The Annals of Statistics
1991, Vol. 19, No. 1, 366–381

COHERENT STATISTICAL INFERENCE AND BAYES THEOREM

BY PATRIZIA BERTI,[1] EUGENIO REGAZZINI[2] AND PIETRO RIGO

Università degli Studi di Firenze and Università "L. Bocconi"

Conditions are given which suffice for the assessment of a coherent inference by means of a Bayesian algorithm, i.e., a suitable extension of the classical Bayes theorem relative to a finite number of alternatives. Under some further hypotheses such inference is shown to be, in addition, coherent in the sense of Heath, Lane and Sudderth. Moreover, a characterization of coherent posteriors is provided, together with some remarks concerning finitely additive conditional probabilities.

0. Introduction. A few recent articles by Heath and Sudderth (1978), Lane and Sudderth (1983) and Regazzini (1987) aim to define the concept of coherent statistical inference. In particular, Regazzini's paper introduces this concept in conformity with de Finetti's theory of conditional previsions, and compares it with Heath, Lane and Sudderth's definition. Actually, all these approaches suit the Bayesian theory of statistical inference. However, unlike the classical treatment of Bayesian methods which resort to σ-additive probabilities only, they, more generally, prescribe the use of finitely additive probabilities. Hence, if the parameter space Θ is finite and the probability of a given observation x is strictly positive, then there exists a unique coherent posterior, which is the one determined through the classical Bayes theorem. In line with standard practice, given an arbitrary Θ, we say that a posterior q_x is assigned by the Bayes theorem, if there exists a nonnegative function l on $\mathscr{X} \times \Theta$ ($\mathscr{X}$= set of all the logically possible observations), such that

$$q_x(B) = \left\{\int_\Theta l(x,\theta)\tau(d\theta)\right\}^{-1} \int_B l(x,\theta)\tau(d\theta), \qquad B \subset \Theta,$$

τ being a prior assigned on Θ. Even if a coherent inference does exist (in fact, a coherent inference, as de Finetti meant it, always exists), it is well known that it need not be assessed through the Bayes theorem. There are even cases in which the Bayes theorem does not hold.

In the light of the above remarks, the main purpose of the present paper is to single out hypotheses from which one can deduce a Bayes theorem yielding

Received April 1989; revised November 1989.

[1]Research partially supported by MPI (40% 1985, Gruppo Nazionale "Processi Stocastici e Calcolo Stocastico").

[2]Research partially supported by MPI (40% 1987, Gruppo Nazionale "Modelli Probabilistici") and by CNR (Progetto Strategico "Statistica dei Processi Aleatori").

AMS 1980 *subject classifications*. Primary 62A15; secondary 60A05.

Key words and phrases. Bayes theorem, coherent conditional probability, conglomerability, coherent posterior, disintegrability, integration with respect to a finitely additive set function.

coherent statistical inferences. Theorem 3.1 is a Bayes theorem which produces coherent inferences in the sense meant by de Finetti. Theorem 3.3 considers additional hypotheses in order to ensure that inferences, obtained by means of the Bayes theorem, are coherent in the sense meant by Heath, Lane and Sudderth. The proof of these statements is based on Theorem 2.2, which provides necessary and sufficient conditions so that a posterior turns out to be coherent in line with de Finetti's approach. From an operative point of view, this proposition provides a criterion which is more useful than the one deducible from Theorem 2.3 of Regazzini (1987). Theorem 2.2 is a direct consequence of a more general proposition due to Rigo (1988), characterizing coherent conditional probabilities in a very general context. Such a characterization is mentioned in Section 1, together with further remarks which provide an up-to-date, concise survey of a few recent studies in de Finetti's theory of coherence.

The analysis developed in Section 3 is useful in a milieu of completely additive laws also. In fact, it allows us to decide whether an inference assessed according to the σ-additive generalizations of the Bayes theorem [see Kallianpur and Striebel (1968)] are coherent in compliance with our Definition 2.1; see Example 3.5 and Section 4.

Finally, we recall that the present paper originates from a drastic revision of an unpublished technical report by the second author [cf. Regazzini (1984)]. We have undertaken this revision in the hope of clarifying connections between coherent statistical inferences and posteriors obtained by means of the Bayes theorem.

1. Preliminaries. The concept recurring more frequently in the present paper is that of probability on a class $\mathscr{K}$ of conditional events. In order to define it, we will use de Finetti's betting scheme as in Regazzini (1985, 1987). According to such a scheme, a person who wants to summarize his degree of belief in each element $E|H$ of $\mathscr{K}$ with a real number, is supposed to be obliged to accept any bet on $E|H$, on the basis of unit prices that he has fixed, and with stakes arbitrarily chosen by an opponent. If these prices avoid sure losses, then they determine a probability on $\mathscr{K}$. More precisely, indicating prices with $P: \mathscr{K} \to \mathbb{R}$, suppose an opponent resolves to put stakes $s_1, \dots, s_n$ on the events $E_1|H_1, \dots, E_n|H_n$ belonging to $\mathscr{K}$, with the proviso that the bet on $E_k|H_k$, $k = 1, \dots, n$, is called off whenever H_k does not come true. Consequently, adopting the useful suggestion of de Finetti that the same symbol that designates an event also designates the indicator of that event, the random quantity

$$(1.1) \qquad G(E_k|H_k, s_k; k = 1, \dots, n) := \sum_{i=1}^{n} s_i\{P(E_i|H_i) - E_i\}H_i$$

represents the gain from a combination of bets, on the events taken into consideration, with stakes $s_1, \dots, s_n$. Roughly speaking, P is said to be a probability on $\mathscr{K}$ if, provided that at least one H_k, $k = 1, \dots, n$, comes true, it avoids sure losses, and this holds for all choices of n, $(s_1, \dots, s_n)$ and $(E_1|H_1, \dots, E_n|H_n)$. After observing that an unconditional event E can be regarded as a conditional one: $E = E|\Omega$, Ω denoting the sure event, we make precise the notion of probability on $\mathscr{K}$.

DEFINITION 1.1. Let $\mathscr{K}$ be a class of conditional events and P a real-valued function on $\mathscr{K}$. Then P is said to be a *probability* on $\mathscr{K}$ if and only if, for every choice of n, $(E_1|H_1, \dots, E_n|H_n)$ in $\mathscr{K}$ and $(s_1, \dots, s_n)$ in $\mathbb{R}^n$, the gain G defined by (1.1) fulfills the inequalities

$$\inf(G|H_0) \le 0 \le \sup(G|H_0),$$

where $H_0 = \bigcup_{i=1}^{n} H_i$, inf and sup are taken with respect to all the elementary events relative to $\{E_i, H_i\colon i = 1, \dots, n\}$ which imply H_0.

It is well known that, if $\mathscr{K} = \{E|H\colon H \in \mathscr{H}^0, E \in \mathscr{E}, \mathscr{H} \subset \mathscr{E}, \mathscr{H}^0 = \mathscr{H} \setminus \{\varnothing\}, \mathscr{H}$ and $\mathscr{E}$ are algebras of events$\}$, then $P\colon \mathscr{K} \to \mathbb{R}$ is a probability if and only if

(a) $P(\cdot|H)$ is a nonnegative, additive function on $\mathscr{E}$, for every H in $\mathscr{H}^0$;
(b) $P(E|H) = 1$ whenever $E \in \mathscr{E}$, $H \in \mathscr{H}^0$ and $H \subset E$;
(c) $P(E \cap H_1|H_2) = P(E|H_1 \cap H_2)P(H_1|H_2)$ whenever $H_2, H_1 \cap H_2$ belong to $\mathscr{H}^0$ and E belongs to $\mathscr{E}$.

When $\mathscr{H}^0 = \{\Omega\}$, P is a probability on $\mathscr{K}$ if and only if it is a *probability charge* on $\mathscr{E}$, i.e., P is a nonnegative, additive function on $\mathscr{E}$ with $P(\Omega) = 1$. We take the opportunity to recall that the term *charge*, in this paper, designates any additive set function with values in $[0, \infty]$, and which vanishes at $\varnothing$.

Generally, if $\mathscr{K}$ does not possess the above structure, conditions (a)–(c) are necessary for a real-valued function to be a probability on $\mathscr{K}$, but they may not be sufficient. The following proposition, which will be frequently used later on, provides a useful necessary and sufficient condition in order that P be a probability on

(1.2) $\mathscr{K} := \{E|H\colon E \in \mathscr{E}, H \in \mathscr{C}, \mathscr{E}$ is an algebra, $\mathscr{C} \subset \mathscr{E}$ and $\varnothing \notin \mathscr{C}\}$.

It has been introduced into de Finetti's theory by Rigo (1988) according to an analogous condition discovered by Császár (1955) in order to represent a Rényi conditional probability space by a suitable family of measures. In any case, such a proposition can be proved by elementary arguments based on Definition 1.1.

THEOREM 1.2. *Let $\mathscr{K}$ be assigned in conformity with* (1.2). *Then P is a probability on $\mathscr{K}$ if and only if*

(α) *for each H in $\mathscr{C}$, $E \to P(E|H)$ is a probability on $\mathscr{E}$;*
(β) *$H \in \mathscr{C}$, $E \in \mathscr{E}$ and $H \subset E \Rightarrow P(E|H) = 1$;*
(γ) *$E_i \in \mathscr{E}$, $H_i \in \mathscr{C}$, $i = 1, \dots, n$, and*

$$H_{n+1} := H_1 \Rightarrow \prod_{i=1}^{n} P(E_i \cap H_{i+1}|H_i) = \prod_{i=1}^{n} P(E_i \cap H_i|H_{i+1}).$$

In a number of statistical problems, $\mathscr{C}$ is the union of the sure event with two of its partitions: the first one generated by the values of an observable sample, the second one generated by the values of a parameter. We are showing that, under such a circumstance, condition (γ) of Theorem 1.2 can be put in a different and, from a certain point of view, more convenient form.

COROLLARY 1.3. *Let Π_1, Π_2 be partitions of Ω included in the algebra $\mathscr{E}$ and*

$$\mathscr{C} = \Pi_1 \cup \{\Omega\} \cup \Pi_2.$$

Then P is a probability on $\mathscr{K}$ if and only if, besides conditions (α) and (β) of Theorem 1.2, it satisfies:

$(\gamma 1)$ $P(E \cap H) = P(E|H)P(H)$ whenever $E \in \mathscr{E}$ and $H \in \mathscr{C}$;
$(\gamma 2)$ condition (γ) of Theorem 1.2 holds whenever $E_i \in \mathscr{E}$, $H_i \in \Pi_r$, $H_{i+1} \in \Pi_s$ and $P(H_i) = 0$, $i = 1, \dots, n$; $r, s = 1, 2$; $r \neq s$.

PROOF. Under conditions (α) and (β), $(\gamma 1)$ and $(\gamma 2)$ are particular cases of (γ). Thus, it is merely to be proved that, provided $\mathscr{C} = \Pi_1 \cup \{\Omega\} \cup \Pi_2$, (γ) follows from $(\gamma 1)$ and $(\gamma 2)$. Let $E_1, \dots, E_n \in \mathscr{E}$, $H_1, \dots, H_n \in \mathscr{C}$ and $H_{n+1} := H_1$. If $P(H_i) > 0$ for all i's, then (γ) is a direct consequence of $(\gamma 1)$. This happens also when $\{H_1, \dots, H_n\}$ includes at least one event with positive probability and at least one event with zero probability. In such a case, indeed, since $H_{n+1} = H_1$, there are k, j with $k \neq j$ such that $P(H_k) = 0$, $P(H_{k+1}) > 0$, $P(H_j) > 0$, $P(H_{j+1}) = 0$ and, consequently,

$$P(E_k \cap H_k | H_{k+1}) = 0, \qquad P\big(E_j \cap H_{j+1} | H_j\big) = 0.$$

Condition (γ) trivially holds also when $E_i \cap H_i \cap H_{i+1} = \varnothing$ for some i's. Hence, it is sufficient to check (γ) only when $P(H_i) = 0$ for all i's and H_i, H_{i+1} belong to different partitions. □

The remaining part of this section is devoted to some remarks concerning integration with respect to a finitely additive set function, in conformity with the exposition of Dunford and Schwartz theory included in Chapter 4 of Bhaskara Rao and Bhaskara Rao (1983). Let $\mathscr{E}$ be an algebra of subsets of Ω and λ a charge on $\mathscr{E}$. After putting

$$\lambda^*(A) = \inf\{\lambda(B) \colon B \supset A,\, B \in \mathscr{E}\}$$

for each $A \subset \Omega$, we say that the sequence of real-valued functions $\{f_n\}$, defined on Ω, converges *hazily* to $f \colon \Omega \to [-\infty, \infty]$ if, for every $\varepsilon > 0$,

$$\lim_{n \to \infty} \lambda^*(\{\omega \in \Omega \colon |f_n(\omega) - f(\omega)| > \varepsilon\}) = 0.$$

Given a partition $\{A_1, \dots, A_n\}$ of Ω in $\mathscr{E}$ and a point $(a_1, \dots, a_n)$ of $\mathbb{R}^n$, $g = \sum_{k=1}^n a_k I_{A_k}$ is said to be a *simple function*. Such a function is said to be λ-integrable if $\lambda(A_k) < \infty$ whenever $a_k \neq 0$ and the integral of g, denoted by

$\int_\Omega g\,d\lambda$, is defined to be the real number $\sum_{k=1}^{n} a_k \lambda(A_k)$. (We adopt the convention that $0 \cdot \infty = 0$.) The generalization of this concept is carried out according to Definition 1.4.

DEFINITION 1.4. The function $f\colon \Omega \to [-\infty, \infty]$ is said to be *λ-integrable* if there exists a sequence $\{f_n\}$ of λ-integrable simple functions converging to f hazily and such that $\lim_{m,n\to\infty} \int_\Omega |f_n - f_m|\, d\lambda = 0$. If f is λ-integrable, the *integral* of f with respect to λ is defined to be the real number $\int_\Omega f d\lambda := \lim_{n\to\infty} \int_\Omega f_n\, d\lambda$. Notice that such a limit exists by virtue of the definition of λ-integrability and that it does not depend on the choice of $\{f_n\}$.

In this paper, a function which can be obtained as the hazy limit of a sequence of simple functions is said to be *λ-measurable*. [Bhaskara Rao and Bhaskara Rao (1983), page 101, designate this type of function with the term of T_1-measurable.] If f is nonnegative, λ-measurable but not λ-integrable then, by definition, $\int_\Omega f d\lambda := \infty$. Moreover, when f is bounded and λ is a probability, λ-measurability and λ-integrability are equivalent conditions. If, for each $\varepsilon > 0$ there exists $k = k(\varepsilon)$ such that $\lambda^*(\{\omega \in \Omega\colon |f(\omega)| > k\}) < \varepsilon$, then f is said to be *λ-smooth*; bounded functions, λ-measurable functions are examples of λ-smooth functions. The following proposition will be used in Section 3.

PROPOSITION 1.5. *Let λ be a charge on $\mathscr{E}$ and h, g λ-measurable functions from Ω to $[0, \infty]$. Set $f(\omega) = h(\omega)/g(\omega)$ if $0 < g(\omega) < \infty$, otherwise, if $g(\omega) = 0$ or $g(\omega) = \infty$, define $f(\omega)$ to be any real number. Then λ-smoothness of $1/g$: $\Omega \to [0, \infty]$ yields λ-measurability of f. (For the definition of $1/g$ we adopt the conventions: $1/0 = \infty$, $1/\infty = 0$.)*

PROOF. Since h, g and $1/g$ are λ-smooth, $\lambda^*(\{\omega\colon g(\omega) = 0 \text{ or } g(\omega) = \infty \text{ or } h(\omega) = \infty\}) = 0$. Thus, it can be assumed that $0 < g < \infty$ and $h < \infty$. If $1/g$ is λ-measurable, then $f = h/g$ is λ-measurable, being a product of λ-measurable functions [cf. Bhaskara Rao and Bhaskara Rao (1983), Corollary 4.4.9]. Hence, it suffices to show that $1/g$ is λ-measurable. Actually we are proving the following condition, equivalent to λ-measurability [cf. Bhaskara Rao and Bhaskara Rao (1983), Theorem 4.4.7], and called *T_2-measurability with respect to λ*: For each $\varepsilon > 0$, there exists a partition $\{F_0, F_1, \dots, F_n\}$ of Ω in $\mathscr{E}$ such that $\lambda(F_0) < \varepsilon$ and $|1/g(\omega_1) - 1/g(\omega_2)| < \varepsilon$ whenever $\omega_1, \omega_2 \in F_i$, $i = 1, \dots, n$.

Fix $\varepsilon > 0$. By λ-smoothness of $1/g$, there is $k \in (0, 1/\sqrt{3}]$ such that $\lambda^*(\{\omega\colon g(\omega) < k\}) < \varepsilon/3$. Then, by definition of λ^*, a set $G_0 \in \mathscr{E}$ can be found such that $G_0 \supset \{\omega\colon g(\omega) < k\}$ and $\lambda(G_0) < 2\varepsilon/3$; further

$$\left|\frac{1}{g(\omega_1)} - \frac{1}{g(\omega_2)}\right| = \frac{|g(\omega_1) - g(\omega_2)|}{g(\omega_1)g(\omega_2)}$$

$$\leq \frac{|g(\omega_1) - g(\omega_2)|}{k^2} \quad \text{for every } \omega_1, \omega_2 \notin G_0.$$

Since g is λ-measurable, or equivalently T_2-measurable with respect to λ, there is a partition $\{H_0, H_1, \dots, H_n\}$ of Ω in $\mathscr{E}$ such that $\lambda(H_0) < \varepsilon k^2$ and $|g(\omega_1) - g(\omega_2)| < \varepsilon k^2$ whenever $\omega_1, \omega_2 \in H_i$, $i = 1, \dots, n$. Setting $F_0 = G_0 \cup H_0$ and $F_i = H_i \setminus G_0$, $i = 1, \dots, n$, the desired partition is obtained. □

The last result of the present section is related to Theorem 1 of Dubins (1975). It deals with a partition of Ω, Π, included in $\mathscr{E}$, and with any algebra $\mathscr{E}_\Pi \subset \mathscr{E}$, whose elements are unions of elements of Π.

THEOREM 1.6. *Let P be a probability on $\mathscr{E}$ and σ a real-valued function on $\mathscr{E} \times \Pi$ such that, for each $h \in \Pi$, $E \to \sigma(E, h)$ is a probability on $\mathscr{E}$ with $\sigma(h, h) = 1$. Denote by γ the restriction of P to $\mathscr{E}_\Pi$. Then*

$$P(E) = \int_\Pi \sigma(E, h)\gamma(dh)$$

holds for all $E \in \mathscr{E}$ if and only if $h \to \sigma(E, h)$ is a γ-measurable function for each $E \in \mathscr{E}$ and

$$\gamma(S) \inf_{h \in S} \sigma(E, h) \le P(E \cap S) \le \gamma(S) \sup_{h \in S} \sigma(E, h) \tag{1.3}$$

for all $S \in \mathscr{E}_\Pi$ and $E \in \mathscr{E}$.

PROOF. In view of triviality of necessity, assume that $\sigma(E, \cdot\,)$ is γ-measurable for each $E \in \mathscr{E}$ and that (1.3) holds. Fix $E \in \mathscr{E}$. Then, if $\{S_1, \dots, S_n\}$ is a finite partition of Ω in $\mathscr{E}_\Pi$, we have

$$\sum_{i=1}^n \gamma(S_i) \inf_{h \in S_i} \sigma(E, h) \le \sum_{i=1}^n P(E \cap S_i) \le \sum_{i=1}^n \gamma(S_i) \sup_{h \in S_i} \sigma(E, h).$$

Thus, the result follows from the obvious equality $P(E) = \sum_{i=1}^n P(E \cap S_i)$, and from the definition of the Stieltjes integral, which, here, coincides with the Dunford–Schwartz integral [cf. Bhaskara Rao and Bhaskara Rao (1983), Section 4.5]. □

2. Coherent inferences. Let $\mathscr{X}$ be the set of all possible outcomes of an experiment and Θ the set of all realizations of a random parameter. $\Omega \subset \mathscr{X} \times \Theta$ designates the collection of the logically possible couples (x, θ); consequently, Ω is to be regarded as the sure event. Moreover, $\mathscr{A}_\mathscr{X}$, $\mathscr{A}_\Theta$ and $\mathscr{A}$ denote algebras of subsets of $\mathscr{X}$, Θ and Ω, respectively. In what follows, it is assumed that $\{x\} \in \mathscr{A}_\mathscr{X}$ for each $x \in \mathscr{X}$, $\{\theta\} \in \mathscr{A}_\Theta$ for each $\theta \in \Theta$, $(A \times B) \cap \Omega \in \mathscr{A}$ for each $A \in \mathscr{A}_\mathscr{X}$ and $B \in \mathscr{A}_\Theta$, and $x \in \mathscr{X}$, $\theta \in \Theta$, $C \in \mathscr{A} \Rightarrow C^x := \{\theta : (x, \theta) \in C\} \in \mathscr{A}_\Theta$, $C_\theta := \{x : (x, \theta) \in C\} \in \mathscr{A}_\mathscr{X}$. It is easy to check that $\Omega^x \ne \varnothing$ and $\Omega_\theta \ne \varnothing$ for $x \in \mathscr{X}$ and $\theta \in \Theta$.

Note: This framework is a revised version of the one adopted in Section 2 by Regazzini (1987); the present modification is required in order that the possible hypothesis that Ω is properly included in $\mathscr{X} \times \Theta$ should make sense. We take the opportunity to point out that condition (c2) at page 851 of the quoted paper is to be restated as follows: *the function* $\theta \to P_\theta(C_\theta)$ *is* τ*-integrable for all* $C \in \mathscr{A}$.

Suppose a sampling model and a prior distribution are assigned. This means that a mapping P_θ, from Θ to the set of all probabilities on $\mathscr{A}_{\mathscr{X}}$ with $P_\theta(\Omega_\theta) = 1$, and a probability τ on $\mathscr{A}_\Theta$ are assigned. It is well known that, under these circumstances, P_θ and τ determine a probability on the events $A|\{\theta\}$, B, where $A \in \mathscr{A}_{\mathscr{X}}$, $B \in \mathscr{A}_\Theta$ and $\theta \in \Theta$; see Theorem 2.2 in Regazzini (1987). Hence, at least one coherent extension of the latter probability to $\mathscr{D} := \{C|H: C \in \mathscr{A}, H = \Omega \text{ or } H = \Omega_\theta \times \{\theta\}, \theta \in \Theta\}$ can be defined; see Theorem 1.4 of Regazzini (1987). Specifically, this means that a probability on $\mathscr{A}$ is available, say π, with $\pi((\mathscr{X} \times B) \cap \Omega) = \tau(B)$ for $B \in \mathscr{A}_\Theta$, and such that

$$(2.1) \qquad Q(C|H) := \begin{cases} P_\theta(C_\theta), & \text{if } C \in \mathscr{A} \text{ and } H = \Omega_\theta \times \{\theta\}, \theta \in \Theta, \\ \pi(C), & \text{if } C \in \mathscr{A} \text{ and } H = \Omega, \end{cases}$$

is a probability on $\mathscr{D}$. In its turn, Q can be coherently extended to $\mathscr{K} = \mathscr{D} \cup \mathscr{F}$, where $\mathscr{F} = \{C|\{x\} \times \Omega^x: x \in \mathscr{X}, C \in \mathscr{A}\}$; let us denote such an extension with Q'.

In a Bayesian approach, any inferential problem is solved by considering the restriction of Q' to $\mathscr{F}$: the posterior distribution on Θ, relative to the assignment of Q. This leads us to Definition 2.1.

DEFINITION 2.1. Let Q be a probability on $\mathscr{D}$ according to (2.1). Then, the restriction to $\mathscr{F}$ of any coherent extension Q' of Q from $\mathscr{D}$ to $\mathscr{K}$ is said to be a *coherent posterior relative to* Q. From now on $\{q_x: x \in \mathscr{X}\}$ will designate any coherent posterior on $\mathscr{A}_\Theta$.

We notice that Theorem 2.3 of Regazzini (1987) states necessary and sufficient conditions in order that a probability on $\mathscr{A}_\Theta$ could be considered as a coherent posterior relative to Q. A more convenient criterion can be deduced from Corollary 1.3. In fact, by identifying $\mathscr{E}$ with $\mathscr{A}$, P with Q', Π_1 with $\{\{x\} \times \Omega^x: x \in \mathscr{X}\}$ and Π_2 with $\{\{\theta\} \times \Omega_\theta: \theta \in \Theta\}$, Corollary 1.3 directly implies the validity of Theorem 2.2.

THEOREM 2.2. *Let* Q *be a probability on* $\mathscr{D}$ *according to* (2.1). *Then* $\{q_x: x \in \mathscr{X}\}$ *is a coherent posterior relative to* Q *if and only if*

(i) *for each* $x \in \mathscr{X}$, q_x *is a probability on* $\mathscr{A}_\Theta$ *such that* $q_x(\Omega^x) = 1$;
(ii) $q_x(B)\pi(\{x\} \times \Omega^x) = \pi((\{x\} \times B) \cap \Omega)$ *for all* x *in* $\mathscr{X}$ *and* B *in* $\mathscr{A}_\Theta$;

$$\text{(iii)} \qquad \begin{aligned} \prod_{i=1}^{m} P_{\theta_i}(\{x_{i+1}\}) q_{x_i}(\{\theta_i\}) &= \prod_{i=1}^{m} P_{\theta_i}(\{x_i\}) q_{x_{i+1}}(\{\theta_i\}), \\ \prod_{i=1}^{m} P_{\theta_i}(\{x_i\}) q_{x_i}(\{\theta_{i+1}\}) &= \prod_{i=1}^{m} P_{\theta_{i+1}}(\{x_i\}) q_{x_i}(\{\theta_i\}), \end{aligned}$$

whenever $x_1, \dots, x_m \in \mathscr{X}$, $\theta_1, \dots, \theta_m \in \Theta$, $x_{m+1} := x_1$, $\theta_{m+1} := \theta_1$, *and* $\pi(\{x_i\} \times \Omega^{x_i}) = \pi(\Omega_{\theta_i} \times \{\theta_i\}) = 0$ *for* $i = 1, \dots, m$, $m \geq 2$.

In order to exclude sure losses, the inferrer has to select his posterior according to Theorem 2.2. Generally, however, there is not only one posterior fulfilling conditions (i)–(iii), but there exist several coherent assignments. For instance, if $P_\theta(\{x\}) = 0$ for all x and θ, and if $\pi(\{x\} \times \Omega^x) = 0$ for all x, then every collection $\{q_x: x \in \mathscr{X}\}$ satisfying requisite (i) is a coherent posterior relative to Q. The "large" set of coherent solutions, which is often available, guarantees the possibility of fitting real situations. On the other hand, just because of this "large" set, one could object that coherence is a weak condition. Perhaps this is technically true, but some brief remarks are in order. The main purpose of coherence is not to identify a "small" set of admissible posteriors so as to avoid the inferrer's troubles; on the contrary, a coherence condition must possess a clear substantial meaning and must be essential for *any* inferential problem. If one wishes to restrict the class of allowable inferences, introducing some additional constraint, the latter must be justified by exhibiting its logical content for any inferential situation. We are afraid that each further constraint, even if sensible for some particular class of problems, runs the risk of being arbitrary in general. See also Regazzini (1987), Section 5.

In most statistical applications, $\{q_x: x \in \mathscr{X}\}$ is assessed through a procedure inspired by the classical Bayes theorem. From the point of view adopted in this paper, however, such a posterior is merely a candidate, and can be assigned only if it is coherent relative to Q. Therefore, it is important to establish when it is actually so. The main purpose of the present paper is the analysis of suitable conditions under which a coherent posterior can be assessed by means of a Bayesian algorithm. This analysis is needed not only because of the generality of the context described in the present section, which does not prescribe σ-additivity; as a matter of fact, if σ-additivity holds, our study is useful to check whether inferences, determined by the usual, well-known versions of the Bayes theorem, are coherent in the sense of Definition 2.1.

3. A Bayes theorem. The first part of the present section answers the question put at the end of the previous one. The second part analyzes conditions under which a posterior assessed by the Bayes theorem is apt to represent an inference, coherent in the sense of Heath, Lane and Sudderth.

There exists at least one probability α on the power set of Θ such that $\alpha(B) = \tau(B)$ for every $B \in \mathscr{A}_\Theta$. Hence, we can define the extension of P_θ and τ to $\mathscr{D}$, according to the rule

$$\pi(C) = \int_\Theta P_\theta(C_\theta)\alpha(d\theta), \qquad C \in \mathscr{A}. \tag{3.1}$$

In fact, it is immediate to check coherence of (2.1) when π is defined according to (3.1). Note also that, if π is assessed through (3.1) and $P_\theta(\{x\}) = 0$ for some fixed x and all θ, then Theorem 2.2 implies that every probability q_x on $\mathscr{A}_\Theta$

such that $q_x(\Omega^x) = 1$ is coherent, independently of the selection of q_y for $y \neq x$. We are assuming the following hypotheses:

(β1) *there are a nonnegative function* l *on* $\mathscr{X} \times \Theta$, *null on* $\mathscr{X} \times \Theta \setminus \Omega$, *and a charge* λ *on* $\mathscr{A}_{\mathscr{X}}$ *such that*

$$P_\theta(A) = \int_A l(x,\theta)\lambda(dx) \quad \text{for all } \theta \in \Theta \text{ and } A \in \mathscr{A}_{\mathscr{X}};$$

(β2) *for each* $x \in \mathscr{X}$, $\theta \to l(x,\theta)$ *is* τ*-measurable.*

Set $\rho(x) = \int_\Theta l(x,\theta)\tau(d\theta)$ for every x in $\mathscr{X}$. We aim at showing that, under conditions (β1) and (β2), there is a coherent posterior $\{q'_x \colon x \in \mathscr{X}\}$, relative to Q with π defined by (3.1), determined through a Bayesian algorithm. To this end, we need to assume that $0 < \rho(x) < \infty$ for at least one $x \in \mathscr{X}$. Otherwise, if $\rho(x) = 0$ or $\rho(x) = \infty$ for every $x \in \mathscr{X}$, our developments, though formally correct, are practically vacuous, since the problem we are dealing with merely does not arise. In this connection, it should also be stressed that our assessment of q'_x when $\rho(x) \notin (0,\infty)$ is only a possible coherent one. For instance, when $\rho(x) = \infty$, every probability q'_x on $\mathscr{A}_\Theta$ such that $q'_x(\Omega^x) = 1$ turns out to be coherent (cf. the proof of Theorem 3.1). We refer to Section 4 for some remarks about the choice of q'_x when $\rho(x) \notin (0,\infty)$.

Let us define q'_x according to the following rule, in which $x \in \mathscr{X}$ and B is any element of $\mathscr{A}_\Theta$:

$$0 < \rho(x) < \infty \Rightarrow q'_x(B) = \int_B l(x,\theta)\tau(d\theta)/\rho(x),$$

$\rho(x) = 0$ and Ω^x is infinite $\Rightarrow q'_x$ is any probability on $\mathscr{A}_\Theta$ such that $q'_x(\{\theta\}) = 0$ for every θ in Ω^x, and $q'_x(\Omega^x) = 1$,

$\rho(x) = 0$, Ω^x is finite and

$$\sum_{\theta \in \Omega^x} l(x,\theta) > 0 \Rightarrow q'_x(B) = \sum_{\theta \in B \cap \Omega^x} l(x,\theta) \Big/ \sum_{\theta \in \Omega^x} l(x,\theta),$$

$\rho(x) = 0$, Ω^x is finite and $\sum_{\theta \in \Omega^x} l(x,\theta) = 0 \Rightarrow q'_x(B) = \text{card}(B \cap \Omega^x)/\text{card}(\Omega^x)$,

$\rho(x) = \infty \Rightarrow q'_x(B) = \tau(B \cap \Omega^x)/\tau(\Omega^x)$.

THEOREM 3.1. *Let* (β1) *and* (β2) *hold, and* Q *be defined on* $\mathscr{D}$ *according to* (2.1) *with* π *assessed through* (3.1). *Then* q' *is a coherent posterior relative to such a* Q.

PROOF. We show that q' satisfies conditions (i)–(iii) of Theorem 2.2. First, it is immediate to check (i). In order to verify (ii), notice that

$$\begin{aligned}\pi((\{x\} \times B) \cap \Omega) &= \int_B P_\theta(\{x\})\alpha(d\theta)\\ &= \int_B [l(x,\theta)\lambda(\{x\})]\alpha(d\theta) = \int_B [l(x,\theta)\lambda(\{x\})]\tau(d\theta)\end{aligned}$$

for every B in $\mathscr{A}_\Theta$. If $\rho(x) < \infty$, setting $B = \Theta$ yields $\pi(\{x\} \times \Omega^x) = \lambda(\{x\})\rho(x)$; hence, being $0 \leq \int_B l(x, \theta)\tau(d\theta) \leq \rho(x)$, condition (ii) easily follows. If $\rho(x) = \infty$, since

$$1 \geq \pi(\{x\} \times \Omega^x) = \int_\Theta [l(x,\theta)\lambda(\{x\})]\tau(d\theta)$$

and

$$\rho(x) = \int_\Theta l(x,\theta)\tau(d\theta) = \infty,$$

to avoid absurd conclusions it must be $\lambda(\{x\}) = 0$. Thus, when $\rho(x) = \infty$, (ii) holds for q'_x (whatever q'_x may be). As far as (iii) is concerned, let $x_k \in \mathscr{X}$, $\theta_k \in \Theta$, $\pi(\{x_k\} \times \Omega^{x_k}) = \pi(\Omega_{\theta_k} \times \{\theta_k\}) = 0$, $k = 1, \dots, m$, and $x_{m+1} := x_1$, $\theta_{m+1} := \theta_1$. If there exists an x_k with $\rho(x_k) > 0$, then, since $\pi(\{x_k\} \times \Omega^{x_k}) = 0$, it must be $\lambda(\{x_k\}) = 0$. In that case, being $P_\theta(\{x_k\}) = l(x_k, \theta)\lambda(\{x_k\}) = 0$, condition (iii) is trivially satisfied. Thus, it can be assumed that $\rho(x_k) = 0$ for $k = 1, \dots, m$. Next, the check of condition (iii) is immediate when there is an x_k such that Ω^{x_k} is infinite or Ω^{x_k} is finite with $\sum_{\theta \in \Omega^{x_k}} l(x_k, \theta) = 0$. Indeed, in the first case $q'_{x_k}(\{\theta\}) = 0$, $\forall\ \theta$ and, consequently, both sides of condition (iii) vanish; in the second one, the same circumstance holds since $P_\theta(\{x_k\}) = l(x_k, \theta)\lambda(\{x_k\}) = 0$. Finally, if Ω^{x_k} is finite and $g(x_k) := \sum_{\theta \in \Omega^{x_k}} l(x_k, \theta) > 0$ for $k = 1, \dots, m$, we obtain

$$\begin{aligned}\prod_{i=1}^m P_{\theta_i}(\{x_{i+1}\})q'_{x_i}(\{\theta_i\}) &= \prod_{i=1}^m \{l(x_{i+1},\theta_i)\lambda(\{x_{i+1}\})\}\{l(x_i,\theta_i)/g(x_i)\} \\ &= \prod_{i=1}^m \{l(x_i,\theta_i)\lambda(\{x_i\})\}\{l(x_{i+1},\theta_i)/g(x_{i+1})\} \\ &= \prod_{i=1}^m P_{\theta_i}(\{x_i\})q'_{x_{i+1}}(\{\theta_i\}),\end{aligned}$$

which proves the first part of condition (iii); the second one can be proved in an analogous way. □

If for each A in $\mathscr{A}_\mathscr{X}$ we put

$$\mu(A) = \pi((A \times \Theta) \cap \Omega),$$

then μ turns out to be a probability on $\mathscr{A}_\mathscr{X}$. According to Heath and Sudderth (1978) and Lane and Sudderth (1983), an inference q is substantially asked to meet the condition

$$\pi(C) = \int_\mathscr{X} q_x(C^x)\mu(dx) \quad \text{for all } C \in \mathscr{A}. \tag{3.2}$$

Connections between this type of coherence and that stated in Definition 2.1 are analyzed in Section 4 of Regazzini (1987). Here, we are providing conditions under which posterior q', assessed by the previous Bayes theorem, meets (3.2). Incidentally, we note that the question makes sense, in general, only if

$\mu^*(\{x: \rho(x) = 0 \text{ or } \rho(x) = \infty\}) = 0$; otherwise, the answer depends on the particular choice of q'_x when $\rho(x) \notin (0, \infty)$. Let us start by characterizing coherent posteriors, in the sense of Definition 2.1, which satisfy (3.2).

LEMMA 3.2. *Let q be a coherent posterior relative to Q, where Q is defined by* (2.1). *Then q satisfies* (3.2) *if and only if $x \to q_x(C^x)$ is μ-measurable for each $C \in \mathscr{A}$ and*

$$\mu(A) \inf_{x \in A} q_x(C^x) \leq \pi((A \times \Theta) \cap C) \leq \mu(A) \sup_{x \in A} q_x(C^x)$$

for all $C \in \mathscr{A}$ and $A \in \mathscr{A}_{\mathscr{X}}$.

PROOF. For $x \in \mathscr{X}$ and $C \in \mathscr{A}$, put $\sigma(C, \{x\} \times \Omega^x) = q_x(C^x)$, and apply Theorem 1.6 to π, σ, $\Pi = \{\{x\} \times \Omega^x: x \in \mathscr{X}\}$ and $\mathscr{E}_\Pi = \{(A \times \Theta) \cap \Omega: A \in \mathscr{A}_{\mathscr{X}}\}$. □

At this point, our goal can be stated in these terms: Find hypotheses under which q' satisfies the conditions of Lemma 3.2. To this end, besides (β1) and (β2) and (3.1), assume

(β3) *α can be determined in such a way that*

$$\int_\Theta \left[\int_{\mathscr{X}} l(x, \theta) I_C(x, \theta) \lambda(dx) \right] \alpha(d\theta)$$
$$= \int_{\mathscr{X}} \left[\int_\Theta l(x, \theta) I_C(x, \theta) \alpha(d\theta) \right] \lambda(dx) \quad \textit{for all } C \in \mathscr{A};$$

(β4) *the function $1/\rho(x) = \{\int_\Theta l(x, \theta) \tau(d\theta)\}^{-1}$ is μ-smooth.* (*For the definition of $1/\rho$ we adopt the conventions*: $1/0 = \infty$, $1/\infty = 0$.)

In a finitely additive framework, (β3) is an analog of Fubini's reduction formula; sufficient conditions for the validity of (β3) have been given by Sinclair (1974) and Thomsen (1978). As far as (β4) is concerned, it is apt to grant the integrability of q' with respect to μ. We note that

$$(\beta 3) \text{ and } (\beta 4) \Rightarrow \mu^*(\{x: \rho(x) = 0 \text{ or } \rho(x) = \infty\}) = 0.$$

We are now able to prove Theorem 3.3.

THEOREM 3.3. *Let* (β1)–(β4) *hold, and Q be defined on $\mathscr{D}$ according to* (2.1) *with π assessed through* (3.1). *Then the coherent posterior q' of Theorem* 3.1 *satisfies* (3.2).

PROOF. We have to prove that q' satisfies conditions of Lemma 3.2. From the second member of (β3) we deduce that the function defined on $\mathscr{X}$ by

$$g_C(x) := \int_\Theta l(x, \theta) I_C(x, \theta) \alpha(d\theta) = \int_{C^x} l(x, \theta) \tau(d\theta)$$

is λ-measurable for every C in $\mathscr{A}$. Moreover, given A in $\mathscr{A}_{\mathscr{X}}$:

$$\mu(A) = \int_{\Theta} P_\theta(A)\alpha(d\theta) = \int_{\mathscr{X}}\left[\int_{\Theta} l(x,\theta)I_A(x)\alpha(d\theta)\right]\lambda(dx) = \int_A \rho(x)\lambda(dx),$$

and μ turns out to be absolutely continuous with respect to λ. Hence, λ-measurability of g_C implies μ-measurability of g_C. In particular, setting $C = \Omega$, $\rho \equiv g_\Omega$ is μ-measurable. Note also that, if $0 < \rho(x) < \infty$, then $q'_x(C^x) = g_C(x)/\rho(x)$. Thus, since ρ and g_C are μ-measurable, and $1/\rho$ is μ-smooth, Proposition 1.5 implies that $x \to q'_x(C^x)$ is μ-measurable. Finally,

$$\begin{aligned}
\mu(A)\inf_{x\in A} q'_x(C^x) &= \Big\{\inf_{x\in A} q'_x(C^x)\Big\}\int_A \rho(x)\lambda(dx)\\
&\le \int_A q'_x(C^x)\rho(x)\lambda(dx)\\
&= \int_A\left[\int_\Theta l(x,\theta)I_C(x,\theta)\alpha(d\theta)\right]\lambda(dx)\\
&= \int_\Theta\left[\int_{\mathscr{X}} l(x,\theta)I_A(x)I_C(x,\theta)\lambda(dx)\right]\alpha(d\theta)\\
&= \int_\Theta P_\theta(A\cap C_\theta)\alpha(d\theta)\\
&= \pi((A\times\Theta)\cap C) \le \Big\{\sup_{x\in A} q'_x(C^x)\Big\}\int_A \rho(x)\lambda(dx)\\
&= \mu(A)\sup_{x\in A} q'_x(C^x).
\end{aligned}$$

□

EXAMPLE 3.4. Let $\mathscr{X} = \Theta = \mathbb{R}$, $\mathscr{A}_{\mathscr{X}} = \mathscr{A}_\Theta = \mathscr{B}$ (= class of Borel subsets of $\mathbb{R}$), λ = Lebesgue measure on $(\mathbb{R}, \mathscr{B})$, $l(x,\theta) = (2\pi)^{-1/2}\exp\{-(x-\theta)^2/2\}$ with $(x,\theta) \in \mathbb{R}^2$, $\Omega = \mathbb{R}^2$, $\mathscr{A} = \mathscr{B}^{(2)}$ (= class of Borel subsets of $\mathbb{R}^2$). Fix $\varepsilon \in (0,1)$; let $\tau = \tau_c + \delta_0^*$ be a probability on $(\mathbb{R}, \mathscr{B})$, where τ_c and δ_0^* are charges on $\mathscr{B}$, τ_c is strongly continuous [Bhaskara Rao and Bhaskara Rao (1983), page 142] and δ_0^* is characterized by a mass $c_2\varepsilon$ concentrated on 0 and by masses $c_1\varepsilon$ and $c_3\varepsilon$ adherent, respectively, to the left and to the right of 0, with $\delta_0^*(\mathbb{R}) = \varepsilon$, $c_i \ge 0$, $\forall\, i$, $\sum_{i=1}^3 c_i = 1$. Motivations for the choice of such a prior are included in Consonni and Veronese (1987). Clearly, conditions (β1) and (β2) hold and $0 < \rho(x) = \int_{\mathbb{R}}(2\pi)^{-1/2}\exp\{-(x-\theta)^2/2\}\tau_c(d\theta) + \varepsilon(2\pi)^{-1/2}\exp\{-x^2/2\}$, $\forall$ $x \in \mathbb{R}$. Therefore, by virtue of Theorem 3.1, $q'_x(B) = \int_B l(x,\theta)\tau(d\theta)/\rho(x)$, $B \in \mathscr{A}_\Theta$, is a coherent posterior for x in $\mathbb{R}$. In particular,

$$q'_x((-\infty,0)) = \left\{\int_{(-\infty,0)} l(x,\theta)\tau_c(d\theta) + c_1\varepsilon l(x,0)\right\}/\rho(x),$$

$$q'_x((-\infty,0]) = q'_x((-\infty,0)) + c_2\varepsilon l(x,0)/\rho(x),$$

$$q'_x((-\infty,\theta]) = \begin{cases} \displaystyle\int_{(-\infty,\theta]} l(x,\theta)\tau_c(d\theta)/\rho(x), & \text{if } \theta < 0,\\[2ex] \displaystyle\left\{\int_{(-\infty,\theta]} l(x,\theta)\tau_c(d\theta) + \varepsilon l(x,0)\right\}/\rho(x), & \text{if } \theta > 0.\end{cases}$$

EXAMPLE 3.5. Let $\mathscr{X} = [-1, 1]$, $\Theta = (0, 1]$, $\mathscr{A}_{\mathscr{X}} = \mathscr{B} \cap [-1, 1]$, $\mathscr{A}_{\Theta} = \mathscr{B} \cap (0, 1]$, $\Omega = \{(x, \theta): 0 < \theta \le 1, -\theta \le x \le \theta\}$, $\mathscr{A} = \mathscr{B}^{(2)} \cap \Omega$; λ = Lebesgue measure on $(\mathscr{X}, \mathscr{A}_{\mathscr{X}})$;

$$l(x, \theta) = \begin{cases} 1/\theta, & \text{if } 0 \le x \le \theta \text{ and } \theta \text{ is any irrational number in } (0,1), \\ 1/\theta, & \text{if } -\theta \le x \le 0 \text{ and } \theta \text{ is any rational number in } (0,1], \\ 0, & \text{elsewhere;} \end{cases}$$

τ = Lebesgue measure on $(\Theta, \mathscr{A}_{\Theta})$. Also here conditions ($\beta$1) and ($\beta$2) hold; as a matter of fact, the example at issue falls within the classical Kolmogorovian approach to probability and, consequently, (β3) and (β4) are satisfied. Clearly, $\rho(x) = 0$ if $x = 1$ or $x \in [-1, 0)$, $\rho(0) = \infty$, $\rho(x) = -\log x$ if $x \in (0, 1)$. For every $x \in (-1, 0)$, select a diffuse probability d_x on $\mathscr{A}_{\Theta}$ such that $d_x([-x, 1] \cap Q) = 1$, Q denoting the set of rational numbers. Then

$$q_x'(B) = \begin{cases} \int_B l(x, \theta)\, d\theta / (-\log x), & \text{if } x \in (0,1), \\ I_B(1), & \text{if } x = 1 \text{ or } x = -1, \\ \tau(B), & \text{if } x = 0, \\ d_x(B), & \text{if } x \in (-1, 0), \end{cases}$$

is a coherent posterior. In addition, such a posterior is easily shown to satisfy (3.2).

EXAMPLE 3.6. Let $\mathscr{X} = [1, \infty) \cap Q$, $\Theta = (0, 1] \cap Q$ and $\Omega = \{(x, \theta): \theta \in \Theta, x = 1 + n\theta, n \in \mathbb{N} \cup \{0\}\}$. Take $\mathscr{A}_{\mathscr{X}}$, $\mathscr{A}_{\Theta}$ and $\mathscr{A}$ as the power sets on $\mathscr{X}$, Θ and Ω, respectively; λ = counting measure; $l(x, \theta) = 1/2^{n+1}$ if $\theta \in \Theta$ and $x = 1 + n\theta$ for some n, and $l(x, \theta) = 0$ if $(x, \theta) \notin \Omega$. For instance, this could be a model for some propagation phenomenon, involving only a fraction $\theta > 0$ of a very large population, and distinguished by a strong dependence among the individual behaviors. It is worth stressing that, fixed $x \in \mathscr{X} \setminus \{1\}$, $l(x, \theta) \to 0$ as $\theta \to 0$. Let τ be a probability on $(\Theta, \mathscr{A}_{\Theta})$ characterized by a mass c_1 adherent to the right of 0 and by masses c_2 and c_3 concentrated, respectively, on the points $\frac{1}{2}$ and 1, with $c_1 \ge 0$, $c_2, c_3 > 0$, $\sum_{i=1}^{3} c_i = 1$. Once again conditions (β_1) and (β_2) hold, with $\rho(1) = \frac{1}{2}$ and $\rho(x) = c_2 l(x, \frac{1}{2}) + c_3 l(x, 1)$ for $x \ne 1$. Thus, $0 < \rho(x) < \infty$ whenever at least one of the points $\frac{1}{2}, 1$ belongs to Ω^x, and $\rho(x) = 0$ otherwise. In the former case, q_x' can be defined according to the usual Bayesian algorithm, while in the latter, being Ω^x not finite, q_x' can be coherently assessed as any probability on $\mathscr{A}_{\Theta}$ adherent to the right of 0.

4. Concluding remarks. This section has been divided into two parts. The first deals with the connections between our previous results and the general version of Bayes theorem obtained, in a Kolmogorovian framework, by Kallianpur and Striebel (1968); the second includes a few comments concerning the selection of q_x' when $\rho(x) = 0$ or $\rho(x) = \infty$.

The quoted authors provide general Bayes theorems which express posteriors as conditional probabilities with respect to a σ-algebra. This circumstance shows that the aim of Kallianpur and Striebel's study differs from ours; hence, a comparison in terms of greater or lesser generality cannot be proposed from a merely logical angle. On the other hand, conditional probability with respect to a σ-algebra is introduced via the analytical concept of a Radon–Nikodym derivative, but its concrete, probabilistic meaning is elusive. Hence, if the conditioning σ-algebra is generated by an observable random entity, then conditional probability with respect to such a σ-algebra is used as conditional probability with respect to any value of the conditioning entity. In compliance with this interpretation of conditional probabilities with respect to σ-algebras, a comparison between our statements and those of Kallianpur and Striebel makes sense from a more basic angle. Such a comparison shows that conditions (β1)–(β4) are weaker than Kallianpur and Striebel's hypotheses and, consequently, the resulting statements determine a wider scale of application for the Bayesian algorithm. This is shown by Examples 3.4 and 3.6. Moreover, such an enlargement of applicability is stated in terms of coherence, and not in a purely formal way. This observation is important since coherence, for example, places restrictions on our freedom of choosing q'_x when $\rho(x) = 0$, whilst no restriction is dictated, in such a case, when the Kolmogorovian approach is adopted. To illustrate this point, let us reconsider Example 3.6 (supposing $c_1 = 0$ so that the example itself falls within the Kolmogorovian setting). When $\Omega^x \cap \{\frac{1}{2}, 1\} = \varnothing$, we are obliged to assign a posterior satisfying conditions of Theorem 2.2. On the contrary, if we follow the classical approach, there is no reason to favour one distinguished probability law rather than another. Note also that, in Example 3.5, when $x \in (-1, 0)$ we are able to select a diffuse probability d_x giving the whole mass to a countable set of rationals. This seems to be a sensible choice; nevertheless, it would not be available in a σ-additive framework.

Leaving apart the relationships between coherence and the Kolmogorovian setting, let us turn to some brief remarks concerning the choice of q'_x when $\rho(x) = 0$ or $\rho(x) = \infty$. Plainly, in such cases, our definition of q'_x is not the only coherent one, and any other coherent assessment is in order. Thus, to make easier the evaluation of q'_x, it can be helpful to attach some meaning to $\rho(x) = 0$ and $\rho(x) = \infty$. An heuristic interpretation is the following. Suppose that $\rho(x) = 0$ and $l(x, \theta) > 0$ for some θ. [Otherwise, if $l(x, \cdot\,) \equiv 0$, $\rho(x)$ is forced to be 0 whatever the prior τ may be.] Since $\rho(x) = 0$ is equivalent to $\tau^*(\{\theta\colon\ l(x, \theta) > \varepsilon\}) = 0$ for every $\varepsilon > 0$, it can be argued that the function $l(x, \cdot\,)$ is "very low" where the prior τ "gives mass." At least in a rough sense, this can be viewed as a sharp disagreement between the prior beliefs and the experimental result. Hence, the inferrer should deeply modify his prior opinions (possibly, selecting q'_x in a different way with respect to our evaluation of Section 3). Generally, however, it is not possible to hint an explicit rule to define q'_x as a function of τ and $l(x, \cdot\,)$, since the evaluation of q'_x depends on the distinctive features of the specific problem under examination. This is why, in Section 3, we have been vague in our assessment of q'_x when $\rho(x) = 0$ and

Ω^x is infinite. Nevertheless, one possibility could be to ask q'_x to be singular with respect to τ, at least when this is meaningful in the investigated problem. This happens, for instance, in Example 3.5 but not necessarily in Example 3.6.

Conversely, if $\rho(x) = \infty$, the same argument applied when $\rho(x) = 0$ leads to a full confirmation of the prior beliefs. Consequently, the inferrer should preserve them, that is, if $\rho(x) = \infty$, he should assess $q'_x(B) = \tau(B \cap \Omega^x)/\tau(\Omega^x)$ for all B in $\mathscr{A}_\Theta$.

Finally, we mention a different approach to assign q'_x when $\rho(x) = 0$. However, the reader ought to be warned that the probability obtained this way does not necessarily meet the requirements suggested so far. The idea is to express q'_x, at least when the necessary conditions are available, as a "limit of truncated Bayes inferences" [cf. Regazzini (1987), Section 3]. Fixed $x \in \mathscr{X}$ with $\rho(x) = 0$, consider the probability

$$(4.1) \qquad q'_x(B) = \int_B l(x,\theta)\gamma(d\theta) \Big/ \int_\Theta l(x,\theta)\gamma(d\theta) \quad \text{for all } B \in \mathscr{A}_\Theta,$$

where γ is a σ-additive charge on $\mathscr{A}_\Theta$ such that $0 < \int_\Theta l(x,\theta)\gamma(d\theta) < \infty$. In addition, suppose $\mathscr{A}_\Theta$ is a σ-algebra and there exists a nondecreasing sequence $\{\Theta_n\} \subset \mathscr{A}_\Theta$ such that $\Theta_n \uparrow \Theta$, $\gamma(\Theta_n)$ and $\int_{\Theta_n} l(x,\theta)\gamma(d\theta)$ belong to $(0,\infty)$ for each n greater than some n_0, and

$$(4.2) \qquad \tau(B) = \lim_{n\to\infty} \gamma(B \cap \Theta_n)/\gamma(\Theta_n) \quad \text{whenever } B \in \mathscr{A}_\Theta \text{ and the limit does exist.}$$

Of course, it is also tacitly assumed that the class of B's for which (4.2) holds is not "too small"; otherwise the link between τ and γ is poor or even null. Under the previous conditions, the inferrer can define q'_x as in (4.1), provided this is a coherent choice (which is always true when γ is diffuse). It is easily shown, by the monotone convergence theorem, that q'_x can be expressed as the limit of the sequence of probabilities obtained, for each $n > n_0$, by applying the Bayes theorem in the space $(\Theta_n, \mathscr{A}_\Theta \cap \Theta_n)$ with the prior probability $\gamma(\cdot \cap \Theta_n)/\gamma(\Theta_n)$. For instance, when $\Theta = \mathbb{N}$, our assessment of q'_x, for the case $\rho(x) = 0$, Ω^x finite and $l(x,\cdot) \neq 0$, can be obtained by the previous device setting $\Theta_n = \{1,\dots,n\}$ and γ = counting measure.

Acknowledgments. This paper has been improved by some remarks of the referees and the Associate Editor. Specifically, the authors are indebted for a suggestion which simplifies the proof of Proposition 1.5 and for a critical discussion of Section 4.

REFERENCES

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

Consonni, G. and Veronese, P. (1987). Coherent distributions and Lindley's paradox. In *Probability and Bayesian Statistics* (R. Viertl, ed.) 111–120. Plenum, New York.

Császár, Á. (1955). Sur la structure des espaces de probabilité conditionnelle. *Acta Math. Acad. Sci. Hung.* **6** 337–361.

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

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

KALLIANPUR, G. and STRIEBEL, C. (1968). Estimation of stochastic systems: Arbitrary system process with additive white noise observation error. *Ann. Math. Statist.* **39** 785–801.

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

REGAZZINI, E. (1984). A Bayes's theorem for finitely additive probabilities and some considerations about the use of improper priors. Technical Report S30/1984, Dip. Matematica, Univ. di Milano.

REGAZZINI, E. (1985). Finitely additive conditional probabilities. *Rend. Sem. Mat. Fis. Milano* **55** 69–89.

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

RIGO, P. (1988). Un teorema di estensione per probabilita' condizionate finitamente additive. *Atti della XXXIV Riunione Scientifica della Societa' Italiana di Statistica* **2** 27–34.

SINCLAIR, G. D. (1974). A finitely additive generalization of the Fichtenholz–Lichtenstein theorem. *Trans. Amer. Math. Soc.* **193** 359–374.

THOMSEN, W. (1978). On a Fubini-type theorem and its applications in game theory. *Math. Operationsforsch. Statist.* **9** 419–423.

ISTITUTO DI MATEMATICA APPLICATA
"GIOVANNI SANSONE"
VIA S. MARTA, 3
50139 FIRENZE
ITALY

ISTITUTO DI METODI QUANTITATIVI
UNIVERSITÀ "L. BOCCONI"
VIA SARFATTI, 25
20136 MILANO
ITALY

DIPARTIMENTO STATISTICO
VIA CURTATONE, 1
50123 FIRENZE
ITALY