The Annals of Statistics
1996, Vol. 24, No. 3, 1235–1249

ON THE EXISTENCE OF INFERENCES WHICH ARE CONSISTENT WITH A GIVEN MODEL[1]

BY PATRIZIA BERTI AND PIETRO RIGO

Università di Firenze

If $\{p_\theta\}$ is a σ-additive statistical model and π a finitely additive prior, then any statistic T is sufficient, with respect to a suitable inference consistent with $\{p_\theta\}$ and π, provided only that $p_\theta(T = t) = 0$ for all θ and t. Here, sufficiency is to be intended in the Bayesian sense, and consistency in the sense of Lane and Sudderth. As a corollary, if $\{p_\theta\}$ is σ-additive and diffuse, then, whatever the prior π, there is an inference which is consistent with $\{p_\theta\}$ and π. Two versions of the main result are also obtained for predictive problems.

1. Introduction and motivation. Suppose that, for each θ in a parameter space Θ, a probability p_θ is assigned on a σ-field $\mathscr{A}_{\mathfrak{X}}$ of subsets of $\mathfrak{X}$. Here, $\mathfrak{X}$ is to be seen as the collection of possible outcomes of some experiment. Call the family $p := \{p_\theta: \theta \in \Theta\}$ a (*statistical*) *model*. Likewise, call an *inference* any family $q := \{q_x: x \in \mathfrak{X}\}$, the q_x being probabilities on a σ-field $\mathscr{A}_\Theta$ of subsets of Θ.

In the Bayesian approach, once a model p is assigned, the goal of inferential analysis lies in assessing an inference q. But also some non-Bayesian procedures, such as confidence intervals, fiducial distributions and some likelihood methods, can be led back, at least formally, to the selection of an inference q. Hence, a main question in statistical inference is: which are the "admissible" inferences once a model is assigned? Or, using a different terminology, which inferences are "consistent" with a given model?

One answer is that an inference q is *consistent* with a model p provided there are probabilities π and m, defined on the power sets of Θ and $\mathfrak{X}$, respectively, such that

$$\int\int \phi(x, \theta) p_\theta(dx) \pi(d\theta) = \int\int \phi(x, \theta) q_x(d\theta) m(dx) \tag{1.1}$$

for every bounded $\phi: \mathfrak{X} \times \Theta \to \mathbb{R}$ which is measurable with respect to (w.r.t.) the product σ-field $\mathscr{A}_{\mathfrak{X}} \otimes \mathscr{A}_\Theta$. Plainly, π and m are finitely additive probabilities and are not forced to be σ-additive. Besides, if (1.1) holds for a particular π, q is also said to be consistent with p and π.

The above notion of consistency has been introduced by Lane and Sudderth (1983). Substantially, it is a weak version of a definition of coherence for

Received May 1994; revised November 1995.
[1] Research supported in part by M.U.R.S.T.
AMS 1991 *subject classifications*. Primary 62A15; secondary 60A05.
Key words and phrases. Cardinality, coherence, consistent inference, diffuse probability, finite additivity, perfect probability, posterior Bayes rule, prediction, sufficient statistic.

statistical inference given by Heath and Sudderth (1978). We think that (1.1) is an interesting condition for at least three (not independent) reasons. First, at least for diffuse models, (1.1) is sufficient (even if not necessary) for *dF-coherence*, that is, for that notion of coherence introduced by Regazzini (1987) developing de Finetti's theory of coherence for conditional probabilities (cf. Section 3). Indeed, in our opinion, dF-coherence is a fundamental requisite for any inference q. Second, while allowing for finite additivity and not imposing measurability constraints, (1.1) captures much of the meaning of the usual Kolmogorov definition of conditional probability (cf. Section 6). Third, (1.1) may also be interpreted as a sort of "second order de Finetti's coherence" [Berti and Rigo (1992)], and, in this sense, (1.1) establishes some link between the theories of de Finetti and Kolmogorov.

However, a serious shortcoming in the definition of consistency is that, given a fixed model p and a "prior" π on the power set of Θ, it may be that there are not an inference q and a probability m satisfying (1.1). Examples of this phenomenon are in Dubins (1975), Prikry and Sudderth (1982) and Heath and Sudderth (1989). [Incidentally, a further drawback is the possible nonuniqueness of the couple q, m satisfying (1.1), even when it exists, but this is generally unavoidable.]

One result in this paper is that, if p_θ is σ-additive and $p_\theta\{x\} = 0$ for all θ and x, then, for *any* prior π, there are an inference q and a probability m satisfying (1.1). In other terms, the above-mentioned shortcoming in the definition of consistency cannot occur for σ-additive, diffuse models. Moreover, the previous result can be strengthened as follows. If p_θ is σ-additive and T is any measurable function on $\mathscr{X}$ such that $p_\theta\{x: T(x) = t\} = 0$ for all θ and t, then, whatever the prior π, condition (1.1) holds for some probability m and inference q such that

$$x_1, x_2 \in \mathscr{X} \quad \text{and} \quad T(x_1) = T(x_2) \quad \Rightarrow \quad q_{x_1} = q_{x_2}. \tag{1.2}$$

As far as consistency is regarded as a satisfactory requisite for an inference q, (1.2) has a nice interpretation in terms of sufficiency (in the Bayesian sense; cf. Section 6). Briefly, whatever the prior π, any statistic T is sufficient, w.r.t. a suitable consistent inference q, provided only that no value t in the range of T has positive probability under the model. This seems in line with both the substantial meaning of sufficiency and the subjective view of probability. Indeed, the assessment of q can be split into two steps. First, a partition of $\mathscr{X}$ is selected, by grouping those experimental outcomes which, according to the inferrer, have the same inferential content. This step precisely amounts to the choice of a (sufficient) statistic T. Subsequently, a probability law on $\mathscr{A}_\Theta$ is attached to every element of the partition. Our result grants that, if consistency is seen as "enough" for q, this procedure can always be performed whenever no element in the partition has positive probability under the model.

The above considerations about sufficiency are particularly meaningful in a predictive framework, that is, when the inferential analysis aims to predict future observable facts. Two versions of the result on sufficient statistics are obtained. In the first, predictive inferences are based on a statistical model and a prior distribution, while in the second they are not.

Two such versions are given in Section 5, after the main result is proved in Section 3. The proof rests on a simple cardinality argument. Section 3 also includes some heuristic comments on the role played by the assumptions of σ-additivity and diffuseness of the model. Moreover, Section 2 contains some preliminary material, Section 4 some examples and Section 6 a few general remarks.

A last note is that the interplay between consistency and sufficiency is also treated, from a different point of view, in Wetzel (1993).

2. Preliminaries. In this section, after introducing some terminology and notation, three lemmas are given. The first two are mere technical facts, to be used in the proofs of subsequent results. Even if quite intuitive, the third has perhaps some autonomous interest.

Let Ω be any set. Throughout this paper, a *probability* is a nonnegative, finitely additive function, defined on some field of subsets of Ω, and assuming value 1 at Ω. Let $\mathscr{P}(\Omega)$ denote the power set of Ω. Given any partition $\mathscr{U}$ of Ω, a *$\mathscr{U}$-strategy* is a function $P(\cdot \mid \cdot)$ on $\mathscr{P}(\Omega) \times \mathscr{U}$ such that, for each $H \in \mathscr{U}$, $P(\cdot|H)$ is a probability on $\mathscr{P}(\Omega)$ with $P(H|H) = 1$.

Let $\mathscr{D}$ be a σ-field of subsets of Ω. An *atom* of $\mathscr{D}$ is the intersection of all the elements of $\mathscr{D}$ including a given point of Ω. Let $\mathscr{U}(\mathscr{D})$ be the collection of atoms of $\mathscr{D}$. Then, $\mathscr{U}(\mathscr{D})$ need not be included in $\mathscr{D}$, but $\mathscr{U}(\mathscr{D})$ is a partition of Ω and every element of $\mathscr{D}$ is union of elements of $\mathscr{U}(\mathscr{D})$. A sufficient condition for $\mathscr{U}(\mathscr{D}) \subset \mathscr{D}$ is that $\mathscr{D}$ is countably generated.

A probability P on $\mathscr{D}$ is said to be *diffuse* whenever $\mathscr{U}(\mathscr{D}) \subset \mathscr{D}$ and $P(H) = 0$ for $H \in \mathscr{U}(\mathscr{D})$. Further, P is said to be *perfect* whenever, for every $\mathscr{D}$-measurable function $f\colon \Omega \to \mathbb{R}$, there is a Borel set $B \subset f(\Omega)$ such that $P(f \in B) = 1$. We remark that, if $(\Omega, \mathscr{D})$ is a *standard space*, that is, Ω is a Borel set in some complete, separable metric space and $\mathscr{D}$ the corresponding Borel σ-field, then $\mathscr{D}$ is countably generated and any σ-additive P on $\mathscr{D}$ is perfect. More generally, this is also true if $(\Omega, \mathscr{D})$ is a *Lusin space*, as defined, for instance, in Blackwell (1955). A last point is that all the integrals in this paper are intended in the sense of Dunford and Schwartz (1958), Chapter 3. We are now in a position to state the three preliminary lemmas.

LEMMA 2.1. *Let A, B be nonempty sets, and for each $a \in A$, let $L(a)$ be a subset of B. If* card $A \leq$ card B, *then there is an injective function $f\colon A \to B$ such that $f(a) \in L(a)$ whenever $a \in A$ and* card $L(a) \geq$ card A.

PROOF. Let $\leq$ be a well ordering on A such that, setting $I(a) = \{x \in A\colon x < a\}$, one has card $I(a) <$ card A for every $a \in A$. Assign also any well ordering on B. For fixed $a \in A$, suppose $f\colon I(a) \to B$ is defined such that

$$f(x) = \begin{cases} \min[B - f(I(x))], & \text{if card } L(x) < \text{card } A, \\ \min[L(x) - f(I(x))], & \text{if card } L(x) \geq \text{card } A. \end{cases} \tag{2.1}$$

Since $\operatorname{card} B \geq \operatorname{card} A > \operatorname{card} I(a) = \operatorname{card} f(I(a))$, the set $B - f(I(a))$ is nonempty. Likewise, if $\operatorname{card} L(a) \geq \operatorname{card} A$, then $L(a) - f(I(a)) \neq \emptyset$. Hence, f can be extended to $I(a) \cup \{a\}$ by setting $f(a) = \min[B - f(I(a))]$ or $f(a) = \min[L(a) - f(I(a))]$ according to whether $\operatorname{card} L(a) < \operatorname{card} A$ or $\operatorname{card} L(a) \geq \operatorname{card} A$. By transfinite induction, there is $f\colon A \to B$ behaving as in (2.1) for every $x \in A$. In particular, such f has the desired properties. □

LEMMA 2.2. *Let $\mathscr{U}$ be a partition of Ω, $P(\cdot \mid \cdot)$ a $\mathscr{U}$-strategy and P a probability on a field $\mathscr{C}$. Setting $\mathscr{C}_0 = \{E \in \mathscr{C}\colon P(E) > 0$ and $E \cap H \neq H$ for each $H \in \mathscr{U}\}$, suppose that*

(2.2) *for every $E \in \mathscr{C}_0$ there is $H \in \mathscr{U}$ such that $P(E \mid H) = 1$.*

Then, there is a probability λ on $\mathscr{P}(\mathscr{U})$ such that

$$P(E) = \int P(E \mid H)\,\lambda(dH) \quad \text{for all } E \in \mathscr{C}.$$

PROOF. Let $\mathscr{Z}$ be the linear space spanned by the indicators of the elements of $\mathscr{C}$. For $X \in \mathscr{Z}$ and $H \in \mathscr{U}$, define $Q(X) = \int X\,dP$ and $Q(X \mid H) = \int X(\omega) P(d\omega \mid H)$. By Theorem 3.1 of Berti and Rigo (1992), it is enough to prove that $Q(X) \leq \sup_H Q(X \mid H)$ for all $X \in \mathscr{Z}$. Fix X in $\mathscr{Z}$. Since $\mathscr{C}$ is a field, X can be expressed as $X = \sum_{i=1}^n a_i I_{E_i}$, where $a_1, \dots, a_n \in \mathbb{R}$ and $\{E_1, \dots, E_n\}$ is a partition of Ω in $\mathscr{C}$. Pick j with $P(E_j) > 0$ and $a_j \geq a_i$ for each i with $P(E_i) > 0$. Using (2.2), it is easily seen that, whether or not $E_j \in \mathscr{C}_0$, there is $H_j \in \mathscr{U}$ with $P(E_j \mid H_j) = 1$. Hence,

$$Q(X) = \sum a_i P(E_i) \leq a_j = Q\big(X \mid H_j\big) \leq \sup_H Q(X \mid H). \qquad \square$$

LEMMA 2.3. *Let $\mathscr{D}$ be a countably generated σ-field, $\mathscr{U}(\mathscr{D})$ the set of atoms of $\mathscr{D}$ and P a probability on $\mathscr{D}$. If P is σ-additive, diffuse and perfect, then $P(E) = 0$ whenever $E \in \mathscr{D}$ and $\operatorname{card}\{H \in \mathscr{U}(\mathscr{D})\colon H \subset E\} < \operatorname{card} \mathbb{R}$.*

PROOF. Let $E \in \mathscr{D}$ with $\operatorname{card}\{H \in \mathscr{U}(\mathscr{D})\colon H \subset E\} < \operatorname{card} \mathbb{R}$. Since $\mathscr{D}$ is countably generated, there is $f\colon \Omega \to \mathbb{R}$ with $\mathscr{D} = f^{-1}(\mathscr{B})$, $\mathscr{B}$ denoting the Borel σ-field on $\mathbb{R}$. Let $B_1 \in \mathscr{B}$ be such that $E = f^{-1}(B_1)$. Since P is perfect, there is $B_2 \in \mathscr{B}$ with $B_2 \subset f(\Omega)$ and $P(f \in B_2) = 1$. Let $B = B_1 \cap B_2$. Noting that $B_1 \cap f(E^c) = \emptyset$, one obtains: $\operatorname{card} B \leq \operatorname{card} f(E) = \operatorname{card}\{H \in \mathscr{U}(\mathscr{D})\colon H \subset E\} < \operatorname{card} \mathbb{R}$. Hence, B must be countable, since this is generally true for any $U \in \mathscr{B}$ such that $\operatorname{card} U < \operatorname{card} \mathbb{R}$ [Parthasarathy (1967), Theorem 2.8, page 12]. Noting that $P \circ f^{-1}$ is σ-additive and diffuse, one has $P(E) = P(f \in B_1) = P(f \in B) = 0$. □

3. The main theorem. We shall now return to the general inferential setting outlined in Section 1. Before stating our main result, some more information is in order, concerning some of the statements made earlier in Section 1.

As already pointed out, if p is diffuse, then consistency of q with p suffices for the dF-coherence of q. This follows, for instance, from Theorem 4.1 of Berti and Rigo (1994). Using Theorem 4.1, it is also easy to produce examples of nonconsistent inferences which are dF-coherent. The notion of dF-coherence has been introduced by Regazzini (1987), by applying to an inferential problem the general idea of de Finetti's coherence. Briefly, dF-coherence amounts to demanding that all the "ingredients" of the problem (in our case, p, q, π and m) are parts of the same conditional probability, the latter to be intended in de Finetti's sense. We refer to Regazzini (1987) and to Berti, Regazzini and Rigo (1991) for a discussion of the underlying ideas, as well as for a comparison with the other notion of coherence, introduced by Heath and Sudderth (1978), which we now briefly recall.

Suppose for a moment that every element p_θ of the model p is extended as a probability on $\mathscr{P}(\mathfrak{X})$. Then, after assessing a prior π on $\mathscr{P}(\Theta)$, one can define

$$m_\pi(A) = \int p_\theta(A)\pi(d\theta) \quad \text{for all } A \subset \mathfrak{X}.$$

If (1.1) holds for some π, with m_π in the place of m, then q is coherent w.r.t. p according to Heath and Sudderth. Thus, coherence in Heath and Sudderth's sense implies consistency. However, unless q satisfies some measurability assumptions, the converse need not be true [Berti and Rigo (1994), Example 3.6].

Let T be a *statistic*, that is, a measurable function from $(\mathfrak{X}, \mathscr{A}_\mathfrak{X})$ into some measurable space $(\mathscr{T}, \mathscr{A}_\mathscr{T})$. In this paper, unless otherwise stated, T is said to be *sufficient for an inference* q whenever "q depends on the data only through T," that is, condition (1.2) holds. See Section 6 for a brief comparison between the usual notions of sufficiency and the present one.

For the reasons explained in Section 1, the following proposition is in line with the meaning of the above definition of sufficiency.

THEOREM 3.1. *Given a model p and a statistic T, suppose that*:

(i) $\operatorname{card} \mathscr{A}_\mathfrak{X} \le \operatorname{card} \mathbb{R}$, $\operatorname{card} \mathscr{A}_\Theta \le \operatorname{card} \mathbb{R}$ *and the σ-field $T^{-1}(\mathscr{A}_\mathscr{T})$ is countably generated*;

(ii) *for each $\theta \in \Theta$, p_θ is a σ-additive and perfect probability on $\mathscr{A}_\mathfrak{X}$, such that $p_\theta(F) = 0$ for all atoms F of $T^{-1}(\mathscr{A}_\mathscr{T})$.*

Then, for any probability π on $\mathscr{P}(\Theta)$, there are an inference q and a probability m on $\mathscr{P}(\mathfrak{X})$ such that:

(a) *for each $x \in \mathfrak{X}$, q_x is a σ-additive probability on $\mathscr{A}_\Theta$*;
(b) *condition* (1.1) *holds for p, q, π and m*;
(c) *condition* (1.2) *holds for q and T*; *that is*, *T is sufficient for q.*

PROOF. For $C \subset \mathfrak{X} \times \Theta$ and $(x, \theta) \in \mathfrak{X} \times \Theta$, let $C_\theta = \{x: (x, \theta) \in C\}$ and $C^x = \{\theta: (x, \theta) \in C\}$. Moreover, in the notation of Lemma 2.2, define $\Omega =$

$\mathfrak{X} \times \Theta$, $\mathscr{C} = \mathscr{A}_{\mathfrak{X}} \otimes \mathscr{A}_{\Theta}$ and $P(C) = \int p_{\theta}(C_{\theta})\pi(d\theta)$ for $C \in \mathscr{C}$. Finally, let $\mathscr{D} = T^{-1}(\mathscr{A}_{\mathscr{T}})$, $\mathscr{U}(\mathscr{D})$ the partition of $\mathfrak{X}$ in the atoms of $\mathscr{D}$ and $L(C) = \{F \in \mathscr{U}(\mathscr{D})$: $(F \times \Theta) \cap C \neq \emptyset\}$ for $C \in \mathscr{C}$. We claim that it is enough to prove that

$$\text{card}\, L(C) \geq \text{card}\, \mathbb{R} \quad \text{for all } C \in \mathscr{C} \text{ with } P(C) > 0. \tag{3.1}$$

Assume indeed that (3.1) holds. By (i), $\text{card}\, \mathscr{C} \leq \text{card}\, \mathbb{R}$, so that Lemma 2.1 implies the existence of an injective function f, from $\{C \in \mathscr{C}\colon P(C) > 0\}$ into $\mathscr{U}(\mathscr{D})$, such that $f(C) \in L(C)$. For each $F \in \mathscr{U}(\mathscr{D})$, select a σ-additive probability on $\mathscr{A}_{\Theta}$, say ν_F, according to the following rule. If F is not in the range of f, ν_F is arbitrary. Otherwise, if C is the unique element in the domain of f with $F = f(C)$, pick $x \in F$ with $C^x \neq \emptyset$ [which is possible since $F \in L(C)$] and take ν_F such that $\nu_F(C^x) = 1$. Now, define q by $q_x = \nu_F$ for every $x \in F$. Plainly, q satisfies (a) and (c). Moreover, (b) holds, too. In fact, let $\mathscr{U} = \{\{x\} \times \Theta\colon x \in \mathfrak{X}\}$ and let $P(\cdot|\cdot)$ be any $\mathscr{U}$-strategy such that

$$P(C|\{x\} \times \Theta) = q_x(C^x) \quad \text{for all } x \in \mathfrak{X} \text{ and } C \in \mathscr{C}.$$

Then $P(\cdot\mid\cdot)$ satisfies (2.2), and Lemma 2.2 implies the existence of a probability m on $\mathscr{P}(\mathfrak{X})$ such that (1.1) holds for all ϕ's of the form $\phi = I_C$, with $C \in \mathscr{C}$. But this is enough, since any bounded $\mathscr{C}$-measurable ϕ is the uniform limit of some sequence of simple functions.

It remains to check (3.1). Fix $C \in \mathscr{C}$ with $\text{card}\, L(C) < \text{card}\, \mathbb{R}$. We are showing that $P(C) = 0$. Let M be the union of those F's such that $F \in L(C)$. Since $C_{\theta} \subset M$ for all θ, it suffices to prove that $p_{\theta}(A) = 0$ whenever $\theta \in \Theta$, $A \in \mathscr{A}_{\mathfrak{X}}$ and $A \subset M$. In fact, in this case, $P(C) = \int p_{\theta}(C_{\theta})\pi(d\theta) = 0$. Fix $\theta \in \Theta$ and $A \in \mathscr{A}_{\mathfrak{X}}$ with $A \subset M$. If $p_{\theta}(A) > 0$, one can define $\mu(E \cap A) = p_{\theta}(E \cap A)/p_{\theta}(A)$ for all $E \in \mathscr{D}$. By (i), $\mathscr{D} \cap A$ is a countably generated σ-field of subsets of A, and its atoms are of the form $F \cap A$ with $F \in \mathscr{U}(\mathscr{D})$ and $F \cap A \neq \emptyset$. By (ii), μ is σ-additive, diffuse and perfect. Hence, denoting by $\mathscr{U}(\mathscr{D} \cap A)$ the set of atoms of $\mathscr{D} \cap A$, one has

$$\begin{aligned} \text{card}\, L(C) &\geq \text{card}\{F \in \mathscr{U}(\mathscr{D})\colon F \cap A \neq \emptyset\} \\ &= \text{card}\, \mathscr{U}(\mathscr{D} \cap A) \geq \text{card}\, \mathbb{R}, \end{aligned} \tag{3.2}$$

where the first inequality depends on $A \subset M$, and the second inequality follows from Lemma 2.3 applied to $(A, \mathscr{D} \cap A, \mu)$. But (3.2) is a contradiction; hence, $p_{\theta}(A) = 0$. $\square$

Some examples concerning Theorem 3.1 are included in Section 4. Here, we give a few remarks. If the singletons $\{t\}$ belong to $\mathscr{A}_{\mathscr{T}}$, the assumption that p_{θ} vanishes on the atoms of $T^{-1}(\mathscr{A}_{\mathscr{T}})$ can be written as $p_{\theta}\{x\colon T(x) = t\} = 0$ for all $t \in \mathscr{T}$. Next, given any σ-field $\mathscr{F}$, a sufficient condition for $\text{card}\, \mathscr{F} \leq \text{card}\, \mathbb{R}$ is that $\mathscr{F}$ is generated by some subfamily $\mathscr{F}_0$ such that $\text{card}\, \mathscr{F}_0 \leq \text{card}\, \mathbb{R}$. Moreover, $T^{-1}(\mathscr{A}_{\mathscr{T}})$ is countably generated whenever $\mathscr{A}_{\mathscr{T}}$ is. Thus, (i) holds, for instance, whenever $\mathscr{A}_{\mathfrak{X}}$, $\mathscr{A}_{\Theta}$ and $\mathscr{A}_{\mathscr{T}}$ are all countably generated. We also recall that (cf. Section 2), if $(\mathfrak{X}, \mathscr{A}_{\mathfrak{X}})$ is a Lusin space and p_{θ} is σ-additive, then p_{θ} is perfect.

Setting $(\mathscr{T}, \mathscr{A}_{\mathscr{T}}) = (\mathfrak{X}, \mathscr{A}_{\mathfrak{X}})$ and $T(x) = x$, we get the following result which is a direct consequence of Theorem 3.1.

COROLLARY 3.2. *Given a model p, suppose that $\mathscr{A}_{\mathfrak{X}}$ is countably generated, card $\mathscr{A}_\Theta \le$ card $\mathbb{R}$ and p_θ is σ-additive, diffuse and perfect for each $\theta \in \Theta$. Then, for any probability π on $\mathscr{P}(\Theta)$, there are an inference q and a probability m on $\mathscr{P}(\mathfrak{X})$ satisfying* (a) *and* (b) *of Theorem* 3.1.

In Corollary 3.2, assuming p σ-additive and diffuse is fundamental. This follows from the results of Dubins (1975), Section 2, and Heath and Sudderth (1989), Section 5. In both cases, for a suitable prior π, there is no inference q consistent with p and π. However, p is diffuse but not σ-additive in the first case, and σ-additive but not diffuse in the second. More generally, in Theorem 3.1, the σ-additivity of p_θ and the diffuseness of $p_\theta \circ T^{-1}$ are key assumptions. In fact, in these hypotheses (and if the sets $\{T = t\}$ are in $\mathscr{A}_{\mathfrak{X}}$), every subset of $\mathfrak{X} \times \Theta$ with positive probability must intersect the strips $\{T = t\} \times \Theta$ for "sufficiently many" values t. This fact enables one to have great freedom in assessing q, and this is crucial in proving Theorem 3.1. Such freedom, instead, is not available when $p_\theta(T = t) > 0$ for some θ and t. For, in that case, there is π such that $m_\pi(T = t) = \int p_\theta(T = t)\pi(d\theta) > 0$. Under such π, the set $\{T = t\}$ has a privileged status, and this determines q on $\{T = t\}$. Indeed, since q is to be constant on $\{T = t\}$, setting $\phi = I_{\{T=t\}\times B}$ in (1.1) yields

$$q_x(B) = \int_B p_\theta(T = t)\pi(d\theta)/m_\pi(T = t) \quad \text{for every } x \in \{T = t\} \text{ and } B \in \mathscr{A}_\Theta.$$

However, with the latter definition of q, (1.1) can fail for some other $\psi \neq \phi$. A further shortcoming can arise if $\{x_0\} \in \mathscr{A}_{\mathfrak{X}}$ and $m_\pi\{x_0\} > 0$ for some $x_0 \in \{T = t\}$. In fact, one is also forced to set

$$q_{x_0}(B) = \int_B p_\theta\{x_0\}\pi(d\theta)/m_\pi\{x_0\},$$

and clearly the two definitions can conflict.

The next lemma gives another version of the ideas underlying Theorem 3.1. It is quite technical, but will be useful in dealing with point estimation problems (see Example 4.4).

LEMMA 3.3. *Suppose that $\Theta = \mathbb{R}$, $\mathscr{A}_\Theta$ is the Borel σ-field,* card $\mathscr{A}_{\mathfrak{X}} =$ card $\mathbb{R}$ *and $d\colon \mathfrak{X} \to \mathbb{R}$ is an $\mathscr{A}_{\mathfrak{X}}$-measurable function. Let p be a model and π a probability on $\mathscr{P}(\Theta)$ such that*

$$\text{(3.3)} \qquad \operatorname{card}\{x \in A: \inf B < d(x) < \sup B\} = \operatorname{card}\mathbb{R},$$

whenever $A \in \mathscr{A}_{\mathfrak{X}}$, $B \in \mathscr{A}_\Theta$ and $\int_B p_\theta(A)\pi(d\theta) > 0$. Then, there are an inference q and a probability m on $\mathscr{P}(\mathfrak{X})$ such that q_x is σ-additive for every $x \in \mathfrak{X}$,

$$\text{(3.4)} \qquad d(x) = \int \theta q_x(d\theta), \qquad \int \theta^2 q_x(d\theta) < +\infty \quad \textit{for every } x \in \mathfrak{X}$$

and

$$\int\int h(x)g(\theta)p_\theta(dx)\pi(d\theta) = \int\int h(x)g(\theta)q_x(d\theta)m(dx) \tag{3.5}$$

for every $h: \mathcal{X} \to \mathbb{R}$ *and* $g: \Theta \to \mathbb{R}$ *bounded and measurable.*

PROOF. Let $\mathscr{V} = \{A \times B: A \in \mathscr{A}_{\mathcal{X}}, B \in \mathscr{A}_\Theta \text{ and } \int_B p_\theta(A)\pi(d\theta) > 0\}$. Under (3.3), Lemma 2.1 implies the existence of an injective $f: \mathscr{V} \to \mathcal{X}$ such that

$$f(A \times B) \in \{x \in A: \inf B < d(x) < \sup B\} \quad \text{for every } A \times B \in \mathscr{V}.$$

Fix $x \in \mathcal{X}$. If x is not in the range of f, take q_x as the unit mass at $d(x)$. Otherwise, if $A \times B$ is the unique element of $\mathscr{V}$ for which $x = f(A \times B)$, let q_x be σ-additive, satisfying (3.4) and such that $q_x(B) = 1$. Clearly, the existence of such q_x depends on $\inf B < d(x) < \sup B$. To prove (3.5), just apply Lemma 2.2 as in the proof of Theorem 3.1, with the only difference that $\mathscr{C}$ is now defined as the field (and not the σ-field) generated by $\{A \times B: A \in \mathscr{A}_{\mathcal{X}}, B \in \mathscr{A}_\Theta\}$. $\square$

4. Some examples. This section includes four examples. The first one does not make a direct use of Theorem 3.1, but can be useful for describing the procedure of assessing an inference q via a suitable statistic T. The other three examples, instead, are based on Theorem 3.1 and Lemma 3.3. In all four examples, $\mathscr{A}_{\mathcal{X}}$, $\mathscr{A}_\Theta$ and $\mathscr{A}_{\mathscr{T}}$ are the Borel σ-fields on $\mathcal{X}$, Θ and $\mathscr{T}$.

EXAMPLE 4.1 (James–Stein estimator). Let $\mathcal{X} = \Theta = \mathbb{R}^k$, $k \geq 3$, and let p_θ be $N(\theta, I)$, I denoting the identity matrix. Let the prior π be translation invariant; that is, $\pi(B + \theta) = \pi(B)$ for every $\theta \in \Theta$ and $B \in \mathscr{A}_\Theta$. Suppose also that, in the inferrer's opinion, the inference q should be affected by the data only through the statistic

$$T(x) = \left(1 - \frac{r(\|x\|^2)}{\|x\|^2}\right)x,$$

where r is some nondecreasing function such that $0 \leq r \leq 2(k-2)$ and $\|\cdot\|$ denotes the Euclidean norm. For instance, such an opinion can be based on decision-theoretic arguments. Indeed, T is minimax as an estimator of θ under the quadratic loss $L(\theta_1, \theta_2) = \|\theta_1 - \theta_2\|^2$ [Baranchik (1970)]. Moreover, by a well-known result of Stein (1955), the usual sufficient statistic in the classical Fisherian sense, that is, $d(x) = x$, is not admissible. In particular, an estimator dominating d is just given by T after setting $r \equiv k - 2$ [James and Stein (1961).]

By choosing T, the inferrer has partitioned $\mathcal{X}$ into subsets having the same inferential content. To assign q, the remaining step is to select a family $\{\nu_t\}$ of probabilities on $\mathscr{A}_\Theta$ and to set $q_x = \nu_t$ whenever $T(x) = t$. Moreover, the resulting q is asked to be consistent with the model p and the prior π. Theorem 3.1 implies that, actually, at least one q of this type always exists.

However, in this particular example, an interesting q can also be obtained explicitly. Define q_x to be $N(T(x), I)$. Then, q_x is centered on $T(x)$ which, in a sense, is a reasonable estimate of θ. Moreover, q is consistent with p and π. In other words, this choice of q makes T the *posterior* Bayes rule, under quadratic loss, w.r.t. a consistent inference.

We prove that q is consistent with p and π. Let q_x^* be $N(x, I)$. Then q^* is coherent with p and π in Heath and Sudderth's sense [Heath and Sudderth (1978), Example 4.1]. Hence, setting $m_\pi(A) = \int p_\theta(A)\pi(d\theta)$ for $A \in \mathscr{A}_{\mathfrak{X}}$, it suffices to show that

$$\int\int \phi(x,\theta) q_x(d\theta) m_\pi(dx) = \int\int \phi(x,\theta) q_x^*(d\theta) m_\pi(dx) \tag{4.1}$$

for every bounded measurable $\phi\colon \mathfrak{X} \times \Theta \to \mathbb{R}$. Let f be the density of an $N(0,I)$. Using that $\|x - T(x)\| \to 0$ as $\|x\| \to +\infty$, it can be verified that

$$\left| \int \phi(x,\theta) q_x(d\theta) - \int \phi(x,\theta) q_x^*(d\theta) \right|$$

$$\leq \sup|\phi| \int |f(\theta - T(x)) - f(\theta - x)|\, d\theta \to 0$$

as $\|x\| \to +\infty$. Since $m_\pi(A) = 0$ if A is compact, (4.1) follows.

EXAMPLE 4.2 (Inferences depending on the median). Let $\mathfrak{X} = \mathbb{R}^n$, $\Theta = \mathbb{R}$ and let p_θ be the probability distribution of a random sample of size n drawn from an $N(\theta, 1)$. Let π be translation invariant. In Example 4.1, we have considered a statistic T inducing a σ-field smaller than the minimal sufficient one, where "sufficient" is intended in the classical Fisherian sense. Now, we deal with a statistic T inducing a different but not smaller σ-field. Let $T(x) = \text{median}(x)$. By Theorem 3.1, there is an inference q, consistent with p and π, such that $q_{x_1} = q_{x_2}$ whenever $\text{median}(x_1) = \text{median}(x_2)$.

EXAMPLE 4.3 (Marginalization paradox). Let $T\colon \mathfrak{X} \to \mathbb{R}$ and $g\colon \Theta \to \mathbb{R}$ be measurable functions, where $\mathfrak{X}$ and Θ are Borel sets in complete separable metric spaces. Let p be a σ-additive model and q an inference. Usually, in marginalization paradoxes, q is the formal posterior of some improper prior. The marginalized model, that is, $p_\theta \circ T^{-1}$, is assumed to depend on θ only through $g(\theta)$. Likewise, the marginalized inference $q_x \circ g^{-1}$ is supposed to depend on x only through $T(x)$. A paradox is claimed in case there is no prior π_0 for $g(\theta)$ which, when combined with the marginalized model $p_\theta \circ T^{-1}$, leads to the marginalized inference $q_x \circ g^{-1}$.

A marginalization paradox, as described above, cannot occur if q is consistent with p and some π. In fact, setting $\pi_0 = \pi \circ g^{-1}$, it can be shown that $q_x \circ g^{-1}$ is consistent with $p_\theta \circ T^{-1}$ and π_0 [Sudderth (1980)]. Suppose now that $p_\theta(T = t) = 0$ for all θ and t. Then Theorem 3.1 grants that, whatever the prior π, it is always possible to avoid the paradox. Indeed, the inferrer can select an inference q such that: (a) q_x depends on x only through $T(x)$, so that, in particular, the same is true for $q_x \circ g^{-1}$; (b) q is consistent with p and π, so that the paradox does not arise.

EXAMPLE 4.4 (Point estimation). Let $\mathcal{X} = \mathbb{R}^n$, $\Theta = \mathbb{R}$ and let $d\colon \mathcal{X} \to \Theta$ be Borel measurable and such that

$$\lim_{\theta \to -\infty} p_\theta(d \geq c) = \lim_{\theta \to +\infty} p_\theta(d \leq c) = 0 \quad \text{for every real } c. \tag{4.2}$$

Thus, d is a possible estimator of θ and it is linked with the model p by condition (4.2). For many choices of p, (4.2) seems to be a reasonable request for an estimator d of θ. Suppose also that p_θ is σ-additive and diffuse for all θ and take the prior π such that $\pi(B) = 0$ for every compact B. Then, in view of Lemma 3.3, there are an inference q and a probability m on $\mathscr{P}(\mathcal{X})$ satisfying conditions (3.4) and (3.5). By (3.4), q_x has mean $d(x)$ and finite variance for every x. Hence, d is the posterior Bayes rule, under quadratic loss, w.r.t. q. To summarize:

If p and π are as above, every d satisfying (4.2) *can be seen as the posterior Bayes rule w.r.t. some suitable inference q which is linked with p and π by* (3.5). Note, however, that, in this finitely additive setting, (3.5) is a weaker condition than consistency.

To prove the existence of q and m, we check condition (3.3) of Lemma 3.3. Fix $A \in \mathscr{A}_{\mathcal{X}}$ and $B \in \mathscr{A}_\Theta$ with $\eta := \int_B p_\theta(A)\pi(d\theta) > 0$. Setting $D = \{x \in A\colon \inf B < d(x) < \sup B\}$, one has to show that $\operatorname{card} D = \operatorname{card} \mathbb{R}$. Using (4.2) and $\eta > 0$, one can find $a > 0$ such that, if $\theta \in B$ and $|\theta| > a$, then $p_\theta(A - D) < \eta/2$. Further, since π vanishes on compact sets,

$$\eta = \int_{B \cap [-a,a]^c} p_\theta(A)\pi(d\theta) \leq \sup\{p_\theta(A)\colon \theta \in B, |\theta| > a\}.$$

Hence, there is θ with $p_\theta(D) > 0$. Since p_θ is σ-additive and diffuse, Lemma 2.3 implies that $\operatorname{card} D = \operatorname{card} \mathbb{R}$.

5. Two predictive versions of Theorem 3.1. Suppose that a couple of experiments are to be performed. What is asked is to predict the outcome of one of the trials conditionally on the outcome of the other. More precisely, denoting by $\mathcal{Y}$ and $\mathcal{X}$ the respective sample spaces, the task is the assignment of a *predictive inference*, that is, a family $v := \{v_x\colon x \in \mathcal{X}\}$ of probabilities on a σ-field $\mathscr{A}_{\mathcal{Y}}$ of subsets of $\mathcal{Y}$.

Let P be a given probability on the product σ-field $\mathscr{A}_{\mathcal{X}} \otimes \mathscr{A}_{\mathcal{Y}}$. Usually, the choice of v is subjected to some "compatibility" conditions with P. In line with the definition of consistency [see also Lane and Sudderth (1984)], we ask P and v to be linked by the equation

$$\int \phi\, dP = \int\int \phi(x,y) v_x(dy) m(dx) \tag{5.1}$$

for some probability m on $\mathscr{P}(\mathfrak{X})$ and every bounded, $\mathscr{A}_{\mathfrak{X}} \otimes \mathscr{A}_{\mathscr{Y}}$-measurable ϕ: $\mathfrak{X} \times \mathscr{Y} \to \mathbb{R}$.

Let T be any statistic such that $\alpha\{x: T(x) = t\} = 0$ for all t, where $\alpha(\cdot) = P(\cdot \times \mathscr{Y})$ denotes the marginal of P on $\mathscr{A}_{\mathfrak{X}}$. Again, our purpose is showing that, under some assumptions on P, there exist a probability m and a predictive inference v satisfying (5.1) and

$$x_1, x_2 \in \mathfrak{X} \quad \text{and} \quad T(x_1) = T(x_2) \quad \Rightarrow \quad v_{x_1} = v_{x_2}. \tag{5.2}$$

The interpretation of (5.2) is the same as that given in Section 1 for (1.2), with $(\mathscr{Y}, \mathscr{A}_{\mathscr{Y}})$ in the place of $(\Theta, \mathscr{A}_{\Theta})$. Now, however, using a statistic T as a preliminary step in the choice of v seems to be particularly "natural" [Cifarelli and Regazzini (1982)].

Two distinct situations about P are considered. In the first, P is assessed in a traditional way, that is, by integrating a model w.r.t. a prior. In the second, P is directly assigned, without passing through the mediation of a parametric statistical model. Even if quite unusual, this latter case is theoretically important. Indeed, as far as the problem is a predictive one, resorting to a parametric model is merely a tool for the analysis, perhaps useful but never conceptually essential.

For the sake of simplicity, in what follows $(\mathfrak{X}, \mathscr{A}_{\mathfrak{X}})$, $(\mathscr{Y}, \mathscr{A}_{\mathscr{Y}})$ and $(\mathscr{T}, \mathscr{A}_{\mathscr{T}})$ are assumed to be standard spaces (cf. Section 2). We are now able to begin with the first case. Let

$$P(E) = \int P_\theta(E)\pi(d\theta) \quad \text{for all } E \in \mathscr{A}_{\mathfrak{X}} \otimes \mathscr{A}_{\mathscr{Y}}, \tag{5.3}$$

where π is a probability on $\mathscr{P}(\Theta)$, and, for each θ, P_θ is a σ-additive probability on $\mathscr{A}_{\mathfrak{X}} \otimes \mathscr{A}_{\mathscr{Y}}$. Moreover, let α_θ and β_θ denote the marginals of P_θ on $\mathscr{A}_{\mathfrak{X}}$ and $\mathscr{A}_{\mathscr{Y}}$, that is, $\alpha_\theta(A) = P_\theta(A \times \mathscr{Y})$ and $\beta_\theta(B) = P_\theta(\mathfrak{X} \times B)$ for $A \in \mathscr{A}_{\mathfrak{X}}$, $B \in \mathscr{A}_{\mathscr{Y}}$ and $\theta \in \Theta$. Then, after selecting some (parametric) inference q, a standard rule for defining v is

$$v_x(B) = \int \beta_\theta(B|x) q_x(d\theta) \quad \text{for all } x \in \mathfrak{X} \text{ and } B \in \mathscr{A}_{\mathscr{Y}}, \tag{5.4}$$

where $\beta_\theta(B|x)$ denotes a regular version of "the conditional probability of B given x and θ."

Even if T is sufficient for q, condition (5.2) can clearly fail for v, because of the dependence on x of $\beta_\theta(B|x)$. This is why in the next proposition it is supposed that, given θ, the coordinates x and y are conditionally independent.

COROLLARY 5.1. *Let P be as in* (5.3), *T a statistic and q a* (*parametric*) *inference. Define v as in* (5.4) *with $\beta_\theta(B)$ in place of $\beta_\theta(B|x)$ and suppose that*:

(j) *q and T satisfy* (1.2), *and q is consistent with π and the marginal model* $\{\alpha_\theta: \theta \in \Theta\}$;

(jj) *for every bounded*, $\mathscr{A}_{\mathfrak{X}} \otimes \mathscr{A}_{\mathscr{Y}}$-*measurable* ϕ, *the function* $(x, \theta) \to \int \phi(x, y)\beta_\theta(dy)$ *is* $\mathscr{A}_{\mathfrak{X}} \otimes \mathscr{A}_\Theta$-*measurable*;

(jjj) *for each* $\theta \in \Theta$, $A \in \mathscr{A}_{\mathfrak{X}}$ *and* $B \in \mathscr{A}_{\mathscr{Y}}$, $P_\theta(A \times B) = \alpha_\theta(A)\beta_\theta(B)$.

Then, v *satisfies* (5.1) *and* (5.2) *for some probability* m *on* $\mathscr{P}(\mathfrak{X})$. *Moreover*, *a sufficient condition for the existence of a* (*parametric*) *inference* q *satisfying* (j) *is*

(j*) $\operatorname{card} \mathscr{A}_\Theta \le \operatorname{card} \mathbb{R}$, *and*, *for each* $\theta \in \Theta$ *and* $t \in \mathscr{T}$, $\alpha_\theta\{x: T(x) = t\} = 0$.

PROOF. Since (1.2) holds for q, v satisfies (5.2). By (j), there is m such that (1.1) holds with α_θ in the place of p_θ. To check (5.1), fix ϕ and define $f(x, \theta) = \int \phi(x, y)\beta_\theta(dy)$. By (jjj) and Fubini's theorem, $\int \phi(x, y)P_\theta(dx, dy) = \int f(x, \theta)\alpha_\theta(dx)$. Thus,

$$\begin{aligned}
&\int\int \phi(x, y) v_x(dy) m(dx) \\
&\quad = \int\int\int \phi(x, y)\beta_\theta(dy) q_x(d\theta) m(dx) \\
&\quad = \int\int f(x, \theta) q_x(d\theta) m(dx) \\
&\quad = \int\int f(x, \theta)\alpha_\theta(dx)\pi(d\theta) \quad \text{by (jj) and consistency of } q \\
&\quad = \int\int \phi(x, y) P_\theta(dx, dy)\pi(d\theta) = \int \phi \, dP.
\end{aligned}$$

Finally, assume (j*) holds. Then, since $(\mathfrak{X}, \mathscr{A}_{\mathfrak{X}})$ and $(\mathscr{T}, \mathscr{A}_{\mathscr{T}})$ are standard spaces and α_θ is σ-additive, the existence of q is a direct consequence of Theorem 3.1. □

EXAMPLE 5.2. Let $\mathfrak{X} = \mathscr{Y} = \Theta = \mathbb{R}^k$, all equipped with the Borel σ-field. Let π be translation invariant and let P_θ be the probability distribution of a random sample of size 2 from an $N(\theta, I)$. When $k \ge 3$, by the techniques of Example 4.1, the predictive inference v could be asked to depend on x only through $T(x) = [1 - r(\|x\|^2)/\|x\|^2]x$, where r is nondecreasing and $0 \le r \le 2(k-2)$. Define q_x to be $N(T(x), I)$. As proved in Example 4.1, q is a consistent (parametric) inferece and clearly it satisfies (1.2). Using such q, v can be calculated by (5.4), where $\beta_\theta(\cdot|x)$ is taken to be $N(\theta, I)$. In particular, v_x turns out to be $N(T(x), 2I)$. By Corollary 5.1, v satisfies (5.1) and (5.2).

We close this section with the second case on P.

THEOREM 5.3. *Let P be a probability on $\mathscr{A}_{\mathfrak{X}} \otimes \mathscr{A}_{\mathscr{Y}}$, α the marginal of P on $\mathscr{A}_{\mathfrak{X}}$ and T a statistic. If α is σ-additive and $\alpha\{x: T(x) = t\} = 0$ for all $t \in \mathscr{T}$, then there are a probability m on $\mathscr{P}(\mathfrak{X})$ and a predictive inference v, satisfying* (5.1) *and* (5.2), *and such that v_x is σ-additive for every $x \in \mathfrak{X}$.*

PROOF. Let $L(E) = \{F: F \text{ atom of } \mathscr{D}, (F \times \mathscr{Y}) \cap E \neq \emptyset\}$, where $E \in \mathscr{A}_{\mathfrak{X}} \otimes \mathscr{A}_{\mathscr{Y}}$ and $\mathscr{D} = T^{-1}(\mathscr{A}_{\mathscr{T}})$. By the same argument used in the proof of Theorem 3.1, it suffices to show that $P(E) = 0$ whenever $\operatorname{card} L(E) < \operatorname{card} \mathbb{R}$. Fix $E \in \mathscr{A}_{\mathfrak{X}} \otimes \mathscr{A}_{\mathscr{Y}}$ with $\operatorname{card} L(E) < \operatorname{card} \mathbb{R}$ and set $A = \{x \in \mathfrak{X}: (x, y) \in E$ for some $y \in \mathscr{Y}\}$. Since $(\mathfrak{X}, \mathscr{A}_{\mathfrak{X}})$ and $(\mathscr{Y}, \mathscr{A}_{\mathscr{Y}})$ are standard spaces, A is an analytic set, and since α is σ-additive, this implies the existence of $A_1, A_2 \in \mathscr{A}_{\mathfrak{X}}$ such that $A_1 \subset A \subset A_2$ and $\alpha(A_1) = \alpha(A_2)$ [cf. Brown and Purves (1973), page 908]. If $\alpha(A_1) > 0$, then, by using the hypotheses on α and applying Lemma 2.3 to $(A_1, \mathscr{D} \cap A_1, \mu)$ where $\mu(\cdot) = \alpha(\cdot \cap A_1)/\alpha(A_1)$, the contradiction $\operatorname{card} L(E) \geq \operatorname{card} \mathbb{R}$ can be obtained. Hence, $\alpha(A_1) = 0$, and since $E \subset A_2 \times \mathscr{Y}$, one obtains $P(E) \leq P(A_2 \times \mathscr{Y}) = \alpha(A_2) = \alpha(A_1) = 0$. □

6. Concluding remarks. Let T be a statistic. In a classical Fisherian setting, whether or not T is sufficient depends on the model only. In a Bayesian framework, T is often said to be sufficient if, whatever the prior π, the posterior of π depends only on T. Under this definition, Bayes sufficiency is also essentially determined by the model only. In particular, if the model is dominated (and if conditional probability is intended in the usual Kolmogorovian sense), then T is classically sufficient if and only if it is Bayes sufficient [cf. Blackwell and Ramamoorthi (1982)].

In the subjective approach to probability, it is more appropriate perhaps to relate the sufficiency of T to the inference q which is actually assessed, and not to other inferences that could be assigned but are not. This is why, in this paper, T is said to be "sufficient for q" provided (1.2) holds for a particular q. In a sense, if the choice of q is a subjective act, only subject to some constraint like (1.1), the sufficiency of T is also reducible to a subjective statement.

To make the above considerations effective, one needs a general result implying that, for "many" choices of T, it is possible to view T as sufficient; that is, there is a consistent inference q which makes T sufficient. This is just Theorem 3.1. Note, however, that Theorem 3.1 is merely an existence result, not to be confused with something like a rule for assessing q.

A last note concerns the notion of conditional probability which underlies (1.1). The interested reader is referred to Heath and Sudderth (1978), Lane and Sudderth (1983), Regazzini (1987) and Berti and Rigo (1992). Here, we only make clear the connections between such a notion and the usual one. Given the model p and the prior π, define $P(C) = \int p_\theta(C_\theta)\pi(d\theta)$ for $C \in \mathscr{A}_{\mathfrak{X}} \otimes \mathscr{A}_\Theta$, where $C_\theta = \{x: (x, \theta) \in C\}$. In a standard setting, π and p_θ are σ-additive and the function $\theta \to p_\theta(A)$ is $\mathscr{A}_\Theta$-measurable for every fixed $A \in \mathscr{A}_{\mathfrak{X}}$. Then, a (standard) inference q is the marginalization on $\mathscr{A}_\Theta$ of a *regular conditional distribution* for P given the sub-σ-field $\{A \times \Theta \colon A \in \mathscr{A}_{\mathfrak{X}}\}$. Hence, q_x is σ-additive for every x, and

$$x \to q_x(B) \text{ is } \mathscr{A}_{\mathfrak{X}}\text{-measurable for every fixed } B \in \mathscr{A}_{\Theta}, \tag{6.1}$$

$$P(C) = \int q_x(C^x) m_\pi(dx) \quad \text{for all } C \in \mathscr{A}_{\mathfrak{X}} \otimes \mathscr{A}_{\Theta}, \tag{6.2}$$

where $C^x = \{\theta \colon (x, \theta) \in C\}$ and $m_\pi(A) = \int p_\theta(A)\pi(d\theta)$ for $A \in \mathscr{A}_{\mathfrak{X}}$.

Loosely speaking, a consistent inference q, that is, an inference of the type studied in this paper, is like a standard inference except q_x is allowed to be finitely additive and the measurability condition (6.1) is not requested. Since (6.1) can fail, (6.2) must assume the weaker form $P(C) = \int q_x(C^x)m(dx)$ for some extension m of m_π to $\mathscr{P}(\mathfrak{X})$. Indeed, in this weaker form, (6.2) is equivalent to (1.1).

It is because of these changes in the standard notion of conditional probability that a statement like Theorem 3.1 becomes available.

Acknowledgments. This paper benefited from some discussions with Eugenio Regazzini and Sonia Petrone.

REFERENCES

BARANCHIK, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.* **41** 642–645.

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

BERTI, P. and RIGO, P. (1992). Weak disintegrability as a form of preservation of coherence. *Journal of the Italian Statistical Society* **1** 161–181.

BERTI, P. and RIGO, P. (1994). Coherent inferences and improper priors. *Ann. Statist.* **22** 1177–1194.

BLACKWELL, D. (1955). On a class of probability spaces. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **2** 1–6. Univ. California Press, Berkeley.

BLACKWELL, D. and RAMAMOORTHI, R. V. (1982). A Bayes but not classically sufficient statistic. *Ann. Statist.* **10** 1025–1026.

BROWN, L. D. and PURVES, R. (1973). Measurable selections of extrema. *Ann. Statist.* **1** 902–912.

CIFARELLI, D. M. and REGAZZINI, E. (1982). Some considerations about mathematical statistics teaching methodology suggested by the concept of exchangeability. In *Exchangeability in Probability and Statistics* (G. Koch and F. Spizzichino, eds.) 185–205. North-Holland, Amsterdam.

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

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

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

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

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 361–379. Univ. California Press, Berkeley.

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

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

PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces*. Academic, New York.

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

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

STEIN, C. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 197–206. Univ. California Press, Berkeley.

SUDDERTH, W. D. (1980). Finitely additive priors, coherence and the marginalization paradox. *J. Roy. Statist. Soc. Ser. B* **42** 339–341.

WETZEL, N. R. (1993). Coherent inferences for multivariate data models. Ph.D. dissertation, School Statist., Univ. Minnesota, Minneapolis.

DIPARTIMENTO DI STATISTICA "G. PARENTI"
UNIVERSITÀ DI FIRENZE
VIALE MORGAGNI 59
50134 FIRENZE
ITALY