The Annals of Statistics
1989, Vol. 17, No. 2, 907–919

COHERENT INFERENCE FROM IMPROPER PRIORS AND FROM FINITELY ADDITIVE PRIORS

BY DAVID HEATH AND WILLIAM SUDDERTH[1]

Cornell University and University of Minnesota

Conditions are given for the formal posterior of an improper prior to be coherent and applied to translation models. An example is given of a proper countably additive statistical model and a finitely additive prior for which there is no posterior.

1. Introduction. A notion of coherence for statistical inference was introduced in a previous paper [6]. It was shown that an inference is coherent if and only if it corresponds to the posterior of a finitely additive prior. A similar result was proved for predictions and predictive inferences in [8]. (See Regazzini [10] for a clear exposition of a different notion of coherence.)

In practice many Bayesians use improper, countably additive priors to represent diffuse prior knowledge rather than finitely additive priors. There are several reasons for this including the relatively easy calculation and the essential uniqueness of the formal posterior of an improper prior and the lack of familiarity with the finitely additive theory. As was shown by examples in [6], the use of an improper prior sometimes results in a coherent inference and sometimes not. The obvious problem is to find an effective criterion for determining when an inference from an improper prior will be coherent.

Bayesians have long justified their use of improper priors by arguing that they can be approximated in some sense by proper priors. A useful discussion is given by Stone [11] who defines a notion of approximation which we adapt for our purposes. Our first result (Theorem 3.1) is that an improper prior leads to a coherent inference if and only if it can be approximated by proper priors in this sense. Even this result is difficult to apply in specific examples. However, it can be used to derive a sufficient condition for coherence which is often easy to verify. This condition is presented in Theorem 3.2 and applied in several examples.

A Bayesian, who seeks to avoid incoherent inferences, might be advised to abandon improper, countably additive priors and use only finitely additive priors. One difficulty with this approach is that not every finitely additive prior has a posterior. Examples of this phenomenon presented heretofore have involved finitely additive conditional distributions for the observations as well as a finitely additive prior. An example is presented below in which the conditional distributions are countably additive with finite support. Thus it can happen

Received October 1986; revised November 1987.
[1]Research supported by NSF Grant DMS-84-21208.
AMS 1980 *subject classifications*. 62A15, 60A05.
Key words and phrases. Coherence, improper priors, finite additivity, Bayesian inference.

that, even for a standard statistical model, a finitely additive prior leads to no inference.

The next section contains the necessary definitions and preliminary results.

2. Preliminaries. For any set S, $P(S)$ denotes the collection of finitely additive probability measures defined on all subsets of S. If f is a bounded, real-valued function defined on S and $\gamma \in P(S)$, then the γ-integral of f will be written $\gamma(f)$, $\int f\,d\gamma$ or $\int f(s)\gamma(ds)$.

Let Θ and X be nonempty sets corresponding to the set of possible states of nature and the set of possible outcomes for a certain experiment, respectively. A statistical *model* p is a mapping which assigns to each $\theta \in \Theta$ an element p_θ of $P(X)$. An *inference* q assigns to each $x \in X$ an element q_x of $P(\Theta)$. (In our earlier paper [6] we did not require each q_x to belong to $P(\Theta)$ and considered the more general notion of a "conditional odds function." We impose the new restriction here in order to simplify the exposition and also because it is a natural requirement recommended by de Finetti [3], page 339.) Thus p is a conditional probability distribution on X given Θ and q is a conditional distribution on Θ given X. Let $\mathbf{B}(\Theta)$ and $\mathbf{B}(X)$ be given σ-fields of subsets of Θ and X, respectively. The model p is called *measurable* if every p_θ is countably additive on $\mathbf{B}(X)$ and if, for every $A \in \mathbf{B}(X)$, the mapping $\theta \to p_\theta(A)$ is $\mathbf{B}(\Theta)$-measurable. A *measurable inference* q is defined similarly. The standard models and inferences of statistics are, of course, measurable.

An inference q might correspond in practice to a system of confidence intervals, a posterior distribution or a fiducial distribution. For an operational interpretation, regard q_x as a conditional odds function used by the statistician to post odds on subsets of Θ after observing x. The inference q is called coherent if it is impossible for a gambler to devise a system based on q, which consists of placing a finite number of bets on subsets of Θ after x is observed and which attains an expected payoff greater than some positive constant for every $\theta \in \Theta$. (See [6] for the precise definition.)

An element π of $P(\Theta)$ will be called a *prior*. A prior π and model p determine a *marginal* $m \in P(X)$ by the formula

$$m(\phi) = \int p_\theta(\phi)\pi(d\theta) \tag{2.1}$$

for bounded functions $\phi\colon X \to R$. Let $\mathbf{B} = \mathbf{B}(\Theta) \times \mathbf{B}(X)$ be the product σ-field on $\Theta \times X$. An inference q is called a *posterior* for the prior π, the model p being understood, if

$$\int\int \phi(\theta, x)p_\theta(dx)\pi(d\theta) = \int\int \phi(\theta, x)q_x(d\theta)m(dx) \tag{2.2}$$

for all bounded, $\mathbf{B}$-measurable functions $\phi\colon \Theta \times X \to R$. In other words, q is a conditional distribution for Θ given X under the measure on $\mathbf{B}$ determined by π and p as defined by the left-hand side of (2.2).

The model p and inference q are called *consistent* if there exist $\pi \in P(\Theta)$ and $m \in P(X)$ such that (2.2) holds for all bounded, $\mathbf{B}$-measurable ϕ.

The following proposition summarizes a few results from [6] and [7].

PROPOSITION 2.1. *The following are equivalent statements about an inference q relative to a given model p.*

(a) *q is coherent.*
(b) *q is the posterior of some prior π.*
(c) *p and q are consistent.*
(d) *For every bounded, real-valued* **B**-*measurable function ϕ on $\Theta \times X$,*

$$\inf_{\theta} p_\theta(\phi_\theta) \le \sup_x q_x(\phi^x),$$

where $\phi_\theta(x) = \phi(\theta, x) = \phi^x(\theta)$.

The results of the proposition are stated as in [6] and [7] for general p and q which are not necessarily measurable. Thus the inner integrals in (2.2), corresponding to $p_\theta(\phi_\theta)$ and $q_x(\phi^x)$, need not be measurable functions of θ and x, respectively. This is the reason why π and m must be defined on all subsets of their respective spaces Θ and X. Now if p and q are measurable, then so are the functions $p_\theta(\phi_\theta)$ and $q_x(\phi^x)$ and we need only specify π and m on $\mathbf{B}(\Theta)$ and $\mathbf{B}(X)$, respectively, for (2.2) to make sense. It is also easy to see that the proposition remains true for measurable p and q if we consider priors and marginals to be defined only on the appropriate σ-fields.

Let $M(\Theta)$ and $M(X)$ be the collections of countably additive measures defined on $\mathbf{B}(\Theta)$ and $\mathbf{B}(X)$, respectively. By an *improper prior* is meant an element π of $M(\Theta)$ such that $\pi(\Theta)$ is infinite. Suppose that, for a given statistical model p, there is a reference measure $\nu \in M(X)$ such that every p_θ is absolutely continuous with respect to ν. Let $f(\cdot|\theta)$ be the density for p_θ. For $x \in X$, define

$$q_x(d\theta) = \frac{f(x|\theta)\pi(d\theta)}{\int f(x|t)\pi(dt)} \tag{2.3}$$

whenever the denominator is finite and not zero and let q_x be an arbitrary fixed element of $P(\Theta)$ otherwise. The inference q is called the *formal posterior* of the improper prior π. If $f(\cdot|\cdot)$ is **B**-measurable and if the denominator above is ν-almost everywhere finite and positive, then q is a measurable inference. Of course, if π is proper and countably additive on $\mathbf{B}(\Theta)$, then the q given by (2.3) is a genuine posterior for π and is coherent by Proposition 1. (Another approach to evaluating improper priors is in Eaton [5].)

3. Approximation by proper priors. Let α and β be measures on $\mathbf{B}(\Theta)$ and define the total variation distance by

$$\|\alpha - \beta\| = \sup\left\{\left|\int \phi\, d\alpha - \int \phi\, d\beta\right| : \sup|\phi| \le 1,\ \phi \in L_\infty(\Theta)\right\}, \tag{3.1}$$

where $L_\infty(\Theta)$ is the space of bounded, real-valued, $\mathbf{B}(\Theta)$-measurable functions on

Θ. Next consider an inference $\tilde{q}$ and a prior $\pi \in P(\Theta)$ which has marginal m and posterior q. Define

$$d_\pi(q, \tilde{q}) = \int \|q_x - \tilde{q}_x\| m(dx), \tag{3.2}$$

which can be thought of as the expected distance between the inferences q and $\tilde{q}$ when the expectation is calculated from the marginal of the prior π.

DEFINITION. The inference $\tilde{q}$ is *approximable by proper priors* (a.p.p.) if

$$\inf d_\pi(q, \tilde{q}) = 0, \tag{3.3}$$

where the infimum is over all π, q such that $\pi \in P(\Theta)$ and q is the posterior of π. If $\tilde{\pi}$ is an improper prior with formal posterior $\tilde{q}$, we say that $\tilde{\pi}$ is *approximable by proper priors* if $\tilde{q}$ is.

As mentioned in the introduction, this notion of approximation was inspired by Stone [11, 12] who did not, however, consider finitely additive priors.

THEOREM 3.1. *An inference $\tilde{q}$ is coherent if and only if it is approximable by proper priors.*

PROOF. If $\tilde{q}$ is coherent, then, by Proposition 2.1, there exists $\pi \in P(\Theta)$ with posterior $q = \tilde{q}$ and $d_\pi(q, \tilde{q}) = 0$.

Suppose now that $\tilde{q}$ is a.p.p. We will use Proposition 2.1(d) to show that $\tilde{q}$ is coherent.

Let $\phi \in L_\infty(\Theta \times X)$ and $\varepsilon > 0$. Set $b = \sup|\phi|$. Choose $\pi \in P(\Theta)$ with posterior q such that

$$d_\pi(q, \tilde{q}) < \varepsilon/b.$$

Then

$$\begin{aligned}
&\left| \int q_x(\phi^x) m(dx) - \int \tilde{q}_x(\phi^x) m(dx) \right| \\
&\quad \le \int |q_x(\phi^x) - \tilde{q}_x(\phi^x)| m(dx) \\
&\quad \le (\sup|\phi|) d_\pi(q, \tilde{q}) < \varepsilon.
\end{aligned}$$

Hence

$$\begin{aligned}
\int p_\theta(\phi_\theta) \pi(d\theta) &= \int q_x(\phi^x) m(dx) \\
&\le \int \tilde{q}_x(\phi^x) m(dx) + \varepsilon
\end{aligned}$$

and consequently,

$$\sup_x \tilde{q}_x(\phi^x) \ge \inf_\theta p_\theta(\phi_\theta) - \varepsilon.$$

Because ε is arbitrary $\tilde{q}$ satisfies Proposition 2.1(d). □

Suppose now that π is an improper prior, $p_\theta(dx) = f(x|\theta)\nu(dx)$ for every $\theta \in \Theta$ and π has formal posterior q as in (2.3). The natural and often used way to attempt an approximation of q by proper priors is to truncate π to a set of finite measure. To be precise, let $K \in \mathbf{B}(\Theta)$ satisfy $0 < \pi(K) < \infty$ and define the *truncation* of π to K as the proper prior π_K where

$$\pi_K(\phi) = \frac{1}{\pi(K)} \int_K \phi(\theta)\pi(d\theta), \qquad \phi \in L_\infty(\Theta). \tag{3.4}$$

Let q^K and m_K be the posterior and marginal determined by π_K, respectively. Formulas (2.1) and (2.3) specialize to give

$$m_K(\psi) = \frac{1}{\pi(K)} \int_K \int \psi(x) p_\theta(dx)\pi(d\theta), \qquad \psi \in L_\infty(X),$$

$$q_x^K(\phi) = \frac{\int_K \phi(\theta) f(x|\theta)\pi(d\theta)}{\int_K f(x|\theta)\pi(d\theta)}, \qquad \phi \in L_\infty(\Theta). \tag{3.5}$$

It seems likely that, for measurable models, whenever q is a.p.p., it can be approximated by truncations. However, we have not proved such a result.

For a certain class of group invariant problems, Stone [11, 12] showed that Haar measure, used as an improper prior, could be approximated in a sense close to the present one by truncations. A similar result was obtained for amenable, locally compact groups in [6]. Suppose $X = \Theta = G$ is such a group and the model p is a generalized translation family $p_\theta(dx) = f(\theta^{-1}x)\nu(dx)$, where ν is right Haar measure. If ν is used as an improper prior, the corresponding inference is

$$q_x(d\theta) \propto f(\theta^{-1}x)\nu(d\theta) \tag{3.6}$$

and is coherent by [6], Theorem 3. Stone [13] has also given examples which illustrate that this inference need not be coherent if G is not amenable.

In general, the criterion of approximability by proper priors seems difficult to apply directly. For example, it follows from the discussion above that, if p is a translation family on the line such as the $N(\theta, 1)$, then Lebesgue measure $d\theta$ gives a coherent inference. However, it remains unclear whether improper priors such as $\theta^2\, d\theta$ will do so. The next result gives a sufficient condition for coherence which allows us to check that the corresponding inference is coherent.

Suppose π is an improper prior with formal posterior q. For each $K \in \mathbf{B}(\Theta)$ such that $0 < \pi(K) < \infty$, define

$$\beta(K) = \int q_x(K^c) m_K(dx). \tag{3.7}$$

Here m_K is the marginal on X determined by the truncated prior π_K. The number $\beta(K)$ is the posterior probability under π that $\theta \notin K$ averaged under the truncation of π to K. More crudely, $\beta(K)$ is the chance that q says $\Theta \notin K$ given that $\Theta \in K$.

THEOREM 3.2. *If*

$$\inf\{\beta(K): 0 < \pi(K) < \infty\} = 0, \tag{3.8}$$

then π *is approximable by proper priors. Indeed, given* $K \in \mathbf{B}(\Theta)$ *with* $0 < \pi(K) < \infty$, $\beta(K) \le d_{\pi_k}(q^K, q) \le 2\beta(K)$.

PROOF. It suffices to prove the inequalities. For the first inequality, notice that, for all x, $q_x^K(K^c) = 0$ and so, by (3.1),

$$\|q_x - q_x^K\| \ge q_x(K^c).$$

The first inequality now follows from (3.2) and (3.7).

To prove the second inequality, let $\phi \in L_\infty(\Theta)$ and $\sup|\phi| \le 1$. The inequality will follow from (3.1), (3.2) and (3.7) once it is shown that

$$|q_x(\phi) - q_x^K(\phi)| \le 2q_x(K^c), \qquad m_K\text{-a.s.} \tag{3.9}$$

To verify (3.9), first use the triangle inequality to see

$$|q_x(\phi) - q_x^K(\phi)| \le \left|\int_{K^c} \phi\, dq_x\right| + \left|\int_K \phi\, dq_x - \int_K \phi\, dq_x^K\right|. \tag{3.10}$$

Because $\sup|\phi| \le 1$, the first term on the right side of (3.10) is obviously bounded by $q_x(K^c)$. To obtain the same bound for the second term on the right side of (3.10), use (2.3) and (3.5) to rewrite it as

$$\left|\int_K \phi\, dq_x^K\right| \left|1 - \frac{\int_K f(x|\theta)\pi(d\theta)}{\int f(x|\theta)\pi(d\theta)}\right| \le q_x^K(K)\frac{\int_{K^c} f(x|\theta)\pi(d\theta)}{\int f(x|\theta)\pi(d\theta)} \tag{3.11}$$

$$= 1 \times q_x(K^c). \qquad \square$$

By Theorems 1 and 2, condition (3.8) is a sufficient condition for the formal posterior q to be coherent. Again we do not know whether it is necessary. The condition can often be checked as will be illustrated in the next section with two examples.

4. Two applications to translation families. In this section, $\Theta = X = R^d$, d-dimensional Euclidean space and $d\theta$ or dx has its usual interpretation as Lebesgue measure. The prior π will be a fixed improper prior

$$\pi(d\theta) = g(\theta)\, d\theta,$$

where the prior "density" g is nonnegative and Borel measurable. The model p is assumed to be a measurable translation family given by a family of densities

$$p_\theta(dx) = f(x - \theta)\, dx,$$

where f is Borel measurable. Assume also that the denominator on the right side of (2.3) is Lebesgue almost everywhere finite and positive so that (2.3), which gives the formal posterior q of π, can be rewritten as

$$h(\theta|x) = \frac{f(x-\theta)g(\theta)}{\int f(x-\phi)g(\phi)\,d\phi}, \tag{4.1}$$

where $h(\theta|x)$ is a density for q_x. Write $|\theta|$ for the Euclidean norm of $\theta \in R^d$ and let π_n be the truncation of π to the ball $B_n = \{\theta: |\theta| \le n\}$. Let q^n be the posterior for π_n and the Bayes formula then gives the density below for q_x^n:

$$h_n(\theta|x) = \frac{f(x-\theta)g(\theta)}{\int_{B_n} f(x-\phi)g(\phi)\,d\phi}, \qquad |\theta| \le n. \tag{4.2}$$

So that (4.2) will be valid, assume $\pi(B_n) < \infty$ for all n. For simplicity assume $\pi(B_n) > 0$ also. However, there is no real loss in generality because we will only need below that $\pi(B_n)$ is positive for n large and this follows from our assumption that $\pi(\Theta) = \infty$.

If the tails of the prior density g grow too rapidly, the inference q need not be coherent even for quite well-behaved translation models.

Example 4.1 (Stone [13]). Suppose $\Theta = X = R^1$, p_θ is $N(\theta, 1)$ and $g(\theta) = \exp(a\theta)$ where $a > 0$. Use (4.1) to see that q_x is $N(x + a, 1)$. In Proposition 2.1(d), take ϕ to be the indicator function of the set $S = \{(\theta, x): \theta < x + a\}$ and notice that

$$p_\theta(\phi_\theta) = p_\theta[\theta - a, \infty] = p_0[-a, \infty] > \tfrac{1}{2},$$

while

$$q_x(\phi^x) = q_x[-\infty, x + a] = q_0[-\infty, a] = \tfrac{1}{2}.$$

Thus q is incoherent.

The critical feature of this example is the exponential growth of the prior density g. The normal model could be replaced by many translation families including, for example, the uniform translation model where p_θ is the uniform distribution on the interval $[\theta, \theta + 1]$. Thus the exponential growth of g is too much even when the p_θ have compact support. Here is a condition which rules out such growth for g.

Growth condition (GC). For every $a > 0$, $\lim_{n\to\infty}(\pi(B_{n+a})/[\pi(B_n)]) = 1$.

Notice that a prior density which behaves asymptotically like a polynomial will satisfy GC.

The next lemma gives another sufficient condition for coherence when π satisfies GC. In its statement m_n denotes the marginal determined by the truncated prior π_n and the model p.

LEMMA 4.1. *Assume* π *satisfies* GC *and let* $a \geq 0$. *Then the following are true.*

(a) $\pi_n(B_{n-a}) \to 1$ *as* $n \to \infty$.
(b) $m_n(B_{n-a}) \to 1$ *as* $n \to \infty$.
(c) q *is coherent if*

$$\lim_{n \to \infty} \frac{1}{\pi(B_n)} \int_{B_{n-a}} \int_{B_n^c} f(x-\theta) g(\theta)\, d\theta\, dx = 0. \tag{4.3}$$

(d) q *is coherent if*

$$\sup_n \int_{B_{n-a}} \int_{B_n^c} g(\theta) f(x-\theta)\, d\theta\, dx < \infty. \tag{4.4}$$

PROOF. (a) This is obvious if $a = 0$ and immediate from GC if $a > 0$.

(b) Let $\varepsilon > 0$. Choose $b > 0$ such that $p_0(B_b) > 1 - \varepsilon$. Then $p_\theta(B_b + \theta) = p_0(B_b) > 1 - \varepsilon$ for all θ. Now calculate:

$$m_n(B_{n-a}) = \int_{B_n} p_\theta(B_{n-a}) \pi_n(d\theta) \geq \int_{B_{n-a-b}} p_\theta(B_{n-a}) \pi_n(d\theta)$$
$$\geq \int_{B_{n-a-b}} p_\theta(B_b + \theta) \pi_n(d\theta) \geq (1-\varepsilon) \pi_n(B_{n-a-b}).$$

(The next to last inequality holds because $B_b + \theta \subseteq B_{n-a}$ for $\theta \in B_{n-a-b}$.) Now use part (a).

(c) and (d) Let $\varepsilon > 0$. By (3.7) and part (b),

$$\beta(B_n) = \int q_x(B_n^c) m_n(dx) < \varepsilon + \int_{B_{n-a}} q_x(B_n^c) m_n(dx)$$

for n sufficiently large. By Theorems 3.1 and 3.2, the coherence of q will be established if we show

$$\lim_{n \to \infty} \int_{B_{n-a}} q_x(B_n^c) m_n(dx) = 0. \tag{4.5}$$

To see this, let f_n be the density for m_n which is given by

$$f_n(x) = \frac{1}{\pi(B_n)} \int_{B_n} f(x-\theta) g(\theta)\, d\theta$$

and use (4.1) to write

$$q_x(B_n^c) = \frac{\int_{B_n^c} f(x-\theta) g(\theta)\, d\theta}{\int f(x-\theta) g(\theta)\, d\theta}.$$

Hence

$$\int_{B_{n-a}} q_x(B_n^c) m_n(dx) \le \frac{1}{\pi(B_n)} \int_{B_{n-a}} \int_{B_n^c} f(x-\theta) g(\theta)\, d\theta\, dx$$

and (4.5) follows from (4.3).

Now (4.5) also follows from (4.4) because $\pi(B_n) \to \infty$. □

The final two conditions of Lemma 4.1 can be viewed as joint growth conditions on the densities for the prior and the model. We will now apply these conditions to two special situations.

THEOREM 4.1. *Suppose $\Theta = X = R^1$. Assume $\pi(d\theta) = g(\theta)\, d\theta$ is an improper prior with g bounded and $p_\theta(dx) = f(x-\theta)\, dx$ is a translation family such that $\int |x| f(x)\, dx < \infty$. Then the formal posterior q is coherent.*

PROOF. Because g is bounded, it satisfies GC and it suffices by Lemma 4.1(d) to show that

$$\int_{-n}^{n} \int_{B_n^c} f(x-\theta)\, d\theta\, dx \le 2E|Z|, \tag{4.6}$$

where Z is a random variable with density $f(x)$.

Use the fact that $-Z$ has density $f(-x)$ and calculate as follows:

$$\begin{aligned}\int_{B_n^c} f(x-\theta)\, d\theta &= \int_n^{\infty} f(x-\theta)\, d\theta + \int_{-\infty}^{n} f(x-\theta)\, d\theta \\ &= \int_{n-x}^{\infty} f(-\theta)\, d\theta + \int_{n+x}^{\infty} f(\theta)\, d\theta \\ &= P[-Z \ge n-x] + P[Z \ge n+x].\end{aligned}$$

Hence,

$$\begin{aligned}\int_{-n}^{n} \int_{B_n^c} f(x-\theta)\, d\theta\, dx &= \int_{-n}^{n} \{P[-Z \ge n-x] + P[Z \ge n+x]\}\, dx \\ &= \int_0^{2n} \{P[-Z \ge y] + P[Z \ge y]\}\, dy \\ &= 2\int_0^{2n} P[|Z| \ge y]\, dy \\ &\le 2E|Z|.\end{aligned}$$

□

THEOREM 4.2. *Suppose $\Theta = X = R^1$ and $p_\theta \sim N(\theta, 1)$. Let $\pi(\theta)$ be an improper prior satisfying* GC *and*

$$g(\theta)/|\theta|^r \to k \quad \text{as } |\theta| \to \infty \tag{4.7}$$

for some positive constants r and k. Then the formal posterior q is coherent.

PROOF. Let f be the density for $p_0 - N(0,1)$. For every $s > 0$, $f(x)$ is $O(|x|^{-s})$. Thus the theorem will follow from the next lemma. □

LEMMA 4.2. *If π is as in Theorem 4.2 and if $f(x) = O(|x|^{-s})$ for some $s > r + 2$, then the formal posterior q is coherent.*

PROOF. The proof is an application of Lemma 4.1(c). First use (4.7) to choose $t > 0$ so that $g(\theta) \geq (k/2)|\theta|^r$ for $|\theta| \geq t$. Then, for $n \geq t$,

$$\pi(B_n) = \int_{-n}^{n} g(\theta)\, d(\theta) \geq k\int_{t}^{n} \theta^r\, d\theta = k'(n^{r+1} - t^{r+1}),$$

where k' is a constant independent of n. Use (4.7) again to see that $g(\theta) \leq 2k|\theta|^r$ for $|\theta|$ sufficiently large. Thus (4.3) and the lemma will be proved once we show that

$$\int_{B_{n-1}}\int_{B_n^c} |\theta|^r |x - \theta|^{-s}\, d\theta\, dx \leq cn^r$$

for some constant $c > 0$ and n sufficiently large. To see this, change the order of integration to get

$$\int_{B_n^c} |\theta|^r \int_{B_{n-1}} |x - \theta|^{-s}\, dx\, d\theta \tag{4.8}$$

and calculate the inner integral for $|\theta| \geq n$:

$$\begin{aligned}\int_{B_{n-1}} |x - \theta|^{-s}\, dx &\leq \int_{B_{n-1}} (|\theta| - |x|)^{-s}\, dx \\ &= 2\int_0^{n-1} (|\theta| - x)^{-s}\, dx \\ &\leq c_1(|\theta| - n + 1)^{-s+1}.\end{aligned}$$

Thus (4.8) is dominated by a constant times

$$\begin{aligned}\int_{B_n^c} |\theta|^r (|\theta| - n + 1)^{-s+1}\, d\theta &= 2\int_n^{\infty} \theta^r (\theta - n + 1)^{-s+1}\, d\theta \\ &= 2\int_1^{\infty} (u + n - 1)^r u^{-s+1}\, du \\ &= 2(n-1)^r \int_1^{\infty} (u/(n-1) + 1)^r u^{-s+1}\, du \\ &\leq 2(n-1)^r \int_1^{\infty} (2u)^r u^{-s+1}\, du \\ &= 2^{r+1}(n-1)^r \int_1^{\infty} u^{r-s+1}\, du \\ &\leq c(n-1)^r.\end{aligned}$$

□

It would be interesting to know whether Theorem 4.2 generalizes to higher dimensions.

Both Theorems 4.1 and 4.2 illustrate that coherence of an inference from an improper prior depends on the relationship between the prior and the model, and not on the prior alone. In fact, given any improper prior $\pi(d\theta) = g(\theta)\,d\theta$, there is a model p for which the formal posterior q is incoherent. For example, if $\Theta = R^1$ and g is locally integrable and everywhere positive, then a simple transformation $\phi = \phi(\theta)$ gives a prior $\pi'(d\phi) = e^{\phi}\,d\phi$ and the normal model of Example 4.1 will lead to an incoherent inference.

5. A measurable model and finitely additive prior for which there is no coherent inference. In contrast to the situation with improper priors, the posterior of a proper prior is, by Proposition 2.1, always coherent. However, it can happen that a finitely additive prior has no posterior for a given model p. Examples of this phenomenon are given in [4] and [9]. In these examples the measures p_θ are only finitely additive. Here is a simple example in which p is measurable.

Let $X = \Theta = Z = \{0, \pm 1, \pm 2, \dots\}$ and let p be the translation model such that

$$p_\theta\{\theta + 1\} = p_\theta\{\theta - 1\} = \tfrac{1}{2}$$

for all θ. Take the prior π to be of the form

$$\pi = (\mu + \nu)/2,$$

where μ is countably additive with support the set A of integers divisible by 4 and ν is purely finitely additive and supported by the set B of integers equal to 2 modulo 4. Thus $\mu(A) = 1$ and $\mu\{n\} > 0$ for $n \in A$; $\nu(B) = 1$ and $\nu\{n\} = 0$ for all n. (This example is related to one of Dubins [4], page 205.)

LEMMA 5.1. *There is no posterior for the prior* π.

PROOF. Assume, to get a contradiction, that π has a posterior q and let m be the corresponding marginal on X. Let O be the set of odd integers. Clearly, $p_\theta(O) = 1$ for $\theta \in E = A \cup B$ and, by (2.1), $m(O) = 1$ also.

The key point is that $q_x(A) = 1$ for all $x \in 0$. To see this, suppose $x = 4n + 1$, write P for the joint distribution and calculate

$$P[\theta = 4n, x = 4n + 1] = \pi\{4n\}p_{4n}\{4n + 1\} = \mu\{4n\}/4.$$

Also,

$$P[\theta = 4n, x = 4n + 1] = m\{4n + 1\}q_x\{4n\} = \mu\{4n\}q_x\{4n\}/4.$$

Hence

$$q_x(A) = q_x\{4n\} = 1.$$

Similarly, if $x = 4n + 3$,

$$q_x(A) = q_x\{4n + 4\} = 1.$$

Thus

$$P(A \times X) = \int q_x(A) m(dx) = 1.$$

But

$$P(A \times X) = \pi(A) = \tfrac{1}{2},$$

a contradiction. □

Our final result will imply the existence of approximate posteriors in the example just given. To formulate the result, let π be a prior and p be a model. Then π and p determine a probability distribution P on $\Theta \times X$ by the formula

$$P(A) = \int p_\theta(A_\theta)\pi(d\theta), \tag{5.1}$$

where $A \subseteq \Theta \times X$, $A_\theta = \{x: (\theta, x) \in A\}$. For $\varepsilon > 0$, an inference q is called an *ε-posterior* for π if

$$\left| P(A) - \int q_x(A^x) m(dx) \right| \le \varepsilon \tag{5.2}$$

for every $A \subseteq \Theta \times X$. Here m is the marginal of P on X as in (2.1) and $A^x = \{\theta: (\theta, x) \in A\}$.

PROPOSITION 5.1. *Let $X = \Theta = (0, \pm 1, \pm 2, \dots\}$ and let p be a translation family so that $p_\theta(A) = p_0(A - \theta)$ for $A \subset X$. If p_0 is countably additive, then every prior π has an ε-posterior for every $\varepsilon > 0$.*

PROOF. We can assume p_0 has finite support. To see this, let $0 < \delta < \frac{1}{2}$ and use the countable additivity of p_0 to choose a finite set $F \subset X$ such that $p_0(F) > 1 - \delta$. Define a new translation family p^δ by taking

$$p_0^\delta(B) = p_0(B \cap F)/p_0(F)$$

for $B \subset X$. Then define $p_\theta^\delta(B) = p_0^\delta(B - \theta)$ and define P^δ by (5.1) with p_θ replaced by p_θ^δ. Then

$$|P^\delta(A) - P(A)| \le 3\delta$$

for every $A \subset \Theta \times X$.

Assume now that $p_0(F) = 1$, where F is finite. Let $L = \{(\theta, x): x + \theta \in F\}$. Notice that the θ-section of L is $L_\theta = F - \theta$ so that, by (5.1), $P(L) = 1$. Next consider the partition of L whose elements are the x-sections $L^x = F - x$. Because each L^x has the same number of elements as F, it follows from Dubins ([4], Proposition 1, page 95) that there is a q satisfying (5.2) for every $A \subset \Theta \times X$. □

The proposition still holds (with the same proof) if the integers are replaced by any countable group. We do not know if it holds for translation families on larger groups such as the real numbers. We also do not know whether the conclusion necessarily holds when every p_θ is countably additive but p is not a translation family. The conclusion need not hold for finitely additive p as is shown in [4] and [9].

Acknowledgment. We want to thank two referees who made a number of useful suggestions and saved us from an erroneous proof of an earlier version of Theorem 4.2.

REFERENCES

[1] ARMSTRONG, T. E. and SUDDERTH, W. D. (1981). Nearly strategic measures. *Pacific J. Math.* **94** 251–257.

[2] DE FINETTI, B. (1972). *Probability, Induction, and Statistics*. Wiley, New York.

[3] DE FINETTI, B. (1975). *Theory of Probability* **2**. Wiley, New York.

[4] DUBINS, L. (1975). Finitely additive conditional probabilities, conglomerability and disintegrations. *Ann. Probab.* **3** 89–99.

[5] EATON, M. L. (1982). A method for evaluating improper prior distributions. In *Statistical Decision Theory and Related Topics III* (S. S. Gupta and J. O. Berger, eds.) **1** 329–351. Academic, New York.

[6] HEATH, D. and SUDDERTH, W. (1978). On finitely additive priors, coherence, and extended admissibility. *Ann. Statist.* **6** 333–345.

[7] LANE, D. and SUDDERTH, W. (1983). Coherent and continuous inference. *Ann. Statist.* **11** 114–120.

[8] LANE, D. and SUDDERTH, W. (1984). Coherent predictive inference. *Sankhyā Ser. A* **46** 166–185.

[9] PRIKRY, K. and SUDDERTH, W. (1982). Singularity with respect to strategic measures. *Illinois J. Math.* **26** 460–465.

[10] REGAZZINI, E. (1987). de Finetti's coherence and statistical inference. *Ann. Statist.* **15** 845–864.

[11] STONE, M. (1965). Right Haar measure for convergence in probability to quasi posterior distributions. *Ann. Math. Statist.* **36** 440–453.

[12] STONE, M. (1970). Necessary and sufficient condition for convergence in probability to invariant posterior distributions. *Ann. Math. Statist.* **41** 1349–1353.

[13] STONE, M. (1976). Strong inconsistency from uniform priors. *J. Amer. Statist. Assoc.* **71** 114–116.

SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
UPSON HALL
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853-7501

SCHOOL OF STATISTICS
270 VINCENT HALL
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MINNESOTA 55455