A CONSISTENCY THEOREM FOR REGULAR CONDITIONAL DISTRIBUTIONS

PATRIZIA BERTI, EMANUELA DREASSI, AND PIETRO RIGO

Abstract. Let $(\Omega, \mathcal{B})$ be a measurable space, $\mathcal{A}_n \subset \mathcal{B}$ a sub-σ-field and μ_n a random probability measure on $(\Omega, \mathcal{B})$, $n \geq 1$. In various frameworks, one looks for a probability P on $\mathcal{B}$ such that μ_n is a regular conditional distribution for P given $\mathcal{A}_n$ for all n. Conditions for such a P to exist are given. The conditions are quite simple when $(\Omega, \mathcal{B})$ is a compact Hausdorff space equipped with the Borel or the Baire σ-field (as well as under similar assumptions). Applications to Gibbs measures and Bayesian statistics are given as well.

1. The problem

Let $(\Omega, \mathcal{B})$ be a measurable space and $\mathbb{P}$ the collection of all probability measures on $\mathcal{B}$. For $B \in \mathcal{B}$ and any map $\mu : \Omega \to \mathbb{P}$, we let $\mu(B)$ denote the function on Ω given by $\omega \mapsto \mu(\omega)(B)$. We also let $\sigma(\mu) = \sigma\{\mu(B) : B \in \mathcal{B}\}$.

Let $P \in \mathbb{P}$ and $\mathcal{A} \subset \mathcal{B}$ a sub-σ-field. A *regular conditional distribution* (r.c.d.), for P given $\mathcal{A}$, is a map $\mu : \Omega \to \mathbb{P}$ such that $\mu(B)$ is a version of $E_P(I_B \mid \mathcal{A})$ for all $B \in \mathcal{B}$. For a r.c.d. to exist, it suffices that P is perfect and $\mathcal{B}$ countably generated.

This note originates from the following question. Given a sub-σ-field $\mathcal{A} \subset \mathcal{B}$ and a map $\mu : \Omega \to \mathbb{P}$ such that $\sigma(\mu) \subset \mathcal{A}$, under what conditions is there $P \in \mathbb{P}$ such that μ is a r.c.d. for P given $\mathcal{A}$? Such a question is easily answered. Once stated, however, it grows quickly into the following new question. Suppose we are given $\{\mathcal{A}_n, \mu_n : n \in I\}$, where $\mathcal{A}_n \subset \mathcal{B}$ is a sub-σ-field,

$$\mu_n : \Omega \to \mathbb{P} \text{ is a map such that } \sigma(\mu_n) \subset \mathcal{A}_n,$$

and $I = \{1, 2, \ldots\}$ or $I = \{1, \ldots, k\}$ for some $k \geq 1$. *Under what conditions is there* $P \in \mathbb{P}$ *such that* μ_n *is a r.c.d. for* P *given* $\mathcal{A}_n$ *for all* $n \in I$? If such a P exists, the μ_n are said to be *consistent*.

We aim to give conditions for the μ_n to be consistent; see Theorems 6 and 7.

Throughout, $\mathbb{M}$ denotes the (possibly empty) set

$$\mathbb{M} = \{P \in \mathbb{P} : \mu_n \text{ is a r.c.d. for } P \text{ given } \mathcal{A}_n \text{ for all } n \in I\}.$$

2. Motivations

Reasonable conditions for consistency, if available, are of potential interest.

As a first (heuristic) example suppose that, for each $n \in I$, expert n declares his/her opinions on a certain phenomenon conditionally on his/her information $\mathcal{A}_n$. This produces a collection of random probability measures $\mu_n : \Omega \to \mathbb{P}$ such that $\sigma(\mu_n) \subset \mathcal{A}_n$. In this framework, most literature focus on how to summarize the

2000 *Mathematics Subject Classification.* 60A05, 60A10, 62F15, 62G99.

Key words and phrases. Gibbs measure, Posterior distribution, Random probability measure, Regular conditional distribution.

experts' opinions μ_n. But a preliminary (and not escapable) question is whether the μ_n are consistent. If they are not, some of the μ_n should be discarded.

Let us turn now to more formal examples.

2.1. **Gibbs measures.** The consistency problem of Section 1 is classical in statistical mechanics. The following is from [3]. Let $X = (X_i : i \in S)$ be a process, with state space $(E, \mathcal{E})$, indexed by the countable set S. In order to assess the probability distribution of X, to be called a *Gibbs measure*, we proceed as follows.

Let $\Omega = E^S$ and $\mathcal{B} = \sigma(X_i : i \in S)$, where $X_i : \Omega \to E$ is the i-th canonical projection. For each finite $\Lambda \subset S$, we assign a random probability measure γ_Λ on $(E^\Lambda, \mathcal{E}^\Lambda)$, measurable with respect to $\sigma(X_i : i \notin \Lambda)$. Precisely, $\gamma_\Lambda(\omega)$ is a probability measure on $(E^\Lambda, \mathcal{E}^\Lambda)$ for each $\omega \in \Omega$ and $\sigma(\gamma_\Lambda) \subset \sigma(X_i : i \notin \Lambda)$. Here, γ_Λ is regarded as the conditional distribution of $(X_i : i \in \Lambda)$ given $(X_i : i \notin \Lambda)$. Indeed, for finite Λ, the Gibbsian formalism of equilibrium statistical mechanics provides a simple and reasonable scheme for selecting γ_Λ. But of course a consistency problem arises for the collection $\{\gamma_\Lambda : \Lambda \text{ finite}\}$.

Precisely, fix any enumeration $\Lambda_1, \Lambda_2, \ldots$ of the finite parts of S and define

$$Z_n = (X_i : i \notin \Lambda_n), \quad \mathcal{A}_n = \sigma\big(Z_n\big), \quad \mu_n(\omega) = \gamma_{\Lambda_n}(\omega) \times \delta_{Z_n(\omega)}.$$

Then, assessing $\{\gamma_\Lambda : \Lambda \text{ finite}\}$ makes sense if and only if the μ_n are consistent.

2.2. **Bayesian inference.** Loosely speaking, given two events A and B, to assess $\text{Prob}(A \mid B)$ is often easier than to evaluate $\text{Prob}(A)$. This vague remark can be useful in Bayesian statistics.

Let $(\mathcal{X}, \mathcal{F})$ and $(\Theta, \mathcal{G})$ be measurable spaces and $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ a (measurable) collection of probabilities on $\mathcal{F}$. Fix a prior probability π on $\mathcal{G}$ and define $m(F) = \int P_\theta(F)\, \pi(d\theta)$ for $F \in \mathcal{F}$. A posterior for $\mathcal{P}$ and π is a (measurable) collection $\mathcal{Q} = \{Q_x : x \in \mathcal{X}\}$ of probabilities on $\mathcal{G}$ satisfying

$$\int_G P_\theta(F)\, \pi(d\theta) = \int_F Q_x(G)\, m(dx) \quad \text{for all } F \in \mathcal{F} \text{ and } G \in \mathcal{G}.$$

The standard Bayes procedure is to select a prior π and to calculate (or to approximate) the posterior $\mathcal{Q}$. To assess π is often very arduous. Sometimes, it is more convenient to avoid the explicit choice of π, and to assign directly a collection $\mathcal{Q} = \{Q_x : x \in \mathcal{X}\}$ of probabilities on $\mathcal{G}$. Here, Q_x describes our opinions on θ when x is observed. In these cases, the standard Bayes scheme is reverted. One first selects some collection $\mathcal{Q}$ of probabilities on $\mathcal{G}$ and then asks whether $\mathcal{Q}$ is consistent with $\mathcal{P}$, in the sense that $\mathcal{Q}$ is the posterior of $\mathcal{P}$ and some prior π.

Examples of this "reverted" Bayes procedure are not unusual. Suppose $\mathcal{Q}$ is the formal posterior of an improper prior, or else it is obtained by some empirical Bayes method. Then, $\mathcal{Q}$ is assessed without explicitly selecting a (proper) prior π. Such $\mathcal{Q}$ may look reasonable or not (there are indeed different opinions). But again, a preliminary (and not escapable) question is whether $\mathcal{Q}$ is consistent with $\mathcal{P}$.

The reverted Bayes procedure agrees with the subjective view of probability. In fact, it has been developed in a coherence framework; see [1], [4], [6], [7] and references therein. However, in a coherence framework, $\mathcal{P}$ and $\mathcal{Q}$ are requested to be consistent under some finitely additive prior.

To investigate the reverted Bayes procedure, *without using finitely additive probabilities* but relying on standard (Kolmogorovian) probability theory, we need exactly the notion of consistency of Section 1.

3. Results

3.1. **Preliminaries.** We begin with a couple of lemmas and a corollary.

Lemma 1. *Let $\mathcal{A}=\sigma(\cup_{n\in I}\mathcal{A}_n)$. The μ_n are consistent if and only if*

$$Q\big(\mu_n(A)=I_A\big)=1 \quad \text{whenever } n\in I \text{ and } A\in\mathcal{A}_n, \tag{1}$$

$$E_Q\big\{\mu_n(B)\big\}=E_Q\big\{\mu_1(B)\big\} \quad \text{whenever } n\in I \text{ and } B\in\mathcal{B}, \tag{2}$$

for some probability measure Q on $\mathcal{A}$. In particular, if Q meets conditions (1)-(2) *and $P(\cdot)=E_Q\{\mu_1(\cdot)\}$, then $P\in\mathbb{M}$ and $P=Q$ on $\cup_{n\in I}\mathcal{A}_n$. Moreover, if each $\mathcal{A}_n$ is countably generated, condition* (1) *can be written as $Q(\Omega_0)=1$, where*

$$\Omega_0=\{\,\mu_n(A)=I_A \ \text{ for all } n\in I \text{ and } A\in\mathcal{A}_n\}.$$

Proof. If the μ_n are consistent, it suffices to let $Q=P|\mathcal{A}$ for some $P\in\mathbb{M}$. Conversely, suppose conditions (1)-(2) hold for some Q. Define $P(B)=E_Q\big\{\mu_1(B)\big\}$ for all $B\in\mathcal{B}$. Fix $n\in I$, $A\in\mathcal{A}_n$ and $B\in\mathcal{B}$. Then, $\mu_n(A\cap B)=I_A\,\mu_n(B)$ on the set $\{\mu_n(A)=I_A\}$. Thus, conditions (1)-(2) yield

$$P(A\cap B)=E_Q\big\{\mu_1(A\cap B)\big\}=E_Q\big\{\mu_n(A\cap B)\,I_{\{\mu_n(A)=I_A\}}\big\}=E_Q\big\{I_A\,\mu_n(B)\big\}.$$

For $B=\Omega$, the above equation reduces to $P(A)=Q(A)$. Hence, $P=Q$ on $\mathcal{A}_n$. Since $\sigma(\mu_n)\subset\mathcal{A}_n$, it follows that $E_P\big\{I_A\,\mu_n(B)\big\}=E_Q\big\{I_A\,\mu_n(B)\big\}=P(A\cap B)$. This proves that $P\in\mathbb{M}$. Finally, suppose the $\mathcal{A}_n$ countably generated and take countable fields $\mathcal{U}_n$ such that $\mathcal{A}_n=\sigma(\mathcal{U}_n)$. Since $\Omega_0=\bigcap_{n\in I}\bigcap_{A\in\mathcal{U}_n}\{\mu_n(A)=I_A\}$, then $\Omega_0\in\mathcal{A}$ and condition (1) amounts to $Q(\Omega_0)=1$. $\square$

In a sense, up to replacing Ω with Ω_0, condition (1) can be assumed to be true whenever the $\mathcal{A}_n$ are countably generated. In this case, in fact, the μ_n are certainly not consistent if $\Omega_0=\emptyset$. Otherwise, if $\Omega_0\neq\emptyset$, they are consistent if and only if there is a probability Q on the trace σ-field $\mathcal{A}\cap\Omega_0$ satisfying condition (2).

Among other things, Lemma 1 answers our initial question, raised in Section 1. Suppose in fact $I=\{1\}$. Since (2) is trivially true, $\mathbb{M}\neq\emptyset$ if and only if condition (1) holds. In turn, condition (1) is automatically true if $\Omega_0\neq\emptyset$ (just let $Q=\delta_\omega$ for some $\omega\in\Omega_0$, so that Lemma 1 implies $\mu_1(\omega)\in\mathbb{M}$). Furthermore, $\Omega_0\neq\emptyset$ is equivalent to (1) if $\mathcal{A}_1$ is countably generated (but not in general).

From now on, whether or not the $\mathcal{A}_n$ are countably generated, it is assumed $\Omega_0=\Omega$ or equivalently

$$\mu_n(\omega)(A)=I_A(\omega) \quad \text{for all } \omega\in\Omega,\ n\in I \text{ and } A\in\mathcal{A}_n. \tag{3}$$

Lemma 2. *Suppose condition* (3) *holds. The μ_n are consistent if and only if*

$$\int\mu_j(\omega)(B)\,Q_j(d\omega)=\int\int\mu_n(x)(B)\,\mu_j(\omega)(dx)\,Q_j(d\omega) \tag{4}$$

for some $j\in I$, some probability Q_j on $\mathcal{A}_j$, and all $n\in I$ and $B\in\mathcal{B}$.

Proof. Suppose the μ_n are consistent. Fix $P\in\mathbb{M}$ and $j\in I$ and let $Q_j=P|\mathcal{A}_j$. Since $P(dx)=\mu_j(\omega)(dx)\,Q_j(d\omega)$, then

$$\int\int\mu_n(x)(B)\,\mu_j(\omega)(dx)\,Q_j(d\omega)=\int\mu_n(x)(B)\,P(dx)=P(B)=\int\mu_j(\omega)(B)\,Q_j(d\omega)$$

for all $n \in I$ and $B \in \mathcal{B}$. Conversely, suppose condition (4) holds for some $j \in I$ and probability Q_j on $\mathcal{A}_j$. Define $P(B) = E_{Q_j}\{\mu_j(B)\}$ for all $B \in \mathcal{B}$. By (3),

$$P(A) = E_{Q_j}\{\mu_j(A)\} = E_{Q_j}\{I_A\} = Q_j(A) \quad \text{for all } A \in \mathcal{A}_j.$$

Let $n \in I$ and $B \in \mathcal{B}$. Since $P = Q_j$ on $\mathcal{A}_j$, condition (4) yields

$$E_P\{\mu_n(B)\} = \int \mu_n(x)(B)\, P(dx) = \int \int \mu_n(x)(B)\, \mu_j(\omega)(dx)\, Q_j(d\omega)$$
$$= \int \mu_j(\omega)(B)\, Q_j(d\omega) = \int \mu_j(\omega)(B)\, P(d\omega) = E_P\{\mu_j(B)\}.$$

An application of Lemma 1 (with $Q = P|\mathcal{A}$) concludes the proof. □

Corollary 3. *Suppose condition* (3) *holds. If there is* $\omega_0 \in \Omega$ *such that*

$$\mu_1(\omega_0)(\cdot) = \int \mu_n(x)(\cdot)\, \mu_1(\omega_0)(dx) \quad \textit{for all } n \in I, \tag{5}$$

the μ_n *are consistent. Moreover, condition* (5) *is equivalent to consistency of the* μ_n *in case* $\mathcal{B}$ *is countably generated and* $\mathcal{A}_n \supset \mathcal{A}_1$ *for all* $n \in I$.

Proof. Just apply Lemma 2 with $j = 1$ and $Q_j = \delta_{\omega_0}$. Next, suppose $\mathcal{B}$ countably generated, $\mathcal{A}_n \supset \mathcal{A}_1$ for all n, and the μ_n are consistent. Fix $P \in \mathbb{M}$ and define $A = \{\omega : \mu_1(\omega)(\cdot) = \int \mu_n(x)(\cdot)\, \mu_1(\omega)(dx) \text{ for all } n \in I\}$. Then,

$$\mu_1(B) = E_P\big(I_B \mid \mathcal{A}_1\big) = E_P\big\{E_P(I_B \mid \mathcal{A}_n) \mid \mathcal{A}_1\big\} = \int \mu_n(x)(B)\, \mu_1(\cdot)(dx), \quad P\text{-a.s.},$$

for fixed $B \in \mathcal{B}$ and $n \in I$. Hence, $\mathcal{B}$ countably generated yields $P(A) = 1$. □

When the μ_n are consistent, various questions on $\mathbb{M}$ arise. A natural one is uniqueness of $P \in \mathbb{M}$. Another question is existence (and possibly uniqueness) of $P \in \mathbb{M}$ such that $P \sim P_0$, where P_0 is a given reference measure.

In general, to find non trivial conditions for uniqueness of $P \in \mathbb{M}$ looks very arduous. For instance, $P(\cdot \mid A) \in \mathbb{M}$ whenever $P \in \mathbb{M}$, $A \in \cap_n \mathcal{A}_n$ and $P(A) > 0$. However, uniqueness conditions are available in particular cases. One such case is that of Gibbs measures; see Chapter 8 of [3]. Here, incidentally, uniqueness of $P \in \mathbb{M}$ is crucial as non uniqueness corresponds to phase transitions.

The second question is connected to equivalent martingale measures.

Proposition 4. *Suppose condition* (3) *holds and* $\cup_{n\in I}\mathcal{A}_n$ *is a field. Fix a probability* P_0 *on* $\mathcal{A} = \sigma(\cup_{n\in I}\mathcal{A}_n)$ *and let* F *be the linear space generated by* $\mu_n(B) - \mu_1(B)$ *for all* $n \in I$ *and* $B \in \mathcal{B}$. *Then, there is* $P \in \mathbb{M}$ *such that* $P \sim P_0$ *on* $\mathcal{A}$ *if and only if*

$$\overline{F - L_\infty^+} \cap L_\infty^+ = \{0\} \tag{6}$$

where $L_\infty = L_\infty(\Omega, \mathcal{A}, P_0)$ *and the closure is in the weak* topology on* L_∞.

Proof. By a result of Kreps [5], condition (6) is equivalent to existence of a probability Q on $\mathcal{A}$ such that

$$Q \sim P_0 \quad \text{and} \quad E_Q(f) = 0 \text{ for all } f \in F. \tag{7}$$

Thus, if $P \in \mathbb{M}$ and $P \sim P_0$ on $\mathcal{A}$, it suffices to note that $Q = P|\mathcal{A}$ meets condition (7). Conversely, under (6), take Q satisfying (7) and define $P(\cdot) = E_Q\{\mu_1(\cdot)\}$. In view of (3), Lemma 1 implies $P \in \mathbb{M}$ and $P = Q$ on $\cup_{n\in I}\mathcal{A}_n$. Since $\cup_{n\in I}\mathcal{A}_n$ is a field, it follows that $P = Q \sim P_0$ on $\mathcal{A}$. □

As an extreme example, if F is finite dimensional, existence of $P \in \mathbb{M}$ such that $P \sim P_0$ on $\mathcal{A}$ reduces to the no arbitrage condition

$$P_0\big(f > 0\big) > 0 \quad \Longleftrightarrow \quad P_0\big(f < 0\big) > 0 \quad \text{for each } f \in F.$$

3.2. **Main results.** Some notation is needed. Let $f : \Omega \to \mathbb{R}$ be a bounded $\mathcal{B}$-measurable function. We write $P(f) = E_P(f) = \int f\,dP$ whenever $P \in \mathbb{P}$. For any map $\mu : \Omega \to \mathbb{P}$, we denote $\mu(f)$ the function on Ω given by

$$\mu(\omega)(f) = \int f(x)\,\mu(\omega)(dx), \quad \omega \in \Omega.$$

A $\mathcal{B}$-*determining class* is a class $\mathcal{S}$ of bounded $\mathcal{B}$-measurable functions $f : \Omega \to \mathbb{R}$ such that, for arbitrary P_1, $P_2 \in \mathbb{P}$,

$$P_1 = P_2 \quad \Longleftrightarrow \quad P_1(f) = P_2(f) \text{ for all } f \in \mathcal{S}.$$

If X is a topological space, $C(X)$ denotes the set of real continuous functions, $\mathfrak{B}(X)$ the Borel σ-field, and $\mathfrak{B}_0(X) := \sigma\big[C(X)\big]$ the Baire σ-field. Say that X is *pseudocompact* if each $f \in C(X)$ is a bounded function. Clearly, a compact space is pseudocompact. Our main tool is the following.

Lemma 5. *Let L be a vector lattice of real functions on a set X. Assume $1 \in L$ and, for any function $f : X \to [0,\infty)$,*

$$f \wedge n \in L \text{ for all } n \geq 1 \quad \Longrightarrow \quad f \in L.$$

Then, each linear positive functional U on L admits the representation $U(f) = \int f\,d\nu$, $f \in L$, for some (unique) measure ν on $\sigma(L)$. Next, suppose every $f \in L$ is bounded and fix a linear subspace $F \subset L$. If $\sup f \geq 0$ for all $f \in F$, there is a probability measure P on $\sigma(L)$ such that $E_P(f) = 0$ for all $f \in F$.

Proof. The first part of the Lemma is Theorem 8 of [2]. We prove the second part. Suppose $f_1 + \lambda_1 = f_2 + \lambda_2$ where f_1, $f_2 \in F$ and λ_1, $\lambda_2 \in \mathbb{R}$. Since F is a linear space and $\sup f \geq 0$ for all $f \in F$, one obtains $\lambda_1 = \lambda_2$ and $f_1 = f_2$. Let G be the linear space generated by F and the constants. Each $g \in G$ admits the representation $g = f + \lambda$, for some unique $f \in F$ and $\lambda \in \mathbb{R}$, and thus one can define $T(g) = T(f + \lambda) = \lambda$. Such a T is a linear functional on G satisfying $T = 0$ on F and $T(g) \leq \sup g$ for all $g \in G$. Since L consists of bounded functions, $f \mapsto \sup f$ is a real sublinear functional on L. By Hahn Banach theorem, T can be extended to a linear functional U on L satisfying $U(f) \leq \sup f$ for all $f \in L$. Finally, as $U(1) = T(1) = 1$, the first part of the Lemma yields $U(f) = E_P(f)$, $f \in L$, for some probability P on $\sigma(L)$. □

Note that, if X is a pseudocompact space and $\mathcal{E}$ a σ-field on X, one can take $L = C(X)$ or $L = \{f \in C(X) : f \text{ is } \mathcal{E}\text{-measurable}\}$ in Lemma 5.

We are now in a position to state our main result.

Theorem 6. *Suppose condition (3) holds. Fix $j \in I$ and a $\mathcal{B}$-determining class $\mathcal{S}$. The μ_n are consistent provided one of the following conditions (a)-(b) holds.*

(a) *Ω is a pseudocompact space and $\mathcal{B} \subset \mathfrak{B}_0(\Omega)$. Further,*

$$\sup_{\omega\in\Omega} h(\omega) \geq 0 \quad \text{and} \quad h \in C(\Omega)$$

for each function $h : \Omega \to \mathbb{R}$ in the linear space generated by

$$\{\mu_n(f) - \mu_j(f) : n \in I,\ f \in \mathcal{S}\}.$$

(b) *There are a pseudocompact space K, a σ-field $\mathcal{K} \subset \mathfrak{B}_0(K)$ and a surjective map $\phi : \Omega \to K$, such that $\mathcal{A}_j = \phi^{-1}(\mathcal{K})$. Further,*

$$\sup_{\omega\in\Omega} h(\omega) \geq 0 \quad \text{and} \quad h \text{ is continuous in the topology induced by } \phi$$

for each function $h : \Omega \to \mathbb{R}$ in the linear space generated by

$$\Big\{ \int \mu_n(x)(f)\, \mu_j(\cdot)(dx) - \mu_j(\cdot)(f) \ :\ n \in I,\ f \in \mathcal{S} \Big\}.$$

(The topology induced by ϕ is $\phi^{-1}(\mathcal{U})$ where $\mathcal{U}$ is the topology on K).

Proof. We first prove (b). Fix $h \in H$, where H is the linear space generated by

$$\int \mu_n(x)(f)\, \mu_j(\cdot)(dx) - \mu_j(\cdot)(f) \quad \text{for all } n \in I \text{ and } f \in \mathcal{S}.$$

Since ϕ is surjective and h measurable with respect to $\mathcal{A}_j = \phi^{-1}(\mathcal{K})$, there is a unique function $k : K \to \mathbb{R}$ such that $h = k \circ \phi$. (Just set $k(a) = h(\omega_a)$ for each $a \in K$, where $\omega_a \in \Omega$ satisfies $\phi(\omega_a) = a$). Also $k \in C(K)$, due to h is continuous in the topology induced by ϕ, and $\sup_{a\in K} k(a) = \sup_{\omega\in\Omega} h(\omega) \geq 0$. Thus, $F := \{k : h = k \circ \phi \text{ for some } h \in H\}$ is a linear subspace of $C(K)$ and $\sup k \geq 0$ for all $k \in F$. By Lemma 5, applied with $X = K$ and $L = C(K)$, there is a probability P on $\sigma(L) = \mathfrak{B}_0(K)$ such that $E_P(k) = 0$ for all $k \in F$. Next, since $\mathcal{A}_j = \phi^{-1}(\mathcal{K})$ with ϕ surjective, each $A \in \mathcal{A}_j$ can be written as $A = \{\phi \in B_A\}$ for some unique $B_A \in \mathcal{K} \subset \mathfrak{B}_0(K)$. Hence, it makes sense to define

$$Q_j(A) = Q_j(\phi \in B_A) = P(B_A) \quad \text{for all } A \in \mathcal{A}_j.$$

Then, Q_j is a probability on $\mathcal{A}_j$ and $Q_j \circ \phi^{-1} = P$ on $\mathcal{K}$. Given $h \in H$,

$$E_{Q_j}(h) = E_{Q_j}(k \circ \phi) = E_P(k) = 0$$

where $k \in F$ and $h = k \circ \phi$. In particular,

$$\int \mu_j(\omega)(f)\, Q_j(d\omega) = \int\int \mu_n(x)(f)\, \mu_j(\omega)(dx)\, Q_j(d\omega) \quad \text{for all } n \in I \text{ and } f \in \mathcal{S}.$$

Since $\mathcal{S}$ is $\mathcal{B}$-determining, an application of Lemma 2 concludes the proof of (b).

Finally, to prove (a), take F the linear space generated by $\mu_n(f) - \mu_j(f)$ for all $n \in I$ and $f \in \mathcal{S}$. Then, (a) follows precisely as (b), by applying Lemma 5 with $X = \Omega$ and $L = C(\Omega)$ and by using Lemma 1 instead of Lemma 2.

□

In real problems, part (a) of Theorem 6 is much more convenient when Ω is pseudocompact. For non pseudocompact Ω, however, part (b) may be useful as well; see Examples 9 and 11.

The connections between Theorem 6 and Proposition 4 should also be stressed. In a sense, the latter is the density-counterpart of the former. Apart from technicalities, in both cases, the main issue is existence of a probability measure with null expectation on a suitable linear space of bounded random variables. This is achieved via Lemma 5, as regards Theorem 6, and by a result of Kreps [5] in case of Proposition 4. Perhaps, Theorem 6 and Proposition 4 could be unified in a single statement, or at least they could be given essentially parallel proofs.

Our last result, suggested by ideas in [3] (see Theorems (4.17) and (4.22)), is tailor-made for Gibbs measures.

Theorem 7. *Suppose condition* (3) *holds,* Ω *is a pseudocompact space, and*

$$\mathcal{B}=\sigma(V) \quad \textit{where} \quad V=\{f\in C(\Omega): f \textit{ is } \mathcal{B}\textit{-measurable}\}.$$

Suppose also that $\mu_n(f)\in C(\Omega)$ *for all* $n\in I$ *and* $f\in V$*, and there is a subsequence* $\{n_j\}$ *satisfying*

(8) *for each* $n\geq 1$*, there is* $j_0\geq 1$ *such that* $\mathcal{A}_{n_j}\subset\mathcal{A}_n$ *for all* $j\geq j_0$.

The μ_n *are consistent if there is* $\omega_0\in\Omega$ *such that*

$$\mu_n(\omega_0)(\cdot)=\int \mu_m(x)(\cdot)\,\mu_n(\omega_0)(dx) \quad \textit{whenever } \mathcal{A}_n\subset\mathcal{A}_m.$$

Moreover, existence of such ω_0 *is equivalent to consistency of the* μ_n *in case* $\mathcal{B}$ *is countably generated.*

Proof. As Ω is pseudocompact, each $f\in V$ is bounded. Since $\sigma(V)=\mathcal{B}$ and V is closed under multiplications, V is a $\mathcal{B}$-determining class. By the latter fact and condition (3), for every $P\in\mathbb{P}$ one obtains

$$P\in\mathbb{M} \quad\Longleftrightarrow\quad E_P\big\{\mu_n(f)\big\}=E_P(f) \text{ for all } n\in I \text{ and } f\in V.$$

Let $W=\{f\in V: 0\leq f\leq 1\}$ and let $[0,1]^W$ be equipped with the product topology. Every $P\in\mathbb{P}$ can be regarded as a map $P: W\to[0,1]$. Since $[0,1]^W$ is compact, $\{\mu_{n_j}(\omega_0): j\geq 1\}$ admits a converging subnet, say $\{\mu_\alpha(\omega_0):\alpha\in D\}$ where D is a suitable directed set. Each $f\in V$ can be written as $f=a+b\,g$ for some a, $b\in\mathbb{R}$ and $g\in W$. Accordingly, one can define

$$U(f)=\lim_\alpha \mu_\alpha(\omega_0)(f)=a+b\,\lim_\alpha\mu_\alpha(\omega_0)(g) \quad \text{for all } f\in V.$$

Such U is a linear positive functional on V with $U(1)=1$. By Lemma 5, applied with $X=\Omega$ and $L=V$, one obtains $U(f)=E_{P_0}(f)$ for all $f\in V$ and some probability P_0 on $\sigma(V)=\mathcal{B}$. Fix $n\in I$ and $f\in V$. Since $\mu_n(f)\in V$ and $\mathcal{A}_\alpha\subset\mathcal{A}_n$ for large α, then

$$E_{P_0}\big\{\mu_n(f)\big\}=\lim_\alpha\mu_\alpha(\omega_0)\big\{\mu_n(f)\big\}=\lim_\alpha\mu_\alpha(\omega_0)(f)=U(f)=E_{P_0}(f).$$

Hence, $P_0\in\mathbb{M}$ and the μ_n are consistent. Finally, suppose $\mathcal{B}$ countably generated and define $A=\{\omega:\mu_n(\omega)(\cdot)=\int\mu_m(x)(\cdot)\,\mu_n(\omega)(dx)$ whenever $\mathcal{A}_n\subset\mathcal{A}_m\}$. Then, $P(A)=1$ for each $P\in\mathbb{M}$, by the same argument in the proof of Corollary 3.

□

Remark 8. In part (a) of Theorem 6, $\mathcal{B}\subset\mathfrak{B}_0(\Omega)$ can be replaced by $\mathcal{B}\subset\mathfrak{B}(\Omega)$ if each probability on $\mathfrak{B}_0(\Omega)$ can be extended to a probability on $\mathfrak{B}(\Omega)$. This is trivially true if Ω is metric ($\mathfrak{B}_0(\Omega)=\mathfrak{B}(\Omega)$ in this case) or if Ω is compact and Hausdorff. The same comment holds for part (b) up to replacing $(\Omega,\mathcal{B})$ with $(K,\mathcal{K})$.

4. Examples

Example 9. Let $\Omega = \mathbb{R}^n \setminus \{0\}$ and $\mathcal{B} = \mathfrak{B}(\Omega)$. Define

$$\phi(\omega) = \frac{\omega}{\|\omega\|}, \quad K = \{\omega : \|\omega\| = 1\}, \quad \mathcal{A}_1 = \phi^{-1}\big(\mathfrak{B}(K)\big), \quad \mathcal{S} = \{f \in C(\Omega) : f \text{ bounded}\}.$$

Define also $\lambda(\omega) = \frac{\max_i |\omega_i|}{\|\omega\|}$, where ω_i is the i-th coordinate of ω, and

$$\mu_1(\omega)(B) = \lambda(\omega) \int_0^\infty I_B[r\,\phi(\omega)] \exp\left(-\lambda(\omega)\, r\right) dr \quad \text{for all } B \in \mathcal{B}.$$

Then, $\mu_1(A) = I_A$ if $A \in \mathcal{A}_1$. Also, if $f \in \mathcal{S}$, it is not hard to see that $\mu_1(f)$ is continuous in the topology induced by ϕ. Thus, in principle, given any collection $\{\mu_n : n \in I,\, n > 1\}$ of random probability measures, consistency of $\{\mu_n : n \in I\}$ can be checked through part (b) of Theorem 6. To fix ideas, suppose $\mathcal{A}_2 \subset \mathcal{B}$ is a sub-σ-field and $\mu_2 : \Omega \to \mathbb{P}$ any map. Then, μ_1 and μ_2 are consistent whenever

$$\sigma(\mu_2) \subset \mathcal{A}_2, \qquad \mu_2(A) = I_A \text{ for } A \in \mathcal{A}_2, \quad \mu_2(f) \in \mathcal{S} \text{ for } f \in \mathcal{S},$$

$$\sup_\omega \lambda(\omega) \int_0^\infty \Big\{\mu_2[r\,\phi(\omega)](f) - f[r\,\phi(\omega)]\Big\} \exp\left(-\lambda(\omega)\, r\right) dr \geq 0 \quad \text{for } f \in \mathcal{S}.$$

Example 10. (Gibbs measures). Let $(\Omega, \mathcal{B}) = (E^S, \mathcal{E}^S)$ where $(E, \mathcal{E})$ is a measurable space and S a countable set. As in Subsection 2.1, select a collection $\gamma = \{\gamma_\Lambda : \Lambda \subset S,\, \Lambda \text{ finite}\}$ where each γ_Λ is a (suitably measurable) random probability measure on $(E^\Lambda, \mathcal{E}^\Lambda)$. Given γ, define $\mathcal{A}_n$ and μ_n as in Subsection 2.1. Then, conditions (3) and (8) are automatically true. Therefore, by Theorem 7, the μ_n are consistent provided

(i) E is a compact space and $\mathcal{E} = \mathfrak{B}_0(E)$;
(ii) $\mu_n(f) \in C(\Omega)$ for all $n \geq 1$ and $f \in V$, where Ω is given the product topology and $V = \{f \in C(\Omega) : f \text{ is } \mathcal{B}\text{-measurable}\}$;
(iii) There is $\omega_0 \in \Omega$ such that $\mu_n(\omega_0)(\cdot) = \int \mu_m(x)(\cdot)\, \mu_n(\omega_0)(dx)$ if $\mathcal{A}_n \subset \mathcal{A}_m$.

However, the present example adds very little to what already known. In fact, by arguments in [3] (see e.g. the Introduction to Chapter 4 and Theorem (4.17)), the μ_n are consistent provided E is a compact metric space and γ a *quasilocal specification*. Now, the quasilocality condition essentially amounts to (ii) and each specification γ satisfies

(iv) $\mu_n(\omega_0)(\cdot) = \int \mu_m(x)(\cdot)\, \mu_n(\omega_0)(dx)$ if $\mathcal{A}_n \subset \mathcal{A}_m$ for all $\omega_0 \in \Omega$;

we refer to [3] for details. Thus, the only contributions of this example are that metrizzability of E can be dropped and (iv) can be weakened into (iii).

Example 11. (Bayesian inference). In the notation of Subsection 2.2, define

$$\Omega = \mathcal{X} \times \Theta, \quad \mathcal{B} = \mathcal{F} \otimes \mathcal{G}, \quad \mathcal{A}_1 = \phi_1^{-1}(\mathcal{F}), \quad \mathcal{A}_2 = \phi_2^{-1}(\mathcal{G}),$$
$$\mu_1(x, \theta) = \delta_x \times Q_x, \quad \mu_2(x, \theta) = P_\theta \times \delta_\theta,$$

where $\phi_1(x, \theta) = x$ and $\phi_2(x, \theta) = \theta$ for all $(x, \theta) \in \mathcal{X} \times \Theta$. Condition (3) trivially holds. Hence, consistency of $\mathcal{Q}$ with $\mathcal{P}$ can be checked by part (b) of Theorem 6 if at least one between $(\mathcal{X}, \mathcal{F})$ and $(\Theta, \mathcal{G})$ is a pseudocompact space equipped with the Baire σ-field. This fact, however, is basically known; see Corollary 3.1 of [6].

Example 12. (Predictive inference). For each $n \in I := \{1, 2, \ldots\}$, a point x_n is observed in a measurable space $(\mathcal{X}_n, \mathcal{F}_n)$. The problem is to make inference on

$(x_{n+1}, x_{n+2}, \ldots)$, conditionally on $(x_1, \ldots, x_n)$, in a sequential framework. Define $\Omega = \prod_{i=1}^{\infty} \mathcal{X}_i$, $\mathcal{B} = \otimes_{i=1}^{\infty} \mathcal{F}_i$ and $\mathcal{A}_n = \sigma(X_1, \ldots, X_n)$, where $X_i(\omega) = x_i$ for all $\omega = (x_1, \ldots, x_i, \ldots) \in \Omega$. Also, for each $n \geq 1$, select a measurable collection

$$\mathcal{P}_n = \{ P_n(\cdot \mid x_1, \ldots, x_n) : (x_1, \ldots, x_n) \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_n \}$$

of laws on $\otimes_{i>n} \mathcal{F}_i$. Measurability means that $(x_1, \ldots, x_n) \mapsto P_n\big(B \mid x_1, \ldots, x_n\big)$ is measurable for fixed $B \in \otimes_{i>n} \mathcal{F}_i$. Each $P_n(\cdot \mid x_1, \ldots, x_n)$ should be regarded as the conditional distribution of $(X_{n+1}, X_{n+2}, \ldots)$ given that $X_1 = x_1, \ldots, X_n = x_n$. Note that, even if a parameter space $(\Theta, \mathcal{G})$ is available, we do not assess any prior on $\mathcal{G}$. To test consistency of $\{\mathcal{P}_n : n \geq 1\}$, define

$$\mu_n(\omega) = \delta_{(x_1, \ldots, x_n)} \times P_n(\cdot \mid x_1, \ldots, x_n) \quad \text{where } \omega = (x_1, \ldots, x_n, \ldots) \in \Omega.$$

Again, condition (3) is trivially true. Thus, by Corollary 3, the μ_n are consistent provided condition (5) holds for some $\omega_0 \in \Omega$. If the $\mathcal{F}_n$ are countably generated, existence of ω_0 satisfying (5) is necessary for consistency as well.

Acknowledgment: This note benefited from the helpful suggestions of two anonymous referees.

References

[1] Berti P., Rigo P. (1996) On the existence of inferences which are consistent with a given model, *Ann. Statist.*, 24, 1235-1249.

[2] Berti P., Rigo P. (2000) Integral representation of linear functionals on spaces of unbounded functions, *Proc. Amer. Math. Soc.*, 128, 3251-3258.

[3] Georgii H.O. (2011) *Gibbs measures and phase transitions (Second Edition)*, de Gruyter Studies in Mathematics, 9.

[4] Heath D., Sudderth W.D. (1978) On finitely additive priors, coherence and extended admissibility, *Ann. Statist.*, 6, 333-345.

[5] Kreps D.M. (1981) Arbitrage and equilibrium in economics with infinitely many commodities, *J. Math. Econ.*, 8, 15-35.

[6] Lane D.A., Sudderth W.D. (1983) Coherent and continuous inference, *Ann. Statist.*, 11, 114-120.

[7] Regazzini E. (1987) de Finetti's coherence and statistical inference, *Ann. Statist.*, 15, 845-864.

Patrizia Berti, Dipartimento di Matematica Pura ed Applicata "G. Vitali", Universita' di Modena e Reggio-Emilia, via Campi 213/B, 41100 Modena, Italy

E-mail address: patrizia.berti@unimore.it

Emanuela Dreassi, Dipartimento di Statistica "G. Parenti", Universita' di Firenze, viale Morgagni 59, 50134 Firenze, Italy

E-mail address: dreassi@ds.unifi.it

Pietro Rigo (corresponding author), Dipartimento di Economia Politica e Metodi Quantitativi, Universita' di Pavia, via S. Felice 5, 27100 Pavia, Italy

E-mail address: prigo@eco.unipv.it