On Finitely Additive Priors, Coherence, and Extended Admissibility

by

David Heath
Cornell University

William Sudderth*
University of Minnesota

Technical Report No. 278
November, 1976

*Research supported by National Science Foundation Grant MCS76-03940

Abstract

A decision maker is seen to be coherent in the sense of de Finetti if, and only if, his probabilities are computed in accordance with some finitely additive prior. If a bounded loss function is specified, then a decision rule is extended admissible (i.e. not uniformly dominated) if, and only if, it is Bayes for some finitely additive prior. However, if an improper, countably additive prior is used, then decisions need not cohere and decision rules need not be extended admissible. Invariant, finitely additive priors are found and their posteriors calculated for a class of problems which include translation parameter problems.

AMS 1970 subject classifications. 62C10, 60A05.

Key words and phrases. Coherence, extended admissibility, finite additivity, decision theory, invariant priors, improper priors.

0. Introduction. The main object of this note is to present a simple formulation of statistical decision problems using finitely additive probabilities as has been recommended by Bruno de Finetti [7] and Leonard J. Savage [13]. One difficulty is that the theory of conditional distributions for finitely additive probabilities is relatively new and still incomplete. (Some major results are in Lester Dubins' paper [4].) Nevertheless, it is possible to characterize coherent conditional odds functions as being the posteriors of finitely additive priors, a result which was proved by Freedman and Purves [8] for finite spaces. There is probably an interesting extension of their theorem in a countably additive setting as well, but we suspect that such a result would be more difficult than Theorem 1 and its corollary. Theorem 1 is proved by a simple separation argument and the same type of argument also yields a characterization of the extended admissible decision rules as being just the Bayes rules, at least in the case of a bounded loss function (Theorem 2). The computation of the Bayes rule for a given finitely additive prior typically requires its posterior and Theorem 3 enables one to find the posterior for a natural prior in a class of generalized translation parameter problems including the Behrens-Fisher problem. The final section treats briefly the relationship of improper, countably additive priors to the finitely additive theory. Many of the improper priors which are commonly employed lead to posteriors which could have been reached from a proper, finitely additive prior. However, this is not the case for every improper prior and such priors can result in incoherence and uniformly inadmissible decision rules (Example 5.2).

1. Preliminaries. Let S be a nonempty set. A probability π on S here means a finitely additive probability measure defined on all subsets of S. Denote by $P(S)$ the collection of all probabilities on S and by $L(S)$ the space of all bounded, real-valued functions with domain S. To each π in $P(S)$ corresponds a unique nonnegative linear functional ℓ on $L(S)$ such that

$$(1.1) \qquad \pi(E) = \ell(1_E)$$

for all $E \subseteq S$. Furthermore, every nonnegative linear functional ℓ on $L(S)$ such that $\ell(1_S) = 1$ determines a π in $P(S)$ by (1.1). Henceforth, such an ℓ is identified with the corresponding π and, as suggested by de Finetti, the indicator function 1_E is identified with the set E. The value of π at a function f will be written $\pi(f)$, $\int f\, d\pi$, or $\int f(s)\pi(ds)$.

The following lemma is a slight improvement of Theorem 1 in [10], and the proof is almost the same.

Lemma 1. *Let $F \subseteq L(S)$. Then (i) there exists a π in $P(S)$ such that $\pi(f) \geq 0$ for all $f \in F$ if, and only if, (ii) every finite, convex combination of functions in F has a nonnegative supremum.*

Proof: Because π is a nonnegative linear functional, it is trivial that (i) implies (ii). Now assume (ii) and give $L(S)$ the sup norm topology. Let C be the set of all functions of the form $a_1f_1 + \ldots + a_nf_n$ where $a_i \geq 0$ and $f_i \in F$ for all i. Define $N = \{f \in L(S): \sup f < 0\}$. Then C and N are convex sets which are disjoint by (ii). Furthermore, the interior of N is nonempty since it contains, for example, the function $f \equiv -1$. By a standard separation theorem ([6], p. 417), there is a nontrivial, continuous linear functional ℓ and a real number r such that $\ell(f) \geq r$ on C and $\ell(f) \leq r$ on N. Because 0 is a limit point of both C and N and ℓ is continuous, the constant r must equal 0. Thus ℓ is a nonnegative functional and the π if (i) can be taken to be $a^{-1}\ell$ where $a = \ell(1)$. $\square$

A π in $P(S)$ can be extended to functions f on S, which are bounded below, by taking the inner integral thus.

(1.2) $$\pi(f) = \sup_n \pi(f \wedge n).$$

Here $f \wedge n$ is the minimum of f with n and n ranges over the set of natural numbers.

2. Coherence and a theorem of Freedman and Purves. Let X and Θ be nonempty sets to be thought of as the set of possible observations and possible states of nature respectively. Let p be a conditional probability on Θ given X, that is, a mapping from Θ into $P(X)$. For $\theta \in \Theta$ and $B \subseteq X$, $p(\theta)(B)$ will sometimes be written $p_\theta(B)$ or $p(B|\theta)$. Consider, informally at first, a game with three participants known as the bookie, the gambler, and the master of ceremonies (MC). The MC selects $\theta \in \Theta$ and then, using the probability $p(\theta)$, selects $x \in X$. Next the MC reveals x to the bookie and the gambler. Then the bookie posts odds on subsets of Θ, after which the gambler places a finite number of bets. Finally, the MC reveals θ and the bookie and the gambler settle up. As shown by Freedman and Purves in [8] for the case when Θ and X are finite, the bookie must post odds consistent with some posterior distribution or else the gambler can attain positive expected winnings for all values of θ. The object of this section is to extend their result to infinite sets.

It is convenient, for later applications, to equip Θ and X with σ-fields $\mathcal{B}_1 = \mathcal{B}(\Theta)$ and $\mathcal{B}_2 = \mathcal{B}(X)$ of subsets. A conditional odds function q is a mapping from $X \times \mathcal{B}_1$ to the unit interval. The interpretation is that, after observing $x \in X$, the bookie posts the odds $q(x, B): 1 - q(x, B)$ on sets $B \in \mathcal{B}_1$. A simple betting system is a pair (A, b) where $A \in \mathcal{B}_1 \times \mathcal{B}_2$ and b is a bounded, real-valued, $\mathcal{B}_2$-measurable function on X. Here the interpretation is that, after observing x, the gambler stakes $b(x)q(x, A^x)$ on the event $A^x = \{\theta:(\theta, x) \in A\}$. The payoff from the bookie to the gambler is

$$\varphi(\theta, x) = b(x)[A(\theta, x) - q(x, A^x)]$$

and the expected payoff, for a given θ, is

$$E(\theta) = \int\varphi(\theta, x)\, p(dx|\theta).$$

In an unconditional setting ([7], p. 87), de Finetti has defined coherence to be the absence of a combination of bets which results in a uniformly positive loss. Similarly, in this note an odds function q is called <u>coherent</u> if there does not exist a finite number of simple betting systems $(A_1, b_1), \ldots, (A_n, b_n)$ with associated expected payoffs $E_1, \ldots, E_n$ such that

$$(2.1)\qquad \inf_{\theta}\{E_1(\theta) + \ldots + E_n(\theta)\} > 0.$$

This definition could be modified by allowing infinitely many bets or by weakening the requirement of uniform positivity in (2.1). Buehler [3] has also explored "preference-reversal" coherence, a notion of coherence formulated without betting.

An element π of $P(\Theta)$ is called a prior. Each prior π together with the conditional probability p determines a <u>marginal</u> $m \in P(X)$ by the formula

$$(2.2)\qquad m(B) = \int p(B|\theta)\pi(d\theta)$$

for $B \subseteq X$.

<u>Theorem 1</u>. <u>A conditional odds function q is coherent if, and only if, there is a prior π such that, for every simple betting system (A, b),</u>

$$(2.3)\qquad \int\int b(x)A(\theta, x)\; p(dx|\theta)\pi(d\theta) = \int b(x)q(x, A^x)m(dx),$$

<u>where m is the marginal of (π, p) on X.</u>

<u>Proof:</u> Apply Lemma 1 with $S = \Theta$ and F the collection of all functions $E(\theta)$ which are expected payoff functions of a simple betting system. Use the fact that F is closed under multiplication by a real constant and also use the definition of m. □

A conditional odds function q can be regarded as a conditional probability on Θ given X if, for each x, $q(\cdot|x) \equiv q(x, \cdot)$ is an element of $P(\Theta)$.

There is good reason for the bookie to select an odds function q which is a conditional probability since otherwise, for those x at which $q(x, \cdot)$ is not in $P(\Theta)$, the gambler can inflict a sure loss ([10], Theorem 5).

A <u>posterior for a prior</u> π (<u>relative to</u> $\mathcal{B}_1 \times \mathcal{B}_2$) is a conditional probability q on Θ given X such that for all bounded, real-valued, $\mathcal{B}_1 \times \mathcal{B}_2$-measurable functions φ on $\Theta \times X$,

$$(2.4) \qquad \iint \varphi(\theta, x) p(dx|\theta)\pi(d\theta) = \iint \varphi(\theta, x) q(d\theta|x) m(dx),$$

where m is the marginal of (π, p) on X.

<u>Corollary 1</u>. <u>A conditional probability</u> q <u>on</u> Θ <u>given</u> X <u>is coherent if, and only if, it is a posterior for some prior</u> π.

The corollary is not as satisfactory as it may seem. For as Lester Dubins [4] has shown, not every prior has a posterior. Thus a question implicit in the corollary and of some independent interest is for which priors do posteriors exist. A special case is treated in Section 4.

3. Decision problems and extended admissibility. To formulate an abstract statistical decision problem, introduce, in addition to the sets X, Θ and conditional probability p, a nonempty action set A and a loss function L which maps $\Theta \times A$ to the nonnegative reals. A (randomized) decision rule δ is a conditional probability on A given X. The risk $r(\delta)$ of a decision rule δ is defined by

$$(3.1) \qquad r(\delta)(\theta) = r(\theta, \delta)$$

$$= \int L'(\theta, \delta(x))p(dx|\theta),$$

where

$$(3.2) \qquad L'(\theta, \gamma) = \int L(\theta, a)\gamma(da)$$

for $\theta \in \Theta$, $x \in X$, $\gamma \in P(A)$. A mapping f from X to A is a pure decision rule. Such an f can also be regarded as a randomized rule if, for each x, $f(x)$ is identified with that element of $P(A)$ which assigns probability one to $\{f(x)\}$.

Notions of admissibility will be formulated relative to a fixed collection D of decision rules and elements of D will be called D-rules. D can be thought of, at present, as the collection of all decision rules, but, in some applications to come, it will be taken to be the set of decision rules which are measurable in an appropriate sense. As usual, a D-rule δ_o is admissible if there does not exist a D-rule δ such that

$$r(\theta, \delta) \leq r(\theta, \delta_o)$$

for all θ with strict inequality holding for some θ. Call δ_o ε-admissible if there does not exist a D-rule δ for which

$$r(\theta, \delta) \leq r(\theta, \delta_o) - \varepsilon$$

for all θ. As in Blackwell and Girshick [2], δ_o is extended admissible if it is ε-admissible for every $\varepsilon > 0$. As is easily seen, δ_o is extended admissible if, and only if,

$$(3.3) \qquad \inf_{\delta} \sup_{\theta} [r(\theta, \delta) - r(\theta, \delta_o)] \geq 0.$$

Obviously, every admissible rule is extended admissible. However, the converse is false, as Example 4.1 (or a simpler example) shows.

A decision rule δ_o is Bayes for a prior π if

$$(3.4) \qquad \pi(r(\delta_o)) \leq \pi(r(\delta))$$

for all decision rules δ.

Theorem 2. Every Bayes rule is extended admissible. If the loss function L is bounded, and the set D is convex, then every extended admissible rule is Bayes.

Proof: The first assertion is a trivial consequence of the definitions. To prove the second, suppose δ_o is extended admissible. Set $S = \Theta$ and take F to be the collection of all functions $r(\delta) - r(\delta_o)$, where δ is a decision rule. The collection D is convex by hypothesis and, consequently, so is F because

$$p_1 r(\delta_1) + \ldots + p_n r(\delta_n) = r(p_1\delta_1 + \ldots + p_n\delta_n)$$

for $p_i \geq 0$, $p_1 + \ldots + p_n = 1$. By (3.3), Lemma 1(ii) and, hence, (i) hold. Obviously, δ_o is Bayes for the π of Lemma 1(i). □

The remainder of this section is devoted to the problem of calculating a Bayes rule for a given prior. The conventional solution begins with the calculation of the posterior distribution. As already mentioned, a posterior need not exist in the general finitely additive setting. However, if there is a posterior, the Bayes rule can be calculated in the usual fashion by minimizing posterior loss.

Two cases will be considered. Until the completion of Lemma 3.1 below, take D to be the set of all decision rules. Measurability restrictions will be imposed later. A posterior q for a prior π is <u>complete</u> if (2.4) holds for all bounded, real-valued functions φ on $\Theta \times X$.

<u>Lemma 3.1</u>. <u>Let</u> q <u>be a complete posterior for the prior</u> π <u>and suppose that, for each</u> $x \in X$, <u>the infimum</u>

$$\inf\{\int L'(\theta, \gamma)q(d\theta|x): \gamma \in P(A)\}$$

<u>is achieved at</u> $\gamma = \delta_o(x)$. <u>Then, if</u> L <u>is bounded,</u> δ_o <u>is Bayes for</u> π.

<u>Proof</u>: Let m be the marginal for (π, p) and let δ be a decision rule. Then

$$\begin{aligned}\pi(r,\delta_o)) &= \int\int L'(\theta, \delta_o(x))p(dx|\theta)\pi(d\theta) \\ &= \int\int L'(\theta, \delta_o(x))q(d\theta|x)m(dx) \\ &\leq \int\int L'(\theta, \delta(x))q(d\theta|x)m(dx) \\ &= \int\int L'(\theta, \delta(x))p(dx|\theta)\pi(d\theta) \\ &= \pi(r(\delta)).\end{aligned}$$

□

Assume from now until the end of Lemma 3.3 that Θ, X, and A are equipped with these σ-fields of their subsets: $\mathcal{B}_1 = \mathcal{B}(\Theta)$, $\mathcal{B}_2 = \mathcal{B}(X)$, and $\mathcal{B}_3 = \mathcal{B}(A)$. Let $C(A)$ be the collection of those probabilities γ in $P(A)$ which are countably additive when restricted to $\mathcal{B}_3$ and identify each such γ with its restriction to $\mathcal{B}_3$. There is a natural σ-field Σ of subsets of $C(A)$; namely, Σ is the least σ-field such that, for every $E \in \mathcal{B}_3$, the mapping $\gamma \to \gamma(E)$ is measurable from $C(A)$ to the unit interval equipped with its usual Borel sets. (This measurable structure on $C(A)$ was explored by Dubins and Freedman [5].) Take D to be the set of decision rules δ which are measurable maps from X into $C(A)$. In more common parlance, D is the collection of regular conditional probabilities on A given X. Assume also that p is a regular conditional probability on X given Θ and that L is $\mathcal{B}_1 \times \mathcal{B}_3$-measurable. A decision problem is called measurable if all the assumptions of this paragraph hold.

Lemma 3.2. *For a measurable decision problem, the function* $(\theta, x) \to L'(\theta, \delta(x))$ *is* $\mathcal{B}_1 \times \mathcal{B}_2$-*measurable for every D-rule* δ.

Proof: Almost immediate from the definition of L' in (3.2) and the Lemma in Section 5 of [15]. □

Lemma 3.3. *Let q be a posterior for* π *relative to* $\mathcal{B}_1 \times \mathcal{B}_2$ *and let* δ_o *be a D-rule. Then each of the following conditions implies its successor.*

(a) For every $x \in X$,

$$\int L'(\theta, \delta_o(x))q(d\theta|x) = \inf_{a\in A} \int L(\theta, a)q(d\theta|x).$$

(b) For every $x \in X$,

$$\int L'(\theta, \delta_o(x))q(d\theta|x) = \inf_{\gamma\in C(A)} \int L'(\theta, \gamma)q(d\theta|x).$$

(c) *If L is bounded, then* δ_o *is Bayes for* π.

Proof: To see that (a) implies (b), calculate as follows:

$$\begin{aligned}\int L'(\theta, \gamma)q(d\theta|x) &= \iint L(\theta, a)\gamma(da)q(d\theta|x)\\ &= \iint L(\theta, a)q(d\theta|x)\gamma(da)\\ &\geq \iint L'(\theta, \delta_o(x))q(d\theta|x)\gamma(da)\\ &= \int L'(\theta, \delta_o(x))q(d\theta|x).\end{aligned}$$

Here the successive lines are by (3.2), by Fubini's theorem, by (a), and obvious, respectively.

The proof that (b) implies (c) is a calculation like the one used in the proof of Lemma 3.1, but it relies on Lemma 3.2 for the measurability of L'. □

Most of the results presented so far require that L be bounded. The following lemma treats a general nonnegative loss function, and, while the result is far from satisfactory, it does find application in Example 4.1.

Lemma 3.4. *Let δ_o be a D-rule which is Bayes for the prior π when the loss function is $L_n = L \wedge n$ for every $n = 1, 2, \ldots$. Suppose also that $\pi(r_n(\delta_o)) \to \pi(r(\delta_o))$ as $n \to \infty$, where r_n is the risk function corresponding to L_n. Then δ_o is Bayes for π in the original decision problem.*

Proof: Let δ be a D-rule and calculate thus.

$$\pi(r(\delta)) \geq \pi(r_n(\delta)) \geq \pi(r_n(\delta_o)) \to \pi(r(\delta_o)).$$

The first inequality is obvious. The second inequality and the convergence are by hypothesis. □

4. Posteriors for translation families. The problem considered in this section is the existence and calculation of the posterior for a (generalized) translation family of distributions when the prior is invariant in an appropriate sense. Assume throughout this section that Θ is a locally compact topological group and that $X = \Theta$. Suppose also that Θ is amenable in the sense that there exists a finitely additively, left-invariant probability on the Borel subsets $\mathcal{B}$ of Θ. The following condition is equivalent to amenability ([9], Theorem 3.6.2) and also proves useful in calculations. Let h be the left Haar measure on $\mathcal{B}$.

> Follner's Condition. For every $\epsilon > 0$ and every compact set $K \subseteq \Theta$, there is a Borel set $U \subseteq \Theta$ with $0 < h(U) < \infty$ and
>
> $$(4.1) \qquad \left|\frac{h(\theta U \Delta U)}{h(U)}\right| < \epsilon$$
>
> for all $\theta \in K$. (Here $A \Delta B = (A - B) \cup (B - A)$.)

Using Follner's Condition, one can find a left-invariant mean π as follows: Consider the collection $\{(\epsilon, K): \epsilon > 0, \ K \subseteq \Theta, \ K \text{ compact}\}$ which is directed under the relation '$\leq$' where $(\epsilon_1, K_1) \leq (\epsilon_2, K_2) \Leftrightarrow \epsilon_1 \geq \epsilon_2$ and $K_1 \subseteq K_2$. For each (ϵ, K), let $U_{(\epsilon, K)}$ satisfy (4.1) and define

$$(4.2) \qquad \pi_{(\epsilon, K)}(B) = \frac{h(U_{(\epsilon, K)} \cap B)}{h(U_{(\epsilon, K)})},$$

for every $B \in \mathcal{B}$. Then $\{\pi_{(\epsilon, K)}\}$ is a net in the space $[0, 1]^{\mathcal{B}}$, and this space is compact when given the product topology. Consequently, there is a subnet $\{\pi_\alpha\}$ (with corresponding sets U_α) which converges to an element π of $[0, 1]^{\mathcal{B}}$. In other words, $\pi_\alpha(B) \to \pi(B)$ for every $B \in \mathcal{B}$, and, hence, $\int f \, d\pi_\alpha \to \int f \, d\pi$ for all bounded, $\mathcal{B}$-measurable f. It follows easily that π is finitely additive and left-invariant. Throughout this section, this π

will be used as a prior. (As pointed out in [9], there are often many such left-invariant means π. Anyone of them will do in what follows.)

If a fixed element p_e of $P(X)$ is "translated" by elements of the group $X = \Theta$, then it generates a conditional probability p on X given Θ which is defined by

$$(4.3) \qquad \int f(x)p_\theta(dx) = \int f(x\theta)p_e(dx)$$

for bounded, $\mathcal{B}$-measurable functions f. Here e is the identity element of the group. Such a conditional probability p is called a <u>translation family</u>. (By taking X to be real n-dimensional space considered as an additive group, one sees that the present definition does include the traditional translation families of distributions. Additional examples are given below.)

An element α of $P(X)$ is <u>tight</u> if $\alpha(B) = \sup\{\alpha(K): K \subseteq B,\ K \text{ compact}\}$ for every $B \in \mathcal{B}$. A tight probability α is easily seen to be countably additive ([1], Exercise 7, p. 11) and, conversely, if X is complete and separable, then every countably additive α is tight ([1], Theorem 1.4). A translation family p is <u>tight</u> if p_e (and, hence, each p_θ) is tight.

<u>Theorem 3</u>. <u>Suppose that</u> p <u>is a tight translation family</u>. <u>Then, under the left-invariant prior</u> π, <u>a posterior is the tight translation family</u> q <u>where</u>

$$(4.4) \qquad \int f(\theta)q_e(d\theta) = \int f(\theta^{-1})p_e(d\theta)$$

<u>for every bounded</u>, $\mathcal{B}$-<u>measurable function</u> f <u>on</u> Θ.

<u>Proof</u>: Let φ be a bounded, $\mathcal{B} \times \mathcal{B}$-measurable function on $\Theta \times X$. The first step of the proof is to establish the following equality:

$$(4.5)\qquad \lim_{\alpha} \frac{1}{h(U_\alpha)} \iint_{U_\alpha} [\varphi(\theta, x\theta) - \varphi(x^{-1}\theta, \theta)]\, h(d\theta)p_e(dx) = 0.$$

To verify (4.5), calculate thus. (Recall that sets are identified with their indicator functions.)

$$(4.6)\qquad \left|\int_{U_\alpha} [\varphi(\theta, x\theta) - \varphi(x^{-1}\theta, \theta)]h(d\theta)\right| = \left|\int_{xU_\alpha} \varphi(x^{-1}\theta, \theta)h(d\theta) - \int_{U_\alpha} \varphi(x^{-1}\theta, \theta)h(d\theta)\right|$$

$$= \left|\int \varphi(x^{-1}\theta, \theta)[(xU_\alpha)(\theta) - U_\alpha(\theta)]h(d\theta)\right|$$

$$\leq \|\varphi\| \int |xU_\alpha - U_\alpha|\, dh$$

$$= \|\varphi\| h(xU_\alpha \Delta U_\alpha),$$

where $\|\varphi\| = \sup\{|\varphi(\theta, x)| : (\theta, x) \in \Theta \times X\}$.

Formula (4.5) now follows easily from (4.1) and the tightness of p_e.

Next calculate again.

$$(4.7)\qquad \iint \varphi(\theta, x)p_e(dx)\Pi(d\theta) = \iint \varphi(\theta, x\theta)p_e(dx)\Pi(d\theta)$$

$$= \lim_{\alpha} \frac{1}{h(U_\alpha)} \iint_{U_\alpha} \varphi(\theta, x\theta)p_e(dx)h(d\theta)$$

$$= \lim_{\alpha} \frac{1}{h(U_\alpha)} \iint_{U_\alpha} \varphi(\theta, x\theta)h(d\theta)p_e(dx)$$

$$= \lim_{\alpha} \frac{1}{h(U_\alpha)} \iint_{U_\alpha} \varphi(x^{-1}\theta, \theta)h(d\theta)p_e(dx)$$

$$= \lim_{\alpha} \frac{1}{h(U_\alpha)} \iint_{U_\alpha} \varphi(x^{-1}\theta, \theta)p_e(dx)h(d\theta)$$

$$= \lim_{\alpha} \frac{1}{h(U_\alpha)} \iint_{U_\alpha} \varphi(x\theta, \theta)q_e(dx)h(d\theta)$$

$$= \lim_{\alpha} \frac{1}{h(U_\alpha)} \iint_{U_\alpha} \varphi(x, \theta) q_\theta(dx) h(d\theta)$$

$$= \iint \varphi(x, \theta) q_\theta(dx) \pi(d\theta)$$

$$= \iint \varphi(\theta, x) q_x(d\theta) \pi(dx).$$

The successive lines in (4.7) are, respectively, by (4.3), definition of π, Fubini's theorem for the countably additive measures p_e and h, (4.6), Fubini again, (4.4), definition of a translation family, definition of π again, and obvious.

Let $B \in \mathcal{B}$ and take φ to be the indicator function of $\Theta \times B$. Then (4.7) becomes

$$\int p_\theta(B) \pi(d\theta) = \pi(B)$$

and, hence, π is the marginal on X. Consequently, it follows from (4.7) that q is a posterior. It is easy to verify that q_e is tight because p_e is. □

<u>Corollary 2</u>. <u>If</u> p <u>is a tight translation family, then the translation family</u> q <u>defined by</u> (4.4) <u>is a coherent conditional probability</u>.

<u>Proof</u>: Use Theorem 3 and Corollary 1 of Section 2. □

In the following examples, ξ is an X-valued random variable with distribution p_e. By (4.4), q_e is the distribution of ξ^{-1}. It follows from (4.3) that p_θ is the distribution of $\xi\theta$ and likewise that qx is the distribution of $\xi^{-1}x$ for every θ and x.

<u>Example 4.1</u>. Let $X = \Theta = R^n$ and regard R^n as an additive group. That R^n is amenable is well-known and easy to check with the aid of Follner's condition. Let p_e be a countably additive probability on $\mathcal{B}$. Think of p_e as the dis-

tribution of an n-dimensional random variable ξ. Then p_θ is the distribution of $\xi + \theta$, q_e is the distribution of $-\xi$ and q_x the distribution of $-\xi + x$. Suppose now that ξ is symmetric. Then $q_x = p_x$. For a specific example, take ξ to be $N(0, I)$; that is, normal with mean zero and covariance the identity matrix. Then, by the above, the posterior distribution given x, is $N(x, I)$ and this posterior is, by Corollary 2, coherent. For this same example, consider the usual estimation problem with $A = R^n$ and $L(\theta, a) = \|\theta - a\|^2$. It is easy to see, with the aid of Lemmas 3.3 and 3.4, that the Bayes rule for the invariate prior Π is $\delta(x) \equiv x$, and so, by Theorem 2, δ is weakly admissible. This δ has a constant risk function and its weak admissibility is thus equivalent, as follows from (3,3), to its being minimax. That δ is minimax follows also from a general theorem of Kiefer [11]. Finally, δ is not admissible for $n \geq 3$ as was shown by Stein [14].

Example 4.2. This example treats the problem of inference from a univariate normal distribution with unknown mean and variance. To formulate the problem in the setting of this section, it is convenient to regard an observation of a random sample as being an observation of the sufficient statistics, namely, the sample mean and variance. Formally, let $X = \Theta = \{(\mu, \sigma^2): \mu \in R', \sigma^2 > 0\}$. The group operation is that of the affine group in one dimension:

$$(\mu, \sigma^2)\cdot(\nu, \tau^2) = (\tau\mu + \nu, (\sigma\tau)^2).$$

In essence, (μ, σ^2) is identified with the mapping $x \to \sigma x + \mu$ and the operation is composition of functions. Notice that $(\mu, \sigma^2)^{-1} = (-\mu\sigma^{-1}, \sigma^{-2})$. The amenability of this group is shown in ([9], pp. 68-69).

Let $Y_1, \ldots, Y_n$ be independent, $N(0, 1)$ variables and let $(\bar{Y}, S^2)$ be the sample mean and variance. Take p_e to be the distribution of $\xi = (\bar{Y}, S^2)$.

Then, for $\theta = (\mu, \sigma^2)$, p_θ is the distribution of the sample mean and variance of a sample of size n from a $N(\mu, \sigma^2)$ variable. Furthermore, q_e is the distribution of $\xi^{-1} = (-\overline{Y}S^{-1}, S^{-2})$. Set $T = -\overline{Y}S^{-1}$. ($\sqrt{n-1}\, T$ has a Student t distribution.) Then the posterior distribution of (μ, σ^2) given $x = (\overline{y}, s^2)$ is q_x, which is the distribution of

$$\xi^{-1}x = (sT + \overline{y}, s^2S^{-2}).$$

In particular, the posterior distribution of μ is specified by the fact that $(\mu - \overline{y})s^{-1}$ has the same distribution as T.

Example 4.3. Consider now independent samples from two normal populations each with unknown means and variances. The notation of the previous example will be used with subscripts 1 and 2 to indicate the two populations. In particular, X_1 and X_2 are taken to be the group of the previous example and the X of this example is taken to be the direct product of X_1 with X_2. Let $\xi = ((\overline{Y}_1, S_1^2), (\overline{Y}_2, S_2^2))$. Then $\xi^{-1} = ((T_1, S_1^{-2}), (T_2, S_2^{-1}))$ and, for $x = ((\overline{y}_1, s_1^2), (\overline{y}_2, s_2^2))$, q_x is the distribution of

$$\xi_x^{-1} = ((s_1T_1 + \overline{y}_1, s_1^2S_1^{-2}), (s_2T_2 + \overline{y}_2, s_2^2S_2^{-2})).$$

In particular, the posterior distribution of the difference of the two means is that of $s_1T_1 + \overline{y}_1 - s_2T_2 - \overline{y}_2$.

The coherence of the posterior distributions of the last two examples is closely related to the fact that, if these distributions are used to calculate confidence intervals for the mean and difference of the means, respectively, then there exists "no negatively biased relevant selection" as was shown by Robinson [12].

5. <u>Improper priors can result in incoherent posteriors and uniformly inadmissible decision rules</u>. The object of this section is to begin an exploration of the relation of finitely additive priors to improper priors. By an <u>improper prior</u> is meant, as usual, a countably additive measure ν on $\mathcal{B}(\Theta)$ such that $\nu(\Theta)$ is infinite. To avoid unnecessary technicalities, assume throughout this section that X, Θ, and A are Borel subsets of some Euclidean space and are equipped with their usual Borel fields of subsets. By a <u>density</u> is here meant a density with respect to Lebesgue measure. Suppose that each of the measures p_θ has a density $f(x|\theta)$ on X and that ν has a density $g(\theta)$ on Θ. Define

$$(5.1) \qquad h(\theta|x) = \frac{f(x|\theta)g(\theta)}{\int_\Theta f(x|t)g(t)dt}$$

whenever the denominator is finite and not zero. By a posterior for ν is meant a regular conditional probability q on Θ given X such that, for each θ, $q(\cdot|x)$ has a density $h(\theta|x)$. If ν were a proper, countably additive prior, then Bayes formula (5.1) would, in fact, define the density of its posterior. In the improper case, (5.1) is often used as a formal device without theoretical foundation. Nevertheless, it sometimes leads to a genuine posterior which could also have been obtained from a proper, finitely additive prior.

<u>Example 5.1</u>. Let $X = \Theta = R^n$ and suppose p_θ is a normal distribution with mean θ and covariance the identity matrix. Let ν be Lebesgue measure on R^n. Then (5.1) yields the same posterior as that obtained from an invariant finitely additive prior as in Example 4.1. (The normal distribution is not essential here and any translation family would work equally well.)

However, as the next example illustrates, the use of (5.1) with an improper prior can result in an incoherent posterior.

<u>Example 5.2</u>. Let $\Theta = X = R^+$, the set of strictly positive real numbers. For every θ, let p_θ be the uniform distribution on the interval $[\theta/2, 3\theta/2]$ so that p_θ has the density

$$\begin{aligned} f(x|\theta) &= \theta^{-1} \quad \text{for} \quad \theta/2 \leq x \leq 3\theta/2, \\ &= 0 \qquad \text{elsewhere.} \end{aligned}$$

Let ν be Lebesgue measure on Θ so that $g(\theta) = 1$ for all θ. Then (5.1) leads to a posterior q with density

$$(5.2) \qquad \begin{aligned} g(\theta|x) &= (\theta \log 3)^{-1} \quad \text{for} \quad 2x/3 \leq \theta \leq 2x, \\ &= 0 \qquad\qquad \text{elsewhere,} \end{aligned}$$

for every x.

Consider the simple betting system A, b where $b \equiv 1$ and $A = \{(\theta, x): \theta \leq x \leq 3\theta/2\}$. Then, for every x, $A^x = \{\theta: 2x/3 \leq \theta \leq x\}$ and

$$q(A^x|x) = \frac{1}{\log 3} \int_{2x/3}^{x} \frac{1}{\theta}\, d\theta = \frac{\log 3 - \log 2}{\log 3}.$$

The corresponding expected loss is

$$\begin{aligned} E(\theta) &= \int [A(\theta, x) - q(A^x|x)]\, p(dx|\theta) \\ &= 1/2 - \frac{\log 3 - \log 2}{\log 3} \\ &> 0. \end{aligned}$$

Hence, q is incoherent.

Improper priors are also used to calculate decision rules. In fact, the Bayes rule for an improper prior is taken to be that rule which minimizes the expected posterior loss if such a rule exists. Example 5.2 can be used to show that the Bayes rule for an improper prior need not be extended admissible even for a measurable decision problem with a bounded loss function.

Example 5.2. (continued). In addition to the structure already specified, introduce $A = R^+$ and $L(\theta, a) = \min\{a^{-1}|\theta - a|, 3\}$.

To compute the Bayes rule f for the improper prior ν, it suffices to find, for each x, that action $a = f(x)$ which minimizes $\int L(\theta, a)g(\theta|x)dx$, where g is given by (5.2). Since the density $g(\cdot|x)$ is concentrated on the interval $I(x) = [2x/3, 2x]$, it is clear that a must lie in $I(x)$ and it follows that $\theta \leq a \leq 3\theta$ and, hence, $L(\theta, a) = a^{-1}|\theta - a| \leq 3$. Thus a must minimize

$$\int a^{-1}|\theta-a|g(\theta|x)dx = \int_{2x/3}^{a}(1-\theta/a)\frac{1}{\theta \log 3}\,d\theta + \int_{a}^{2x}(\theta/a - 1)\frac{1}{\theta \log 3}\,d\theta.$$

By differentiating, one can see that $a = 4x/2$, the midpoint of $I(x)$.

Consider now any decision rule of the form $f_\beta(x) = \beta^{-1}x$ where $2/3 \leq \beta^{-1} \leq 2$ so that $f(x) \in I(x)$ for each x. Then its risk is

$$r(\theta, f_\beta) = \int_{\theta/2}^{3\theta/2} \beta x^{-1}|\theta - \beta^{-1}x|\theta^{-1}dx$$

$$= 2\{\beta[(\log(2/\sqrt{3}) - 1) + \log\beta] + 1\}.$$

The risk does not depend on θ, but only on the constant β. The unique β which minimizes $r(\theta, f_\beta)$ is easily seen to be $\beta_o = 2/\sqrt{3}$. Every other f_β, including the Bayes rule for ν which is $f_{4/3}$, is uniformly dominated by f_{β_o} and, therefore, is not extended admissible.

An obvious problem, which is suggested by Example 5.2, is to determine which improper priors do lead to coherent posteriors and extended admissible decision rules. This problem is, in view of Corollary 1 and Theorem 2, almost the same as that of finding those improper priors whose posteriors could also be obtained from proper, finitely additive priors.

References

1. Billingsley, Patrick (1968). Convergence of Probability Measures. Wiley, New York.

2. Blackwell, D., and M. A. Girschick (1954). Theory of Games and Statistical Decisions. Wiley, New York.

3. Buehler, R. J. (1976). "Coherent preferences," Ann. Statist. 4, 1051-1064.

4. Dubins, Lester E. (1975). "Finitely additive conditional probabilities, conglomerality, and disintegrations," Ann. Prob. 3, 89-99.

5. Dubins, Lester and David Freedman (1964). "Measurable sets of measures," Pacific J. of Math. 14, 1211-1222.

6. Dunford, Nelson and Jacob J. Schwartz (1958). Linear Operators, Part I: General Theory. Interscience, New York.

7. de Finetti, Bruno (1975). Theory of Probability, v. 1, Wiley, New York.

8. Freedman, David A. and Roger A. Purves (1969). "Bayes' method for bookies," Ann. Math. Statist. 40, 1177-1186.

9. Greenleaf, Frederick P. (1969). Invariant Means on Topological Groups. Van Nostrand, New York.

10. Heath, David C. and William D. Sudderth (1972). "On a theorem of de Finetti, oddsmaking, and game theory," Ann. Math. Statist. 43, 2072-2077.

11. Kiefer, J. (1957). "Invariance, minimax sequential estimation, and continuous time processes," Ann. Math. Statist. 28, 573-601.

12. Robinson, G. K. (1976). "Properties of Students' t and of the Behrens-Fisher solution to the two means problem," Ann. Statist. 4, 963-971.

13. Savage, Leonard J. (1972). The Foundations of Statistics. Dover, New York.

14. Stein, C. (1956). "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," Proc. of 3rd Berkeley Symp. on Math. Stat. and Prob., v. 1, pp. 197-206. University of Calif. Press, Berkeley.

15. Sudderth, William D. (1971). "On measurable gambling problems," Ann. Math. Statist. 42, 260-269.