The Annals of Mathematical Statistics
1969, Vol. 40, No. 3, 953–969

UPPER AND LOWER PROBABILITY INFERENCES FOR FAMILIES OF HYPOTHESES WITH MONOTONE DENSITY RATIOS

BY A. P. DEMPSTER[1]
Harvard University

0. Summary. Models of the second kind, described in Section 2, are applied to define inferences about single parameter families having monotone density ratio structure. A definition and some properties of this structure are given in Section 3 along with basic general formulas for inferences. Mixture families always have the monotone density ratio property, and some theory and illustrations are given in Section 4. Location and scale parameter families sometimes have the property, as shown and illustrated in Section 5 where the examples consist of uniform and normal location parameters and exponential and normal scale parameters.

1. Introduction. The aim of this paper is to apply a general theory of statistical inference [1], [2], [3], [4], [5] to some single parameter families of hypotheses which have often been used in the past to illustrate other theories of inference. The discussion is limited to *models of the second kind* [1], [5] which apply when the observations represent a sample from an unknown member of a specified family of multinomial populations. But as remarked in [1] and illustrated in Section 4, 5, continuous observables can be unambiguously included as limiting cases of multinomial observables.

Uncertain knowledge or information about the true or actual point of some sample space is represented by a distribution of random subsets of the sample space, which in turn induces [2] upper and lower probabilities for events defined as fixed subsets of the sample space, and upper and lower expected values for real-valued functions defined over the sample space. When the random subsets are single point subsets, the model specifies an ordinary distribution over the sample space so that the upper and lower bounds coincide. Correspondingly [5], the approach to inference may be viewed as a generalization of Bayesian inference, which reduces to ordinary Bayesian inference when the prior information assumes the form of an ordinary prior distribution. In generalized Bayesian inference, the prior information may consist simply of a restriction to a parametric family of hypotheses or may consist of such a restriction together with a prior distribution over the parametric space. The former type is that included directly in models of the second kind, while the latter or more general types are not explicitly considered in this paper. See Section 2 for an elaboration of the discussion of prior knowledge.

Received 12 September 1968.
[1] This work was facilitated by a grant from the National Science Foundation (GP-8774). Part of the work was done while the author was a Fellow of the The John Simon Guggenheim Memorial Foundation.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

The families of hypotheses considered here are single parameter families with monotone density ratio, as defined in Section 3. It follows that the inferences are expressible in terms of random closed intervals for which some elementary theory has been developed in [4]. Single parameter families are not of much practical interest, for excepting very small sample sizes the differences among various approaches to inference tend to be slight. The promise of the new methods lies in multiparameter situations, where older approaches to inference give more trouble. The present paper explores foothills before tackling more difficult peaks.

2. Models of the second kind: a review. Consider multinomial sampling with k specified categories, for some fixed positive integer $k \geqq 2$. The k categories will be indexed by the set

$$K = \{1, 2, \cdots, k\}, \tag{2.1}$$

for convenience, although each model of the second kind treats the k categories symmetrically under all permutations. Denoting a population frequency in category i by $\pi(i)$ for $i \varepsilon K$, any complete set of population frequencies is specified by a vector

$$\boldsymbol{\pi} = [\pi(1), \pi(2), \cdots, \pi(k)] \tag{2.2}$$

belonging to the simplex

$$\Pi = \{\boldsymbol{\pi}; \pi(i) \geqq 0 \;\forall\; i \varepsilon K, \quad \textstyle\sum_{i \varepsilon K} \pi(i) = 1\}. \tag{2.3}$$

The true population frequency vector $\boldsymbol{\pi}$, while not precisely known, will be assumed to lie in a given parametric family

$$\boldsymbol{\pi}(\theta) = [\pi(1, \theta), \pi(2, \theta), \cdots, \pi(k, \theta)] \tag{2.4}$$

for $\theta \varepsilon \Theta$. This familiar type of assumption has an unfamiliar role in generalized Bayesian inference; specifically, the assumption expresses a special form of prior information. The general form is expressed by a specified distribution of random subsets of Π, and the special case is the distribution which assigns probability unity to the single subset

$$\Pi(\Theta) = \{\boldsymbol{\pi}(\theta); \theta \varepsilon \Theta\}. \tag{2.5}$$

Such prior information says that, although it is certain that $\boldsymbol{\pi}$ is somewhere in $\Pi(\Theta)$, there is no further information at all about θ in the sense that every proper nonempty subset of $\Pi(\Theta)$ has upper probability unity and lower probability zero. An important technical property of this informationless prior knowledge is that the corresponding inferences may be combined with nontrivial prior information according to a simple rule of combination [2], [5] to produce the inferences appropriate to the nontrivial prior information. In other words, the inferences described in this paper may be viewed as inferences based on sample information plus parametric assumption only, but ready for combination with any particular user's more detailed prior knowledge within $\Pi(\Theta)$. In particular

[4], [5], the user whose prior knowledge is represented by a distribution over the single point subsets of $\Pi(\Theta)$, i.e. by an ordinary prior distribution, will be led back to ordinary Bayesian inferences.

A distinctive feature of the models of generalized Bayesian inference is a space U whose points are conceived as being in one-one correspondence with the members of the population being sampled. The points of U are not directly observable but are logically related to the observation space K and the unrestricted parameter space Π. In the case of models of the second kind, the space U is taken to be a mathematical copy of Π, i.e.,

$$U = \{\mathbf{u}; u(i) \geqq 0 \,\forall\, i \,\varepsilon\, K, \textstyle\sum_{i\varepsilon K} u(i) = 1\} \tag{2.6}$$

where

$$\mathbf{u} = [u(1), u(2), \cdots, u(k)]' \tag{2.7}$$

corresponds to a general point of the population. The operation of drawing a single random member of the population is viewed as governed by a uniform distribution over the simplex U, and the member $\mathbf{u}$ thus drawn is supposed to be compatible with points $[j, \boldsymbol{\pi}] \,\varepsilon\, K \times \Pi$ if and only if

$$\pi(i)u(j) \leqq \pi(j)u(i) \,\forall\, i \,\varepsilon\, K. \tag{2.8}$$

There are two equivalent ways to fit the assumptions of the previous sentence into an abstract framework. The first, adopted as standard in [4], [5], is to interpret (2.8) as specifying a multivalued mapping from U to $K \times \Pi$, i.e., as defining a subset of points $[j, \boldsymbol{\pi}] \,\varepsilon\, K \times \Pi$ for each point $\mathbf{u} \,\varepsilon\, U$. Accordingly, the uniform distribution over U induces a distribution of random subsets of $K \times \Pi$ which in turn defines upper and lower probabilities for fixed subsets of $K \times \Pi$. The second viewpoint, mentioned briefly in [5], is to interpret the model as specifying a distribution over a partition of $U \times K \times \Pi$. Here, the subsets defining the partition may be indexed by $\mathbf{u} \,\varepsilon\, U$, the subset corresponding to a given $\mathbf{u}$ consists of triples $[\mathbf{u}, j, \boldsymbol{\pi}]$ satisfying (2.8), and the distribution over subsets is induced by the uniform distribution over U. The second way to conceptualize the model is important because it shows that acceptance of the model amounts to putting one's trust in a certain marginal distribution. See the author's reply to discussants in [5].

An essential feature of the new theory of inference is a concept of conditioning [4], [5]. When the model is conditioned on a fixed $\boldsymbol{\pi}$ and conditional probabilities are computed for the categories in K to be observed in a single drawing, then $\boldsymbol{\pi}$ has the desired interpretation as the vector of probabilities associated with the elements of K. On the other hand, if $\boldsymbol{\pi}$ remains unknown while an observation $j \,\varepsilon\, K$ is recorded, then the suggested inferences about π are defined by conditioning the model on the subset $\{j\} \times \Pi(\Theta)$ of $K \times \Pi$, where $\{j\}$ is the subset of K consisting of the single point j. These inferences are expressible through a distribution of random subsets of $\Pi(\Theta)$, or through the equivalent distribution of random subsets of Θ. Suppose that $\Omega(j, \mathbf{u})$ denotes the random subset of θ

determined via (2.8) by the random $\mathbf{u} \varepsilon U$ when j is fixed. Only those $\mathbf{u}$ for which $\Omega(j, \mathbf{u})$ is not empty contribute to the relevant conditional distribution with fixed j. The remainder of Section 2 sketches the derivation of an explicit formula for $\Pr(\Omega(j, \mathbf{u}) \supset \Lambda)$ according to this relevant distribution, where Λ denotes a fixed subset of Θ.

Events expressible by conditions on the random subset $\Omega(j, \mathbf{u})$ are also expressible by conditions on the random point $\mathbf{u} \varepsilon U$. Thus, if $\{\theta\}$ denotes the fixed subset of Θ consisting of the single point θ, then the condition $\Omega(j, \mathbf{u}) \supset \{\theta\}$ is also expressible as $u \varepsilon U(j, \theta)$ where $U(j, \theta)$ consists of points satisfying (2.8) with $\boldsymbol{\pi}(\theta)$ in the role of $\boldsymbol{\pi}$. $U(j, \theta)$ is the subsimplex of U determined by replacing the vertex whose jth coordinate is unity with $\boldsymbol{\pi}(\theta)$, while leaving the remaining $k - 1$ vertices the same. It follows that

$$\Pr(\Omega(j, \mathbf{u}) \supset \{\theta\}) = C\pi(j, \theta) \tag{2.9}$$

where C is a normalizing factor which enters because the restriction of $\boldsymbol{\pi}$ to $\Pi(\Theta)$ means that $\mathbf{u}$ is restricted in the conditional model to $\bigcup_{\theta\varepsilon\Theta} U(j, \theta)$. Specifically, C^{-1} is the $(k - 1)$-dimensional volume of $\bigcup_{\theta\varepsilon\Theta} U(j, \theta)$. More generally, the condition $\Omega(j, \mathbf{u}) \supset \Lambda$ is also expressible as

$$\mathbf{u} \varepsilon \bigcup_{\theta\varepsilon\Lambda} U(j, \theta) \tag{2.10}$$

from which it follows [1] that

$$\begin{aligned} \Pr(\Omega(j, \mathbf{u}) \supset \Lambda) &= C[\textstyle\sum_{i=1}^{k} \sup_{\theta\varepsilon\Lambda} \pi(i, \theta)/\pi(j, \theta)]^{-1} && \text{if } \inf_{\theta\varepsilon\Lambda} \pi(j, \theta) > 0 \\ &= 0 && \text{otherwise.} \end{aligned} \tag{2.11}$$

The explanation of (2.11) is (i) that clearly $\bigcap_{r\varepsilon\Lambda} U(j, \theta)$ has volume zero if any component $U(j, \theta)$ or limit of such components, has volume zero, while (ii) if all such subsimplexes of U have positive content, then their intersection is a similar simplex of positive content whose volume is easily computed [1] to yield (2.11).

Topological considerations have been ignored in the preceding discussion, but it is clear that (2.11) holds when Λ is a closed set according to some topology on Θ and the mapping $\theta \to \boldsymbol{\pi}(\theta)$ is continuous and one-one. Such conditions, although not very tight, suffice for the examples of this paper.

3. Families of hypotheses with monotone density ratio. Henceforth it will be assumed that the parameter θ is real-valued on $\alpha \leqq \theta \leqq \beta$, allowing $\alpha = -\infty$ and $\beta = +\infty$ as limiting cases, and that the corresponding hypotheses $\boldsymbol{\pi}(\theta)$ trace a continuous curve in the simplex Π as θ runs from α to β. Such a family of hypotheses will be said to have a *monotone density ratio* if for all $i, j \varepsilon K$ the ratio $\pi(i, \theta)/\pi(j, \theta)$ is a monotone nonincreasing or nondecreasing function of θ on $\alpha \leqq \theta \leqq \beta$ ignoring values of θ for which $\pi(i, \theta) = \pi(j, \theta) = 0$. Within the class of families of hypotheses with continuous $\boldsymbol{\pi}(\theta)$, the subclass consisting of families with monotone density ratio has an important property relative to the corresponding models of the second kind.

LEMMA 3.1. *The region* $\Omega(j, \mathbf{u})$ *defined in Section 2 is either empty or a closed interval for all* $j \varepsilon K$ *and all* $\mathbf{u}$ *interior to* U *if and only if the family of hypotheses* $\pi(\theta)$ *for* $\alpha \leqq \theta \leqq \beta$ *has a monotone density ratio.*

PROOF. First suppose that $\pi(\theta)$ has a monotone density ratio. Consider a fixed $j \varepsilon K$ and a fixed $\mathbf{u}$ interior to U for which $\Omega(j, \mathbf{u})$ is not empty. If $\theta \varepsilon \Omega(j, \mathbf{u})$, then $\pi(j, \theta) > 0$, because (2.8) together with $\pi(j, \theta) = 0$ would imply $u(j) = 0$ against the assumption that $\mathbf{u}$ is interior to U. Thus $\Omega(j, \mathbf{u})$ consists of those θ for which

$$\pi(i, \theta)/\pi(j, \theta) \leqq u(i)/u(j) \qquad \text{for} \quad i \varepsilon K. \tag{3.1}$$

Setting

$$\theta_* = \liminf \{\theta; \theta \varepsilon \Omega(j, \mathbf{u})\} \qquad \text{and} \qquad \theta^* = \limsup \{\theta; \theta \varepsilon \Omega(j, \mathbf{u})\}, \tag{3.2}$$

it follows, since $\pi(\theta)$ is continuous in θ, that θ_* and θ^* both belong to $\Omega(j, \mathbf{u})$ while no θ outside the range $\theta_* \leqq \theta \leqq \theta^*$ does so. To assume that some θ on $\theta_* < \theta < \theta^*$ is outside $\Omega(j, \mathbf{u})$ would be to assume that (3.1) fails for some $i \varepsilon K$, implying a contradiction of the monotonicity of $\pi(i, \theta)/\pi(j, \theta)$ as θ ranges over $\theta_*, \theta, \theta^*$. Thus

$$\Omega(j, \mathbf{u}) = \{\theta; \theta_* \leqq \theta \leqq \theta^*\}. \tag{3.3}$$

Conversely, suppose that $\Omega(j, \mathbf{u})$ is a closed interval or empty for all $j \varepsilon K$ and all $\mathbf{u}$ interior to U. If at the same time the monotone density ratio property were to fail, then there would exist $i, j \varepsilon K$ and $\theta_1 < \theta < \theta_2$ such that $\pi(i, \theta)/\pi(j, \theta)$ lies outside the closed interval determined by $\pi(i, \theta_1)/\pi(j, \theta_1)$ and $\pi(i, \theta_2)/\pi(j, \theta_2)$. But then the point $\mathbf{u}$ whose coordinates are proportional to $\max \{\pi(i, \theta_1)/\pi(j, \theta_1), \pi(i, \theta_2)/\pi(j, \theta_2)\}$ for $i = 1, 2, \cdots, k$ would be such that θ_1 and θ_2 belong to $\Omega(j, \mathbf{u})$ while θ did not, thus contradicting the hypothesis that $\Omega(j, \mathbf{u})$ is a closed interval. Note that the $\mathbf{u}$ found here may lie on the boundary of U, but by continuity some neighboring but interior $\mathbf{u}$ would have the same property.

In view of Lemma 3.1, inferences about θ in the case of a family with monotone density ratio depend on the theory of random closed intervals [4]. In particular, if Λ is taken to be the closed interval $[\theta_1, \theta_2]$, then $\Pr(\Omega(j, \mathbf{u}) \supset \Lambda)$ as in (2.11) may be written

$$H(\theta_1, \theta_2) = \Pr(T_1 \leqq \theta_1, T_2 \geqq \theta_2) \tag{3.4}$$

where $[T_1, T_2]$ denotes the random closed interval $\Omega(j, \mathbf{u})$. It is clear from (3.4) that the function $H(\theta_1, \theta_2)$ defined for $\alpha \leqq \theta_1 \leqq \theta_2 \leqq \beta$ determines the full distribution of the random closed interval $[T_1, T_2]$. The task of computing $H(\theta_1, \theta_2)$ from (2.11) is simplified in the case of a family with monotone density ratio by the fact that for each $i \varepsilon K$ the supremum appearing in (2.11) is attained either for $\theta = \theta_1$ or $\theta = \theta_2$. More explicit representations of $H(\theta_1, \theta_2)$ will be given below, depending on a second property of families with monotone density ratio:

LEMMA 3.2. *Suppose that* $\pi(\theta)$ *for* $\alpha \leqq \theta \leqq \beta$ *represents a family of hypotheses* *with monotone density ratio. Then* $\pi(\theta)$ *induces a partial ordering on the elements of* K. *Specifically,* K *may be expressed as the union of mutually exclusive subsets* $K_1, K_2, \cdots, K_r$ *for some* r *on* $1 \leqq r \leqq k$, *where, if* i *and* j *belong to a common subset* K_s, *then*

$$\pi(i, \theta_1)\pi(j, \theta_2) = \pi(i, \theta_2)\pi(j, \theta_1) \tag{3.5}$$

for all θ_1, θ_2 *on* $\alpha \leqq \theta_1 \leqq \theta_2 \leqq \beta$, *while, if* $i \varepsilon K_s$ *and* $j \varepsilon K_t$ *with* $s < t$, *then*

$$\pi(i, \theta_1)\pi(j, \theta_2) \geqq \pi(i, \theta_2)\pi(j, \theta_1) \tag{3.6}$$

for all θ_1, θ_2 *on* $\alpha \leqq \theta_1 \leqq \theta_2 \leqq \beta$ *with strict inequality holding for some pair* θ_1, θ_2.

PROOF. The subsets defining the partition of K may be constructed to satisfy (3.5) by grouping into equivalence classes those $j \varepsilon K$ which define the same likelihood function of θ, where likelihood functions $C\pi(j, \theta)$ are regarded as determined only up to an arbitrary constant multiplier. If K' and K'' denote a pair of distinct equivalence classes, then, in view of the monotone density ratio property, *either* (3.5) holds for $i \varepsilon K'$ and $j \varepsilon K''$ with strict inequality for some pair θ_1, θ_2 *or* the analogous property holds with the direction of inequality reversed. Accordingly, it may be said that $K' < K''$ in the first case and that $K' > K''$ in the second case. It is easily checked that the transitivity property holds for this ordering and hence that it is a complete ordering of the equivalence classes defined by (3.5). It remains only to label the equivalence classes $K_1, K_2, \cdots, K_r$ in accordance with their established order.

In view of Lemma 3.2 and (2.11), the function $H(\theta_1, \theta_2)$ determined by a single observation $j \varepsilon K_s$ may be written

$$\begin{aligned} H(\theta_1, \theta_2) &= C[\textstyle\sum_{t<s} \sum_{i \varepsilon K_t} \pi(i, \theta_1)/\pi(j, \theta_1) + \sum_{i \varepsilon K_s} (1) \\ &\qquad + \textstyle\sum_{t>s} \sum_{i \varepsilon K_t} \pi(i, \theta_2)/\pi(j, \theta_2)]^{-1} \\ &\qquad\qquad \text{if } \pi(j, \theta) > 0 \quad \text{on} \quad \theta_1 \leqq \theta \leqq \theta_2 \\ &= 0 \qquad \text{otherwise,} \end{aligned} \tag{3.7}$$

or

$$\begin{aligned} H(\theta_1, \theta_2) &= C[\textstyle\sum_{t<s} k_t R_{ts}(\theta_1) + k_s + \sum_{t>s} k_t R_{ts}(\theta_2)]^{-1} \\ &\qquad\qquad \text{if } \pi(j, \theta) > 0 \quad \text{on} \quad \theta_1 \leqq \theta \leqq \theta_2 \\ &= 0 \qquad \text{otherwise,} \end{aligned} \tag{3.8}$$

where k_t denotes the number of points in K_t and $R_{ts}(\theta)$ denotes the ratio $\pi(i, \theta)/\pi(j, \theta)$ common to $i \varepsilon K_t$ and $i \varepsilon K_s$.

If the integer labels $1, 2, \cdots, k \varepsilon K$ are arranged to have natural order consistent with the order of their associated equivalence classes $K_1, K_2, \cdots, K_r$, then (3.7) or (3.8) may also be written

$$\begin{aligned} H(\theta_1, \theta_2) &= C[(\Pi(j, \theta_1)/\pi(j, \theta_1)) + ((1 - \Pi(j, \theta_2))/\pi(j, \theta_2))]^{-1} \\ &\qquad\qquad \text{if } \pi(j, \theta) > 0 \quad \text{on} \quad \theta_1 \leqq \theta \leqq \theta_2 \\ &= 0 \qquad \text{otherwise,} \end{aligned} \tag{3.9}$$

where

$$\Pi(j, \theta) = \textstyle\sum_{i \leqq j} \pi(i, \theta), \qquad \forall j \in K, \tag{3.10}$$

expresses the cumulative distribution function of the discrete distribution $\pi(\theta)$.

Although formula (3.9) is more compact than (3.7) or (3.8), it is less illuminating. For example, it is clear from (3.7) or (3.8) that $H(\theta_1, \theta_2)$ does not depend on θ_1 if $j \in K_1$ and does not depend on θ_2 if $j \in K_r$. In the former case, a random closed interval $[T_1, T_2]$ whose distribution is specified as in (3.4) has a fixed T_1, while in the latter case it has a fixed T_2. Another pair of closely related properties of $H(\theta_1, \theta_2)$ follow directly from (3.7) or (3.8). First, the function $H(\theta_1, \theta_2)$ is the same for all j in the same equivalence class. Second, if the model were altered by pooling the categories defining each equivalence class, and reporting only the observed equivalence class rather than the original category j, the function $H(\theta_1, \theta_2)$ would be altered only in its normalizing constant C. Thus, it is sufficient for inference purposes to know the equivalence class of an observation, not the observation itself. Moreover, it is permissible without affecting inferences to make an initial simplification in any model with monotone density ratio by pooling the categories within equivalence classes K_s. These properties extend immediately to samples of size n, and imply in particular that the inferences depend only on $n_1, n_2, \cdots, n_r$ where n_s denotes the number of sample observations in K_s.

Since the intersection of n closed intervals is either a closed interval or is empty, the rule of combination [1], [2], [5] for passing from a single observation $j \in K$ to a sample of n observations $j_1, j_2, \cdots, j_n \in K$ yields inferences which are also expressible in terms of a random closed interval. The function $H(\theta_1, \theta_2)$ which characterizes the inferences based on a sample of size n is found up to a scalar multiplier by multiplying the $H(\theta_1, \theta_2)$ functions of the observations taken singly, i.e., the resultant $H(\theta_1, \theta_2)$ function is expressible as

$$\begin{aligned} H(\theta_1, \theta_2) &= C' \textstyle\prod_{h=1}^{n} [(\Pi(j_h, \theta_1)/\pi(j_h, \theta_1)) + ((1 - \Pi(j_h, \theta_2))/\pi(j_h, \theta_2))]^{-1} \\ &\qquad \text{if } \pi(j_h, \theta) > 0 \quad \text{for } \theta_1 \leqq \theta \leqq \theta_2, 1 \leqq h \leqq n, \\ &= 0 \quad \text{otherwise,} \end{aligned} \tag{3.11}$$

where C' is a normalizing constant.

The computation [4] of upper and lower probabilities from (3.11) directs attention to certain derived functions. Setting $\theta_1 = \theta_2 = \theta$ in (3.11) yields the likelihood function

$$H(\theta, \theta) = C' \textstyle\prod_{h=1}^{n} \pi(j_h, \theta). \tag{3.12}$$

If the bivariate cdf $H(\theta_1, \theta_2)$ defines an absolutely continuous bivariate distribution, then the partial derivative $H_1(\theta_1, \theta_2) = (\partial/\partial\theta_1)H(\theta_1, \theta_2)$ is also of interest at $\theta_1 = \theta_2 = \theta$, and it is easily shown that

$$H_1(\theta, \theta) = C' \textstyle\sum_{l=1}^{n} [D(j_l, \theta) \prod_{h=1, h \neq l}^{n} \pi(j_h, \theta)] = H(\theta, \theta) \sum_{l=1}^{n} E(j_l, \theta) \tag{3.13}$$

where

$$D(j, \theta) = \Pi(j, \theta)(d/d\theta)\pi(j, \theta) - \pi(j, \theta)(d/d\theta)\Pi(j, \theta) \tag{3.14}$$

and

$$E(j, \theta) = D(j, \theta)/\pi(j, \theta). \tag{3.15}$$

Note that $D(j, \theta) \equiv E(j, \theta) \equiv 0$ when $j \varepsilon K_1$. As indicated in [4], upper and lower probability inferences in the absolutely continuous case are expressible as

$$P^*([\theta_1, \theta_2]) = \int_{\theta_1}^{\theta_2} H_1(\theta, \theta)\, d\theta + H(\theta_1, \theta_1) \tag{3.16}$$

and

$$P_*([\theta_1, \theta_2]) = \int_{\theta_1}^{\theta_2} H_1(\theta, \theta)\, d\theta - H(\theta_2, \theta_2) + H(\theta_1, \theta_2). \tag{3.17}$$

The normalizing constant C' follows from $P^*([\alpha, \beta]) = 1$.

There are alternatives to formulas (3.13) through (3.17) which involve $H_2(\theta_1, \theta_2) = (\partial/\partial\theta_2)H(\theta_1, \theta_2)$ rather than $H_1(\theta_1, \theta_2)$. The alternatives need not be presented in detail since they follow by changing the sign of the parameter and applying the original formulas.

In most applications the integrals appearing in (3.16) or (3.17) will not be analytically tractable, and numerical approximations will be required. From the second line of (3.13), the integrals are formally the same as those involved in an ordinary Bayesian analysis with the role of the prior density played by $\sum_1^n E(j_l, \theta)$. For moderately large n, it will often happen that the likelihood factor has approximately the form of a normal density; specifically,

$$\log H(\theta, \theta) \doteq \log C'' - \tfrac{1}{2}nI(\hat{\theta})[\theta - \hat{\theta}]^2, \tag{3.18}$$

where C'' is a normalizing constant, $\hat{\theta}$ is the maximum likelihood estimator of θ and $I(\theta)$ is Fisher's information function. For extremely large n, the factor $\sum_1^n E(j_l, \theta)$ will be effectively constant over the narrow region about $\hat{\theta}$ over which $H(\theta, \theta)$ varies significantly, and the size of the factor $\sum_1^n E(j_l, \theta)$ in (3.13) will be such as effectively to eliminate the difference between upper and lower probabilities in (3.16) and (3.17). For extremely large n, therefore, the inferences will be practically indistinguishable from Bayesian inferences. For moderately large n, it may be possible to retain (3.18) while approximately $\log \sum_1^n E(j_l, \theta)$ by a linear or quadratic expansion about $\hat{\theta}$, and thence to obtain approximate expressions for upper and lower probability inferences requiring only tables of the normal cumulative and density functions.

Illustrations will be presented in Section 4, 5 where K is replaced by the real line or more generally by ordinary p-space R^p, and the family of discrete densities $\pi(\theta)$ is replaced by a family of densities $f(\mathbf{x}, \theta)$ defining absolutely continuous distributions over R^p. The upper and lower probability inferences derived for such illustrations should be conceived as limits of upper and lower probability inferences defined for an approximating sequence of finite multinomial families of distributions. In fact, however, the inferences will be derived from the obvious limiting forms of formula (3.11) and its consequences. In these early stages of development of generalized Bayesian inference, it seems unnecessary to justify rigorously this substitution of one limiting operation by another, but it may be worthwhile to reiterate the heuristic basis of the substitution, first mentioned in [1]. Consider a sequence of finite partitions $K^{(i)}$ of R^p and the corresponding sequence of families of distributions $\pi^{(i)}(\theta)$ over $K^{(i)}$ induced by the family $f(\mathbf{x}, \theta)$ over R^p. Suppose that the partitions $K^{(i)}$ increase in fineness in such a way that sums over $K^{(i)}$ approach integrals over R^p. It is plausible then that the corresponding sequence of expressions (2.15) tends to a similar expression with the sum over $K^{(i)}$ replaced by an integral over R^p and with the discrete densities $\pi^{(i)}(\theta)$ replaced by limiting $f(\mathbf{x}, \theta)$. From (2.11), the chain of development leads to (3.7) and thence to (3.11) and formulas for inferences. Note that the definition of monotone density ratio extends directly to $f(\mathbf{x}, \theta)$, and that Lemma 3.2 extends directly. As before, the equivalence classes coming from Lemma 3.2 may be pooled, meaning that the observation space R^p may sometimes be replaced by a simpler partition of itself.

4. Mixture families. If

$$\mathbf{p} = [p(1), p(2), \cdots, p(k)] \quad \text{and} \quad \mathbf{q} = [q(1), q(2), \cdots, q(k)]$$

are distinct given points in Π, then the *mixture family*

$$\pi(\theta) = \theta\mathbf{p} + (1 - \theta)\mathbf{q}, \qquad \text{for } 0 \leqq \theta \leqq 1, \tag{4.1}$$

is obviously a family with montone density ratio. It will be supposed, in order to eliminate irrelevant categories, that at least one of $p(i)$ and $q(i)$ is nonzero for all $i \varepsilon K$. The equivalence classes K_s of Lemma 3.2 are sets of $i \varepsilon K$ for which the ratio $p(i)/q(i)$ is constant, and the implied ordering of the K_s is the same as the numerical ordering of the associated ratios $p(i)/q(i)$. It will be assumed that the elements of K are arranged to possess an order consistent with that of the K_s, i.e., an order such that

$$p(1)/q(1) \leqq p(2)/q(2) \leqq \cdots \leqq p(k)/q(k). \tag{4.2}$$

To apply (3.11), it is convenient to set

$$P(j) = \textstyle\sum_{i \leqq j} p(i) \qquad \text{and} \quad Q(j) = \textstyle\sum_{i \leqq j} q(i), \tag{4.3}$$

so that

$$\Pi(j, \theta) = \theta P(j) + (1 - \theta)Q(j) \tag{4.4}$$

for $0 \leqq \theta \leqq 1$ and $j \varepsilon K$. Formula (3.13) reduces to

$$D(j, \theta) = p(j)Q(j) - q(j)P(j) \tag{4.5}$$

for $j \varepsilon K$. Since $D(j, \theta)$ does not depend on θ, the first form of $H_1(\theta, \theta)$ in (3.13) shows that $H_1(\theta, \theta)$ is a linear combination of the likelihood functions defined by omitting the observations one at a time, where each such likelihood function is a product of $n - 1$ monomials in θ. The integrals in (3.15) and (3.16) do not in general have simple analytic expressions, but for moderate n it would be easy to compute sequentially, adding observations one at a time, the coefficients in the polynomial $H_1(\theta, \theta)$ and thence to compute values of the integrals. For larger n, asymptotic approximations as discussed in Section 3 may be used.

EXAMPLE 1. Suppose that $k = 4$ and that the 4 categories arise from a 2×2 contingency table

$$\pi = \begin{bmatrix} \pi(1) & \pi(4) \\ \pi(3) & \pi(2) \end{bmatrix} \tag{4.6}$$

whose margins $\xi = \pi(1) + \pi(4)$ and $\eta = \pi(1) + \pi(3)$ are assumed fixed and known. Without loss of generality it may be assumed that $0 \leqq \xi \leqq \eta \leqq \frac{1}{2}$. The family of hypotheses with fixed margins has the form (4.1) where

$$\mathbf{p} = \begin{bmatrix} 0 & \xi \\ \eta & 1 - \xi - \eta \end{bmatrix} \quad \text{and} \quad \mathbf{q} = \begin{bmatrix} \xi & 0 \\ \eta - \xi & 1 - \eta \end{bmatrix}. \tag{4.7}$$

The arguments in (4.6) are arranged to satisfy (4.2).

If $n(i)$ denotes the number of sample observations in cell i for $i = 1, 2, 3, 4$, then (3.11) specializes to

$$\begin{aligned} H(\theta_1, \theta_2) &= C'[(1 - \theta_2)\xi]^{n(1)} \cdot [(((1 + \xi - \eta) - 2\theta_1\xi)/((1 - \eta) - \theta_1\xi)) \\ &\quad + (((\eta - \xi) + 2\theta_2\xi)/((1 - \eta) - \theta_2\xi))]^{-n(2)} \\ &\quad \cdot[((1 - \theta_1\xi)/((\eta - \xi) + \theta_1\xi)) + (\theta_2\xi/((\eta - \xi) + \theta_2\xi))]^{-n(3)} \\ &\quad \cdot [\theta_1\xi]^{n(4)}, \end{aligned} \tag{4.8}$$

while (3.12) simplifies to

$$H(\theta, \theta) = C'[\xi(1 - \theta)]^{n(1)}[(1 - \eta) - \theta\xi]^{n(2)}[(\eta - \xi) + \theta\xi]^{n(3)}[\theta\xi]^{n(4)}, \tag{4.9}$$

and (3.13) yields

$$\begin{aligned} H_1(\theta, \theta) = H(\theta, \theta)[&(\xi(1 - \xi - \eta)/((1 - \eta) - \theta\xi)) \\ &+ (\xi(1 + \eta - \xi)/((\eta - \xi) + \theta\xi)) + (\xi/\theta\xi)]. \end{aligned} \tag{4.10}$$

Inferences follow directly, but, as remarked above, analytic expressions for the required integrals are not generally available. An exception occurs when $\xi = \eta = \frac{1}{2}$. In this case categories 1 and 2 form an equivalence class K_1 while categories 3 and 4 form an equivalence class K_2. Pooling categories leads immediately to the binomial model which has been treated elsewhere [1], [4].

EXAMPLE 2. If the finite observation space K is replaced by R^p, then the mixture model should specify a pair of density functions $f(\mathbf{x})$ and $g(\mathbf{x})$ for $\mathbf{x} \varepsilon R^p$, while (4.1) becomes

$$\psi(\mathbf{x}, \theta) = \theta f(\mathbf{x}) + (1 - \theta)g(\mathbf{x}) \quad \text{for} \quad 0 \leqq \theta \leqq 1. \tag{4.11}$$

The equivalence classes determined by Lemma 3.2 may be labeled by $y =$

$\log [f(\mathbf{x})/g(\mathbf{x})]$, i.e., points $\mathbf{x}$ with a common y belong to the same class and the order of the classes agrees with the order of y values. Of course, any monotone strictly increasing function z of y would serve as well for describing equivalence classes. The model may be reduced to a consideration of z alone, so that (4.11) is replaced by

$$\psi^*(z, \theta) = \theta f^*(z) + (1 - \theta)g^*(z) \qquad \text{for } 0 \leqq \theta \leqq 1, \tag{4.12}$$

where for simplicity it will be assumed that the induced distributions of z are absolutely continuous with densities $\psi^*(z, \theta)$. Similarly, (4.4) is replaced by

$$\Psi^*(z, \theta) = \theta F^*(z) + (1 - \theta)G^*(z) \qquad \text{for } 0 \leqq \theta \leqq 1 \tag{4.13}$$

where

$$\Psi^*(z, \theta) = \int_{-\infty}^{z} \psi^*(u, \theta)\, du. \tag{4.14}$$

Formulas (3.11) through (3.18) apply directly with observations $j_1, j_2, \cdots, j_n$ replaced by $z_1, z_2, \cdots, z_n$, with $\pi(j_l, \theta)$ replaced by $\psi^*(z_l, \theta)$, with $\Pi(j_l, \theta)$ replaced by $\Psi^*(z_l, \theta)$, and with $D(j, \theta)$ and $E(j, \theta)$ replaced by

$$D(z, \theta) = f^*(z)G^*(z) - g^*(z)F^*(z) \tag{4.15}$$

and

$$E(z, \theta) = D(z, \theta)/\psi^*(z, \theta). \tag{4.16}$$

In particular, f and g could be taken to be multivariate normal distributions with different means but common nonsingular covariances. Then y would be a linear function of $\mathbf{x}$ and z could be chosen by a further linear transformation so that

$$\begin{aligned} f^*(z) &= \phi(z - \Delta), & g^*(z) &= \Phi(z + \Delta), \\ F^*(z) &= \phi(z - \Delta), & G^*(z) &= \Phi(z + \Delta), \end{aligned} \tag{4.17}$$

for some $\Delta > 0$, where ϕ and Φ denote the density and cumulative distribution functions of the standard $N(0, 1)$ distribution.

5. Location or scale parameter families. A location parameter family of distributions on the line may be represented by the family of density functions $f(x - \theta)$ on $-\infty < x < \infty$, where θ is a location parameter on $-\infty < \theta < \infty$ and f or F specify a given absolutely continuous distribution on the line. This location parameter family has a monotone density ratio with the ordering of equivalence classes (Lemma 3.2) agreeing with the ordering of observables if and only if for each $t > 0$ the ratio $f(x + t)/f(x)$ is a monotone nonincreasing function of x for all x such that the ratio is not of the form 0/0. Under this condition, formula (3.9) yields

$$\begin{aligned} H(\theta_1, \theta_2) &= C[(F(x - \theta_1)/f(x - \theta_1)) + ((1 - F(x - \theta_2))/f(x - \theta_2))]^{-1} \\ &\qquad \text{if } f(x - \theta) > 0 \quad \text{on } \theta_1 \leqq \theta \leqq \theta_2, \\ &= 0 \quad \text{otherwise,} \end{aligned} \tag{5.1}$$

where x denote a single observation.

Under exponentiation a location parameter family becomes a scale parameter family, so that the theory of the preceding paragraph may also be applied to scale parameter families. It is convenient, however, to have the monotonicity condition and the formula for $H(\theta_1, \theta_2)$ expressed directly in terms of the density $g(x)$ and cumulative $G(x)$ on $x > 0$ which determine the scale parameter family with densities $\theta^{-1}g(x\theta^{-1})$ for $\theta > 0$. This scale parameter family has a monotone density ratio with ordering of equivalence classes agreeing with ordering of observables if and only if for each $u > 1$ the ratio $g(xu)/g(x)$ is a monotone non-increasing function of x for all x such that the ratio is not of the form 0/0. The analog of (5.1) is

$$H(\theta_1, \theta_2) = C[(G(x\theta_1^{-1})/\theta_1^{-1}g(x\theta_1^{-1})) + ((1 - G(x\theta_2^{-1}))/\theta_2^{-1}g(x\theta_2^{-1}))]^{-1} \quad \text{if} \quad g(x\theta^{-1}) > 0 \quad \text{for} \quad \theta_1 \leqq \theta \leqq \theta_2, \qquad = 0 \quad \text{otherwise,} \tag{5.2}$$

for a single observation $x > 0$.

Two examples of location parameter families and two examples of scale parameter families will now be introduced.

Example 3. *Uniform location parameter family.* Take

$$f(x) = 1 \quad \text{for} \quad 0 \leqq x \leqq 1, \qquad = 0 \quad \text{otherwise.} \tag{5.3}$$

When $0 < t \leqq 1$,

$$\begin{aligned} f(x + t)/f(x) &= +\infty \quad &\text{if} \quad -t \leqq x < 0 \\ &= 1 &\text{if} \quad 0 \leqq x \leqq 1 - t \\ &= 0 &\text{if} \quad t - 1 < x \leqq 1, \end{aligned} \tag{5.4}$$

and is of the form 0/0 for other values of x. When $t > 1$,

$$\begin{aligned} f(x + t)/f(x) &= +\infty \quad &\text{if} \quad -t \leqq x \leqq 1 - t \\ &= 0 &\text{if} \quad 0 \leqq x \leqq 1, \end{aligned} \tag{5.5}$$

and is of the form 0/0 for other values of x. The required monotonicity property is therefore satisfied. From (5.1)

$$\begin{aligned} H(\theta_1, \theta_2) &= C[1 + \theta_2 - \theta_1]^{-1} \quad &\text{for} \quad x - 1 \leqq \theta_1 \leqq \theta_2 \leqq x, \\ &= 0 &\text{otherwise.} \end{aligned} \tag{5.6}$$

The $H(\theta_1, \theta_2)$ function appropriate for a sample of n observations $x_1, x_2, \cdots, x_n$ is the product of the $H(\theta_1, \theta_2)$ functions yielded by the individual observations, namely

$$\begin{aligned} H(\theta_1, \theta_2) &= C'[1 + \theta_2 - \theta_1]^{-n} \quad &\text{for} \quad x_M - 1 \leqq \theta_1 \leqq \theta_2 \leqq x_m \\ &= 0 &\text{otherwise,} \end{aligned} \tag{5.7}$$

where x_m and x_M denote respectively the minimum and maximum of the n sample observations.

Considered as a bivariate cumulative distribution function, $H(\theta_1, \theta_2)$ in (5.7) assigns all its probability to the closed triangle $x_M - 1 \leqq \theta_1 \leqq \theta_2 \leqq x_m$ which is illustrated in Figure 1. The measure is conveniently considered as a sum of four measures. Writing

$$d = x_m - (x_M - 1), \tag{5.8}$$

the first component measure consists of

$$C'(1 + d)^{-n} \tag{5.9}$$

located at the point $\theta_1 = x_M - 1$, $\theta_2 = x_m$. The second component consists of

$$C'[1 - (1 + d)^{-n}] \tag{5.10}$$

distributed continuously along the line segment $\theta_1 = x_M - 1$, $x_M - 1 \leqq \theta_2 \leqq x_m$ according to the density function

$$C'(n(1 + \theta_2 - x_M + 1)^{-n-1}. \tag{5.11}$$

The third component is the mirror image of the second, along the line segment $x_M - 1 \leqq \theta_1 \leqq x_m$, $\theta_2 = x_m$, while the fourth component consists of

$$C'[n d - 1 + (1 + d)^{-n}] \tag{5.12}$$

distributed continuously over the triangle $x_M - 1 \leqq \theta_1 \leqq \theta_2 \leqq x_m$ according to the density function

$$C'n(n + 1)(1 + \theta_2 - \theta_1)^{-n-2}. \tag{5.13}$$

Summing the four components and setting the total probability to unity, one finds

$$C'(n d + 1)^{-1}. \tag{5.14}$$

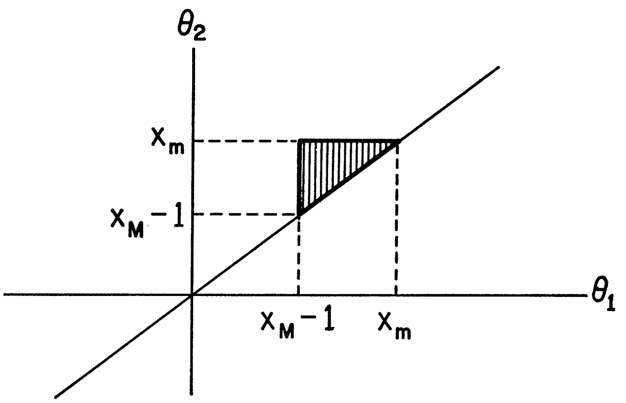

Fig. 1. The shaded triangle represents the region of closed intervals $[\theta_1, \theta_2]$ which carry posterior probability one in Example 3.

As illustrations of the inferences implied by (5.7) it is easily checked that

$$P^*(\theta \leqq t) = (n(t - x_M + 1) + 1)/(n\,d + 1) \tag{5.15}$$

while

$$P_*(\theta \leqq t) = n(t - x_M + 1)/(n\,d + 1) \tag{5.16}$$

for $x_M - 1 \leqq t \leqq x_m$.

Example 3 has the interesting feature that, unlike the other examples of this paper, the differences between meaningful upper probabilities and their corresponding lower probabilities do not tend to zero as sample size increases. This phenomenon is associated with superefficient estimation. Since for given θ

$$\delta = n\,d \tag{5.17}$$

has in the limit as $n \to \infty$ a standard exponential distribution, it might be expected that various exponential approximations would be satisfactory for large n. For example, (5.9) is approximately $C' \exp(-\delta)$, and setting $\phi = n(\theta - x_M + 1)$ one may approximate (5.11) and (5.13) by

$$C' \exp(-\theta_2) \tag{5.18}$$

for $0 \leqq \phi_2 \leqq \delta$ and

$$C' \exp(-[\phi_2 - \phi_1]) \tag{5.19}$$

for $0 \leqq \phi_1 \leqq \phi_2 \leqq \delta$, respectively. Correspondingly, (5.15) and (5.16) may be written

$$P^*(\phi \leqq h) = (h + 1)/(\delta + 1) \tag{5.20}$$

and

$$P_*(\phi \leqq h) = h/(\delta + 1) \tag{5.21}$$

for $0 \leqq h \leqq \delta$, respectively, exactly for any n.

Example 4. *Normal location parameter family.* If $f(x)$ is taken to be $\phi(x) = 2\pi^{-\frac{1}{2}} \exp(-\frac{1}{2}x^2)$, the ratio $f(x + t)/f(x)$ becomes $\exp(-xt - \frac{1}{2}t^2)$ which is monotone decreasing in x for all $t > 0$. Given a sample $x_1, x_2, \cdots, x_n$ the inferences about θ are determined from (3.11) which takes the specific form

$$H(\theta_1, \theta_2) = C \prod_{i=1}^{n} [\exp(\tfrac{1}{2}(x_i - \theta_1)^2)\Phi(x_i - \theta_1) + \exp(\tfrac{1}{2}(x_i - \theta_2)^2)\{1 - \Phi(x_i - \theta_2)\}]^{-1}. \tag{5.22}$$

Since the distribution defined by (5.22) is absolutely continuous over the half plane $-\infty < \theta_1 \leqq \theta_2 < \infty$, formulas (3.16) and (3.17) may be used for the calculation of upper and lower probabilities. The functions $H(\theta_1, \theta_2)$, $H(\theta, \theta)$, and $H_1(\theta, \theta)$ appearing in (3.16) and (3.17) are given respectively by (5.22),

$$H(\theta, \theta) = C'' \exp(-\tfrac{1}{2}(n - 1)s^2 - \tfrac{1}{2}n[\bar{x} - \theta]^2), \tag{5.23}$$

and

$$H_1(\theta, \theta) = H(\theta, \theta) \sum_{i=1}^n E(x_i, \theta), \tag{5.24}$$

where

$$E(x, \theta) = \phi(x - \theta) + [x - \theta]\Phi(x - \theta), \tag{5.25}$$

and $\bar{x}$ and s^2 denote the conventional sufficient statistics

$$\bar{x} = n^{-1} \sum_{i=1}^n x_i \qquad \text{and} \quad s^2 = (n-1)^{-1} \sum_{i=1}^n [x_i - \bar{x}]^2 \tag{5.26}$$

which do not, however, determine the inferences.

The integrations required by (3.16) and (3.17) could be carried somewhat further analytically, but the results are not very elegant, nor are they practically convenient since n bivariate normal table lookups are indicated for each integral. Numerical approximations will be required for practical use, but these are not developed here.

EXAMPLE 5. *Exponential scale parameter family.* When $g(x) = 1 - G(x) = \exp(-x)$, formula (5.2) becomes

$$H(\theta_1, \theta_2) = C[\{\theta_2 - \theta_1\} + \theta_1 e^{x/\theta_1}]^{-1} \tag{5.27}$$

on $0 \leqq \theta_1 \leqq \theta_2 < \infty$. More generally, given a sample $x_1, x_2, \cdots, x_n$, the inferences are determined by

$$H(\theta_1, \theta_2) = C' \prod_{i=1}^n [\{\theta_2 - \theta_1\} + \theta_1 e^{x_i/\theta_1}]^{-1}, \tag{5.28}$$

for $0 \leqq \theta_1 \leqq \theta_2 < \infty$. As always

$$H(\theta, \theta) = C'\theta^{-n} \exp(-\sum_{i=1}^n x_i/\theta) \tag{5.29}$$

gives the ordinary likelihood function of θ. The function $H_1(\theta, \theta)$ required for (3.16) and (3.17) is

$$\begin{aligned} H_1(\theta, \theta) &= H(\theta, \theta) \sum_{i=1}^n (1/\theta)[((x_i/\theta) - 1) + e^{-x_i/\theta}] \\ &= C'\theta^{-n-1} e^{-t/\theta}[((t/\theta) - n) + \sum_{i=1}^n e^{-x_i/\theta}] \end{aligned} \tag{5.30}$$

where $t = \sum_1^n x_i$ denotes the ordinary sufficient statistic which no longer completely determines the inferences.

Regarded as a density function, $H_1(\theta, \theta)$ is a linear combination of densities of random variables which are scaled inverses of random variables with gamma densities. Thus, setting

$$C_n(u) = (1/\Gamma(n)) \int_0^u v^{n-1} e^{-v}\, dv, \tag{5.31}$$

it is easily checked that

$$\begin{aligned} &\int_{\theta_1}^{\theta_2} H_1(\theta, \theta)\, d\theta \\ &\quad = C'[\Gamma(n+1)t^{-n}\{C_{n+1}(t/\theta_1) - C_{n+1}(t/\theta_2)\} \\ &\qquad - n\Gamma(n)t^{-n}\{C_{n+1}(t/\theta_1) - C_{n+1}(t/\theta_2)\} \\ &\qquad + \sum_{i=1}^n \Gamma(n)(t + x_i)^{-n}\{C_n((t + x_i)/\theta_1) - C_n((t + x_i)/\theta_2)\}]. \end{aligned} \tag{5.32}$$

The normalizing constant C' follows from the relation $1 = P^*([0, \infty]) = \int_0^\infty H_1(\theta, \theta)\, d\theta$, so that (5.32) may finally be given as

$$\begin{aligned} \int_{\theta_1}^{\theta_2} H_1(\theta, \theta)\, d\theta = [\{C_{n+1}(t/\theta_1) - C_{n+1}(t/\theta_2)\} - \{C_n(t/\theta_1) - C_n(t/\theta_2)\} \\ + \textstyle\sum_{i=1}^n \{C_n((t + x_i)/\theta_1) - C_n((t + x_i)/\theta_2)\} \\ \cdot (1 + (x_i/t))^{-n}]/[\textstyle\sum_{i=1}^n (1 + (x_i/t))^{-n}]. \end{aligned} \tag{5.33}$$

From (5.28), (5.29), and (5.33) together with the general formulas (3.16) and (3.17), expressions for $P^*([\theta_1, \theta_2])$ and $P_*([\theta_1, \theta_2])$ may be immediately written down.

These inferences may be compared with the inferences yielded by a formal Bayesian analysis with prior pseudodensity θ^{-1}. For the latter,

$$P^*([\theta_1, \theta_2]) = P_*([\theta_1, \theta_2]) = C_n(t/\theta_1) - C_n(t/\theta_2). \tag{5.34}$$

It is clear that for large n the inferences become indistinguishable. For small n, (5.33) offers as a tradeoff for increased complexity, the use of a logical system which does not demand the introduction of a necessarily somewhat arbitrary prior distribution.

EXAMPLE 6. *Normal scale parameter family.* Take for $g(x)$ and $G(x)$, $\psi(x)$ and $\Psi(x)$ defined as

$$\psi(x) = 2\phi(x) = (2/\pi)^{\frac{1}{2}} \exp(-\tfrac{1}{2}x^2) \tag{5.35}$$

and

$$\Psi(x) = 2\Phi(t) - 1 = \int_0^x \psi(u)\, du. \tag{5.36}$$

Since $\psi(xu)/\psi(x) = \exp(-\frac{1}{2}[u^2 - 1]x)$ is monotone decreasing in x for fixed $u > 1$, formula (5.2) may be applied directly. More generally, formulas (3.11) through (3.17) apply directly to a sample $x_1, x_2, \cdots, x_n$. In particular

$$H(\theta, \theta) = C'''\theta^{-n} \exp(-\tfrac{1}{2} \textstyle\sum_{i=1}^n x_i^2/\theta) \tag{5.37}$$

and

$$\begin{aligned} H_1(\theta, \theta) &= H(\theta, \theta) \textstyle\sum_{i=1}^n E(x, \theta) \\ &= C'''\theta^{-n-1} \exp(-\tfrac{1}{2} \textstyle\sum_{i=1}^n x_i^2/\theta)[\textstyle\sum_{i=1}^n \{(x_i/\theta)\psi(x_i/\theta) \\ &\quad + [(x_i/\theta)^2 - 1]\Psi(x_i/\theta)\}]. \end{aligned} \tag{5.38}$$

As with the normal location parameter example, integration of (5.38) will require numerical techniques as yet undeveloped.

REFERENCES

[1] DEMPSTER, A. P. (1966). New methods for reasoning towards posterior distributions based on sample data. *Ann. Math. Statist.* **37** 355–374.

[2] DEMPSTER, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.* **38** 325–339.

[3] DEMPSTER, A. P. (1967). Upper and lower probability inferences based on a sample from a finite univariate population. *Biometrika* **54** 515–528.

[4] DEMPSTER, A. P. (1968). Upper and lower probabilities generated by a random closed interval. *Ann. Math. Statist.* **39** 957–966.

[5] DEMPSTER, A. P. (1968). A generalization of Bayesian inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **30** 205–247.