The Annals of Statistics
1994, Vol. 22, No. 3, 1177–1194

COHERENT INFERENCES AND IMPROPER PRIORS[1]

BY PATRIZIA BERTI AND PIETRO RIGO
Università di Firenze

Formal posteriors for improper priors are investigated in connection with coherence, both in the sense of Regazzini and of Heath and Sudderth. Those priors π which are linked with the improper prior γ by the relation $\pi(B) = 0$ whenever $\gamma(B) < +\infty$ are studied in particular. Moreover, a characterization of the inferences which are coherent according to Heath and Sudderth is found, and several examples, exhibiting several phenomena, are given.

1. Introduction. When facing a Bayesian inferential problem, a standard practice is to use so-called *improper priors*. Roughly speaking, this means that the inferrer does not explicitly state his initial opinions by assessing a probability on the parameter space Θ. Rather, he selects an unbounded measure γ on Θ and declares the inference

$$q_x(B) = \frac{\int_B l(x, \theta)\gamma(d\theta)}{\int_\Theta l(x, \theta)\gamma(d\theta)}, \qquad B \subset \Theta, \tag{1.1}$$

for each sample observation x for which the denominator of (1.1) is finite and positive, $l(x, \cdot)$ denoting the likelihood of x.

In spite of its popularity, the substantial meaning of the aforementioned procedure is not totally clear. Indeed, when dealing with improper priors, different authors intend different things. This circumstance can clearly give rise to misunderstandings. A related shortcoming is that improper priors are sometimes used mechanically, without uniquely specifying their role within the problem, and this can lead to (presumed) paradoxes.

This kind of difficulty does not occur when the inferential analysis is based on the notion of coherence. As far as coherence is concerned, it is immaterial that the inference (1.1) is linked in a certain way with the improper prior γ. The only fundamental fact is whether (1.1) is coherent with respect to the other "ingredients" of the problem (i.e., the sampling model, the prior distribution etc.).

Since the inferrer has not assessed any prior probability on Θ, however, one such ingredient is not available. Hence, to test coherence of (1.1), some probability π on Θ is to be selected. Plainly, any choice of π is acceptable and leads on to some conclusion on the coherence of (1.1). Yet, it seems convenient to pay particular attention to those π's which are linked, in some reasonable sense, with γ. In other terms, it is plausible to suppose that γ is able to give us some rough (and partial) indication about the inferrer's initial opinions.

Received June 1991; revised December 1993.
[1]Research partially supported by M.U.R.S.T. "Coerenza ed inferenza statistica predittiva."
AMS 1991 *subject classifications.* Primary 62A15; secondary 60A05.
Key words and phrases. Bayesian inference, coherence, finite additivity, improper priors.

In most of this paper, π and γ are assumed to satisfy

$$\gamma(B) < +\infty \quad \Rightarrow \quad \pi(B) = 0, \tag{1.2}$$

for each $B \subset \Theta$ in the domain of γ. Under this hypothesis, (1.1) is investigated in connection with two distinct notions of coherence for statistical inference: one proposed by Heath and Sudderth (1978), the other by Regazzini (1987).

Assuming γ is not independent of the inferrer's opinions, condition (1.2) is able to grasp something of the latter. As a classic example, take $\Theta = \mathbb{R}$ and let γ be the Lebesgue measure. In that case, when B is compact, the "ratio" between $\gamma(B)$ and $\gamma(B^c)$ is finite versus infinite; if π is asked to preserve such a ratio, then imposing some form of (1.2) seems unavoidable. Looking back at the general case, we also note that if γ is supposed to be σ-finite, under (1.2) there is a countable partition $\{B_n\}$ of Θ such that $\pi(B_n) = 0$ for each n. Thus, (1.2) implies that π is not σ-additive. Actually, in this paper all probability laws are intended to be finitely additive.

A further remark concerns the two above-quoted concepts of coherence. A discussion of the underlying ideas, as well as a comparison between them, is beyond the aims of this paper. The interested reader is referred to Heath and Sudderth (1978), Lane and Sudderth (1983), Regazzini (1987), Berti, Regazzini and Rigo (1991) and Brunk (1991). Here we only point out that, in our opinion, an inference is acceptable only if coherent in Regazzini's sense. Indeed, Regazzini's definition has been obtained by applying to the inferential setting a strengthened version of de Finetti's coherence principle. On the other hand, coherence in the sense of Heath and Sudderth is generally not fundamental in order to regard an inference as acceptable. Nevertheless, if viewed as an optional requisite, Heath and Sudderth's coherence is surely interesting and is important for many classes of problems.

Finally, this paper is organized as follows. Section 2 is devoted to some terminology and notation. Section 3 includes a characterization of the inferences which are coherent according to Heath and Sudderth. In Section 4 it is proved that if π fulfills (1.2), then the inference (1.1) is coherent in Regazzini's sense. Moreover, at least for discrete models, a partial converse is obtained. Section 5 deals with the issue of identifying those priors π which are "permissible" for (1.1) when the latter is coherent according to Heath and Sudderth. More precisely, conditions are given amounting to the existence of a prior π such that the following hold: (i) π makes (1.1) a coherent inference; (ii) π satisfies (1.2), or π is (in some suitable sense) a limit of truncations of γ or π is σ-additive. It is also proved that, under some assumptions, a necessary condition for (i) is that (1.2) holds.

Moreover, throughout the paper several examples are given. In particular, in connection with coherence in Heath and Sudderth's sense, it is shown that, unlike what is stated in Lane and Sudderth [(1983), Proposition 2.1], consistency of an inference does not imply coherence. It is also proved that (1.1) can result in a coherent inference without admitting any prior of one of the types listed in (ii).

2. Terminology and notation. In this paper, given any set S, the term *probability* is meant as a nonnegative, finitely additive set function, defined on some field $\mathfrak{F}$ of subsets of S, and assuming value 1 at S. A probability μ on $\mathfrak{F}$ is said to be *diffuse* whenever the singletons $\{s\}$ are in $\mathfrak{F}$ and $\mu(\{s\}) = 0 \ \forall \ s \in S$. Let $\mathfrak{U} \subset \mathfrak{F} - \{\emptyset\}$. A *conditional probability* on $\mathfrak{F} \times \mathfrak{U}$ is a function P: $\mathfrak{F} \times \mathfrak{U} \to \mathbb{R}$ such that the following hold: (i) $\forall \ H \in \mathfrak{U}$, $P(\cdot \mid H)$ is a probability on $\mathfrak{F}$ with $P(H \mid H) = 1$; (ii) $P(E \cap F \mid H) = P(E \mid F \cap H)P(F \mid H)$ whenever $E, F \in \mathfrak{F}$ and $F \cap H, H \in \mathfrak{U}$.

Next, the same symbol that designates a subset of S designates its indicator as well. Thus, if $E \subset S$, then E also denotes the function on S assuming value 1 on E and value 0 out of E. Moreover, $\mathcal{P}(S)$ designates the class of all subsets of S, and $L(\mathfrak{F})$ denotes the family of bounded, $\mathfrak{F}$-measurable functions f: $S \to \mathbb{R}$.

Let us turn now to the notation specifically connected with an inferential problem. As usual, $\mathfrak{X}$ and Θ are nonempty sets to be regarded, respectively, as the collection of possible outcomes of an experiment and the collection of possible realizations of a random parameter. Also, $\mathfrak{X}$, Θ and $\mathfrak{X} \times \Theta$ are equipped with σ-fields of subsets $\mathcal{A}_{\mathfrak{X}}$, $\mathcal{A}_{\Theta}$ and $\mathcal{A}$; in particular, $\mathcal{A}$ coincides with the product σ-field $\mathcal{A}_{\mathfrak{X}} \otimes \mathcal{A}_{\Theta}$.

If C is a subset of $\mathfrak{X} \times \Theta$, then $C^x := \{\theta: (x, \theta) \in C\}$ and $C_\theta := \{x: (x, \theta) \in C\}$ denote the sections of C with respect to (w.r.t.) $x \in \mathfrak{X}$ and $\theta \in \Theta$. Similarly, given a function ϕ on $\mathfrak{X} \times \Theta$, ϕ^x and ϕ_θ are the functions, on Θ and $\mathfrak{X}$, respectively, defined by $\phi^x(\theta) = \phi(x, \theta)$ and $\phi_\theta(x) = \phi(x, \theta)$. Let p_θ and q_x be probabilities defined, respectively, on $\mathcal{P}(\mathfrak{X})$ and $\mathcal{P}(\Theta)$ (or, possibly, on $\mathcal{A}_{\mathfrak{X}}$ and $\mathcal{A}_{\Theta}$). Then the integrals of ϕ_θ and ϕ^x w.r.t. p_θ and q_x are also written as $p_\theta(\phi_\theta)$ and $q_x(\phi^x)$, that is,

$$p_\theta(\phi_\theta) = \int \phi(x, \theta) p_\theta(dx) \quad \text{and} \quad q_x(\phi^x) = \int \phi(x, \theta) q_x(d\theta).$$

Finally, the above integrals, as well as all the integrals included in this paper, are to be intended in the sense of Dunford and Schwartz [(1958), Chapter 3]. However, we point out that, in the particular case where the integrand f is measurable and the involved measure ν is σ-additive, $\int f \, d\nu$ can be equivalently regarded as a Lebesgue integral.

3. Some results on coherence in the sense of Heath and Sudderth. The main result of this section is a characterization of the inferences which are coherent according to Heath and Sudderth (in short, HS-coherent inferences). In addition, some remarks are given together with a counterexample.

Let us start by introducing the notion of HS-coherence [Heath and Sudderth (1978)]. A (*sampling*) *model* is a mapping p on Θ which associates a probability p_θ on $\mathcal{P}(\mathfrak{X})$ with each θ in Θ; likewise, an *inference* is a function q on $\mathfrak{X}$ which associates a probability q_x on $\mathcal{P}(\Theta)$ with each x in $\mathfrak{X}$. Moreover, any probability π on $\mathcal{P}(\Theta)$ is said to be a *prior*. For a fixed model p, each prior π induces a probability r on $\mathcal{P}(\mathfrak{X} \times \Theta)$ given by $r(C) = \int p_\theta(C_\theta)\pi(d\theta)$ for $C \subset \mathfrak{X} \times \Theta$. In its turn, r induces a probability m on $\mathcal{P}(\mathfrak{X})$, also called the *marginal of* π *on* $\mathfrak{X}$, by setting

$$m(A) = r(A \times \Theta) = \int p_\theta(A)\pi(d\theta) \quad \forall\, A \subset \mathfrak{X}. \tag{3.1}$$

We are now able to state the concept of HS-coherence.

DEFINITION 3.1. An inference q is *HS-coherent*, w.r.t. a given model p, if there exists a prior π such that

$$\int\int \phi(x,\theta)p_\theta(dx)\pi(d\theta) = \int\int \phi(x,\theta)q_x(d\theta)m(dx) \quad \forall\, \phi \in L(\mathcal{A}), \tag{3.2}$$

where m is defined by (3.1). Moreover, if (3.2) holds, then q is said to be a *posterior for* π and/or π is said to be a *prior for* q.

From now on, it is assumed that a model p has been assigned, so that any statement is to be referred to the fixed model p. In addition, we denote by $\mathcal{V}$ the class of HS-coherent inferences. The following (drastic) example shows that $\mathcal{V} \neq \emptyset$; note, however, that, for a fixed prior π, it may be that a posterior q for π does not exist.

EXAMPLE 3.2. Pick $t \in \Theta$ and take the prior as $\pi = \delta_t$ and the inference as $q_x = \delta_t$ for every x, δ_t denoting the unit mass at t. Then, the marginal m of π on $\mathfrak{X}$ is given by $m(A) = \int p_\theta(A)\pi(d\theta) = p_t(A)$ for each $A \subset \mathfrak{X}$. Hence

$$\begin{aligned}\int q_x(\phi^x)m(dx) &= \int \delta_t(\phi^x)p_t(dx)\\ &= \int \phi(x,t)p_t(dx) = p_t(\phi_t) = \int p_\theta(\phi_\theta)\pi(d\theta) \quad \forall\, \phi \in L(\mathcal{A}).\end{aligned}$$

The following theorem provides a criterion for testing whether an inference is the posterior for a certain type of prior.

THEOREM 3.3. *Let $\{q^n\}$ be a sequence of elements of $\mathcal{V}$; for each n, let π_n be a prior for q^n and let m_n be the marginal of π_n on $\mathfrak{X}$. Given an inference q, the following statements are equivalent*:

(i) $\inf_n \int [q^n_x(\phi^x) - q_x(\phi^x)]m_n(dx) \le 0 \quad \forall\, \phi \in L(\mathcal{A})$;

(ii) *there is a probability μ on $\mathcal{P}(\mathbb{N})$ such that*

$$\int\int \big[q^n_x(\phi^x) - q_x(\phi^x)\big]m_n(dx)\mu(dn) = 0 \quad \forall\, \phi \in L(\mathcal{A});$$

(iii) *there is a probability μ on $\mathcal{P}(\mathbb{N})$ such that q is a posterior for the prior $\pi(\cdot) = \int \pi_n(\cdot)\mu(dn)$.*

PROOF. [(i) $\Rightarrow$ (ii).] Let F be the set of functions $f: \mathbb{N} \to \mathbb{R}$ defined by

$$f(n) = \int \big[q_x^n(\phi^x) - q_x(\phi^x)\big] m_n(dx) \quad \forall\, n \in \mathbb{N},$$

for ϕ varying in $L(\mathcal{A})$. Clearly, F is a linear space of bounded functions and, by (i), $\inf f \leq 0 \;\forall\, f \in F$. Hence, by a suitable version of the separating hyperplane theorem [Heath and Sudderth (1978), Lemma 1], there is a probability μ on $\mathcal{P}(\mathbb{N})$ such that $\int f\, d\mu \leq 0 \;\forall\, f \in F$. Since $f \in F$ implies $-f \in F$, it must be that $\int f\, d\mu = 0 \;\forall f \in F$, and this amounts to (ii).

[(ii) $\Rightarrow$ (iii).] Let $\pi(B) = \int \pi_n(B)\mu(dn) \;\forall\, B \subset \Theta$. Then π is a probability on $\mathcal{P}(\Theta)$, and $\int g(\theta)\pi(d\theta) = \int\int g(\theta)\pi_n(d\theta)\mu(dn)$ for every bounded $g: \Theta \to \mathbb{R}$. Accordingly, the marginal of π on $\mathcal{X}$ is given by

$$m(A) = \int p_\theta(A)\pi(d\theta) = \int\int p_\theta(A)\pi_n(d\theta)\mu(dn) = \int m_n(A)\mu(dn) \quad \forall\, A \subset \mathcal{X}.$$

Fix $\phi \in L(\mathcal{A})$. Then

$$\int p_\theta(\phi_\theta)\pi(d\theta) = \int\int p_\theta(\phi_\theta)\pi_n(d\theta)\mu(dn) = \int\int q_x^n(\phi^x) m_n(dx)\mu(dn), \tag{3.3}$$

where the second equality depends on the fact that q^n is a posterior for π_n. On the other hand, since $\int f(x)m(dx) = \int\int f(x)m_n(dx)\mu(dn)$ for every bounded $f: \mathcal{X} \to \mathbb{R}$, one obtains

$$\begin{aligned} \int q_x(\phi^x)m(dx) &= \int\int q_x^n(\phi^x)m_n(dx)\mu(dn) \\ &\quad - \int\int \big[q_x^n(\phi^x) - q_x(\phi^x)\big] m_n(dx)\mu(dn). \end{aligned} \tag{3.4}$$

Hence, comparing (3.3) with (3.4), condition (ii) implies that q is a posterior for π.

[(iii) $\Rightarrow$ (i).] Since it is trivial that (ii) $\Rightarrow$ (i), it suffices to prove that (iii) $\Rightarrow$ (ii), and this follows from (3.3) and (3.4). □

As a consequence of Theorem 3.3, it is possible to state the following characterization of HS-coherent inferences.

COROLLARY 3.4. *Let q be an inference. Then q is in $\mathcal{V}$ if and only if there is a sequence $\{q^n\} \subset \mathcal{V}$ such that condition* (i) *of Theorem* 3.3 *holds.*

The remaining part of this section includes a few remarks concerning some results of Heath, Lane and Sudderth. Recently, Heath and Sudderth [(1989), Theorem 3.1] have stated that an inference q is HS-coherent if and only if there is a sequence $\{q^n\} \subset \mathcal{V}$ such that

$$\lim_n \int \sup_g \big|q_x^n(g) - q_x(g)\big| m_n(dx) = 0, \tag{3.5}$$

where g varies in $\{f \in L(\mathcal{A}_\Theta)\colon \sup|f| \leq 1\}$ and, as usual, m_n is the marginal on $\mathcal{X}$ of some prior π_n for q^n. Now, once a particular sequence $\{q^n\} \subset \mathcal{V}$ is fixed, it is easily shown that (3.5) implies condition (i) of Theorem 3.3 but the converse is not true. In this sense, in order to test HS-coherence, condition (i) seems to be a more flexible tool than condition (3.5). Indeed, while (i) implies that q is a posterior for $\pi(\cdot) = \int \pi_n(\cdot)\mu(dn)$, for some particular μ, (3.5) implies that q is a posterior for *any* prior in a certain class.

COROLLARY 3.5. *Let q be an inference. If a sequence $\{q^n\} \subset \mathcal{V}$ exists such that* (3.5) *holds, then $\pi(\cdot) = \int \pi_n(\cdot)\mu(dn)$ is a prior for q, for every diffuse probability μ on $\mathcal{P}(\mathbb{N})$.*

PROOF. Let F be as in the proof of Theorem 3.3. Under (3.5), $\lim f(n) = 0$ $\forall f \in F$. Thus, for a fixed diffuse μ, condition (ii) of Theorem 3.3 is satisfied for such μ. □

We also note that Corollary 3.5 implies (the nontrivial part of) the result of Heath and Sudderth, so that the latter actually holds. Yet, their proof is not technically correct. To make this point more precise, we need one more definition. An inference q is said to be *consistent* with a model p if there are probabilities m on $\mathcal{P}(\mathcal{X})$ and π on $\mathcal{P}(\Theta)$ such that $\int p_\theta(\phi_\theta)\pi(d\theta) = \int q_x(\phi^x)m(dx)\ \forall \phi \in L(\mathcal{A})$, or, equivalently, if

$$\inf_\theta p_\theta(\phi_\theta) \leq \sup_x q_x(\phi^x) \quad \forall \phi \in L(\mathcal{A}). \tag{3.6}$$

In a 1983 paper, after having introduced the notion of consistency, Lane and Sudderth [(1983), Proposition 2.1] claim that consistency of q with p is equivalent to HS-coherence of q (w.r.t. p). In their 1989 paper, Heath and Sudderth show that (3.5) implies (3.6) and, basing themselves on the quoted 1983 statement, conclude that (3.5) implies HS-coherence of q. However, unless suitable measurability constraints hold, consistency of q with p does not imply HS-coherence of q. The latter assertion is proved by the following example.

EXAMPLE 3.6 (Consistency of q with p does not imply HS-coherence of q w.r.t. p). Given the finite sets $\mathcal{X} = \{x_1, \dots, x_4\}$ and $\Theta = \{\theta_1, \theta_2\}$, define p and q as follows:

$$p_{\theta_1} = \tfrac{1}{2}(\delta_{x_1} + \delta_{x_2}), \qquad p_{\theta_2} = \tfrac{1}{2}(\delta_{x_3} + \delta_{x_4}), \qquad q_{x_1} = q_{x_3} = \delta_{\theta_1}, \qquad q_{x_2} = q_{x_4} = \delta_{\theta_2},$$

where δ_t denotes the unit mass at t. Suppose now that $\mathcal{A}_\mathcal{X}$ and $\mathcal{A}_\Theta$ are taken to be $\mathcal{A}_\mathcal{X} = \{\emptyset, \mathcal{X}, \{x_1, x_2\}, \{x_3, x_4\}\}$ and $\mathcal{A}_\Theta = \mathcal{P}(\Theta)$. If $C = A \times B$, where $A = \{x_1, x_2\}$ or $A = \{x_3, x_4\}$ and $B = \{\theta_1\}$ or $B = \{\theta_2\}$, there is an x such that $q_x(C^x) = 1$. Since every $\phi \in L(\mathcal{A})$ is constant on such sets $A \times B$, it must be that $\sup_x q_x(\phi^x) = \sup \phi$. Therefore (3.6) holds, that is, q is consistent with p. On the other hand, let π be any probability on $\mathcal{P}(\Theta)$ and let m be the marginal of π on $\mathcal{X}$. If π is a prior for

q, then setting $C = \{x_1, x_2\} \times \{\theta_1\}$ yields

$$\pi(\{\theta_1\}) = \int p_\theta(C_\theta)\pi(d\theta) = \int q_x(C^x)m(dx) = m(\{x_1\}) = \tfrac{1}{2}\pi(\{\theta_1\}),$$

so that $\pi(\{\theta_1\}) = 0$. However, for $C = \{x_3, x_4\} \times \{\theta_2\}$, the same argument gives $\pi(\{\theta_2\}) = 0$, a contradiction. Hence, q is not HS-coherent w.r.t. p.

It can be easily shown that if the function $x \to q_x(B)$ is $\mathcal{A}_\mathfrak{X}$-measurable for all $B \in \mathcal{A}_\Theta$, then consistency of q with p implies HS-coherence of q (w.r.t. p). In other terms, consistent measurable inferences are HS-coherent. Indeed, asking measurability for q is not an actual constraint in most practical problems.

4. Coherence, in Regazzini's sense, of a formal posterior for an improper prior. In the rest of this paper, contrary to Section 3, the various probability laws (i.e., p_θ, q_x, π, r and m) are intended to be defined on the σ-fields $\mathcal{A}_\mathfrak{X}$, $\mathcal{A}_\Theta$ and $\mathcal{A}$, and not necessarily on $\mathcal{P}(\mathfrak{X}), \mathcal{P}(\Theta)$ and $\mathcal{P}(\mathfrak{X} \times \Theta)$. In addition, it is assumed that

$$\{x\} \in \mathcal{A}_\mathfrak{X} \quad \text{and} \quad \{\theta\} \in \mathcal{A}_\Theta \quad \forall\, (x, \theta) \in \mathfrak{X} \times \Theta.$$

Let γ be an *improper prior*, that is, a σ-additive and σ-finite measure $\gamma\colon \mathcal{A}_\Theta \to [0, \infty]$ such that $\gamma(\Theta) = +\infty$. Let us suppose that the model p is "dominated," in the sense that there are both a finitely additive measure $\lambda\colon \mathcal{A}_\mathfrak{X} \to [0, \infty]$ and a function $l\colon \mathfrak{X} \times \Theta \to [0, \infty)$ such that

$$p_\theta(A) = \int_A l(x, \theta)\lambda(dx) \quad \forall\, \theta \in \Theta,\ \forall\, A \in \mathcal{A}_\mathfrak{X}. \tag{4.1}$$

Suppose also that, for every fixed $x \in \mathfrak{X}$, the function $\theta \to l(x, \theta)$ is $\mathcal{A}_\Theta$-measurable.

Under these assumptions, define

$$\rho(x) = \int l(x, \theta)\gamma(d\theta) \quad \forall\, x \in \mathfrak{X}.$$

Then, after setting $D = \{x\colon 0 < \rho(x) < +\infty\}$, any inference q such that

$$q_x(B) = \frac{1}{\rho(x)} \int_B l(x, \theta)\gamma(d\theta) \quad \forall\, x \in D,\ \forall\, B \in \mathcal{A}_\Theta \tag{4.2}$$

will be called *a formal posterior* (for γ). Plainly, it is tacitly assumed that $D \neq \emptyset$.

This section is essentially concerned with dF-coherence, that is coherence in Regazzini's sense [Regazzini (1987)], of a formal posterior. In particular it is shown that, provided that the inferrer's initial opinions are linked with γ in a certain manner, there always exists a dF-coherent formal posterior.

Let us introduce the notion of dF-coherence. Let p be a model [not necessarily the dominated model (4.1)], π a prior on $\mathcal{A}_\Theta$ and r any probability on $\mathcal{A}$. Briefly, an inference q is dF-coherent, w.r.t. (p, π, r), if all the "ingredients" of the problem (i.e., p, π, r and q) are restrictions of the same conditional probability on $\mathcal{A} \times \mathcal{A}^0$, where $\mathcal{A}^0 = \mathcal{A} - \{\emptyset\}$. In this sense, a dF-coherent inferrer is actually working with a conditional probability on $\mathcal{A} \times \mathcal{A}^0$. We refer to Regazzini (1987) for the precise definition, since we need only the following characterization, which is a slightly different version of Theorem 2.2 of Berti, Regazzini and Rigo (1991).

THEOREM 4.1. *Let p be a model, π a prior on $\mathcal{A}_\Theta$ and r any probability on $\mathcal{A}$. An inference q is dF-coherent, w.r.t. (p, π, r), if and only if the following hold:*

(a) $$\pi(B) = r(\mathfrak{X} \times B) \qquad \forall\, B \in \mathcal{A}_\Theta;$$

(b) $$p_\theta(A) r(\mathfrak{X} \times \{\theta\}) = r(A \times \{\theta\}) \quad \forall\, \theta \in \Theta,\ \forall\, A \in \mathcal{A}_{\mathfrak{X}};$$

(c) $$q_x(B) r(\{x\} \times \Theta) = r(\{x\} \times B) \quad \forall\, x \in \mathfrak{X},\ \forall\, B \in \mathcal{A}_\Theta;$$

(d) $$\prod_{i=1}^{n} p_{\theta_i}(\{x_{i+1}\}) q_{x_i}(\{\theta_i\}) = \prod_{i=1}^{n} p_{\theta_i}(\{x_i\}) q_{x_{i+1}}(\{\theta_i\}),$$
$$\prod_{i=1}^{n} p_{\theta_i}(\{x_i\}) q_{x_i}(\{\theta_{i+1}\}) = \prod_{i=1}^{n} p_{\theta_{i+1}}(\{x_i\}) q_{x_i}(\{\theta_i\}),$$

$\forall\, x_1, \ldots, x_n \in \mathfrak{X}$, $\forall\, \theta_1, \ldots, \theta_n \in \Theta$, $x_{n+1} = x_1$, $\theta_{n+1} = \theta_1$ *and* $r(\{x_i\} \times \Theta) = r(\mathfrak{X} \times \{\theta_i\}) = 0\ \forall\, i$.

Let π be a prior on $\mathcal{A}_\Theta$. After selecting a probability α on $\mathcal{P}(\Theta)$ such that $\alpha(B) = \pi(B)$ for $B \in \mathcal{A}_\Theta$, define

$$r(C) = \int p_\theta(C_\theta)\alpha(d\theta) \quad \forall\, C \in \mathcal{A}. \tag{4.3}$$

Further, let q be the formal posterior such that, for $x \notin D$,

$$q_x(B) = \begin{cases} \dfrac{r(\{x\} \times B)}{r(\{x\} \times \Theta)}, & \text{if } r(\{x\} \times \Theta) > 0, \\ d_x(B), & \text{if } r(\{x\} \times \Theta) = 0, \end{cases} \tag{4.4}$$

where $B \in \mathcal{A}_\Theta$ and d_x is any diffuse probability on $\mathcal{A}_\Theta$.

PROPOSITION 4.2. *Let γ be an improper prior, p the model* (4.1), *π a prior on $\mathcal{A}_\Theta$, r the probability* (4.3) *and q the inference defined by* (4.2) *and* (4.4). *A sufficient condition for q to be dF-coherent, w.r.t. (p, π, r), is that*

$$\forall\, B \in \mathcal{A}_\Theta, \qquad \gamma(B) < +\infty \quad \Rightarrow \quad \pi(B) = 0. \tag{4.5}$$

PROOF. By Theorem 4.1, it suffices to verify conditions (a)–(d). First, (a) and (b) are direct consequences of (4.3). Second, (c) trivially holds for $x \notin D$.

Let $x \in D$. Then $l(x,\cdot)$ is γ-integrable, and (4.5) implies that any γ-integrable function g is such that $\int g\, d\pi = 0$. Hence

$$r(\{x\} \times \Theta) = \int p_\theta(\{x\})\alpha(d\theta) = \int \lambda(\{x\}) l(x,\theta)\alpha(d\theta) = \lambda(\{x\}) \int l(x,\theta)\pi(d\theta) = 0,$$

so that (c) is satisfied. Third, to check (d), fix $x_1, \dots, x_n$ with $r(\{x_i\} \times \Theta) = 0\ \forall\, i$. If there is an index i with $x_i \notin D$, then, since q_{x_i} is diffuse, (d) is trivially true. On the other hand, if $x_i \in D\ \forall\, i$, then (d) follows by a direct calculation. □

At least for discrete models, Proposition 4.2 has a partial converse.

PROPOSITION 4.3. *Let Θ be a topological space, let $\mathcal{A}_\Theta$ be a σ-field including the compact sets, and let p, π, r and q be as in Proposition* 4.2 [*without asking* (4.5)]. *Suppose there is an $x_0 \in D$ such that $\lambda(\{x_0\}) > 0$ and $l(x_0,\cdot)$ is continuous and strictly positive. Then a necessary condition for q to be dF-coherent, w.r.t.* (p, π, r), *is that $\pi(K) = 0$ for every compact K.*

PROOF. Since $\lambda(\{x_0\}) > 0$ and $l(x_0,\theta)\lambda(\{x_0\}) = p_\theta(\{x_0\}) \le 1$, $l(x_0,\cdot)$ is bounded and hence π-integrable. Let $a = \int l(x_0,\theta)\pi(d\theta)$. Since $l(x_0,\cdot)$ is continuous and strictly positive, it suffices to prove that $a = 0$. Suppose on the contrary that $a \in (0,+\infty)$, and define $\pi^*(B) = a^{-1}\int_B l(x_0,\theta)\pi(d\theta)\, \forall B \in \mathcal{A}_\Theta$. Then, since $\lambda(\{x_0\}) > 0$, condition (c) of Theorem 4.1 implies that $\pi^* = q_{x_0}$. In particular, since $x_0 \in D$, π^* is σ-additive. Decompose π as $\pi = \pi_1 + \pi_2$, where π_1 is σ-additive and π_2 is a *pure* finitely additive measure [cf. Bhaskara Rao and Bhaskara Rao (1983), Theorem 10.2.1]. Let $\pi_2^*(B) = a^{-1}\int_B l(x_0,\theta)\pi_2(d\theta)$. If $\pi_2^*(\Theta) > 0$, then π_2^* is easily shown to be non-σ-additive. Hence, σ-additivity of π^* implies $\pi_2^*(\Theta) = 0$, that is, π^* has the density $a^{-1}l(x_0,\cdot)$ w.r.t. the σ-additive measure π_1. Thus $\int[1/l(x_0,\theta)]\pi^*(d\theta) = a^{-1}\pi_1(\Theta) < +\infty$, while $\int[1/l(x_0,\theta)]q_{x_0}(d\theta) = +\infty$, a contradiction. Hence, $a = 0$. □

EXAMPLE 4.4. Let p be an exponential family, so that

$$l(x,\theta) = \exp\left(\sum_{i=1}^{k} a_i(\theta)b_i(x) - c(\theta)\right),$$

where the a_i's and c are $\mathcal{A}_\Theta$-measurable. Given any improper prior γ, let q be a formal posterior. By Proposition 4.2, q is dF-coherent w.r.t. (p,π,r), whenever π satisfies (4.5) [and q_x is defined as in (4.4) for $x \notin D$]. Further, suppose $\mathcal{A}_\Theta$ is the Borel σ-field on some $\Theta \subset \mathbb{R}^m$, the a_i's and c are continuous and $\lambda(\{x\}) > 0$ for all x. Then, in view of Proposition 4.3, the only possibility to make q dF-coherent is that π vanishes on compact sets.

Before leaving this section, we introduce a certain class of priors satisfying condition (4.5). Let μ be a diffuse probability on $\mathcal{P}(\mathbb{N})$, and let $\{V_n\}$ be a sequence of elements of $\mathcal{A}_\Theta$ such that $V_n \uparrow \Theta$ and $0 < \gamma(V_n) < +\infty$. Then, setting

$$\pi(B) = \int \frac{\gamma(B \cap V_n)}{\gamma(V_n)}\mu(dn) \quad \forall\, B \in \mathcal{A}_\Theta \tag{4.6}$$

yields a probability π on $\mathcal{A}_\Theta$. Furthermore, since μ is diffuse, one obtains

$$(4.7)\qquad \pi(B) = \lim_n \frac{\gamma(B \cap V_n)}{\gamma(V_n)} \quad \text{whenever } B \in \mathcal{A}_\Theta \text{ and the limit does exist.}$$

In particular, since γ is σ-additive and $\gamma(\Theta) = +\infty, \pi$ fulfills (4.5). Incidentally, (4.7) also shows that Proposition 4.2 generalizes some results of Regazzini [(1987), Section 3].

The next example shows that not every π satisfying (4.5) admits representation (4.6).

EXAMPLE 4.5. Let $\Theta = \mathbb{R}$, let $\mathcal{A}_\Theta$ be the Borel σ-field, let γ be the Lebesgue measure and let $\mathcal{C} = \{B \in \mathcal{A}_\Theta\colon \gamma(B) < +\infty \text{ or } \gamma(B^c) < +\infty\}$. It is easily checked that $\mathcal{C}$ is a field and that, setting $\widetilde{\pi}(B) = 0$ or $\widetilde{\pi}(B) = 1$ according to $\gamma(B) < +\infty$ or $\gamma(B) = +\infty, \widetilde{\pi}$ turns out to be a probability on $\mathcal{C}$. Hence, $\widetilde{\pi}$ has an extension π to $\mathcal{A}_\Theta$ which is still a 0–1 probability. Plainly, (4.5) holds for π. Yet, representation (4.6) fails whatever the sequence $\{V_n\}$ may be. In fact, given $\{V_n\}$, there is a $B \in \mathcal{A}_\Theta$ such that $\gamma(B \cap V_n) = (1/2)\gamma(V_n)\ \forall\, n$, so that $\int [\gamma(B \cap V_n)/\gamma(V_n)]\mu(dn) = 1/2 \neq \pi(B)$ for each probability μ on $\mathcal{P}(\mathbb{N})$.

A result about the π in (4.6) is included in the next section. Here we note that, assuming that γ has some connection with the inferrer's opinions, such a π is a reasonable candidate to describe the latter. Actually, besides satisfying (4.5), π is, in some sense, a "limit" of truncations of γ. Thus, by suitably choosing these truncations (i.e., the sequence $\{V_n\}$), it is sometimes possible to reproduce some features of γ which are seen as significant for the problem at hand. For an example, see Heath and Sudderth [(1978), Section 4].

5. In search of a prior for a formal posterior. Throughout this section, q always denotes a formal posterior. According to HS-coherence, q is "permissible" if and only if it is the posterior for some prior. When this happens, however, there is usually more than one prior for q. Indeed, it is quite possible that every probability on $\mathcal{A}_\Theta$ is a prior for q [see also Brunk (1991), page 842].

EXAMPLE 5.1. Let $\mathcal{X} = [0, \infty)$ and $\Theta = \{0, 1, 2, \dots\}$; let $\mathcal{A}_\mathcal{X}$ be the Borel σ-field and $\mathcal{A}_\Theta = \mathcal{P}(\Theta)$; let λ be the Lebesgue measure; let $l(x, \theta) = 1$ if $x \in [\theta, \theta + 1)$, and $l(x, \theta) = 0$ otherwise. Taking γ as the counting measure yields $q_x = \delta_{[x]}$, where $\delta_{[x]}$ denotes the unit mass concentrated on the integer part of x. Hence, for fixed $\theta \in \Theta$ and $\phi \in L(\mathcal{A})$,

$$\int q_x(\phi^x)p_\theta(dx) = \int_{[\theta,\,\theta+1)} q_x(\phi^x)p_\theta(dx) = \int_{[\theta,\,\theta+1)} \phi(x, \theta)p_\theta(dx) = p_\theta(\phi_\theta),$$

and this clearly implies that q is a posterior for π whatever π is.

In real problems, the situation is not so extreme as in Example 5.1, and the possible priors for q are often bound to belong to some particular class of probabilities on $\mathcal{A}_\Theta$. Hence, it is useful to give conditions implying either the existence or the nonexistence of priors of a given form. The present section is just concerned with the identification of those priors which are behind an HS-coherent q. Specifically, we are interested in three kinds of priors: those satisfying (4.5); those of the form (4.6); and σ-additive priors.

Henceforth, we assume the following hypotheses:

$$D \in \mathcal{A}_{\mathfrak{X}} \text{ and } \lambda(D^c) = 0; \tag{5.1}$$

$$\lambda \text{ is } \sigma\text{-additive and } \sigma\text{-finite}; \tag{5.2}$$

$$l\colon \mathfrak{X} \times \Theta \to [0, \infty) \text{ is } \mathcal{A}\text{-measurable}. \tag{5.3}$$

Condition (5.1) is clearly needed in order to define q uniquely, up to λ-null sets. Conditions (5.2) and (5.3) are essentially for applying Fubini's theorem, which, as will become apparent, is a basic tool in dealing with HS-coherence of q. Note also that, under (5.1)–(5.3), there are no measurability problems.

As noted in Section 1, according to us an inference is acceptable precisely when it is dF-coherent. Consequently, before analyzing any other requisite of q, we must check its dF-coherence. The following corollary of Theorem 4.1 shows that if q is a posterior for some prior π and if q_x is suitably defined for $x \notin D$, then q is dF-coherent w.r.t. p, π and a particular r.

COROLLARY 5.2. *If q is a posterior for some prior π, and if q_x is a diffuse probability on $\mathcal{A}_\Theta$ for $x \notin D$, then q is dF-coherent w.r.t. (p, π, r), where $r(C) = \int p_\theta(C_\theta)\pi(d\theta)\ \forall\, C \in \mathcal{A}$.*

After these preliminaries, we begin to give conditions which amount to the existence of priors satisfying (4.5).

PROPOSITION 5.3. *There is a prior π for q such that $\gamma(B) < +\infty \Rightarrow \pi(B) = 0$ if and only if*

$$\forall\, \varepsilon > 0,\ \forall\, \phi \in L(\mathcal{A}), \qquad \gamma\left(\left\{\theta\colon p_\theta(\phi_\theta) - \int q_x(\phi^x)p_\theta(dx) \le \varepsilon\right\}\right) = +\infty. \tag{5.4}$$

PROOF. Let $\mathcal{Z}$ be the set of functions $f\colon \Theta \to \mathbb{R}$ defined by

$$f(\theta) = p_\theta(\phi_\theta) - \int q_x(\phi^x)p_\theta(dx),$$

for ϕ varying in $L(\mathcal{A})$; π is a prior for q precisely when $\int f\,d\pi = 0\ \forall\, f \in \mathcal{Z}$. Suppose now that there are ϕ and ε such that (5.4) fails, and denote by f the element of $\mathcal{Z}$ associated to ϕ. If π is such that $\gamma(B) < +\infty \Rightarrow \pi(B) = 0$, then $\pi(\{\theta\colon f(\theta) \le \varepsilon\}) = 0$, and hence $\int f\,d\pi \ge \varepsilon$. Conversely, assume that (5.4) holds. Let $M = \{f + g\colon f \in \mathcal{Z}, g \in L(\mathcal{A}_\Theta), g \text{ is } \gamma\text{-integrable}\}$, and let $T \equiv 0$ on M. If $\sup h \ge 0\ \forall\, h \in M$, then the Hahn–Banach theorem implies that T can be extended to $L(\mathcal{A}_\Theta)$ as a positive linear normed functional. Then the restriction π of T to $\mathcal{A}_\Theta$ is a prior for q such that $\gamma(B) < +\infty \Rightarrow \pi(B) = 0$. Fix $\varepsilon > 0$, $f \in \mathcal{Z}$ and $g \in L(\mathcal{A}_\Theta)$, where g is γ-integrable. Let $B = \{\theta: f(\theta) \geq -\varepsilon\}$. Since $\mathcal{Z}$ is a linear space, (5.4) gives $\gamma(B) = +\infty$, and this implies that g cannot be uniformly negative on B. Thus

$$\sup(f+g) \geq \sup_B (f+g) \geq \sup_B (-\varepsilon + g) \geq -\varepsilon. \qquad \square$$

We now give three results implying that, under some assumptions, the possible priors for q are to be sought among those satisfying (4.5), or at least some form of (4.5). The first one is an immediate consequence of Proposition 4.3 and Corollary 5.2, and can be useful when dealing with discrete models. The other two arise from the observation that, for fixed $x \in D$, q_x is absolutely continuous w.r.t. γ, in the sense that $q_x(B) \to 0$ as $\gamma(B) \to 0$. Hence, it is natural to investigate what happens when such absolute continuity is uniform in $x \in D$.

COROLLARY 5.4. *Let Θ be a topological space, and let $\mathcal{A}_\Theta$ be a σ-field including the compact sets. Suppose there is an $x_0 \in D$ such that $\lambda(\{x_0\}) > 0$ and $l(x_0, \cdot)$ is continuous and strictly positive. Then a necessary condition for q to be a posterior for some prior π is that $\pi(K) = 0$ for every compact K.*

PROPOSITION 5.5. *If l is strictly positive and, for any fixed $B \in \mathcal{A}_\Theta$ with $\gamma(B) < +\infty$,*

$$\lim_{\gamma(V) \to 0,\, V \subset B} \sup_{x \in D} q_x(V) = 0, \tag{5.5}$$

then a necessary condition for q to be a posterior for some prior π is that (4.5) *holds.*

PROPOSITION 5.6. *Let Θ be a σ-compact subset of $\mathbb{R}^k$, and let $\mathcal{A}_\Theta$ be the Borel σ-field. Suppose that l is strictly positive, q satisfies* (5.5) *when B is compact, γ is finite on compact sets and there is an $A \in \mathcal{A}_\mathfrak{X}$ with $\lambda(A) > 0$ and $\{\theta: p_\theta(A) \geq \varepsilon\}$ has compact closure for every $\varepsilon > 0$. Then a necessary condition for q to be a posterior for some prior π is that $\pi(K) = 0$ for every compact K.*

In order to prove Proposition 5.5, fix a prior π for q. Since $\pi(V) = \int q_x(V) m(dx)$, where m is the marginal of π on $\mathfrak{X}$, condition (5.5) clearly implies that π is absolutely continuous w.r.t. γ on B, for every B such that $\gamma(B) < +\infty$. But then π is σ-additive on all such B's, so that it is enough to prove the following lemma.

LEMMA 5.7. *Let l be strictly positive. If π is a prior for q which is σ-additive on B, for every $B \in \mathcal{A}_\Theta$ such that $\gamma(B) < +\infty$, then $\pi(B) = 0$ for all such B's.*

PROOF. As in the proof of Proposition 4.3, decompose π as $\pi = \pi_1 + \pi_2$, where π_1 is σ-additive and π_2 is *purely* finitely additive. Suppose that $\pi_1(\Theta) > 0$. Under this assumption, define $\widehat{\pi}(\cdot) = \pi_1(\cdot)/\pi_1(\Theta)$, $g(x) = \int l(x,\theta)\widehat{\pi}(d\theta)$, $G = \{x: 0 < g(x) < +\infty\}$ and $\widehat{q}_x(B) = [1/g(x)] \int_B l(x,\theta)\widehat{\pi}(d\theta)$ for B in $\mathcal{A}_\Theta$ and $x \in G$. By

Fubini's theorem, it is easily verified that $\widehat{\pi}$ is a prior for $\widehat{q}$. Now, let $B \in \mathcal{A}_\Theta$ with $\gamma(B) < +\infty$ and $A \in \mathcal{A}_\mathfrak{X}$ with $m_2(A) = 0$, where $m_2(F) = \int p_\theta(F)\pi_2(d\theta)$ for $F \in \mathcal{A}_\mathfrak{X}$. Since π is σ-additive on $\mathcal{B}$, $\pi_2(B) = 0$. After fixing A and B in such a way, define the σ-additive measures r_1 and r_2 on $\mathcal{A}$ by

$$r_1(C) = \int_A q_x(B \cap C^x)\widehat{m}(dx), \qquad r_2(C) = \int_A \widehat{q}_x(B \cap C^x)\widehat{m}(dx) \quad \forall\, C \in \mathcal{A},$$

where $\widehat{m}$ is the marginal of $\widehat{\pi}$ on $\mathfrak{X}$. Using $m_2(A) = \pi_2(B) = 0$, $\widehat{q}$ is a posterior for $\widehat{\pi}$ and q is a posterior for π, one obtains $r_1 = r_2$. In particular, since l is strictly positive,

$$\text{(5.6)} \quad \gamma(B)\int_A \frac{g(x)}{\rho(x)}\lambda(dx) = \int \frac{1}{l(x,\theta)} r_1(dx,d\theta) = \int \frac{1}{l(x,\theta)} r_2(dx,d\theta) = \widehat{\pi}(B)\lambda(A).$$

If $\lambda(A) \in (0,+\infty)$, (5.6) is absurd, since B can be chosen such that $\gamma(B)$ is arbitrarily large. Hence, if $\lambda(A) \in (0,+\infty)$, it must be that $\pi_1(\Theta) = 0$ and the proof is concluded. We show that, actually, there is an A with $m_2(A) = 0$ and $\lambda(A) \in (0,+\infty)$. Since λ is σ-finite, it is easily checked that there is an A with $\lambda(A) \in (0,+\infty)$ and $\int_A \rho\, d\lambda < +\infty$. Then $\gamma(\{\theta: p_\theta(A) \geq \varepsilon\}) < +\infty$ for every $\varepsilon > 0$, for otherwise $\int_A \rho(x)\lambda(dx) = \int p_\theta(A)\gamma(d\theta) = +\infty$. Thus $\pi_2(\{\theta: p_\theta(A) \geq \varepsilon\}) = 0$ for every $\varepsilon > 0$, so that $m_2(A) = 0$. □

As far as Proposition 5.6 is concerned, its proof essentially coincides with that of Proposition 5.5, by considering compact sets instead of sets B with $\gamma(B) < +\infty$.

EXAMPLE 5.8. Let $\mathfrak{X} = \Theta = \mathbb{R}$, let $\mathcal{A}_\mathfrak{X} = \mathcal{A}_\Theta$ be the Borel σ-field and let γ be the Lebesgue measure. Suppose that $l > 0$, $\lambda(D^c) = 0$ and, for $x \in D$, q_x is symmetrical and unimodal around its median m_x with $q_x((m_x - \varepsilon, m_x + \varepsilon)) \leq q_{x_0}((m_{x_0} - \varepsilon, m_{x_0} + \varepsilon))$ for all sufficiently small $\varepsilon > 0$ and some $x_0 \in D$. For instance, this is the case when $x_0 = 0$, λ is the Lebesgue measure and $l(x,\theta) = f(x-\theta)$, with $f > 0$ even and unimodal. Then (5.5) holds. In fact,

$$\begin{aligned} q_x(V) &\leq q_x\left(\left(m_x - \frac{\gamma(V)}{2}, m_x + \frac{\gamma(V)}{2}\right)\right) \\ &\leq q_{x_0}\left(\left(m_{x_0} - \frac{\gamma(V)}{2}, m_{x_0} + \frac{\gamma(V)}{2}\right)\right) \to 0 \quad \text{as } \gamma(V) \to 0. \end{aligned}$$

Thus, by Proposition 5.5, only probabilities satisfying (4.5) can be priors for q.

EXAMPLE 5.9. Let p be a univariate regular exponential family as defined in Diaconis and Ylvisaker (1979). In particular, Θ, $\mathfrak{X} \subset \mathbb{R}$, Θ is an open interval and

$$l(x,\theta) = \exp\bigl(x\theta - M(\theta)\bigr).$$

Assume further that $\lambda(\mathfrak{X}^0) > 0$, where $\mathfrak{X}^0$ is the interior of $\mathfrak{X}$. Let γ be such that $\lambda(D^c) = 0$ and γ is strictly positive on nonempty open sets. Proposition 5.6 applies to this example, so that the only candidates to be priors for q are probabilities vanishing on compact sets. Indeed, since $\lambda(\mathfrak{X}^0) > 0$, there is a compact $A \subset \mathfrak{X}^0$ with $\lambda(A) > 0$. Further, for fixed $x \in \mathfrak{X}^0$, $l(x,\theta) \to 0$ as θ approaches the boundary of Θ [Diaconis and Ylvisaker (1979), page 273]. Hence, $\{\theta: p_\theta(A) \geq \varepsilon\}$ is compact for every $\varepsilon > 0$. Next, let $V \subset [a,b] \subset \Theta$. Then, for $x \geq 0$,

$$\begin{aligned} q_x(V) &\leq \frac{\exp(xb)\int_V \exp\big(-M(\theta)\big)\gamma(d\theta)}{\int_b^{\sup\Theta} \exp\big(x\theta - M(\theta)\big)\gamma(d\theta)} \\ &\leq \frac{\int_V \exp\big(-M(\theta)\big)\gamma(d\theta)}{\int_b^{\sup\Theta} \exp\big(-M(\theta)\big)\gamma(d\theta)} \to 0 \quad \text{as } \gamma(V) \to 0, \end{aligned}$$

and likewise for $x < 0$. Hence, (5.5) holds when B is compact.

Let us now analyze the existence of priors of the form (4.6). To this purpose, fix a sequence $\{V_n\} \subset \mathcal{A}_\Theta$ such that $0 < \gamma(V_n) < +\infty$ and $V_n \uparrow \Theta$. For every n, define

$$\pi_n(B) = \frac{\gamma(B \cap V_n)}{\gamma(V_n)} \qquad \forall\, B \in \mathcal{A}_\Theta,$$

$$q_x^n(B) = \frac{\int_{B\cap V_n} l(x,\theta)\gamma(d\theta)}{\int_{V_n} l(x,\theta)\gamma(d\theta)} \quad \forall\, x \in D_n,\ \forall\, B \in \mathcal{A}_\Theta,$$

where $D_n = \{x\colon 0 < \int_{V_n} l(x,\theta)\gamma(d\theta) < +\infty\}$.

In view of Fubini's theorem, it is easily shown that q^n is a posterior for π_n. The latter fact, together with Theorem 3.3, suggests the following statement.

PROPOSITION 5.10. *There is a probability μ on $\mathcal{P}(\mathbb{N})$ such that q is a posterior for $\pi(\cdot) = \int \pi_n(\cdot)\mu(dn)$ if and only if*

$$\text{(5.7)} \quad \inf_n \frac{1}{\gamma(V_n)} \int \Big[q_x\big(V_n\phi^x\big) - q_x(V_n)q_x(\phi^x)\Big]\rho(x)\lambda(dx) \leq 0 \quad \forall\, \phi \in L(\mathcal{A}).$$

Moreover, if

$$\text{(5.8)} \qquad \lambda\Big(\big\{x \in D\colon 0 < q_x(V_n) < 1\big\}\Big) > 0 \quad \forall\, n \in \mathbb{N},$$

then μ is forced to be diffuse.

PROOF. After some algebra, (5.7) can be written as

$$\inf_n \int \big[q_x^n(\phi^x) - q_x(\phi^x)\big]m_n(dx) \leq 0 \quad \forall\, \phi \in L(\mathcal{A}),$$

where m_n is the marginal of π_n on $\mathfrak{X}$. Thus, since q^n is a posterior for π_n, the first part of the proposition follows from Theorem 3.3. [Notice that, in Theorem 3.3, the various probability laws are defined on $\mathcal{P}(\mathfrak{X})$ or $\mathcal{P}(\Theta)$, while here such laws are defined on $\mathcal{A}_{\mathfrak{X}}$ or $\mathcal{A}_\Theta$; however, because of the measurability assumptions holding in this section, this difference can be easily overcome.] To prove the second part, define

$$f_n(j) = \int \big[q_x^j(V_n) - q_x(V_n)\big] m_j(dx) \quad \forall\, j, n \in \mathbb{N}.$$

If q is a posterior for π, then $\int f_n(j)\mu(dj) = 0$. Moreover, $f_n(j) \geq 0$ and, under (5.8), $\gamma(V_n)f_n(n) = \int q_x(V_n)q_x(V_n^c)\rho(x)\lambda(dx) > 0$. Hence,

$$\mu\big(\{n\}\big) \leq \frac{1}{f_n(n)} \int f_n(j)\mu(dj) = 0. \qquad \square$$

Informally, condition (5.7) states that the integral of the posterior covariance between ϕ^x and $V_n/\gamma(V_n)$, w.r.t. the measure $\rho(x)\lambda(dx)$, cannot be uniformly positive.

As is easily proved, a sufficient condition for (5.7) is

$$\inf_n \int q_x\big(V_n^c\big) m_n(dx) = 0, \tag{5.9}$$

where m_n is the marginal of π_n on $\mathfrak{X}$. Condition (5.9) has been introduced by Heath and Sudderth [(1989), Theorem 3.2], who have shown that it suffices for HS-coherence of q. We do not know whether (5.9) is also necessary for (5.7), but we suspect it is not. Nevertheless, (5.9) is very useful in practice, since it avoids the checking of (5.7) for every particular ϕ.

A further remark concerns the possibility of taking μ to be diffuse. Suppose that (5.8) fails, that is, $\lambda(\{x \in D\colon 0 < q_x(V_n) < 1\}) = 0$ for some n. Then it is easily shown that

$$m_n\Big(\big\{x \in D \cap D_n\colon q_x(B) = q_x^n(B)\ \forall\, B \in \mathcal{A}_\Theta\big\}\Big) = 1, \tag{5.10}$$

and since q^n is a posterior for π_n, (5.10) implies that q is a posterior for π_n, too. In short, when q is a posterior for $\pi(\cdot) = \int \pi_n(\cdot)\mu(dn)$, either (5.8) holds and μ is bound to be diffuse or (5.8) fails and q is a posterior for a σ-additive prior (precisely for some truncation π_n of γ). This circumstance leads to our third goal in this section: the analysis of σ-additive priors. Let

$$E = \big\{(x,\theta)\colon l(x,\theta) > 0\big\}.$$

PROPOSITION 5.11. *Let π be a σ-additive probability on $\mathcal{A}_\Theta$, and let m be the marginal of π on $\mathfrak{X}$. Suppose $\mathcal{A}_\Theta$ is countably generated. Then q is a posterior for π if and only if there is a set $H \in \mathcal{A}_{\mathfrak{X}}$ such that $m(H) = 1$ and*

$$\frac{\pi(E^x \cap B)}{\pi(E^x)} = \frac{\gamma(E^x \cap B)}{\gamma(E^x)} \quad \forall\, B \in \mathcal{A}_\Theta,\ \forall\, x \in H. \tag{5.11}$$

PROOF. Let $g(x) = \int l(x,\theta)\pi(d\theta)$, $G = \{x\colon 0 < g(x) < +\infty\}$ and

$$q'_x(B) = \frac{1}{g(x)} \int_B l(x,\theta)\pi(d\theta) \quad \forall\, x \in G,\ \forall\, B \in \mathcal{A}_\Theta.$$

If (5.11) holds, then $q_x(B) = q'_x(B)\ \forall\, x \in H \cap G \cap D$, $\forall\, B \in \mathcal{A}_\Theta$. Since $m(H \cap G \cap D) = 1$ and since q' is a posterior for π, it follows that q is a posterior for π. Conversely, suppose that π is a prior for q. The proof can be divided into two parts: (i) q' is proved to coincide with q a.e. w.r.t. m; (ii) it is checked that (i) implies (5.11).

(i) Since q and q' are both posteriors for π, for every fixed B in $\mathcal{A}_\Theta$ one obtains

$$\int_A q_x(B)m(dx) = \int_A q'_x(B)m(dx) \quad \forall\, A \in \mathcal{A}_{\mathfrak{X}}.$$

Hence, there is an $H(B) \in \mathcal{A}_{\mathfrak{X}}$ such that $m(H(B)) = 1$ and $q_x(B) = q'_x(B)$ for $x \in H(B)$. Let $\{B_n\}$ be a countable field generating $\mathcal{A}_\Theta$, and let $H = G \cap D \cap (\cap_n H(B_n))$. Then, $m(H) = 1$ and

$$q_x(B_n) = q'_x(B_n) \qquad \forall\, x \in H,\ \forall\, n \in \mathbb{N}. \tag{5.12}$$

Since $H \subset D \cap G$, q_x and q'_x are σ-additive for every $x \in H$, and consequently (5.12) yields $q_x(B) = q'_x(B)\ \forall\, B \in \mathcal{A}_\Theta$, $\forall\, x \in H$.

(ii) For fixed $B \in \mathcal{A}_\Theta$, let $h(x,\theta) = B(\theta)/l(x,\theta)$ for $(x,\theta) \in E$, and $h(x,\theta) = 1$ otherwise. Then, for $x \in D \cap G$, $q'_x(h^x) = \pi(B \cap E^x)/g(x)$ and $q_x(h^x) = \gamma(B \cap E^x)/\rho(x)$. Consequently, (i) implies that $\pi(B \cap E^x)\rho(x) = \gamma(B \cap E^x)g(x)\ \forall\, x \in H$, and since $m(H) = 1$, (5.11) follows. □

The following corollary of Proposition 5.11, whose technical proof is omitted, gives conditions equivalent to the existence of a prior for q which is a truncation of γ.

COROLLARY 5.12. *Let $\mathcal{A}_\Theta$ be countably generated. Then q is a posterior for a prior of the form $\pi(\cdot) = \gamma(\cdot \cap V)/\gamma(V)$, for some $V \in \mathcal{A}_\Theta$ with $0 < \gamma(V) < +\infty$, if and only if there is an $H \in \mathcal{A}_{\mathfrak{X}}$ such that $0 < \gamma(\{\theta\colon p_\theta(H) = 1\}) < +\infty$ and $\gamma(\{\theta\colon 0 < p_\theta(H) < 1\}) = 0$.*

Some remarks concerning Proposition 5.11 are in order. First, a necessary condition for q to be the posterior for a σ-additive prior is

$$\lambda\Big(\big\{x\colon \gamma(E^x) < +\infty\big\}\Big) > 0.$$

In a sense γ, restricted to the relevant region $E^x = \{\theta\colon l(x,\theta) > 0\}$, does not act as an improper prior for every x in a set of positive λ-measure. Second, as noted in the proof, if q is a posterior for the σ-additive prior π, then q must coincide with q' a.e. w.r.t. m, where q' is the inference obtained by π using the standard Bayes theorem. Third, condition (5.11) means that π and γ induce the same conditional probability P on $\mathcal{A}_\Theta \times \mathfrak{U}$, where $\mathfrak{U} = \{E^x\colon x \in H\}$ and $P(B \mid E^x) = \gamma(B \cap E^x)/\gamma(E^x)$. In this connection note also that, for $x \in H$, one is really working with $P(\cdot \mid E^x)$, that is,

$$q_x(B) = \frac{\int_B l(x,\theta)P\big(d\theta \mid E^x\big)}{\int l(x,\theta)P\big(d\theta \mid E^x\big)} = q'_x(B).$$

We close the paper with a result concerning Proposition 5.10. Such a proposition refers to a fixed sequence $\{V_n\}$, $V_n \uparrow \Theta$, but nothing is said about the possibility of finding $\{V_n\}$ in such a way that the proposition at issue applies and $\pi(\cdot) = \int [\gamma(\cdot \cap V_n)/\gamma(V_n)]\mu(dn)$ is a prior for q. Plainly, when q is not HS-coherent, no sequence $\{V_n\}$ can work. The ensuing question is: provided q is HS-coherent, is there a sequence $\{V_n\}$, $V_n \uparrow \Theta$, such that π is a prior for q? The following example answers the above question negatively and, actually, shows a little more; indeed, an HS-coherent q is found which is neither the posterior of a prior satisfying (4.5) nor the posterior of a σ-additive prior. We also notice that the example substantially answers a conjecture of Heath and Sudderth [(1989), page 911].

EXAMPLE 5.13. Let $\mathfrak{X} = \Theta = (0, \infty)$, let $\mathcal{A}_{\mathfrak{X}} = \mathcal{A}_\Theta$ be the Borel σ-field and let $\gamma = \lambda$ be the Lebesgue measure. Setting $f(x) = x^{-1} + 1$ if $x \in (0,1)$ and $f(x) = x^{-x} + 1$ if $x \geq 1$, let E be the union of the sets C_1 and C_2 defined by

$$C_1 = \Big\{(x,\theta)\colon x \in (0,\infty),\ \theta \in \big(1, f(x)\big)\Big\}, \qquad C_2 = \bigcup_{n=1}^{\infty} [n, n+1) \times \big(2^{-n}, 2^{1-n}\big].$$

Moreover, let p_θ be uniform on E_θ, that is, $l(x,\theta) = E(x,\theta)/\lambda(E_\theta)$. We show that (i) q is HS-coherent, (ii) no π satisfying (4.5) is a prior for q and (iii) no σ-additive π is a prior for q. In particular, no π of the form $\pi(\cdot) = \int [\gamma(\cdot \cap V_n)/\gamma(V_n)]\mu(dn)$, with $V_n \uparrow \Theta$, is a prior for q. In fact, by Proposition 5.10 and the subsequent comments, if q has a prior of the previous form, then either π satisfies (4.5) (because μ is diffuse) or there is a σ-additive prior (precisely a truncation of γ).

(i) Let $\pi_n(B) = 2^{n-1}\gamma(B \cap (0, 2^{1-n}))$ and let

$$q_x^n(B) = \frac{\int_B l(x,\theta)\pi_n(d\theta)}{\int l(x,\theta)\pi_n(d\theta)} \quad \forall\, B \in \mathcal{A}_\Theta,\ \forall\, n \in \mathbb{N},\ \forall\, x \geq n.$$

It is easily shown that $m_n((0,n)) = 0$, where m_n is the marginal of π_n on $\mathfrak{X}$. Let μ be a diffuse probability on $\mathcal{P}(\mathbb{N})$. Since q^n is a posterior for π_n, it suffices to check condition (ii) of Theorem 3.3 for $\{q^n\}$, $\{m_n\}$, q and μ. Fix $\phi \in L(\mathcal{A})$. Since

$$q_x^n(\phi^x) = \frac{\int_0^1 \phi(x,\theta) l(x,\theta)\, d\theta}{\int_0^1 l(x,\theta)\, d\theta} \quad \text{for } x \geq n,$$

and since $q_x((1, f(x))) \to 0$, as $x \to +\infty$, then $[q_x^n(\phi^x) - q_x(\phi^x)] \to 0$ as $x \to +\infty$. Consequently, being $m_n([n,\infty)) = 1$ and μ diffuse, condition (ii) of Theorem 3.3 follows. Notice that, setting $V_n = (0, 2^{1-n})$, then $V_n \downarrow \emptyset$ (instead of $V_n \uparrow \Theta$). Hence, since μ is diffuse, q is a posterior for a prior π such that $\pi((0,\varepsilon)) = 1\ \forall\, \varepsilon > 0$; in particular, neither is π σ-additive nor does π satisfy (4.5).

(ii) Let $C = \{(x,\theta)\colon \theta > 2,\ 0 < x < 1/2(\theta - 1)\}$. Then $p_\theta(C_\theta) = \frac{1}{2}\ \forall\, \theta > 2$ while $q_x(C^x) \leq \frac{1}{4}\ \forall\, x$. Hence, no π satisfying (4.5) can work.

(iii) Let π be a σ-additive prior for q. By Proposition 5.11, there is an $H \in \mathcal{A}_{\mathfrak{X}}$, $m(H) = 1$, such that (5.11) holds. If $\pi((0,1)) < 1$, then $m((0,\varepsilon)) > 0\ \forall\, \varepsilon > 0$

and, consequently, $H \cap (0, \varepsilon) \neq \emptyset$. By σ-additivity of π, there is a $t > 1$ with $\pi([1,t)) > 0$. Since $[1,t) \subset E^x$ for x sufficiently small, (5.11) gives $\pi([1,t)) \leq \gamma([1,t))/\gamma(E^x) \to 0$ as $x \to 0$, which is a contradiction. Hence, $\pi((0,1)) = 1$. By σ-additivity of π, there is an ε with $\pi((\varepsilon, 1)) > 0$. Let n be such that $2^{-n} < \varepsilon \leq 2^{1-n}$ and $C = (1, n+1) \times (2^{-n}, 1)$. Then

$$\int p_\theta(C_\theta)\pi(d\theta) = m\big((1,n+1)\big) > \sup_x q_x(C^x) m\big((1,n+1)\big) \geq \int q_x(C^x) m(dx),$$

contrary to the assumption that π is a prior for q. Thus, there are no σ-additive priors for q.

Acknowledgments. This paper has been improved by suggestions of the referees and an Associate Editor.

REFERENCES

BERTI, P., REGAZZINI, E. and RIGO, P. (1991). Coherent statistical inference and Bayes theorem. *Ann. Statist.* **19** 366–381.

BHASKARA RAO, K. P. S. and BHASKARA RAO, M. (1983). *Theory of Charges.* Academic, London.

BRUNK, H. (1991). Fully coherent inference. *Ann. Statist.* **19** 830–849.

DIACONIS, P. and YLVISAKER, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7** 269–281.

DUNFORD, N. and SCHWARTZ, J. (1958). *Linear Operators, Part 1: General Theory.* Interscience, New York.

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

HEATH, D. and SUDDERTH, W. (1989). Coherent inference from improper priors and from finitely additive priors. *Ann. Statist.* **17** 907–919.

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

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

DIPARTIMENTO STATISTICO
UNIVERSITÀ DEGLI STUDI DI FIRENZE
VIALE MORGAGNI 59
50134 FIRENZE
ITALY