set of feasible decisions (or hypotheses as they will be called now) may be found in an extremely efficient manner. With MPE testing and partial prior information $\mathcal{P}$, the set of feasible hypotheses K is given by

$$K=\bigcup_{P\in\mathcal{P}}\{\theta_k: p_j(x)P_j\leq p_k(x)P_k, \quad \text{for all } j\}.$$

Two types of partial prior information, labeled $\mathcal{P}_1$ and $\mathcal{P}_2$, and the corresponding methods for finding K will now be developed separately.

A. $\mathcal{P}_1$-Type Information

Suppose the "odds" for θ_1 being true are known to lie between 1 in 10 and 1 in 1000; then $10^{-3}\leq P_1/(1-P_1)\leq 10^{-2}$ or $1/1001\leq P_1\leq 1/11$. In general any such "odds" information can be converted into upper and lower bound constraints on the prior probabilities; in other situations the bounds may be given directly. So the first type of partial prior information to be considered is given in the general form:

$$\mathcal{P}_1=\{\boldsymbol{P}: \mathbf{0}\leq\boldsymbol{a}\leq\boldsymbol{P}\leq\boldsymbol{b}\leq\mathbf{1}, \mathbf{1}^T\boldsymbol{P}=1\}, \tag{7}$$

where the consistency condition, $\mathbf{1}^T\boldsymbol{a}\leq 1\leq\mathbf{1}^T\boldsymbol{b}$, must of course be satisfied.

First a test for the feasibility of one particular hypothesis θ_k is stated, followed by a characterization of K which allows the description of a simple method for obtaining K, given partial prior information $\mathcal{P}_1$ and an observation x. By defining

$$\begin{aligned}\alpha_i&=\max\left\{a_i, 1-\sum_{j\neq i}b_j\right\}\\ \beta_i&=\max\left\{b_i, 1-\sum_{j\neq i}a_j\right\}\end{aligned} \tag{8}$$

it is simple to verify that the constraints for $\mathcal{P}_1$ as given by (7) are reduced in a redundancy-removal step to

$$\mathcal{P}_1=\{\boldsymbol{P}: \boldsymbol{\alpha}\leq\boldsymbol{P}\leq\boldsymbol{\beta}, \mathbf{1}^T\boldsymbol{P}=1\}. \tag{9}$$

Lemma 1: Given $\mathcal{P}_1$ as in (9), $\boldsymbol{p}(\cdot)$ and x, define $\boldsymbol{\rho}^k$ by

$$\begin{aligned}\rho_j^k&=\min\{\beta_j, p_k(x)\beta_k/p_j(x)\}, \quad \text{for } p_j(x)>0\\ &=\beta_j, \quad \text{for } p_j(x)=0.\end{aligned}$$

Then hypothesis θ_k is feasible (i.e., $\theta_k\in K$)

$$\Leftrightarrow \boldsymbol{\rho}^k\geq\boldsymbol{\alpha} \quad \text{and} \quad \mathbf{1}^T\boldsymbol{\rho}^k\geq 1.$$

Proof: a) Necessity: suppose $\theta_k\in K$. Then there exists $\boldsymbol{P}$ subject to $\boldsymbol{\alpha}\leq\boldsymbol{P}\leq\boldsymbol{\beta}$, $\mathbf{1}^T\boldsymbol{P}=1$ and $p_j(x)P_j\leq p_k(x)P_k$, for all j. From the definition of $\boldsymbol{\rho}^k$ it follows that $\boldsymbol{P}\leq\boldsymbol{\rho}^k$, whence $\boldsymbol{\alpha}\leq\boldsymbol{\rho}^k$ and $\mathbf{1}^T\boldsymbol{\rho}^k\geq 1$. b) Sufficiency: suppose $\boldsymbol{\rho}^k\geq\boldsymbol{\alpha}$ and $\mathbf{1}^T\boldsymbol{\rho}^k\geq 1$. If $\mathbf{1}^T\boldsymbol{\rho}^k=1$, set $\boldsymbol{P}=\boldsymbol{\rho}^k$. Otherwise, set

$$\boldsymbol{P}=\left[(\mathbf{1}^T\boldsymbol{\rho}^k-1)\boldsymbol{\sigma}^k+(1-\mathbf{1}^T\boldsymbol{\sigma}^k)\boldsymbol{\rho}^k\right]/\mathbf{1}^T(\boldsymbol{\rho}^k-\boldsymbol{\sigma}^k)$$

where $\sigma_k^k=\beta_k$, $\sigma_j^k=\alpha_j$, for all $j\neq k$. [One may verify using (8) $\mathbf{1}^T\boldsymbol{\sigma}^k\leq 1$, so that $\boldsymbol{P}$ is well-defined and is a convex combination of $\boldsymbol{\sigma}^k$ and $\boldsymbol{\rho}^k$]. It is now easy to check that $\boldsymbol{P}\in\mathcal{P}$ and $p_j(x)P_j\leq p_k(x)P_k$, for all j, whence hypothesis θ_k is feasible.

The following theorem characterizes K; the conditions for the theorem are as for Lemma 1 (including the defini-

tion of ρ^k), except that it is assumed for notational convenience and with no loss of generality that $p_1(x)\beta_1 \geq p_2(x) > \beta_2 \geq \cdots \geq p_n(x)\beta_n$.

Theorem 2: Find the largest k subject to

$$p_k(x)\beta_k \geq \max_i p_i(x)\alpha_i.$$

If $\mathbf{1}^T\rho^k \geq 1$, $K = \{\theta_1, \theta_2, \cdots, \theta_k\}$. If $\mathbf{1}^T\rho^k < 1$, find the largest l subject to $\mathbf{1}^T\rho^l \geq 1$; then $1 \leq l < k$ and $K = \{\theta_1, \theta_2, \cdots, \theta_l\}$.

Proof: Note that $i \leq j \Rightarrow \rho^i \geq \rho^j$ by our ordering assumption above. Now $\rho^k \geq \alpha$ while it is not true that $\rho^{k+1} \geq \alpha$, so that if $\mathbf{1}^T\rho^k \geq 1$, it follows from Lemma 1 that $\theta_j \in K \Leftrightarrow j \leq k$. The result for $\mathbf{1}^T\rho^k < 1$ may be checked in an identical manner.

Theorem 2 presents an efficient method for obtaining the set of feasible estimates K. The method is similar in spirit to that for finding the set of feasible estimates for MAP estimation, to be presented in a subsequent paper.

B. $\mathcal{P}_2$-Type Information

Suppose the partial prior information arises from pairwise comparisons between parameters, with statements such as "treatment 1 is at least ten times as likely to cure a certain disease than is treatment 2," i.e., $P_1 \geq 10P_2$. So the second type of information to be considered in this section is given in the general form:

$$\mathcal{P}_2 = \{P: P_j \leq l_j^i P_i, i,j = 1, \cdots, n, \mathbf{P} \geq \mathbf{0}, \mathbf{1}^T\mathbf{P} = 1\}.$$

As for $\mathcal{P}_1$-type information, redundancies in the information are first deleted. Define

$$\lambda_j^i = \max_{\mathcal{P}_2} (P_j/P_i).$$

Efficient computation of the λ_j^i is discussed in the Appendix. This definition ensures that $\lambda_j^i \leq l_j^i$, from which it follows that

$$\mathcal{P}_2 = \{P: P_j \leq \lambda_j^i P_i, i,j = 1, \cdots, n; \mathbf{P} \geq \mathbf{0}, \mathbf{1}^T\mathbf{P} = 1\}.$$

It is also quickly checked that

$$\lambda_j^i \leq \lambda_k^i \lambda_j^k. \tag{10}$$

This fact will be used in proving the theorem below.

Theorem 3: Given $\mathcal{P}_2$, $p(\cdot)$ and x, then if $\lambda_j^i > 0$, for all i,j:[2] hypothesis θ_k is feasible

$$\Leftrightarrow p_k(x)\lambda_k^j \geq p_j(x), \qquad \text{for all } j \neq k. \tag{11}$$

Proof: Hypothesis θ_k is feasible

$$\Leftrightarrow P \in \mathcal{P}_2 \text{ with } P_k p_k(x) \geq P_j p_j(x), \qquad \text{for all } j \neq k$$

$$\Rightarrow \max_{\mathcal{P}_2} \frac{P_k}{P_j} \geq \frac{p_j(x)}{p_k(x)}, \qquad \text{for all } j \neq k \tag{12}$$

$$\Leftrightarrow \lambda_k^j \geq \frac{p_j(x)}{p_k(x)}, \qquad \text{for all } j \neq k.$$

[2]The assumption that $\lambda_j^i > 0$, for all i,j is very reasonable. If $\lambda_j^i = 0$, $P_j = 0$, for all $P \in \mathcal{P}_2$, so that the parameter θ_j may effectively be ignored, i.e., it should not have been included in the parameter set originally.

To reverse the above implication chain, all we need to show is that $\lambda_k^j = \max_{\mathcal{P}_2}(p_k/P_j)$ is achieved simultaneously for all $j \neq k$ by a single $P \in \mathcal{P}_2$. By defining P as

$$P_j = \frac{(1/\lambda_k^j)}{\sum_i (1/\lambda_k^i)}, \qquad \text{for all } j,$$

it follows from (10) that $P \in \mathcal{P}_2$, and further, since $\lambda_k^k = 1$, that $P_k/P_j = \lambda_k^j$, for all j.

In summary, to find K for $\mathcal{P}_2$-type information the constraints for $\mathcal{P}_2$ are reduced prior to observation by calculating λ_j^i using the method presented in the Appendix. After the observation, feasibility or otherwise of each hypothesis may be determined by checking at most $n-1$ of the inequalities (11).

VI. A Sequential Testing Procedure

Sequential decisionmaking procedures always involve two elements: a stopping rule and a decision rule. The observations $x_1, x_2, \cdots, x_t$ are taken sequentially until the stopping rule, depending on $x^t(x_1, x_2, \cdots, x_t)$, says that enough observations have been taken and that the decision rule may be applied to x^t to reach the final decision. The usual Bayesian type of stopping rule [1], [2] is derived by assuming a cost in taking a further observation; the only other standard stopping rule is Wald's sequential probability ratio test (SPRT) [1], [2] for two hypotheses which is based on more classical lines in that no prior probabilities are assumed and the levels of the test are chosen by consideration of the probabilities of error, given that one or the other parameter is the true one. The new stopping rule proposed below is an entirely natural consequence of the model for decisionmaking with partial prior information developed in Section III; for MPE testing it is shown that with arbitrary $\mathcal{P}$, application of the stopping rule involves checking at most $2n-2$ inequalities, which means that in practical situations where observations are made at a rapid rate (e.g., in some signal detection problems), the stopping rule can be run on-line. This efficiency of implementing the stopping rule is its greatest advantage when compared with Bayesian stopping rules which are very difficult to calculate for $n > 2$ (multiple hypotheses testing). The rule is also compared with the SPRT in the case of two hypotheses.

When placed in a decisionmaking situation with partial prior information, indecision may arise, as modeled in Section III, in the form of a set of feasible decisions $K(x^t, \mathcal{P})$. It is sensible to attempt to overcome this indecision by taking further observations in the hope that more information can be obtained and to stop taking further observations if the indecision has been removed, i.e., if one decision is preferred above all others. In other words there follows directly from the idea of a set of feasible decisions, an obvious stopping rule:

$$\text{Stop iff} \quad K(x^t, \mathcal{P}) = \{d_k\}. \tag{13}$$

The principal advantage of the stopping rule (13) is its simplicity, both conceptually and, as shown below, computationally for MPE testing. Of course, (13) is also relatively simple to compute when $\mathcal{P}$ is given by linear inequalities—after each observation, at most n linear programs would need to be solved, as mentioned in Section III.

Details of the rule (13) for MPE hypotheses testing are presented in the following theorem.

Theorem 4—Criterion for $K=\{\theta_k\}$ for MPE testing: Assume $\mathcal{P}$ is closed.[3] Define $\alpha_k=\min_{\mathcal{P}} P_k$ and if $\alpha_k>0$, $\lambda_j^k=\max_{\mathcal{P}}(P_j/P_k)$. Then

i) if $\alpha_k>0$,

$$K=\{\theta_k\} \Leftrightarrow p_k(x^t)>\lambda_j^k p_j(x^t), \quad \text{for all } j\neq k, \quad (14)$$

ii) if $\alpha_k=0$, $K\neq\{\theta_k\}$.

Proof: Hypothesis θ_k is the only feasible hypothesis

$$\Leftrightarrow p_k(x^t)P_k>p_j(x^t)P_j, \qquad \text{for all } j\neq k, \text{ for all } \boldsymbol{P}\in\mathcal{P}.$$

i) If $\alpha_k>0$, $P_k>0$, for all $\boldsymbol{P}\in\mathcal{P}$, so

$$K=\{\theta_k\} \Leftrightarrow p_k(x^t)>p_j(x^t)P_j/P_k, \qquad \text{for all } j\neq k, \text{ for all } \boldsymbol{P}\in\mathcal{P}$$

$$\Leftrightarrow p_k(x^t)>\lambda_j^k p_j(x^t), \qquad \text{for all } j\neq k.$$

ii) $\alpha_k=0 \Leftrightarrow$ there exists $\boldsymbol{P}\in\mathcal{P}$ with $P_k=0$

$$\Leftrightarrow \text{there exists } \boldsymbol{P}\in\mathcal{P} \text{ with } p_k(x^t)P_k\leq p_j(x^t)P_j, \qquad \text{for all } j$$

$$\Leftrightarrow K\neq\{\theta_k\}.$$

The application of this theorem in implementing the stopping rule is straightforward. Note that it is reasonable to expect that the partial prior information will exclude the possibility of zero probabilities, in which case $\alpha_k>0$ for all k. Prior to observation the values of λ_j^k are calculated (see Lemma 5 below) according to the definition in the theorem. After the tth observation x_t, the quantities $p_j(x^t)=p_j(x_1,\cdots,x_t)$ are determined and used to test the inequalities (14) to see if there is only one feasible hypothesis, in which case no more observations are made. It is not difficult to show that at each stage, at most $2n-2$ of the inequalities, (14) needs to be tested; it is also interesting to note that the inequalities can be written in terms of the log-likelihood functions, $\log[p_k(x^t)/p_j(x^t)]$, which would be of computational benefit if the $p_j(x^t)$ belonged to the exponential family of distributions. Although not critical to the implementation of the stopping rule, the following lemma states that the quantities λ_j^k can be calculated by maximizing a linear functional over a projection of $\mathcal{P}$.

[3]The assumption that $\mathcal{P}$ is closed ensures that α_k and λ_j^k are attained by some $\boldsymbol{P}\in\mathcal{P}$. If this were not the case the strictness of the inequalities in (14) would be relaxed and ii) would also need alteration.

Lemma 5: Under the conditions of Theorem 4, if $\alpha_k>0$,

$$\lambda_j^k=\max R_j$$

subject to $R_k=1$ and

$$\boldsymbol{R}\in\{\mu\boldsymbol{P}: \mu\geq 0, \boldsymbol{P}\in\mathcal{P}\}.$$

The proof is omitted. When $\mathcal{P}$ is given by linear inequalities as in (5), this lemma shows that λ_j^k can be calculated by solving linear programs.

The following corollary to Theorem 4 ties together the present sequential procedure with Wald's sequential probability ratio test (SPRT). The SPRT is a sequential procedure for binary hypothesis testing with independent identically distributed (i.i.d.) observations, characterized by levels α and β. The stopping rule is to stop unless

$$\alpha\leq\sum_{i=1}^{t}\log\frac{p_2(x_i)}{p_1(x_i)}\leq\beta.$$

Being straightforward, the proof of the corollary is omitted.

Corollary 6: The following stopping rules are equivalent:

i) Wald's SPRT with levels α and β;
ii) rule (13) for MPE testing, with partial prior information $\mathcal{P}$ such that $\lambda_2^1=e^{-\alpha}$ and $\lambda_1^2=e^{\beta}$, e.g.,

$$\mathcal{P}=\{\boldsymbol{P}: (1+e^{\beta})^{-1}\leq P_2\leq(1+e^{\alpha})^{-1}, P_1+P_2=1\}.$$

There is much known about the SPRT, especially with regards to expected stopping times and probabilities of error given the true hypothesis, so Corollary 6 enables some known results to be applied to the present stopping rule in the binary hypothesis testing case when the observations are i.i.d.; furthermore, when there are more than two hypotheses, it is readily seen that the stopping rule (14) involves only a succession of pairwise comparisons, so the results for the SPRT could no doubt be extended to this multiple hypothesis testing situation as well.

The general stopping rule (13) was derived with no conditions being imposed on the distributions of the observations. However, some conditions will be required if the procedure is to stop in a finite number of steps with probability one (no matter what the true parameter is); a sufficient condition for this is that the observations are identically and independently distributed, and more relaxed conditions could, of course, be formulated.

As an addendum to this section, it is shown that the general framework developed here for sequential observations readily caters for the case when the parameter space (regarded as the state space) is allowed to undergo transition with known probability, and observations depend only on the current state. (The situation being considered is therefore a partially observable, finite state discrete time Markov process—see [17], [18] for some applications and control of such processes.) If the cost of making a decision at a particular time t depends on previous parameter

values $\theta_1, \cdots, \theta_t$, then it is possible to calculate $D_k(x^t)$ at each time instant and thus find $K(x^t, \mathcal{P})$; unfortunately, $D_k(x^t)$ can not be simply expressed in terms of $D_k(x^{t-1})$ and must be completely recalculated at each time point. The more interesting case which will be considered here occurs when the cost of a decision at time t depends only on the present state θ_t. It is possible to recursively update the partial prior information $\mathcal{P}$ for the initial state, to partial "prior" information $\mathcal{P}_2(x^1), \mathcal{P}_3(x^2), \cdots, \mathcal{P}_t(x^{t-1})$ for the states at times $2, 3, \cdots, t$ (details follow for the case when $\mathcal{P}$ is given by linear inequalities). When the costs depend only on the present state, the prior information for the present state is sufficient to determine the set of feasible decisions:

$$K(x^t, \mathcal{P}) = \bigcup_{P \in \mathcal{P}_t(x^{t-1})} \{d_k: D_k(x_t)P \geq 0\} \tag{15}$$

where

$$D_k(x_t) = (C - \mathbf{1}c_k^T) \operatorname*{diag}_{\theta_t} \{p(x_t|\theta_t)\}$$

and $C = [c_{ij}]$ with c_{ij} equal to the cost of choosing decision d_i at time t, given that $\theta_t = j$.

As promised above, details for recursively updating the prior information are now presented. The prior information at time t is

$$\mathcal{P}_t(x^{t-1}) = \{\Pr(\theta_t|x^{t-1}): \Pr(\theta_1) \in \mathcal{P}\};$$

the "posterior" information at time t is

$$\mathcal{Q}_t(x^t) = \{\Pr(\theta_t|x^t): \Pr(\theta_1) \in \mathcal{P}\}.$$

Introducing the notation for the transition probability matrix $\Pi = [\pi_{ij}]$ where $\pi_{ij} = \Pr(\theta_t = i|\theta_{t-1} = j)$, the prior probability $\boldsymbol{P}_t$ where $P_{t,j} = \Pr(\theta_t = j|x^{t-1})$, and the posterior probability $\boldsymbol{Q}_t$ where $Q_{t,j} = \Pr(\theta_t = j|x^t)$, leads to the following recursive relations:

$$\boldsymbol{P}_t = \Pi \boldsymbol{Q}_{t-1} \tag{16}$$

$$\boldsymbol{Q}_t = \frac{\operatorname*{diag}_{\theta_t} \{p(x_t|\theta_t)\} \boldsymbol{P}_t}{\mathbf{1}^T \operatorname*{diag}_{\theta_t} \{p(x_t|\theta_t)\} \boldsymbol{P}_t}. \tag{17}$$

When $\mathcal{P}$ is specified by the linear inequalities (5) the updated information is specified by the following linear inequalities:

$$\mathcal{P}_t(x^{t-1}) = \{\boldsymbol{P}: A_t\boldsymbol{P} \geq \mathbf{0}, \mathbf{1}^T\boldsymbol{P} = 1, \boldsymbol{P} \geq \mathbf{0}\}$$

$$\mathcal{Q}_t(x^t) = \{\boldsymbol{Q}: B_t\boldsymbol{Q} \geq \mathbf{0}, \mathbf{1}^T\boldsymbol{Q} = 1, \boldsymbol{Q} \geq \mathbf{0}\}$$

where the relations (16) and (17) imply:

$$A_t = B_{t-1}\Pi^{-1}, \qquad (\text{with } A_1 = A)$$

$$B_t = A_t \operatorname*{diag}_{\theta_t} \{1/p(x_t|\theta_t)\}.$$

(The cases when Π is singular or some of the $p(x_t|\theta_t)$ are zero are readily handled.)

It is interesting to note that the idea of posterior information introduced above allows an alternative to (15) to be written, viz.

$$K(x^t, \mathcal{P}) = \bigcup_{Q \in \mathcal{Q}_t(x^t)} \{d_k: (C - \mathbf{1}c_k^T)\boldsymbol{Q} \geq \mathbf{0}\}.$$

Although suppressed in the notation, time dependence of the costs C, observation probabilities $p(x_t|\theta_t)$ and transition probabilities Π is permitted; in particular, the case when the process is controlled is catered for, provided that the controls applied until time t are known. A note of caution is in order here—when the control strategy depends on the prior probabilities, the above framework does not consider the investigation of the uncertainty in controls produced by uncertainty in prior probabilities for the initial state; in this situation, uncertainty in the state probabilities produces uncertainty in the controls and thus in the transition probabilities; this uncertainty is then multiplied by the uncertainty in the state probabilities. The effect of multiplying linear inequality uncertainties by linear inequality uncertainties has been investigated elsewhere in the literature, in a somewhat different context [19].

VII. Conclusion

With precise measurements almost any standard decisionmaking procedure (Bayesian or classical) is adequate; with less precise measurements any inference is doubtful, unless our prior knowledge is good. This paper has developed from this common-sense notion and the formulation of the problem of finding the set of feasible decisions corresponding to given partial prior information is its principal innovation. The partial prior information is specified as a set of prior probabilities $\mathcal{P}$; the definition of a set of feasible decisions comes from a simple generalization of the standard Bayesian decisionmaking framework. It has been argued that $\mathcal{P}$ may often be given by linear inequalities in which case computationally attractive methods, involving a finite number of linear programs, exist for the determination of the feasible decisions. In the case of MPE testing when $\mathcal{P}$ is given by special forms of constraints, explicit representations for the set of feasible decisions have been obtained. A natural consequence of the above work is a new stopping rule for sequential testing. For MPE testing with arbitrary $\mathcal{P}$, this rule only involves testing a small number of linear inequalities after each observation.

Perhaps the most serious criticism which could be leveled at the approach taken in this paper is that it relies on being able to quantify the prior knowledge as an exact specification of a set of probabilities. Nevertheless the definition of partial prior information adopted in this paper does seem more realistic than some other available definitions [7], [8]; there are undoubtedly many practical decisionmaking problems which can be better modeled by the present approach than by other available methods. Furthermore, the requirements here are not as strict as those for the Bayesian approach which insist on the precise assignment of prior probabilities. In some sense the

specification of $\mathcal{P}$ may be regarded as the first stage in a "blurring out" of the probability assignments. It follows that the results of this paper, instead of modeling vague prior information and consequent indecision, could be used for performing efficient sensitivity analyses (with respect to the prior probabilities) of the standard Bayesian decision rules. A study of sensitivity to general parameter variation (including prior probabilities) has been undertaken in [20].

In a subsequent paper the authors will extend the concepts of this paper to the case when the parameter set is not necessarily finite, i.e., to the estimation case. Special methods for finding the set of feasible estimates will be developed for minimum variance (conditional mean, MMSE), median, and MAP procedures.

Appendix

This appendix is devoted to establishing an efficient algorithm for removing redundancies in $\mathcal{P}_2$-type information, as required in Section V. Recall that $\mathcal{P}_2$-type information is given by

$$\mathcal{P}_2 = \{\boldsymbol{P}: P_j \leqslant l_j^i P_i,\ i,j=1,\cdots,n;\ \boldsymbol{P} \geqslant \boldsymbol{0};\ \mathbf{1}^T\boldsymbol{P}=1\}$$

where l_i^i is taken as 1 and undefined values of l_j^i are taken as $+\infty$. Redundancy was removed by defining $\lambda_j^i = \max_{\mathcal{P}_2}(P_j/P_i)$, so the algorithm presented here is for the calculation of λ_j^i; the algorithm may be thought of as a combination of dynamic programming and doubling ideas.

Algorithm: Define

$$\lambda_j^i(0) = l_j^i, \qquad \text{for all } i,j,$$

and for $r \geqslant 0$:

$$\lambda_j^i(r+1) = \min_k \lambda_k^i(r)\lambda_j^k(r), \qquad \text{for all } i,j.$$

Then stop when either

$$\lambda_j^i(s+1) = \lambda_j^i(s), \qquad \text{for all } i,j \tag{A.1}$$

or

$$\lambda_k^k(s+1) < 1, \qquad \text{for some } k. \tag{A.2}$$

Two claims are made and proved below: the first states that the algorithm stops in a small number of steps; the second says that when the algorithm stops, the redundancies in the $\mathcal{P}_2$ constraints have been removed.

Claim 1: The algorithm stops for some

$$s \leqslant \lceil \log_2(n-1) \rceil$$

(where $\lceil x \rceil$ is the smallest integer larger than x).

Claim 2: If (A.1) holds, $\lambda_j^i = \lambda_j^i(s)$. If (A.2) holds, $P_k = 0$, for all $\boldsymbol{P} \in \mathcal{P}_2$, so that either $\lambda_k^j = 0$, for all j, or $\mathcal{P}_2$ is empty (equivalently: the $\mathcal{P}_2$ constraints are inconsistent).

Before proving these claims, some preliminary facts are needed.

Fact 1: $\lambda_j^i(r) = \min l_{k(1)}^{k(0)} l_{k(2)}^{k(1)} \cdots l_{k(t)}^{k(t-1)}$, where $t = 2^r$, $k(0) = i$, $k(t) = j$, and the minimization is taken over all $k(1),\cdots,k(t-1) \in \{1,2,\cdots,n\}$.

Fact 2: $\lambda_j^i(r)$ can be written as a "minterm" product—a product with a minimal number of terms:

$$\lambda_j^i(r) = l_{k(1)}^{k(0)} l_{k(2)}^{k(1)} \cdots l_{k(t)}^{k(t-1)},$$

where $1 \leqslant t \leqslant 2^r$, $k(0) = i$, $k(t) = j$.

Fact 3: $\lambda_j^i(r+1) \leqslant \lambda_j^i(r) \leqslant l_j^i$, for all i,j.

Fact 4: If, in the minterm product for $\lambda_j^i(r)$ when $i \neq j$, the parameters $k(0),k(1),\cdots,k(t)$ are not all distinct, then $\lambda_k^k(r) < 1$ for some k.

Only the last fact here is not self-evident. If, in the minterm product for $\lambda_j^i(r)$, $k(p) = k(q) = k$ say, where $0 \leqslant p < q \leqslant t$, then $\lambda_j^i(r) < l_{k(1)}^{k(0)} \cdots l_{k(p)}^{k(p-1)} l_{k(q+1)}^{k(q)} \cdots l_{k(t)}^{k(t-1)}$ since the minterm product for $\lambda_j^i(r)$ was chosen. Thus

$$\begin{aligned} 1 &> l_{k(p+1)}^{k(p)} \cdots l_{k(q)}^{k(q-1)} \\ &= l_{k(p+1)}^{k(p)} \cdots l_{k(q)}^{k(q-1)} l_k^k l_k^k \cdots l_k^k \\ &\geqslant \lambda_k^k(r) \quad \text{(using Fact 1).} \end{aligned}$$

Proof of Claim 1: Take $s = \lceil \log_2(n-1) \rceil$ and suppose (A.1) does not hold, i.e., there exists i,j such that $\lambda_j^i(s+1) < \lambda_j^i(s)$. It will be shown that in this case (A.2) must hold.

If $i = j$,

$$\begin{aligned} \lambda_i^i(s+1) &< \lambda_i^i(s) \\ &\leqslant 1 \quad \text{(using Fact 3)} \end{aligned}$$

so that (A.2) holds.

If $i \neq j$, consider the minterm product for $\lambda_j^i(s+1)$. According to Fact 4, either $k(0),k(1),\cdots,k(t)$ are distinct or (A.2) holds. If $k(0),k(1),\cdots,k(t)$ are distinct, then $t \leqslant n-1 \leqslant 2^s$ so that the minterm expression for $\lambda_j^i(s+1)$ has fewer than 2^s terms in it. Hence, using Fact 1, $\lambda_j^i(s) \leqslant \lambda_j^i(s+1)$, which is a contradiction. Thus (A.2) holds.

A further easily checked fact is needed for the proof of the second claim.

Fact 5: $P_j \leqslant \lambda_j^i(r) P_i$, for all $\boldsymbol{P} \in \mathcal{P}_2$

Proof of Claim 2: a) Suppose (A.1) holds. For a given j, define

$$P_i = \frac{1/\lambda_j^i(s)}{\sum_k \left[1/\lambda_j^k(s)\right]}, \qquad \text{for all } i.$$

Then

$$\begin{aligned} P_p &= \frac{1/\lambda_j^p(s)}{\sum_k \left[1/\lambda_j^i(s)\right]} \\ &\leqslant \frac{\lambda_p^i(s)/\lambda_j^i(s)}{\sum_k \left[1/\lambda_j^k(s)\right]} \quad \left[\text{using (A.1)}\right] \\ &= \lambda_p^i(s) P_i \\ &\leqslant l_p^i P_i, \qquad \text{for all } p \quad \left[\text{by Fact 3}\right]. \end{aligned}$$

Thus $\boldsymbol{P} \in \mathcal{P}_2$. Also $\lambda_j^i(s) = P_j/P_i$, so from Fact 5, it follows that

$$\lambda_j^i(s) = \max_{\mathcal{P}_2}(P_j/P_i) = \lambda_j^i.$$

b) If (A.2) holds,

$$P_k \leqslant \lambda_k^k(s+1)P_k, \qquad \text{for all } \boldsymbol{P} \in \mathcal{P}_2$$
$$\Rightarrow P_k = 0, \qquad \text{for all } \boldsymbol{P} \in \mathcal{P}_2$$
$$\Rightarrow \text{either } \lambda_k^j = 0, \qquad \text{for all } j$$
$$\text{or } \mathcal{P}_2 \text{ is empty.}$$

REFERENCES

[1] S. Zacks, *The Theory of Statistical Inference*. NY: Wiley, 1971.

[2] T. S. Fergusson, *Mathematical Statisitcs: A Decision Theoretic Approach*. NY: Academic, 1967.

[3] H. L. van Trees, *Detection, Estimation and Modulation*, Part 1. NY: Wiley, 1968.

[4] E. L. Lehmann, *Testing Statistical Hypotheses*. NY: Wiley, 1959.

[5] G. A. Barnard and D. R. Cox, Eds., *The Foundations of Statistical Inference*. London: Methuen, 1962.

[6] J. R. Blum and J. Rosenblatt, "On partial *a priori* information in statistical inference," *Ann. Math. Stat.*, vol. 38, pp. 1671–1678, 1967.

[7] H. Kudō, "On partial prior information and the property of parametric sufficiency," in *Proc. Fifth Berkeley Symp. Math. Stat. Prob.*, vol. 1, pp. 251–265, 1967.

[8] J. L. Hodges, Jr. and E. L. Lehmann, "The use of previous experience in reaching statistical decisions," *Ann. Math. Stat.*, vol. 23, pp. 396–407, 1952.

[9] E. T. Jaynes, "Information theory and statistical mechanics," *Phys. Rev.*, vol. 106, no. 4, pp. 620–630, 1957.

[10] E. T. Jaynes, "Prior probabilities," *IEEE Trans. Syst. Sci. Cybern.*, vol. SSC-4, no. 3, pp. 227–241, 1968.

[11] R. L. Kashyap, "Prior probability and uncertainty," *IEEE Trans. Info. Theory*, vol. IT-17, no. 6, Nov. 1971.

[12] R. W. Johnson, "Comments on 'Prior probability and uncertainty,'" *IEEE Trans. Info. Theory*, vol. IT-25, no. 1, Jan. 1979.

[13] L. A. Zadeh, "Fuzzy sets," *Info. & Control*, vol. 8, no. 3, pp. 338–353, 1965.

[14] S. R. Watson, J. J. Weiss, and M. L. Donnell, "Fuzzy decision analysis," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 1–9, Jan. 1979.

[15] R. Jain, "Decisionmaking in the presence of fuzziness and uncertainty," in *Proc. IEEE Conf. on Decision and Control*, vol. 2, pp. 1318–1323, 1977.

[16] J. M. Potter, "Hypothesis testing and estimation with partial prior information," M.E. thesis, Univ. Newcastle, N.S.W., Australia, 1978.

[17] R. D. Smallwood and E. J. Sondik, "The optimal control of partially observable Markov processes over a finite horizon," *Oper. Res.*, vol. 21, pp. 1071–1088, 1973.

[18] E. J. Sondik, "The optimal control of partially observable Markov processes over the infinite horizon: discounted costs," *Oper. Res.*, vol. 26, no. 2, pp. 282–304, March–April 1978.

[19] B. R. Barmish and J. Sankaran, "The propagation of parametric uncertainty via polytopes," *IEEE Trans. Automat. Contr.*, vol. AC-24, no. 2, pp. 346–349, Apr. 1979.

[20] C. C. White III, "Multi-parametric sensitivity in decisionmaking under uncertainty," *Comp. & Biomed. Res.*, vol. 12, pp. 125–130, 1979.