

FULLY COHERENT INFERENCE¹

BY H. D. BRUNK

Oregon State University

In a general setting in which prior distributions that may take on the value ∞ are admitted, an inference based on a posterior for a prior, μ , that is “minimally compatible” with the inference is shown to have a strong property of expectation consistency, that implies a corresponding property of coherence: A nonnegative expected payoff function for a gambler’s strategy is necessarily 0 almost everywhere (μ). In the converse direction, under appropriate regularity conditions involving continuity of the sampling distribution and of the inference, a weaker version of coherence implies that the inference is based on a posterior distribution.

1. Introduction. Dawid and Stone (1972, 1973) introduce and study the concept “expectation consistency.” Let \mathcal{X} denote the domain of a random observable X and let Θ be a set of “states of nature”; each determines a probability distribution for X . A statistician is given the sampling distribution that has, for $\theta \in \Theta$, density $p_{X|\theta}(\cdot|\theta)$ with respect to a prescribed measure, and is required to select, for each $x \in \mathcal{X}$, an inference: a probability distribution over Θ conditioned on the observed value x of X . Let the inference have density $q_{\theta|X}(\cdot|x)$ with respect to a prescribed measure.

The statistician suffers a prescribed loss $g(\theta, x)$ when the actual value θ is later determined. Loosely and informally, the inference q is *expectation consistent* if no loss function whose expectation according to q given x is 0 for every $x \in \mathcal{X}$ can have an expected value according to p given θ that is strictly positive for every $\theta \in \Theta$. Instances of such loss functions occur in connection with discussions of coherence of inference in terms of gambling systems [Heath and Sudderth (1978); Lane and Sudderth (1983); Buehler (1976); Freedman and Purves (1969); Cornfield (1969)]. One imagines a consulting statistician with a client who describes an experiment that leads to a space \mathcal{X} of possible experimental results x . The statistician helps the client describe a family of models, indexed by a parameter space Θ . Each model prescribes a probability distribution $p(\cdot|\theta)$ for the random element X of \mathcal{X} . Now the client asks the statistician to provide a system q of inference that assigns “credibilities” to subsets of the parameter space, conditioned on observed $x \in \mathcal{X}$. The client demands that the statistician be willing to test q against a gambler, as follows.

The master of ceremonies (MC) selects $\theta \in \Theta$ and chooses $x \in \mathcal{X}$ according to $p(\cdot|\theta)$.

Received August, 1986; revised April 1990.

¹Research supported by NSF Grant DMS-84-03562.

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

Key words and phrases. Coherence, expectation consistent inference

The bookie (statistician) knows Θ , \mathcal{X} and $p(\cdot|\cdot)$, and selects a strategy (inference) $q(\cdot|\cdot)$: For $A \subseteq \Theta$ and $x \in \mathcal{X}$, $q(A|x)$ is proportional to the bookie's conditional probability for A given x . For $D \subseteq \Theta \times \mathcal{X}$ and $x \in \mathcal{X}$, let D^x denote the x -section of D ; then $q(D^x|x)$ is proportional to the bookie's conditional probability for D given x ; a refinement is described in Section 2.

The gambler knows Θ , \mathcal{X} , $p(\cdot|\cdot)$ and $q(\cdot|\cdot)$, and determines a strategy.

A strategy for the gambler (GS) consists of a finite collection of simple betting systems. Using a simple betting system, the gambler selects two sets, C and D for which the gambler disagrees with the bookie's assessment of odds and sets the amount of the wager accordingly. The gambler's net gain $g(\theta, x)$ is an example of such a loss function—for the statistician—as occurs in the discussion of expectation consistency. [Detailed definitions are given in Section 3. The inspiration for this particular formulation of a simple betting system comes from Armstrong and Sudderth (1989).]

1.1. EXAMPLE. [A similar example appears in Lane and Sudderth (1983), Example 3.2.] The sampling experiment is to draw successively, with replacement, n marbles from a box that has proportion θ of red marbles and proportion $(1-\theta)$ of green marbles. We set $\Theta := [0, 1]$ and $\mathcal{X} := \{0, 1, 2, \dots, n\}$ (numbers of red marbles that will be drawn). The sampling distribution $p(\cdot|\theta)$ is the binominal distribution with parameters n and θ . The bookie considers the following inference:

$$\begin{aligned} q(\cdot|0) &\text{ assigns probability 1 to } \theta = 0, \\ q(\cdot|n) &\text{ assigns probability 1 to } \theta = 1 \end{aligned}$$

and, for $x \in \{1, 2, \dots, n - 1\}$,

$$q(\cdot|x) \text{ assigns probability 1 to } \theta = \frac{1}{2}.$$

The assignments $q(\cdot|x)$, $x \in \mathcal{X}$, do not seem reasonable, yet q is a posterior for a prior distribution that puts all its probability mass on 0 and 1, and one who knew that either the marbles are all red or they are all green could not be considered to exhibit incoherent behavior while using q . But if one is to take q seriously as an inference in this example, one must consider $0, \frac{1}{2}$ and 1 as the only possible values of θ ; that is, one should have $\Theta := \{0, \frac{1}{2}, 1\}$. In other terms, for a statistician who wishes seriously to consider q as an inference, the trouble with a prior π that assigns probability mass 1 to $\{0,1\}$ is that it is not compatible with q : For $x \in \{1, 2, \dots, n - 1\}$, $q(\{\frac{1}{2}\}|x) > 0$ and $p(x|\frac{1}{2}) > 0$, but $\pi(\{\frac{1}{2}\}) = 0$. In the present paper we shall be interested in prior-inference pairs for which the prior is compatible with the inference according to definitions given in Section 3.

1.2. HSLS coherence. Heath and Sudderth (1978) and Lane and Sudderth (1983) (HSLS) consider a general context in which (Θ, \mathcal{F}) and $(\mathcal{X}, \mathcal{G})$ are measurable spaces: \mathcal{F} is a σ -field of subsets of Θ and \mathcal{G} is a

σ -field of subsets of \mathcal{X} . Let $g(\theta, x)$ denote the payoff function, the gambler's net gain corresponding to values $x \in \mathcal{X}$ and $\theta \in \Theta$, and let $f(\theta)$ denote the expected payoff function, the gambler's expected net gain or the statistician's net loss, calculated according to the probability distribution $p(\cdot|\theta)$ (see Section 3). Then q is HSLs-coherent (given p) if for every gambler's strategy (GS)

$$\inf_{\theta} f(\theta) \leq 0.$$

Equivalently, q is HSLs-coherent provided that the only strategies for the gambler that assure $f(\theta) \geq 0$ for all θ have $\inf_{\theta} f(\theta) = 0$:

$$(HSLS) \quad f(\theta) \geq 0 \text{ for all } \theta \Rightarrow \inf_{\theta} f(\theta) = 0.$$

If there is a prior distribution on Θ that has q as posterior for it and p , then q is HSLs-coherent. In particular, the q described in Example 1.1 is HSLs-coherent.

1.3. *Cornfield coherence.* In Cornfield (1969) we find a stronger version of coherence, though Cornfield deals with finite Θ only. For him, the bookie is coherent if there is a GS such that $f(\theta) \geq 0$ for all θ , and $\exists \theta^*$ such that $f(\theta^*) > 0$.

So q is *Cornfield-coherent* if for every GS either $\exists \theta_*$ such that $f(\theta_*) < 0$, or $f(\theta) \leq 0$ for all θ ; or again, q is Cornfield-coherent provided that

$$(CC) \quad f(\theta) \geq 0 \text{ for all } \theta \Rightarrow f(\theta) \equiv 0.$$

Cornfield shows that if the bookie takes for q the posterior for a prior π that assigns a positive probability to each θ , the bookie is coherent. This is the property of full coherence as defined in a more general context in Section 3. A prior π is "minimally compatible" with an inference q if $\pi(A) > 0$ whenever $q(A|x) > 0$ for all x in a set $B \in \mathcal{S}$ such that $p(B|\theta) > 0$ for all $\theta \in A$. Thus when Θ is countable and a prior π assigns positive probability to each $\theta \in \Theta$, π is minimally compatible with every inference. In Theorem 3.5 we show that if π is minimally compatible with q and q is a posterior for π and p , then $(C - \pi) f(\theta) \geq 0$ for all $\theta \Rightarrow f(\theta) \equiv 0$ a.s. (π).

The debt owed by the present paper to investigations of Heath and Sudderth and of Lane and Sudderth is obvious. But we broaden the scope of the investigations along lines suggested by structures studied by Rényi (1970) and by Dawid and Stone (1972). On the one hand the bookie would like to know that if K is a subset of Θ and if the gambler restricts bets to subsets of $K \times \mathcal{X}$, then there is a point of K where the payoff function is nonpositive (but the bookie can have more, by using a posterior for a prior that is compatible with its posterior; see Theorem 3.5). On the other hand, we wish to extend the domain of applicability of the results by accommodating both conditional distributions over Θ , such as those studied by Rényi (1970) and, like Dawid and Stone, priors generated by possibly infinite measures. Thus a prior μ on a measurable space (Θ, \mathcal{F}) need not assign finite measure to Θ ; rather, for certain subsets K of Θ of positive, finite measure μ determines a probabil-

ity measure π_K by the prescription

$$\pi_K(A) := \frac{\mu(AK)}{\mu(K)}.$$

The gambler selects one from among a family of such sets K and restricts bets to sets in $K \times \mathcal{X}$. Of course, in many applications, Θ will itself be such a set K .

Section 2 introduces notation and terminology.

In Section 4 a weak form of coherence, “coherence for bets independent of x ,” is introduced. In appropriate contexts, it implies that inference is determined by a posterior distribution. Theorem 4.9 is closely related to Corollary 3.1 of Lane and Sudderth (1983). Corollary 4.13 is formulated in terms of densities.

Section 5 includes some supplementary remarks.

2. Notation and terminology. Our most general setting has (Θ, \mathcal{F}) and $(\mathcal{X}, \mathcal{S})$ as arbitrary measurable spaces: \mathcal{F} and \mathcal{S} are σ -fields of subsets of the parameter space Θ and the observable space \mathcal{X} , respectively. A *sampling model* p is a function $\mathcal{S} \times \Theta \rightarrow [0, 1]$ such that $p(\cdot|\theta)$ is a probability measure on \mathcal{S} for each $\theta \in \Theta$, and for each $B \in \mathcal{S}$, $p(B|\cdot): \Theta \rightarrow [0, 1]$ is \mathcal{F} -measurable. That is, p is a transition probability [Neveu (1965), Chapter 3] or kernel [Bauer (1981)]. We denote by $\mathcal{F} * \mathcal{S}$ the σ -field generated by rectangles $A \times B$ with $A \in \mathcal{F}$ and $B \in \mathcal{S}$. If φ is a function $\Theta \times \mathcal{X} \rightarrow \mathbb{R}$, its θ - and x -sections are denoted by $\varphi_\theta: \mathcal{X} \rightarrow \mathbb{R}$ and $\varphi^x: \Theta \rightarrow \mathbb{R}$, defined by $\varphi_\theta(x) := \varphi^x(\theta) := \varphi(\theta, x)$. (The symbol $:=$ indicates that the left member is *defined* to be equal to the right member.) We follow de Finetti in using $p(\cdot|\theta)$ to denote not only the measure on \mathcal{S} , but also the expectation functional it determines:

$$p(\varphi|\theta) := \int \varphi_\theta(x) p(dx|\theta);$$

similarly for other measures. Also, the same symbol will be used for a set and its indicator. For example, if $A \in \mathcal{F} * \mathcal{S}$, then $A_\theta(x) = A^x(\theta) = A(\theta, x) = 1$ if $(\theta, x) \in A$; otherwise each is 0.

We shall interpret “inference” somewhat more broadly than is usual. As a technical term, an inference here is a measure used to generate conditional inferences, in the same way that a measure determines a Rényi system of conditional probabilities. Thus an *inference* q is also a transition measure or kernel: It is a function $\mathcal{F} \times \mathcal{X} \rightarrow [0, \infty]$ such that $q(A|\cdot): \mathcal{X} \rightarrow [0, \infty]$ is \mathcal{S} -measurable for each $A \in \mathcal{F}$, $q(\cdot|x)$ is a measure on \mathcal{F} for each $x \in \mathcal{X}$ and these measures are uniformly σ -finite. For $x \in \mathcal{X}$, $A \in \mathcal{F}$ and $K \in \mathcal{F}$ such that $0 < q(K|x) < \infty$ we interpret

$$q_K(A|x) := \frac{q(AK|x)}{q(K|x)}$$

as the “credibility” that the inference q associates with A , subject to $\theta \in K$. [Note that to multiply q by a positive, measurable function $\mathcal{X} \rightarrow \mathbb{R}$ does not affect the ratio $q(AK|x)/q(K|x)$.] If $0 < q(\Theta|x) < \infty$ for $x \in \mathcal{X}$, then $q(\cdot|x)/q(\Theta|x)$ is an inference in the sense in which the term is used in Heath and Sudderth (1978) and in Lane and Sudderth (1983).

We shall refer to a measure π on \mathcal{F} as *supported* on a set $K \in \mathcal{F}$ provided that $\pi(A) = 0$ when $A \in \mathcal{F}$ and $AK = \emptyset$.

For $K \in \mathcal{F}$, a probability measure π_K supported on K and a sampling model p determine a probability measure r_K on $\mathcal{F} * \mathcal{G}$, supported on $K \times \mathcal{X}$, such that for $A \in \mathcal{F}$ and $B \in \mathcal{G}$, we have

$$r_K(A \times B) = \int_{KA} p(B|\theta)\pi_K(d\theta).$$

The probability π_K is one of the marginals of r_K , the other being m_K , defined for $B \in \mathcal{G}$ by

$$m_K(B) := r_K(\Theta \times B) = r_K(K \times B).$$

A σ -finite measure μ on \mathcal{F} and a sampling model p determine a σ -finite measure r on $\mathcal{F} * \mathcal{G}$ [Ash (1972), Theorem 2.6.2] such that for $A \in \mathcal{F}$ and $B \in \mathcal{G}$, we have $r(A \times B) = \int_A p(B|\theta)\mu(d\theta)$. The measure $\mu [= r(\cdot \times \mathcal{X})]$ is one of the marginals of r , the other being m , defined for $B \in \mathcal{G}$ by $m(B) := r(\Theta \times B)$.

2.1. DEFINITIONS. Let q be an inference, let $K \in \mathcal{F}$ be such that $q(K|x) < \infty$ for $x \in \mathcal{X}$ and let π_K be a probability measure supported on K . Define the conditional inference $q_K: \mathcal{F} \times \mathcal{X} \rightarrow \mathbb{R}$ by

$$q_K(C|x) := \begin{cases} q(CK|x)/q(K|x), & \text{if } q(K|x) > 0, \\ 0, & \text{if } q(K|x) = 0 \end{cases}$$

[in which case also $q(CK|x) = 0$ for $C \in \mathcal{F}$]. Then q is a *posterior* for π_K and p provided that for $A \in \mathcal{F}$ and $B \in \mathcal{G}$,

$$\begin{aligned} r_K(A \times B) &:= \int_{\Theta} \int_{\mathcal{X}} K(\theta) A(\theta) B(x) p(dx|\theta) \pi_K(d\theta) \\ (2.2) \qquad &= \int_K \int_{\mathcal{X}} B(x) q_K(A|x) p(dx|\theta) \pi_K(d\theta) \\ &= \int_{\mathcal{X}} B(x) q_K(A|x) m_K(dx). \end{aligned}$$

It follows that for every bounded \mathcal{G} -measurable function $\psi: \mathcal{X} \rightarrow \mathbb{R}$ and every bounded $\mathcal{F} * \mathcal{G}$ -measurable function φ with support on $K \times \mathcal{X}$,

$$(2.3) \qquad \int_K \int_{\mathcal{X}} \psi(x) [\varphi(\theta, x) - q_K(\varphi^x|x)] p(dx|\theta) \pi_K(d\theta) = 0.$$

One verifies that if $K' \in \mathcal{F}$, $K' \subseteq K$, and if q is a posterior for p and π_K , it is

also a posterior for the normalized restriction π' of π_K to K' : For $A \in \mathcal{F}$, $\pi'(A) := \pi_K(AK')/\pi_K(K')$.

In Section 4 both Θ and \mathcal{X} will be locally compact spaces; \mathcal{F} , \mathcal{G} and $\mathcal{F} * \mathcal{G}$ are then taken to be the classes of Baire sets. Let K be a compact subset of Θ . One can use the Riesz representation theorem to show that then if \mathcal{X} is σ -compact, a sufficient condition that q be a K -posterior for π_K and p is that

$$(2.3') \quad \int_K \int_{\mathcal{X}} \psi(x)[h(\theta) - q_K(h|x)]p(dx|\theta)\pi_K(d\theta) = 0$$

for continuous functions $\psi: \mathcal{X} \rightarrow \mathbb{R}$ with compact support and for functions $h: \Theta \rightarrow \mathbb{R}$ with continuous restriction to K , where for $A \in \mathcal{F}$, $\pi_K(A) := \mu(AK)/\mu(K)$.

2.4. DEFINITIONS. Let μ be a σ -finite measure on \mathcal{F} and let \mathcal{K} be a family of members of \mathcal{F} such that $0 < \mu(K) < \infty$ for $K \in \mathcal{K}$. Then q is a \mathcal{K} -posterior for μ and p provided that for every $K \in \mathcal{K}$, q_K is a K -posterior for π_K and p , where $\pi_K: \mathcal{F} \rightarrow \mathbb{R}$ is defined by

$$\pi_K(C) := \frac{\mu(KC)}{\mu(K)}$$

for $C \in \mathcal{F}$. If μ is a σ -finite measure on \mathcal{F} , q is a posterior for μ and p if it is a \mathcal{K}' -posterior, where \mathcal{K}' is the class of all sets K in \mathcal{F} such that $0 < \mu(K) < \infty$.

2.5. PROPOSITION. Let $(\Theta, \mathcal{F}, \lambda)$ and $(\mathcal{X}, \mathcal{G}, \nu)$ be measure spaces. Let μ be a measure on \mathcal{F} that is absolutely continuous with respect to λ and set $h := d\mu/d\lambda$. For $\theta \in \Theta$, let $p(\cdot|\theta)$ have positive bounded density $\rho(\cdot|\theta)$ with respect to ν and for $x \in \mathcal{X}$ let $q(\cdot|x)$ have density $\sigma(\cdot|x)$ with respect to λ given by

$$(2.6) \quad \sigma(\theta|x) = j(x)h(\theta)\rho(x|\cdot)$$

where j is a positive, bounded, measurable function $\mathcal{X} \rightarrow \mathbb{R}$. Then q is a posterior for μ and p . If $K \in \mathcal{F}$ and $0 < \mu(K) < \infty$, then $0 < q(K|x) < \infty$ for $x \in \mathcal{X}$.

The proof is straightforward.

3. Expectation consistency and coherence. Let (Θ, \mathcal{F}) and $(\mathcal{X}, \mathcal{G})$ be measurable spaces, p a sampling model and q an inference, as defined in Section 2. For $K \in \mathcal{F}$ such that $q(K|x) < \infty$ for $x \in \mathcal{X}$, a simple K -betting system is a triple (C, D, s) , where $\{C, D\} \subseteq \mathcal{F} * \mathcal{G}$, $C^x \cup D^x \subseteq K$ for $x \in \mathcal{X}$ and $s: \mathcal{X} \rightarrow \mathbb{R}$ is bounded and \mathcal{G} -measurable. One thinks of C and D as sets such that for at least some $x \in \mathcal{X}$ the gambler disagrees with the bookie's assessment of odds $q(C^x|x):q(D^x|x)$, $s(x)$ is the amount of the wager (positive, negative or 0) set by the gambler, who makes a conditional payment of $s(x)q_K(C^x|x)$ for the privilege of playing. If it develops that $(\theta, x) \notin D$, the

conditional payment is returned, while if $(\theta, x) \in C$, the gambler is paid $s(x)q_K(D^x|x)$. The gambler's net gain when θ is announced will then be

$$(3.1) \quad g(\theta, x) := s(x)[C(\theta, x)q_K(D^x|x) - D(\theta, x)q_K(C^x|x)].$$

Note that if $\theta \notin K$, then $C(\theta, x) = D(\theta, x) = 0$. If for some $x \in \mathcal{X}$ the gambler's probabilities for C^x and D^x are, respectively, $P(C^x|x)$ and $P(D^x|x) > 0$, the expected gain for the gambler, given x , is the product of $s(x)q_K(D^x|x)P(D^x|x)$ by

$$\frac{P(C^x|x)}{P(D^x|x)} - \frac{q_K(C^x|x)}{q_K(D^x|x)}.$$

So when the gambler thinks the bookie has prescribed unrealistically small odds $q_K(C^x|x)/q_K(D^x|x)$, $s(x)$ will be assigned a positive value and $s(x) < 0$ for those x for which the gambler considers $q_K(C^x|x)/q_K(D^x|x)$ too large. A finite collection of simple K -betting systems $\{(C_i, D_i, s_i), i = 1, 2, \dots, n\}$ is a K -strategy for the gambler. Let g_i be the gambler's net gain from the i th simple K -betting system and let f be the gambler's expected net gain given θ :

$$f(\theta) := p(g_\theta|\theta)$$

where $g(\theta, x) := \sum_{i=1}^n g_i(\theta, x)$. The function g may be regarded as an instance of such a loss—for the statistician—as plays a role in the discussion by Dawid and Stone (1972, 1973) of expectation consistency. Loosely and informally, expectation consistency excludes the possibility that expected loss according to q given x be 0 for each $x \in \mathcal{X}$, while expected loss according to p given θ is positive for all θ .

In the special case considered by Dawid and Stone (1972, 1973), θ and \mathcal{X} are countable and \mathcal{K} is the family of finite subsets of Θ .

The definitions of full coherence and full expectation consistency to follow make use of the concept of compatibility. Let $K \in \mathcal{F}$. Suppose there are sets $A \in \mathcal{F}$ and $B \in \mathcal{G}$ such that $q(AK|x) > 0$ [equivalently, $q_K(A|x) > 0$] for $x \in B$ and $p(B|\theta) > 0$ for $\theta \in A$. One would not consider using such an inference q in the presence of p unless one felt that it was stochastically possible that $(\theta, x) \in A \times B$. But if π is a probability on \mathcal{F} , supported on K , with $\pi(A) = 0$, and r is the probability distribution on $\mathcal{F} * \mathcal{G}$ determined by p and π , then $r(A \times B) = 0$, so that one would not consider such a probability π to reflect adequately opinions of a bookie using q . There are other possible compatibility conditions that could be of interest, so we refer to the present condition as *minimal compatibility*.

3.2. DEFINITION. Let $K \in \mathcal{F}$ and let π be a probability measure on \mathcal{F} , supported on K . Then π is *minimally K -compatible with q* (in the presence of p) provided that $\pi(A) > 0$ for every $A \in \mathcal{F}$ for which there exists $B \in \mathcal{G}$ such that $p(B|\theta) > 0$ for $\theta \in A$ and $q(AK|x) > 0$ for all $x \in B$. If $\mathcal{K} \subseteq \mathcal{F}$ and if μ is a σ -finite measure on \mathcal{F} such that $0 < \mu(K) < \infty$ for $K \in \mathcal{K}$, then μ is *minimally \mathcal{K} -compatible with q* provided that $\pi_K := \mu/\mu(K)$ is minimally K -compatible with q for every $K \in \mathcal{K}$.

Where densities are available as in Proposition 2.5, a prior π is minimally compatible with a posterior whose conditional density σ is given by (2.6).

Note that in Example 1.1, if $K := \Theta$ and π concentrates its probability on $\{0, 1\}$, then q is a posterior for π and p ; but π is *not* minimally K -compatible with q .

Loosely, q is *fully K -expectation consistent* with p if, when expected loss according to q given x is 0 for each $x \in \mathcal{X}$, a nonnegative expected loss according to p given θ is necessarily 0 almost surely, with respect to some probability supported on K that is *minimally compatible with q* .

3.3. DEFINITION. Let $\mathcal{K} \subseteq \mathcal{F}$. An inference q is *fully \mathcal{K} -expectation consistent* with p if for every $K \in \mathcal{K}$ there is a probability π supported on K and minimally K -compatible with q , such that for every bounded $\mathcal{F}^* \mathcal{G}$ -measurable g with the properties

$$(i) \quad q(g^x|x) = 0 \quad \text{for all } x \in \mathcal{X}$$

and

$$(ii) \quad f(\theta) := p(g_\theta|\theta) \geq 0 \quad \text{for all } \theta \in K,$$

one has $f(\theta) = 0$ for π -almost all θ .

Note that for a gambler's K -strategy the bookie's net loss g satisfies (i) above. An inference q is *fully \mathcal{K} -coherent* if for every $K \in \mathcal{K}$ there is a probability π supported on K and minimally K -compatible with q , such that for every gambler's K -strategy for which (ii) holds, one has $f(\theta) = 0$ for π -almost all θ .

The following proposition is a consequence of (2.2), (2.3) and Definition 3.2.

3.4. PROPOSITION. Let $K \in \mathcal{F}$, $K_1 \in \mathcal{F}$, $K_1 \subseteq K$ and let q be an inference that is a K -posterior for p and a K -prior π such that $0 < q(K_1|x) < \infty$ for $x \in \mathcal{X}$. Then q is also a K_1 -posterior for the normalized restriction π_1 of π to K_1 : For $A \in \mathcal{F}$,

$$\pi_1(A) := \frac{\pi(AK_1)}{\pi(K_1)}.$$

If π is minimally K -compatible with q , then π_1 is minimally K_1 -compatible with q .

Note that (2.2) with $A = K_1$ and $B = \mathcal{X}$ implies that $\pi(K_1) > 0$, since

$$\int_{\mathcal{X}} \frac{q(K_1|x)}{q(K|x)} p(dx|\theta) > 0 \quad \text{for all } \theta \in \Theta.$$

3.5. THEOREM. Let $\mathcal{K} \subseteq \mathcal{F}$ and let q be a \mathcal{K} -posterior for p and a σ -finite measure μ on \mathcal{F} such that $0 < \mu(K) < \infty$ for $K \in \mathcal{K}$; and let μ be minimally

\mathcal{K} -compatible with q . Then q is fully \mathcal{K} -expectation consistent with p . A fortiori, q is fully \mathcal{K} -coherent.

It follows from Proposition 3.4 that then q is fully K -expectation consistent with p for every $K \in \mathcal{F}$ that is part of some member of \mathcal{K} , such that $0 < q(K|x) < \infty$ for $x \in \mathcal{X}$.

PROOF OF THEOREM 3.5. Let $K \in \mathcal{K}$ and let g be bounded and measurable and satisfy (i) and (ii) of Definition 3.3. For $\theta \in \Theta$, define $f(\theta) := p(g_\theta|\theta)$. Since q is a \mathcal{K} -posterior for p and μ , it follows from (2.3) that

$$\begin{aligned} \int_K f(\theta)\pi_K(d\theta) &= \int_K \int_{\mathcal{X}} g(\theta, x)p(dx|\theta)\pi_K(d\theta) \\ &= \int_K \int_{\mathcal{X}} q_K(g^x|x)p(dx|\theta)\pi_K(d\theta) = 0, \end{aligned}$$

where $\pi_K(A) := \mu(AK)/\mu(K)$ for $A \in \mathcal{F}$. Thus $f \geq 0$ on K implies $f = 0$ a.e. (π_K). By hypothesis, π_K is minimally K -compatible with q , so that q is fully \mathcal{K} -expectation consistent with p . \square

In Example 1.1, there is no family \mathcal{K} that includes a set $K \in \mathcal{F}$ that contains $\{0, \frac{1}{2}, 1\}$, such that q is fully \mathcal{K} -expectation consistent with p . To see this, note first that if $\{0, \frac{1}{2}, 1\} \subseteq K$, a probability supported on K that is minimally K -compatible with q must assign positive probability to each of the singletons $\{0\}$, $\{\frac{1}{2}\}$ and $\{1\}$. On the other hand, suppose $\{0, \frac{1}{2}, 1\} \subseteq K \in \mathcal{F}$. Then for $x \in \mathcal{X}$, $q_K(\cdot|x) = q(\cdot|x)$ since $q(K|x) = 1$ for $x \in \mathcal{X}$. Set $C := \{\frac{1}{2}, 0\}$, $D := K \times \mathcal{X}$, $s(0) := 1$ and $s(x) := 0$ for $x \neq 0$. Then $q(g^x|x) = 0$ for $x \in \mathcal{X}$,

$$\begin{aligned} g(\theta, x) &= s(x)[C(\theta, x)q(D^x|x) - D(\theta, x)q(C^x|x)], \\ g(\tfrac{1}{2}, 0) &= q(K|0) \geq q(\{0\}|0) = 1, \\ g(\theta, 0) &= 0 \quad \text{for } \theta \neq \tfrac{1}{2}, \\ g(\theta, x) &= 0 \quad \text{for } x > 0, \theta \in \Theta. \end{aligned}$$

So $f(\theta) := p(g_\theta|\theta) = 0$ if $\theta \neq \frac{1}{2}$, but $f(\frac{1}{2}) = (\frac{1}{2})^n$. Since $f(\theta) \geq 0$ for $\theta \in \Theta$ and since, if π is minimally K -compatible with q , $\pi(\{\frac{1}{2}\})$ must be positive, there is no probability π minimally K -compatible with q , such that $f(\theta) = 0$ for π -almost all θ . That is, q is not fully \mathcal{K} -expectation consistent with p , no matter what \mathcal{K} may be, so long as it has a member K that includes $\{0, \frac{1}{2}, 1\}$. In particular, if \mathcal{K} has the single member $\Theta := [0, 1]$, q is not fully \mathcal{K} -expectation consistent.

3.6. EXAMPLE. Let $\Theta := (0, 1)$ and, as in Example 1.1, let n be a positive integer. Set $\mathcal{X} := \{0, 1, \dots, n\}$ and for $\theta \in \Theta$, let $p(\cdot|\theta)$ be binomial with parameters n and θ . Θ is locally compact and σ -compact in the usual topology of the reals and \mathcal{X} is compact in the discrete topology. For $\theta \in \Theta$, $p(\cdot|\theta)$ has

density

$$\rho(x|\theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

with respect to counting measure ν on \mathcal{X} . Let λ denote Lebesgue measure and let $\mathcal{K} := \{K_1, K_2, \dots\}$ where $K_1 \subseteq K_2 \subseteq \dots$, $\bigcup_{n=1}^\infty K_n = \Theta$, $\lambda(K_1) > 0$, and where for $n = 1, 2, \dots$, K_n is a compact subset of $(0, 1)$. Let μ have density with respect to λ on $(0, 1)$ given by

$$\frac{d\mu}{d\lambda}(\theta) := \frac{1}{\theta(1 - \theta)}, \quad 0 < \theta < 1.$$

For $x \in \mathcal{X}$, let $q(\cdot|x)$ have density $\sigma(\cdot|x)$ with respect to λ , given by

$$\sigma(\theta|x) := \theta^{x-1} (1 - \theta)^{n-x-1}, \quad 0 < \theta < 1,$$

so that q is a \mathcal{K} -posterior for μ and p . Since μ and q have positive densities with respect to λ , μ is minimally \mathcal{K} -compatible with q . By Theorem 3.5, q is fully and strongly \mathcal{K} -expectation consistent. In particular, if f is a nonnegative expected payoff function for a K -strategy with $K \in \mathcal{K}$, then $f = 0$ a.e. (μ) and hence a.e. (λ) (cf. Example 4.14).

Lane and Sudderth (1983) observe that q is not Θ -coherent; of course Θ is not contained in K_n for any n .

3.7. EXAMPLE. Set $\mathcal{X} := \Theta := \mathbb{R}$ and let each of \mathcal{F} and \mathcal{S} be the class of Borel subsets of \mathbb{R} . For $B \in \mathcal{S}$ and $\theta \in \mathbb{R}$ define

$$p(B|\theta) := \frac{1}{\sqrt{2\pi}} \int_B \exp\left[-(x - \theta)^2/2\right] dx.$$

Let μ denote Lebesgue measure on \mathbb{R} , and for $A \in \mathcal{F}$ and $x \in \mathbb{R}$ set

$$q(A|x) := \frac{1}{\sqrt{2\pi}} \int_A \exp\left[-(\theta - x)^2/2\right] dx.$$

Let \mathcal{K} denote the family of sets $K \in \mathcal{F}$ such that $0 < \mu(K) < \infty$. Lane and Sudderth (1983) remark that q is coherent, but not derivable from a proper countably additive prior on $\Theta := \mathbb{R}$. But q is a \mathcal{K} -posterior for μ and p , and m is minimally \mathcal{K} -compatible with q . For arbitrary $K \in \mathcal{K}$, if a K -strategy for the gambler produces a nonnegative expected payoff function, it is 0 a.e. (μ).

4. Expectation consistency implies Bayes. In Section 3 we gave conditions under which a Bayes inference is fully expectation consistent and coherent. In the present section we obtain results in the converse direction for cases in which Θ and \mathcal{X} are locally compact spaces. The σ -fields \mathcal{F} , \mathcal{S} and $\mathcal{F} * \mathcal{S}$ are the classes of Baire sets. The key tool is Lemma 4.1, that follows. It is essentially Lemma 1 of Heath and Sudderth (1978) [cf. also Heath and Sudderth (1972); Pierce (1973)].

4.1. LEMMA. Let S be a set and $L(S)$ the normed linear space of all bounded functions $S \rightarrow \mathbb{R}$, with the supremum norm. Let L be a subspace of $L(S)$ that contains the constant function $\mathbf{1}$ [defined by $1(s) := 1$ for $s \in S$]. Let $F \subseteq L$. Then the following are equivalent:

(i) There is a positive linear functional l on L such that $l(f) \geq 0$ for $f \in F$ and $l(\mathbf{1}) > 0$.

(ii) Every nonnegative combination of functions in F has a nonnegative supremum.

PROOF. To prove that (i) implies (ii), suppose the contrary, that there exist a positive integer n , a positive number ε and, for $i = 1, 2, \dots, n$, a positive number a_i and a function f_i in F such that

$$\sum_{i=1}^n a_i f_i(s) < -\varepsilon \quad \text{for all } s \in S.$$

Then

$$l\left(\sum_{i=1}^n a_i f_i(s)\right) \leq l(-\varepsilon \mathbf{1}(s)) = -\varepsilon l(\mathbf{1}) < 0,$$

a contradiction. The proof that (ii) implies (i) is based on the separating hyperplane theorem [cf., e.g., Pryce (1973), Theorem 8.14, page 119]. \square

4.2. COROLLARY. Let Θ be locally compact and let $K \subseteq \Theta$ be compact. Let F be a vector space of bounded Baire functions $\Theta \rightarrow \mathbb{R}$ whose restrictions to K have nonnegative suprema. Then there is a Baire probability measure π_K , supported on K , such that $\pi_K(hK) = 0$ for all Baire $h \in F$ that have continuous restriction $h|_K$ to K .

PROOF. This follows from Lemma 4.1 and the Riesz representation theorem. \square

Let \mathcal{K} be a family of compact subsets of the locally compact space Θ such that $q(K|x) < \infty$ for $x \in \mathcal{X}$ and $K \in \mathcal{K}$. For $K \in \mathcal{K}$, for $x \in \mathcal{X}$ and for bounded Baire $\varphi: \Theta \times \mathcal{X} \rightarrow \mathbb{R}$, $q_K(\varphi|x)$ is defined by (cf. Definition 2.1)

$$q_K(\varphi|x) := \begin{cases} q(\varphi^x K|x) / q(K|x), & \text{if } q(K|x) > 0, \\ 0, & \text{if } q(K|x) = 0. \end{cases}$$

4.3. DEFINITIONS. For fixed $K \in \mathcal{K}$, the sampling model p is K -continuous if for bounded continuous $h: \mathcal{X} \rightarrow \mathbb{R}$ the restriction to K of $p(h|\cdot): \Theta \rightarrow \mathbb{R}$ is continuous. The inference q is K -continuous if for bounded Baire $h: \Theta \rightarrow \mathbb{R}$ with continuous restriction to K , $q_K(h|\cdot): \mathcal{X} \rightarrow \mathbb{R}$ is continuous.

For given p and q and a given compact subset K of Θ , let $\mathcal{H} = \mathcal{H}(p, q, K)$ denote the class of functions $\Theta \rightarrow \mathbb{R}$ of form

$$(4.4) \quad K(\theta) \int_{\mathcal{X}} s(x) [h(\theta) - q_K(h|x)] p(dx|\theta),$$

where $s: \mathcal{X} \rightarrow \mathbb{R}$ is a continuous function with compact support and $H: \Theta \rightarrow \mathbb{R}$ is a bounded, nonnegative Baire function with continuous restriction to K .

4.5. LEMMA. *Let K be a compact subset of Θ , let p be a K -continuous sampling model and let q be a K -continuous inference. If every finite linear combination of function in $\mathcal{H}(p, q, K)$ has nonpositive infimum in K , then there is a probability distribution π_K on \mathcal{F} , supported on K , such that q_K is a posterior for p and π_K .*

PROOF. Let F denote the vector space of finite linear combinations of functions in $\mathcal{H}(p, q, K)$. By the remark preceding Definition 2.4, it suffices to show that there is a probability measure π_K supported on K such that $\int_K f(\theta) \pi_K(d\theta) = 0$ for every function $f \in F$. By hypothesis, for $f \in F$ we have $\inf\{f(\theta): \theta \in K\} \leq 0$. Since $f \in F$ implies $-f \in F$, also $\sup\{f(\theta): \theta \in K\} \geq 0$. By Corollary 4.2 there is a probability measure π_K on \mathcal{F} , supported on K , such that $\pi_K(jK) = 0$ for every function $j \in F$ whose restriction to K is continuous. But since p and q are K -continuous, every function in F has continuous restriction to K . Thus $\int_K f(\theta) \pi_K(d\theta) = 0$ for $f \in F$, completing the proof. \square

In Section 3 we gave conditions under which inference based on a posterior distribution has a rather strong property of expectation consistency. That property implies the analogously defined property of coherence. In the converse direction, under appropriate regularity conditions, an apparently weaker form of coherence of expectation consistency is seen to imply that the inference is based on a posterior distribution. Let \mathcal{K} be a family of sets belonging to \mathcal{F} . For a positive integer n and a set $K \in \mathcal{K}$, we consider a K -strategy $\{(C_i, D_i, s_i), i = 1, 2, \dots, n\}$ (Section 3) such that, for $i = 1, 2, \dots, n$, $D_i := K$ and C_i is independent of x : for some $C'_i \in \mathcal{F}$, $C_i(\theta, x) := C'_i(\theta)$ for $\theta \in \Theta$, $x \in \mathcal{X}$. For such a strategy the bookie's loss from the i th simple bet independent of x is

$$(4.6) \quad g_i(\theta, x) = K(\theta) s_i(x) [C'_i(\theta) - q_K(C'_i|x)],$$

the total loss from the K -strategy is

$$(4.7) \quad g(\theta, x) = \sum_{i=1}^n g_i(\theta, x)$$

and the expected loss given θ is

$$f(\theta) := p(g_\theta|\theta).$$

This motivates the following definition.

4.8. DEFINITION. An inference q is *weakly K -expectation consistent with p* if the validity of (4.6) for every bounded \mathcal{L} -measurable $s_i(\cdot)$ and every $C_i' \in \mathcal{F}$, $i = 1, 2, \dots, n$, implies

$$\inf\{f(\theta) : \theta \in K\} \leq 0,$$

and *weakly \mathcal{K} -expectation consistent with p* if weakly K -expectation consistent with p for every $K \in \mathcal{K}$. Such an inference may also be termed *weakly K -coherent* (*\mathcal{K} -coherent*) for bets independent of x .

For coherence as defined in Heath and Sudderth (1978) (with $K := \Theta$) the sets C_i are allowed to depend on x , so that weak Θ -coherence is at least apparently weaker than coherence, since the class of available loss functions g is smaller for weak Θ -coherence.

4.9. THEOREM. Let Θ and \mathcal{X} be locally compact and σ -compact. Let K be a compact subset of Θ and let p and q be K -continuous. If q is weakly K -expectation consistent, then q is a K -posterior for p and a K -prior π_K .

PROOF. It suffices to show that q_K satisfies the hypothesis of Lemma 4.5. A finite linear combination of functions in $\mathcal{H}(p, q, K)$ is a sum of such functions. Approximating the functions h that appear in (4.4) by nonnegative simple functions, one finds that the linear combination is the expected payoff function for a gambler's K -strategy of bets independent of x . Since q is weakly K -coherent for bets independent of x , the infimum is nonpositive. \square

Note that on setting $K = \Theta$ we have a conclusion of Corollary 3.1 of Lane and Sudderth (1983) with the weaker coherence hypothesis.

For most stochastic models in common use, given a sampling distribution p and an inference q , there is at most one prior π such that q is a posterior for π and p . An obvious counterexample has

π : an arbitrary probability measure on \mathcal{F} ;

for $\theta \in \Theta$, $p(\cdot|\theta)$ is the probability measure degenerate at θ : $p(B|\theta) = B(\theta)$ for $B \in \mathcal{L}$;

for $x \in \mathcal{X}$, $q(\cdot|x)$ is the probability measure degenerate at x : $q(A|x) = A(x)$ for $A \in \mathcal{F}$.

(The author is indebted to D. L. Hanson for a more interesting example involving disjoint rectangles in $\Theta \times \mathcal{X}$). But it is an essential hypothesis of Proposition 4.11 that follows, that p and q fail to admit multiple K -priors, according to the following definition.

4.10. DEFINITION. For $K \in \mathcal{F}$, a sampling distribution, p and an inference q fail to admit multiple K -priors provided that there is at most one K -prior π_K such that q_K is a K -posterior for π_K and p .

The following lemma gives sufficient conditions in order that p and q fail to admit multiple marginals.

4.11. LEMMA. *Let $(\Theta, \mathcal{F}, \lambda)$, $(\mathcal{X}, \mathcal{G}, \nu)$ and $(\Theta \times \mathcal{X}, \mathcal{F} * \mathcal{G}, \lambda \times \nu)$ be measure spaces. Let a joint probability distribution on $\Theta \times \mathcal{X}$ have marginal densities $\delta^\pi: \Theta \rightarrow \mathbb{R}$ with respect to λ and $\delta^m: \mathcal{X} \rightarrow \mathbb{R}$ with respect to ν , conditional density $\rho(\cdot|\theta): \mathcal{X} \rightarrow \mathbb{R}$ with respect to ν for $\theta \in \Theta$, and conditional density $\sigma(\cdot|x): \Theta \rightarrow \mathbb{R}$ with respect to λ for $x \in \mathcal{X}$. Set $\Theta^+ := \{\theta \in \Theta | \delta^\pi(\theta) > 0\}$. If $\rho(x|\theta) > 0$ for $(\theta, x) \in \Theta^+ \times \mathcal{X}$, then there is (to within λ -equivalence) no other marginal density $\delta_1^\pi: \Theta \rightarrow \mathbb{R}$ with respect to λ such that ρ and σ are conditional densities. That is, such a marginal density δ_1^π must be a version of δ^π .*

The proof is given in the appendix.

The term “bunch” in the statement of Proposition 4.12 is used in Rényi’s (1970) sense: a *bunch* of subsets of Θ is a class closed under union, to which the empty set does not belong, such that Θ is a countable union of sets in the class.

4.12. PROPOSITION. *Let Θ and \mathcal{X} be locally compact and σ -compact. Let \mathcal{K} be a bunch of compact subsets of Θ and let (K_n) be an expanding sequence of sets in \mathcal{K} such that $\Theta = \bigcup_{n=1}^\infty K_n$. For $K \in \mathcal{K}$ let q satisfy $0 < q(K|x) < \infty$ for $x \in \mathcal{X}$, let p and q be K -continuous and fail to admit multiple priors and let q be weakly \mathcal{K} -expectation consistent. Then there is a \mathcal{K} -prior μ such that q is a \mathcal{K} -posterior for μ and p .*

PROOF. By Theorem 4.9, for each $K \in \mathcal{K}$ there is a K -prior π_K such that q is a K -posterior for p and π_K . Let $\{K, K'\} \subseteq \mathcal{K}$, $K \subseteq K'$. By Proposition 3.4, q is a K -posterior for p and the normalized restriction of $\pi_{K'}$ to K . Since p and q do not admit multiple marginals, π_K is itself the normalized restriction to K of $\pi_{K'}$: for $A \in \mathcal{F}$, $\pi_K(A) = \pi_{K'}(AK)/\pi_{K'}(K)$. The system $P(\cdot|\cdot): \mathcal{F} \times \mathcal{K}, (A, K) \mapsto P(A|K) := \pi_K(A)$ is a conditional probability [Rényi (1970)]. By Theorem 2.2.1 of Rényi there is a σ -finite measure μ on \mathcal{F} such that $0 < \mu(K) < \infty$ and $\pi_K(A) = P(A|K) = \mu(AK)/\mu(K)$ for $A \in \mathcal{F}$, $K \in \mathcal{K}$. The measure μ is determined uniquely up to a constant factor, and q is a \mathcal{K} -posterior for p and μ . \square

Corollary 4.13 gives regularity conditions which imply that a weakly expectation consistent inference is a posterior.

4.13. COROLLARY. *Let Θ and \mathcal{X} be locally compact and σ -compact. Let \mathcal{K} be a bunch of compact subsets of Θ . For $\theta \in \Theta$ let $p(\cdot|\theta)$ have positive density $\rho(\cdot|\theta)$ with respect to a measure ν on \mathcal{G} . For $x \in \mathcal{X}$ let $q(\cdot|x)$ have positive*

density $\sigma(\cdot|x)$ with respect to a measure λ on \mathcal{F} , such that $0 < q(K|x) < \infty$ and $0 < \lambda(K) < \infty$ for $K \in \mathcal{K}$. Let $\rho(\cdot|\cdot): \mathcal{X} \times \Theta \rightarrow \mathbb{R}$ and $\sigma(\cdot|\cdot): \Theta \times \mathcal{X} \rightarrow \mathbb{R}$ be bounded and measurable $\mathcal{F} \times \mathcal{G}$. For $x \in \mathcal{X}$, let $\rho(x|\cdot): \Theta \rightarrow \mathbb{R}$ be continuous, and for $\theta \in \Theta$, let $\sigma(\cdot|\theta): \mathcal{X} \rightarrow \mathbb{R}$ be continuous.

(a) For compact $K \subseteq \Theta$, p and q are K -continuous, so that if q is weakly K -expectation consistent then q is a K -posterior for p and a prior π .

(b) If q is a weakly \mathcal{K} -expectation consistent then there is a \mathcal{K} -prior, μ , such that

- (i) q is a posterior for p and μ ;
- (ii) μ is minimally \mathcal{K}' -compatible with q , where

$$\mathcal{K}' := \{K \in \mathcal{F} | 0 < \mu(K) < \infty\} \supseteq \mathcal{K};$$

- (iii) q is fully \mathcal{K}' -expectation consistent with p and fully \mathcal{K}' -coherent.

The proof of Corollary 4.13, which adapts the proof of an analogous theorem of Dawid and Stone (1972) to the present context, is given in the appendix.

4.14. EXAMPLE. As in Example 3.6, n is a positive integer $\Theta := (0, 1)$, $\mathcal{X} := \{0, 1, \dots, n\}$, and for $\theta \in \Theta$, $p(\cdot|\theta)$ is the binomial distribution with parameters n and θ . Θ is locally compact and σ -compact in the usual topology of \mathbb{R} , and \mathcal{X} is compact in the discrete topology. For $\theta \in \Theta$, $p(\cdot|\theta)$ has density

$$\rho(x|\theta) := \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

with respect to counting measure ν on \mathcal{X} . Let \mathcal{K} be a bunch of compact subsets of Θ that contains an expanding sequence (K_n) such that $\bigcup_{n=1}^\infty K_n = \Theta$. By Corollary 4.13, if, for $x \in \mathcal{X}$, q has positive density $\sigma(\cdot|x)$ with respect to Lebesgue measure on $(0, 1)$ such that $0 < q(K|x) < \infty$ for $K \in \mathcal{K}$ and if q is weakly \mathcal{K} -expectation consistent with p , then there is a \mathcal{K} -prior μ , minimally compatible with q such that q is a posterior for p and μ , fully \mathcal{K}' -expectation consistent with p and fully \mathcal{K}' -coherent, where $\mathcal{K}' := \{K \in \mathcal{F} | 0 < \mu(K) < \infty\} \supseteq \mathcal{K}$.

4.15. EXAMPLE. As in Example 3.7, let $\Theta := \mathcal{X} := \mathbb{R}$ and let each of \mathcal{F} and \mathcal{G} be the class of Baire (= Borel) sets in \mathbb{R} . Again, for $B \in \mathcal{G}$ and $\theta \in \mathbb{R}$, let

$$p(B|\theta) := \frac{1}{\sqrt{2\pi}} \int_B \exp\left[-(x - \theta)^2/2\right] dx.$$

Let each of λ and ν be Lebesgue measure on \mathbb{R} , so that

$$\rho(x|\theta) = \frac{1}{\sqrt{2\pi}} \exp\left[-(x - \theta)^2/2\right] > 0.$$

Let \mathcal{K} be a bunch of compact subsets of Θ that contains an expanding sequence (K_n) such that $\bigcup_{n=1}^\infty K_n = \Theta$. By Corollary 4.13, if, for $x \in \mathcal{X}$, q has positive density $\sigma(\cdot|x)$ with respect to Lebesgue measure on \mathbb{R} such that $0 < q(K|x) < \infty$ for $K \in \mathcal{K}$ and if q is weakly \mathcal{K} -expectation consistent with p , then there is a \mathcal{K} -prior μ , minimally compatible with q such that q is a posterior for p and μ , fully \mathcal{K}' -expectation consistent with p and fully \mathcal{K}' -coherent, where $\mathcal{K}' := \{K \in \mathcal{F} | 0 < \mu(K) < \infty\} \supseteq \mathcal{K}$.

5. Remarks. In the papers by Heath and Sudderth (1978) and by Lane and Sudderth (1983) a simple betting system (an HSLs system) has $K := \Theta$, $C \in \mathcal{F} * \mathcal{S}$ and $D := \Theta \times \mathcal{X} \setminus C$. The class of HSLs systems is (when $K = \Theta$) intermediate between the two classes introduced in the present paper. A simple Θ -betting system dependent of x (Section 4) is an HSLs-system; in turn, an HSLs system is a simple Θ -betting system as described at the beginning of Section 3. So the gambler is given least latitude in the case of bets independent of x , and most latitude in the case of the systems described in Section 3. These latter systems are essentially equivalent to those described by Regazzini (1987).

Suppose one admits gambler's strategies that consist of countably many simple K -betting systems [cf. Skala (1986)]. Assume that for $x \in \mathcal{X}$ the partial sums $\sum_{i=1}^n g_i^x$ are dominated by a q -integrable function and that $\sum_{i=1}^\infty g_i(\cdot, \cdot)$ converges on $\Theta \times \mathcal{X}$ to a function $g(\cdot, \cdot)$ dominated by a p -integrable function $\mathcal{X} \rightarrow \mathbb{R}$, uniformly on Θ . Since $q(g_i^x|x) = 0$ for $x \in \mathcal{X}$ and for all i , we have (cf. Definition 3.2)

$$(i) \quad q(g^x|x) = 0 \quad \text{for } x \in \mathcal{X};$$

also

$$(ii) \quad g(\theta, x) = 0 \quad \text{for } (\theta, x) \in K^c \times \mathcal{X}.$$

Thus full K -expectation consistency implies full K -coherence even when the gambler's strategy may consist of countably many (sufficiently regular) simple K -betting systems.

APPENDIX

PROOF OF LEMMA 4.11. Suppose the contrary, that another joint probability distribution has the same conditional densities ρ and σ . Then there are versions δ_1^π and δ_1^m of its marginal densities such that for $(\theta, x) \in \Theta \times \mathcal{X}$,

$$\delta_1^\pi(\theta)\rho(x|\theta) = \delta_1^m(x)\sigma(\theta|x)$$

and

$$\delta^\pi(\theta)\rho(x|\theta) = \delta^m(x)\sigma(\theta|x).$$

Since $\delta^\pi(\theta) > 0$ and $\rho(x|\theta) > 0$ for $(\theta, x) \in (\Theta^+ \times \mathcal{X})$, also $\delta^m(x) > 0$ and

$\sigma(\theta|x) > 0$. Dividing each member of the first equation by the corresponding member of the second, we find that

$$\delta_1^\pi(\theta)/\delta^\pi(\theta) = \delta_1^m(x)/\delta^m(x) \quad \text{for } (\theta, x) \in \Theta^+ \times \mathcal{X}.$$

Since δ_1^m is a probability density, there exists $x_0 \in \mathcal{X}$ such that $\delta_1^m(x_0) > 0$. Then for all $\theta \in \Theta^+$,

$$\delta_1^\pi(\theta)/\delta^\pi(\theta) = \delta_1^m(x_0)/\delta^m(x_0) > 0.$$

Thus the ratio $\delta_1^\pi(\theta)/\delta^\pi(\theta)$ is constant on Θ^+ ; also $\delta^\pi(\theta) = 0$ for $\theta \notin \Theta^+$. Since numerator and denominator are probability densities with respect to λ , they must coincide a.e. (λ). \square

PROOF OF COROLLARY 4.13. (a) Let Baire $h: \mathcal{X} \rightarrow \mathbb{R}$ be bounded so that there is a positive number M such that $|h(x)| \leq M$ for $x \in \mathcal{X}$. Let $\theta_0 \in K$. Then for $\theta \in K$,

$$|p(h|\theta) - p(h|\theta_0)| \leq M \int |\rho(x|\theta) - \rho(x|\theta_0)| \nu(dx).$$

By Scheffé's theorem the right member converges to 0 as $\theta \rightarrow \theta_0$, establishing that p is K -continuous. Let Baire $h: \Theta \rightarrow \mathbb{R}$ have continuous restriction to K . By hypothesis $\lambda(K) > 0$ and $\sigma(\theta|x) > 0$ so that $q(K|x) > 0$ for $x \in \mathcal{X}$. Fix $x_0 \in \mathcal{X}$. For $x \in \mathcal{X}$, we have

$$\begin{aligned} & q_K(hK|x) - q_K(hK|x_0) \\ &= \int_K h(\theta) [\sigma(\theta|x) - \sigma(\theta|x_0)] \lambda(d\theta) / q(K|x) \\ &+ \left[q(K|x)^{-1} - q(K|x_0)^{-1} \right] \int_K h(\theta) \sigma(\theta|x_0) \lambda(d\theta). \end{aligned}$$

Applying Scheffé's theorem, we find that $q(K|x) \rightarrow q(K|x_0)$ as $x \rightarrow x_0$. Thus the second term on the right converges to 0 as $x \rightarrow x_0$. Another application of Scheffé's theorem shows that the first term on the right also converges to 0 as $x \rightarrow x_0$, completing the proof of the assertion that q is K -continuous. Conclusion (a) now follows from Theorem 4.9.

(b)(i) We begin by using Lemma 4.11 to show that for $K \in \mathcal{K}$, p and q fail to admit multiple K -marginals. It will then follow from Proposition 4.12 that there is a \mathcal{K} -prior μ such that q is a \mathcal{K} -posterior for p and μ . A brief supplementary argument will complete the proof of (b)(i).

In order to use Lemma 4.11 to conclude that, for $K \in \mathcal{K}$, p and q fail to admit multiple K -marginals, it suffices to show that the K -prior π_K , guaranteed by (a), is absolutely continuous with respect to λ and the corresponding marginal m_K is absolutely continuous with respect to ν . Define the measure

r_K on $\mathcal{F} * \mathcal{S}$ by

$$r_K(A \times B) := \int_{KA} p(B|\theta)\pi_K(d\theta) \quad \text{for } A \in \mathcal{F}, B \in \mathcal{S}, \text{ or}$$

$$(A1) \quad r_K(A \times B) = \int_{KA} \left[\int_B \rho(x|\theta)\nu(dx) \right] \pi_K(d\theta)$$

$$= \int_B \left[\int_{KA} \rho(x|\theta)\pi_K(d\theta) \right] \nu(dx)$$

and let m_K be the other marginal: for $B \in \mathcal{S}$, $m_K(B) := r_K(K \times B)$. Since q is a K -posterior for p and π_K , on setting $\psi_K := 1/q(K|x)$ for $x \in \mathcal{X}$ we have

$$r_K(A \times B) = \int_B q_K(A|x)m_K(dx)$$

$$(A2) \quad = \int_B \psi_K(x) \int_{KA} \sigma(\theta|x)\lambda(d\theta)m_K(dx)$$

$$= \int_A K(\theta) \int_B \sigma(\theta|x)\psi_K(x)m_K(dx)\lambda(d\theta) \quad \text{for } A \in \mathcal{F}, B \in \mathcal{S}.$$

But from (A1) we have, for $B \in \mathcal{S}$,

$$m_K(B) = \int_B \left[\int_K \rho(x|\theta)\pi_K(\theta) \right] \nu(dx),$$

so that m_K is absolutely continuous with respect to ν . Also, from (A2), for $A \in \mathcal{F}$,

$$\pi_K(A) = \int_A K(\theta) \left[\int_{\mathcal{X}} \psi_K(x)\sigma(\theta|x)m_K(dx) \right] \lambda(d\theta),$$

so that π_K is absolutely continuous with respect to λ . We conclude from Proposition 4.12 that there is a \mathcal{K} -prior μ such that q is a \mathcal{K} -posterior for μ and p .

The measure μ provided by Proposition 4.12, is determined uniquely to within a constant factor; a version with $\mu(K_1) = 1$ is given for $A \in \mathcal{F}$, $K \in \mathcal{K}$, $K \supseteq K_1$ by $\mu(AK) = \pi_K(AK)/\pi_K(K_1)$. For $A \in \mathcal{F}$,

$$\mu(A) = \lim_n \mu(AK_n) = \lim_n \pi_{K_n}(AK_n)/\pi_{K_n}(K_1).$$

Since, for $n \in \mathbb{N}$, π_{K_n} is absolutely continuous with respect to λ , also $\lambda(A) = 0$ implies $\mu(A) = 0$ and μ is absolutely continuous with respect to λ . Let $h := d\mu/d\lambda$. For $K \in \mathcal{K}$ and $A \in \mathcal{F}$,

$$\mu(AK) = \pi_K(A)\mu(K)$$

and for an appropriate version of δ_K^π ,

$$K(\theta)h(\theta) = \mu(K)\delta_K^\pi(\theta) \quad \text{for } \theta \in K.$$

Set $K^+ := \{\theta \in K | \delta_K^\pi(\theta) > 0\}$. In order to apply Lemma 4.11 in the present

context, we must replace $\sigma(\theta|x)$ in Lemma 4.11 by the present conditional density

$$(dq_K/d\lambda)(\theta|x) = \psi_K(x)\sigma(\theta|x).$$

We have, for appropriate versions of $\delta_K^\pi := d\pi_K/d\lambda$ and $\delta_K^m := dm_K/d\nu$,

$$\delta_K^\pi(\theta)\rho(x|\theta) = \delta_K^m(x)\psi_K(x)\sigma(\theta|x),$$

each member being a version of $dr_K/d(\lambda \times \nu)$. Since the left member is positive for $\theta \in K^+$ and $x \in \mathcal{X}$, also $\delta_K^m(x) > 0$ for $x \in \mathcal{X}$. By hypothesis, $\psi_K(x) > 0$ and $\sigma(\theta|x) > 0$ for $\theta \in \Theta$ and $x \in \mathcal{X}$, so that $\delta_K^\pi(\theta) > 0$ for $\theta \in K$. Then for $\theta \in K$,

$$h(\theta) = K(\theta)h(\theta) = \mu(K)\delta_K^\pi(\theta) > 0,$$

$$h(\theta)\rho(x|\theta) = \mu(K)\delta_K^\pi(\theta)\rho(x|\theta) = \mu(K)\delta_K^m(x)\psi_K(x)\sigma(\theta|x)$$

for $\theta \in K$, $x \in \mathcal{X}$, and

$$\sigma(\theta|x) = h(\theta)\rho(x|\theta)/[\mu(K)\delta_K^m(x)\psi_K(x)] \quad \text{for } x \in \mathcal{X}, \theta \in K.$$

Since $\Theta = \bigcup_{n=1}^\infty K_n$, for $\theta \in \Theta$ there is a $K \in \mathcal{K}$ such that $\theta \in K$. Since \mathcal{K} is a bunch, if $K_1 \in \mathcal{K}$ and $K_2 \in \mathcal{K}$, then $K_1 \cup K_2 \in \mathcal{K}$. It follows that $\mu(K)\delta_K^m(x)\psi_K(x)$ is independent of K in \mathcal{K} for $x \in \mathcal{X}$, so that the hypotheses of Proposition 2.5 are satisfied, and q is a posterior for μ and p .

(b) (ii) and (iii). We recall that to say that q is a posterior for μ and p is to say that q is a \mathcal{K} -posterior for μ and p , where \mathcal{K} is the class of all sets K in \mathcal{F} such that $0 < \mu(K) < \infty$. As we noted above, $h(\theta) > 0$ for $\theta \in K \in \mathcal{K}$, which implies that $h(\theta) > 0$ for $\theta \in \Theta$. We observe as follows that μ is minimally \mathcal{K}' -compatible with q . Let $K \in \mathcal{K}'$, $x \in \mathcal{X}$, $A \in \mathcal{F}$ and let $q_K(A|x)$ be positive. Then $\int_{AK} \sigma(\theta|x)\lambda(d\theta) > 0$. Hence $\lambda(AK) > 0$; and since $h > 0$, also $\mu(AK) > 0$ and $\pi_K(A) = \mu(AK)/\mu(K) > 0$.

By Theorem 3.5, q is fully \mathcal{K}' -expectation consistent with p and fully \mathcal{K}' coherent. \square

Acknowledgment. The author acknowledges gratefully many useful suggestions by the referees.

REFERENCES

- ARMSTRONG, T. E. and SUDDERTH, W. (1989). Locally coherent rates of exchange. *Ann. Statist.* **17** 1394–1408.
- ASH, R. B. (1972). *Real Analysis and Probability*. Academic, New York.
- BAUER, H. (1981). *Probability Theory and Elements of Measure Theory*. Academic, New York.
- BUEHLER, R. J. (1976). Coherent preferences. *Ann. Statist.* **4** 1051–1064.
- CORNFIELD, J. (1969). The Bayesian outlook and its application. *Biometrics* **25** 617–657.
- DAWID, A. P. and STONE, M. (1972). Expectation consistency of inverse probability distributions. *Biometrika* **59** 486–489.
- DAWID, A. P. and STONE, M. (1973). Expectation consistency and generalized Bayes inference. *Ann. Statist.* **1** 478–485.
- FREEDMAN, D. and PURVES, R. (1969). Bayes' method for bookies. *Ann. Math. Statist.* **40** 1177–1186.

- HEATH, D. and SUDDERTH, W. (1972). On a theorem of de Finetti, oddsmaking and game theory. *Ann. Math. Statist.* **43** 2072–2077.
- HEATH, D. and SUDDERTH, W. (1978). On finitely additive priors, coherence, and extended admissibility. *Ann. Statist.* **6** 333–345.
- LANE, D. A. and SUDDERTH, W. D. (1983). Coherent and continuous inference. *Ann. Statist.* **11** 114–120.
- NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probability*. Holden-Day, San Francisco.
- PIERCE, D. A. (1973). On some difficulties in a frequency theory of inference. *Ann. Statist.* **1** 241–250.
- PRYCE, J. D. (1973). *Basic Methods of Linear Functional Analysis*. Hutchinson University Library, London.
- REGAZZINI, E. (1987). De Finetti's coherence and statistical inference. *Ann. Statist.* **15** 845–864.
- RÉNYI, A. (1970). *Foundations of Probability*. Holden-Day, San Francisco.
- SKALA, H. J. (1988). On σ -additive priors, σ -coherence, and the existence of posteriors. In *Risk, Decision and Rationality* (B. R. Munier, ed.) 563–574. Reidel, Dordrecht.

DEPARTMENT OF STATISTICS
KIDDER HALL 44
OREGON STATE UNIVERSITY
CORVALLIS, OREGON 97331-4606