A notion of conditional probability and some of its consequences

Patrizia Berti[1] · Emanuela Dreassi[2] · Pietro Rigo[3]

Abstract

An alternative notion of conditional probability (say AN) is discussed and investigated. If compared with the usual notion (regular conditional distributions), AN gives up the measurability constraint but requires a properness condition. An existence result for AN is provided. Also, some consequences of AN are pointed out, with reference to Bayesian statistics, exchangeability and compatibility.

Keywords Bayesian inference · Conditional probability · Disintegrability · Regular conditional distribution · Strategy · Sufficiency

Mathematics Subject Classification 60A05 · 60A10 · 91B06 · 62F15

JEL Classification C11 · C44 · C6

✉ Emanuela Dreassi
  emanuela.dreassi@unifi.it

  Patrizia Berti
  patrizia.berti@unimore.it

  Pietro Rigo
  pietro.rigo@unipv.it

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

[2] Dipartimento di Statistica, Informatica, Applicazioni "G. Parenti", Università degli Studi di Firenze, viale Morgagni 59, 50134 Firenze, Italy

[3] Pietro Rigo, Dipartimento di Matematica "F. Casorati", Universita' di Pavia, via Ferrata 1, 27100 Pavia, Italy

1 Introduction

This note is split into two parts: The first (Sect. 2) deals with conditional probability, from a general point of view, while the second (Sects. 3 and 4) highlights some consequences of adopting an alternative notion of conditional probability.

Let us call SN the standard notion of conditional probability (i.e., regular conditional distributions) and AN the alternative notion quoted above. Roughly speaking, AN is obtained from SN giving up the measurability constraint and adding a properness condition. As easily expected, this has both advantages and disadvantages. One major drawback of AN is that essential uniqueness is lost. This is certainly disappointing, but possibly not so crucial in the subjective view of probability. As to the advantages, AN allows to overcome various paradoxes occurring with SN. This is because, thanks to properness, one is actually conditioning on events (and not on sub-σ-fields, as it happens under SN).

Finally, among the possible consequences of AN, we focus on those related to Bayesian statistics, exchangeability and compatibility.

2 Conditional probability

In the sequel, $(\Omega, \mathcal{A}, P)$ is a probability space, $\mathcal{G} \subset \mathcal{A}$ a sub-σ-field, and

$$Q = \{Q(\omega) : \omega \in \Omega\}$$

a collection of probability measures on $\mathcal{A}$. We denote by $Q(\omega, A)$ the value of $Q(\omega)$ at $A \in \mathcal{A}$. Also, $\sigma(Q)$ is the σ-field on Ω generated by the maps $\omega \mapsto Q(\omega, A)$ for all $A \in \mathcal{A}$.

In this notation, Q is a *regular conditional distribution* (r.c.d.) given $\mathcal{G}$ if

(*a*) $\sigma(Q) \subset \mathcal{G}$;
(*b*) $P(A \cap B) = \int_B Q(\omega, A) \, P(\mathrm{d}\omega)$ for all $A \in \mathcal{A}$ and $B \in \mathcal{G}$.

An r.c.d. can fail to exist. However, it exists and is a.s. unique under reasonable conditions, such as $\mathcal{A}$ countably generated and P perfect; see, e.g., Jirina (1954).

(We recall that P is perfect if, for each $\mathcal{A}$-measurable $f : \Omega \to \mathbb{R}$, there is $I \in \mathcal{B}(\mathbb{R})$ such that $I \subset f(\Omega)$ and $P(f \in I) = 1$. If Ω is separable metric and $\mathcal{A} = \mathcal{B}(\Omega)$, perfectness is equivalent to tightness.)

This is the standard notion of conditional probability, based on Kolmogorov's axioms and adopted almost universally. Indeed, apart from rare exceptions, a conditional probability is meant as an r.c.d.

Using r.c.d.'s, however, one is conditioning on a σ-field and not on a specific event. What does it mean? What is the information provided by a σ-field? According to the usual naive interpretation, the information provided by $\mathcal{G}$ is:

(*) *For each event $B \in \mathcal{G}$, it is known whether B is true or false*.

Attaching interpretation (*) to r.c.d.'s is quite dangerous.

Example 1 (Continuous time processes) Let $X = \{X_t : t \geq 0\}$ be a real-valued process on $(\Omega, \mathcal{A}, P)$, adapted to a filtration $\{\mathcal{F}_t : t \geq 0\}$, and let $\mathcal{X}$ be the set of all functions from $[0, \infty)$ into $\mathbb{R}$. Define $\mathcal{N} = \{A \in \mathcal{A} : P(A) = 0\}$ and suppose that

$$\mathcal{N} \subset \mathcal{F}_0 \quad \text{and} \quad \{X = x\} \in \mathcal{N} \quad \text{for each } x \in \mathcal{X}.$$

Even if very usual for continuous time processes, the above assumption conflicts with (*). In fact, since $\{X = x\} \in \mathcal{F}_0$ for each $x \in \mathcal{X}$, interpretation (*) would imply that the actual X-path is already known at time 0. See also (Berti and Rigo 2008, Example 3).

Example 2 (Borel–Kolmogorov paradoxes) Let X and Y be random variables on $(\Omega, \mathcal{A}, P)$ such that $\{X = x\} = \{Y = y\}$ for some x and y. Using r.c.d.'s, the conditional probability given $X = x$ is taken to be $P(\cdot \mid X = x) = Q_X(\omega)$, where Q_X is an r.c.d. given $\sigma(X)$ and $\omega \in \Omega$ is such that $X(\omega) = x$. Similarly, $P(\cdot \mid Y = y) = Q_Y(\omega)$ where Q_Y is an r.c.d. given $\sigma(Y)$ and $Y(\omega) = y$. But since X and Y are different, it may be that $P(\cdot \mid X = x) \neq P(\cdot \mid Y = y)$ even if $\{X = x\} = \{Y = y\}$.

Example 3 (Properness) For interpretation (*) to make sense, Q should be *everywhere proper*, in the sense that

$$Q(\omega) = \delta_\omega \text{ on } \mathcal{G} \text{ for each } \omega \in \Omega.$$

In that case,

$$B = \{\omega \in \Omega : Q(\omega, B) = 1\} \in \sigma(Q) \quad \text{for each } B \in \mathcal{G},$$

so that $\mathcal{G} = \sigma(Q)$. Also, $\sigma(Q)$ is countably generated whenever $\mathcal{A}$ is countably generated. Thus, Q fails to be everywhere proper if $\mathcal{A}$ is countably generated, but $\mathcal{G}$ is not. A weaker notion of properness is

$$Q(\omega) = \delta_\omega \text{ on } \mathcal{G} \text{ for each } \omega \in B_0 \tag{1}$$

where $B_0 \in \mathcal{G}$ and $P(B_0) = 1$. But even condition (1) typically fails unless $\mathcal{G}$ is countably generated. In fact, condition (1) holds if and only if $\mathcal{G} \cap B_0$ is countably generated for some $B_0 \in \mathcal{G}$ with $P(B_0) = 1$; see Berti and Rigo (2007).

A seminal paper on properness is Blackwell and Dubins (1975). Other related references are Berti and Rigo (2007), Berti and Rigo (2008), Maitra and Ramakrishnan (1988).

To make interpretation (*) effective, the notion of r.c.d. is to be modified. We first recall that, for each $\omega \in \Omega$, the $\mathcal{G}$-atom including ω is

$$H(\omega) = \bigcap_{\omega \in B \in \mathcal{G}} B.$$

We also let

$$\Pi = \{H(\omega) : \omega \in \Omega\}.$$

Note that Π is a partition of Ω and each element of $\mathcal{G}$ is a union of elements of Π.

Say that Q is a *strategy* given $\mathcal{G}$, or a $\mathcal{G}$-strategy, if

(a*) $Q(x) = Q(y)$ whenever $x, y \in \Omega$ and $H(x) = H(y)$;
(b*) There is a probability measure $\widehat{P}$ on $\sigma(Q)$ such that

$$P(A) = \int Q(\omega, A) \, \widehat{P}(d\omega) \quad \text{for all } A \in \mathcal{A};$$

(c) Q is everywhere proper, i.e., $Q(\omega) = \delta_\omega$ on $\mathcal{G}$ for each $\omega \in \Omega$.

The above notion of $\mathcal{G}$-strategy is inspired to Blackwell and Dubins (1975) while the term "strategy" is borrowed from Dubins (1975).

Some obvious properties of $\mathcal{G}$-strategies are collected in the next lemma.

Lemma 4 *Let Q be a $\mathcal{G}$-strategy. Then, $\mathcal{G} \subset \sigma(Q)$ and*

$$P(A \cap B) = \int_B Q(\omega, A) \, \widehat{P}(d\omega) \quad \text{for all } A \in \mathcal{A} \text{ and } B \in \mathcal{G}.$$

In particular, $\widehat{P} = P$ on $\mathcal{G}$. Moreover, $\widehat{P} = P$ on $\mathcal{A} \cap \sigma(Q)$ provided $\Pi \subset \mathcal{G}$.

Proof Let $B \in \mathcal{G}$. By (c), $B = \{\omega : Q(\omega, B) = 1\} \in \sigma(Q)$. Further,

$$P(A \cap B) = \int Q(\omega, A \cap B) \, \widehat{P}(d\omega) = \int_B Q(\omega, A) \, \widehat{P}(d\omega) \quad \text{for each } A \in \mathcal{A},$$

where the first equality is by (b^*) and the second by (c). Finally, suppose $\Pi \subset \mathcal{G}$ and fix $A \in \mathcal{A} \cap \sigma(Q)$. By (a^*), each element of $\sigma(Q)$ is a union of $\mathcal{G}$-atoms. Hence,

$$Q(\omega, A) \geq Q(\omega, H(\omega)) = \delta_\omega(H(\omega)) = 1 \quad \text{whenever } \omega \in A.$$

Similarly, $Q(\omega, A) = 0$ if $\omega \notin A$. Therefore,

$$P(A) = \int Q(\omega, A) \, \widehat{P}(d\omega) = \int 1_A(\omega) \, \widehat{P}(d\omega) = \widehat{P}(A).$$

$\square$

By Lemma 4, a $\mathcal{G}$-strategy Q satisfies condition (b) whenever $\sigma(Q) \subset \mathcal{G}$. Generally, however, $\omega \mapsto Q(\omega, A)$ is not $\mathcal{G}$-measurable (or even $\mathcal{A}$-measurable) and cannot be integrated against P. This is the reason for a mixing measure $\widehat{P}$ is involved in condition (b^*).

Condition (a^*) is a weaker version of (a) (it is in fact a consequence of (a)). Roughly speaking, the motivation of (a^*) is that, conditionally on $\mathcal{G}$, one is actually observing

an element of the partition Π rather than a point of Ω. Thus, x and y provide the same information if $H(x) = H(y)$.

Essentially, a $\mathcal{G}$-strategy depends on $\mathcal{G}$ only through its atoms. In particular, if $\mathcal{G}$ includes the singletons, then $\Pi = \{\{\omega\} : \omega \in \Omega\} \subset \mathcal{G}$ and the only $\mathcal{G}$-strategy is $Q(\omega) = \delta_\omega$ on $\mathcal{A}$ for all $\omega \in \Omega$. As an example, take $\mathcal{G} = \{A \in \mathcal{A} : P(A) \in \{0, 1\}\}$ and suppose that $\{\omega\} \in \mathcal{A}$ and $P(\{\omega\}) = 0$ for every $\omega \in \Omega$. Then, $\mathcal{G}$ includes the singletons, so that $Q(\omega) = \delta_\omega$ is the only $\mathcal{G}$-strategy, while an r.c.d. given $\mathcal{G}$ is $Q(\omega) = P$. As a further example, take another sub-σ-field $\mathcal{F} \subset \mathcal{A}$. If $\mathcal{F}$ has the same atoms as $\mathcal{G}$ and $\Pi \subset \mathcal{F} \cap \mathcal{G}$, then Q is an $\mathcal{F}$-strategy if and only if is a $\mathcal{G}$-strategy.

A last remark is that $\mathcal{G}$-strategies are not uniquely determined by P. In particular, they are not essentially unique. This is technically a drawback, as well as a major difference with r.c.d.'s. However, in the subjective view of probability, non uniqueness is possibly not so crucial. In a sense, just as the choice of P is subjective, the choice of Q (once P is given) can be seen as a subjective act as well.

Let us turn now to existence issues. Recall that $(\Omega, \mathcal{A})$ is a *standard space* if Ω is a Borel subset of a Polish space and $\mathcal{A} = \mathcal{B}(\Omega)$. For an r.c.d. given $\mathcal{G}$ to exist, it suffices that $(\Omega, \mathcal{A})$ is a standard space. Instead, for a $\mathcal{G}$-strategy to exist, one needs conditions on both $(\Omega, \mathcal{A})$ and $\mathcal{G}$. The next statement is a translation of some results from Berti and Rigo (1999), Berti and Rigo (2002) concerning existence of disintegrations.

Theorem 5 *Let*

$$G = \big\{(x, y) \in \Omega \times \Omega : H(x) = H(y)\big\}.$$

There is a $\mathcal{G}$-strategy provided $(\Omega, \mathcal{A})$ is a standard space and at least one of the following conditions is satisfied:

  (i) *G is a co-analytic subset of $\Omega \times \Omega$;*
 (ii) *G is an analytic subset of $\Omega \times \Omega$ and all but countably many elements of Π are F_σ or G_δ.*

Proof In view of (Berti and Rigo 1999, Theorem 2) and (Berti and Rigo 2002, Theorem 8) under (i) or (ii), P admits a σ-additive disintegration on the partition Π. This means that, under (i) or (ii), there is a pair (α, β) such that:

• $\alpha(\cdot \mid H)$ is a probability measure on $\sigma(\mathcal{A} \cup \Pi)$ such that $\alpha(H|H) = 1$ for each $H \in \Pi$;
• β is a probability measure on $\sigma(\alpha)$, where $\sigma(\alpha)$ is the σ-field over Π generated by the maps $H \mapsto \alpha(A|H)$ for all $A \in \mathcal{A}$;
• $P(A) = \int_\Pi \alpha(A|H)\, \beta(dH)$ for all $A \in \mathcal{A}$.

Given such (α, β), to obtain a $\mathcal{G}$-strategy, it suffices to let

$$Q(\omega, A) = \alpha\big(A \mid H(\omega)\big) \quad \text{for all } \omega \in \Omega \text{ and } A \in \mathcal{A}.$$

In fact, Q meets (a^*) and (c) (to check (c), just recall that each member of $\mathcal{G}$ is a union of elements of Π). To prove (b^*), for each $S \subset \Pi$, denote by S^* the subset of

Ω obtained as the union of the elements of S. Then, $\sigma(Q) = \{S^* : S \in \sigma(\alpha)\}$. Thus, letting $\widehat{P}(S^*) = \beta(S)$, one trivially obtains

$$P(A) = \int_{\Pi} \alpha(A|H)\,\beta(dH) = \int_{\Omega} Q(\omega, A)\,\widehat{P}(d\omega) \quad \text{for all } A \in \mathcal{A}.$$

$\square$

Theorem 5 implies that a $\mathcal{G}$-strategy exists whenever $(\Omega, \mathcal{A})$ is a standard space and G is a Borel subset of $\Omega \times \Omega$. This happens in several meaningful situations, including the cases where $\mathcal{G}$ is a tail or a symmetric sub-σ-field. In these cases, thus, a $\mathcal{G}$-strategy is available while a proper r.c.d. fails to exist in general; see Blackwell and Dubins (1975) and Example 7.

To close this section, it would be nice to exhibit an example where a $\mathcal{G}$-strategy fails to exist. If $\Pi \subset \mathcal{G}$ and $(\Omega, \mathcal{A})$ is a standard space, however, such example is not available under the usual axioms of set theory (the so called ZFC set theory). Take in fact $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}([0, 1])$, and consider the assertion:

"*For every Borel partition Ψ of $[0, 1]$, the Lebesgue measure on $\mathcal{A}$ admits a strategy given $\sigma(\Psi)$*".

Then, as shown by Dubins and Prikry (1995, Theorem 2), such an assertion is undecidable in ZFC, in the sense that the assertion and its negation are both consistent with ZFC.

Incidentally, as regards existence and nonexistence of $\mathcal{G}$-strategies, things are quite different in a finitely additive framework; see, e.g., Dubins (1975) and Prikry and Sudderth (1982).

3 Bayesian statistical inference

Let $(\mathcal{X}, \mathcal{E})$ and $(\Theta, \mathcal{F})$ be measurable spaces to be regarded, respectively, as the sample space and the parameter space. For the sake of simplicity, the $\mathcal{E}$-atoms are assumed to be the singletons. A *statistical model* is a measurable collection

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\}$$

of probability measures on $\mathcal{E}$, where measurability means that $\theta \mapsto P_\theta(A)$ is $\mathcal{F}$-measurable for each $A \in \mathcal{E}$. A *prior* is a probability measure on $\mathcal{F}$.

Roughly speaking, the problem is to make inference on the parameter θ given the data x. To this end, in the notation of Sect. 2, one lets

$$(\Omega, \mathcal{A}) = (\mathcal{X} \times \Theta, \mathcal{E} \otimes \mathcal{F})$$

and takes $\mathcal{G}$ to be the sub-σ-field of $\mathcal{A}$ generated by the data, namely

$$\mathcal{G} = \{A \times \Theta : A \in \mathcal{E}\}.$$

Since the $\mathcal{E}$-atoms are the singletons, the partition of Ω in the $\mathcal{G}$-atoms is

$$\Pi = \big\{\{x\} \times \Theta : x \in \mathcal{X}\big\}.$$

Also, given a statistical model $\mathcal{P}$ and a prior ν, the reference probability measure P on $\mathcal{A}$ is

$$P(C) = \int \int 1_C(x, \theta) \, P_\theta(\mathrm{d}x) \, \nu(\mathrm{d}\theta) \quad \text{for all } C \in \mathcal{A}.$$

In this framework, a *posterior* is a conditional probability for P given $\mathcal{G}$. Thus, technically, how to define a posterior depends on the adopted notion of conditional probability. Let

$$\mathcal{Q} = \{Q_x : x \in \mathcal{X}\}$$

be a collection of probability measures on $\mathcal{F}$, and let $\sigma(\mathcal{Q})$ be the σ-field over $\mathcal{X}$ generated by the maps $x \mapsto Q_x(B)$ for all $B \in \mathcal{F}$.

As noted in Sect. 2, a conditional probability is usually meant as an r.c.d. In that case, $\mathcal{Q}$ is a posterior provided

$$\sigma(\mathcal{Q}) \subset \mathcal{E} \quad \text{and} \quad P(C) = \int \int 1_C(x, \theta) \, Q_x(\mathrm{d}\theta) \, m_\nu(\mathrm{d}x) \quad \text{for all } C \in \mathcal{A}$$

where

$$m_\nu(A) = P(A \times \Theta) = \int P_\theta(A) \, \nu(\mathrm{d}\theta) \quad \text{for all } A \in \mathcal{E}.$$

Instead, if a conditional probability is meant as a strategy, $\mathcal{Q}$ is a posterior whenever

$$P(C) = \int \int 1_C(x, \theta) \, Q_x(\mathrm{d}\theta) \, m(\mathrm{d}x), \quad C \in \mathcal{A}, \tag{2}$$

where m is any probability measure on $\sigma(\mathcal{Q})$. Note that Lemma 4 yields $m = m_\nu$ on $\mathcal{E}$, so that m is actually an extension of m_ν.

Therefore, the class of posteriors becomes larger if conditional probabilities are meant as strategies and not as r.c.d.'s. Indeed, for $\mathcal{Q}$ to be a posterior, equation (2) is enough and no measurability constraints are required to $\mathcal{Q}$. This fact has some consequences.

In the next result, a posterior is actually meant as a $\mathcal{G}$-strategy, namely a collection $\mathcal{Q} = \{Q_x : x \in \mathcal{X}\}$ of probability measures on $\mathcal{F}$ satisfying equation (2) for some m.

Theorem 6 *Let $\mathcal{P}$ be a statistical model, ν a prior probability on $\mathcal{F}$ and $Y : (\mathcal{X}, \mathcal{E}) \to (\mathcal{Y}, \mathcal{H})$ a measurable map. Suppose:*

- *card $(\mathcal{E}) \le$ card $(\mathbb{R})$, card $(\mathcal{F}) \le$ card $(\mathbb{R})$, and $\mathcal{H}$ is countably generated and includes the singletons;*

- *P_θ is a perfect probability measure such that $P_\theta(Y = y) = 0$ for all $\theta \in \Theta$ and $y \in \mathcal{Y}$.*

Then, there is a posterior $\mathcal{Q} = \{Q_x : x \in \mathcal{X}\}$ such that

$$x_1, \ x_2 \in \mathcal{X} \quad and \quad Y(x_1) = Y(x_2) \quad \Rightarrow \quad Q_{x_1} = Q_{x_2}.$$

Proof Two known facts are to be recalled. Let $(D, \mathcal{D}, \mu)$ be any probability space.

  (j) If $\mathcal{D}$ is countably generated, μ is perfect and $\mu(F) = 0$ for each $\mathcal{D}$-atom F, then the collection of $\mathcal{D}$-atoms has the cardinality of the continuum; see (Berti and Rigo 1996, Lemma 2.3);
  (jj) Let Γ be a class of probability measures on $\mathcal{D}$ and Σ the σ-field over Γ generated by the maps $\gamma \mapsto \gamma(D)$ for all $D \in \mathcal{D}$. Suppose $\mu(D) = \int_\Gamma \gamma(D) \, \beta(d\gamma)$ for all $D \in \mathcal{D}$, where β is a finitely additive probability on Σ. Then, β is σ-additive provided each $\gamma \in \Gamma$ is 0-1-valued; see Theorem 11 and Example 15 of Berti and Rigo (2018).

Next, recall that $(\Omega, \mathcal{A}) = (\mathcal{X} \times \Theta, \mathcal{E} \otimes \mathcal{F})$ and define

$$\mathcal{V} = \{C \in \mathcal{A} : P(C) > 0\} \quad \text{and}$$
$$L(C) = \big\{y \in \mathcal{Y} : (x, \theta) \in C \text{ for some } (x, \theta) \in \Omega \text{ with } Y(x) = y\big\} \quad \text{where } C \subset \Omega.$$

This proof is split into two parts: First, we prove the theorem under the assumption

$$\text{card}\,(L(C)) \geq \ \text{card}\,(\mathbb{R}) \quad \text{for all } C \in \mathcal{V}, \tag{3}$$

and then, we show that (3) is actually true.

Suppose condition (3) holds. Then, since $\text{card}\,(\mathcal{A}) \leq \text{card}\,(\mathbb{R})$, one obtains

$$\text{card}\,(L(C)) \geq \ \text{card}\,(\mathcal{A}) \geq \ \text{card}\,(\mathcal{V}) \quad \text{for all } C \in \mathcal{V}.$$

Hence, there is an injective map $f : \mathcal{V} \to \mathcal{Y}$ such that $f(C) \in L(C)$ for each $C \in \mathcal{V}$; see (Berti and Rigo 1996, Lemma 2.1). For each $y \in \mathcal{Y}$, select a probability measure U_y on $\mathcal{F}$ as follows: If y is not in the range of f, define $U_y = \delta_{\theta_0}$ where $\theta_0 \in \Theta$ is arbitrary. Otherwise, if $y = f(C)$ for some (unique) $C \in \mathcal{V}$, take $(x, \theta) \in C$ with $Y(x) = y$ and set $U_y = \delta_\theta$. For $x \in \mathcal{X}$, define also

$$Q_x = U_{Y(x)} \quad \text{and} \quad T_x(C) = Q_x\{\theta \in \Theta : (x, \theta) \in C\} \quad \text{for all } C \in \mathcal{A}.$$

Then, T_x is a probability measure on $\mathcal{A}$ such that $T_x\big(\{x\} \times \Theta\big) = 1$. Further,

$$\text{for each } C \in \mathcal{V}, \text{ there is } x \in \mathcal{X} \text{ such that } T_x(C) = 1.$$

By (Berti and Rigo 1996, Lemma 2.2) and the above condition, there is a finitely additive probability m_0 on the power set of $\mathcal{X}$ such that

$$\int \int 1_C(x, \theta) \, Q_x(d\theta) \, m_0(dx) = \int T_x(C) \, m_0(dx) = P(C) \quad \text{for all } C \in \mathcal{A}.$$

Let $Q = \{Q_x : x \in \mathcal{X}\}$ and let m be the restriction of m_0 on $\sigma(Q)$. Then, m is σ-additive because of (jj). Therefore, Q is a posterior such that $Q_{x_1} = Q_{x_2}$ whenever $Y(x_1) = Y(x_2)$. This concludes the first part of the proof.

Finally, we prove (3). It suffices to show that $P(C) = 0$ whenever $C \in \mathcal{A}$ and $\text{card}(L(C)) < \text{card}(\mathbb{R})$. Fix one such C and take $A \in \mathcal{E}$ with

$$A \subset \bigcup_{y \in L(C)} \{x \in \mathcal{X} : Y(x) = y\}.$$

Let $\mathcal{D} = A \cap \sigma(Y) = \{A \cap B : B \in \sigma(Y)\}$. Since $\mathcal{H}$ is countably generated, $\sigma(Y)$ is countably generated, which in turn implies that $\mathcal{D}$ is countably generated. Toward a contradiction, suppose $P_\theta(A) > 0$ for some $\theta \in \Theta$. Then, one can define

$$\mu(A \cap B) = \frac{P_\theta(A \cap B)}{P_\theta(A)} \quad \text{for all } B \in \sigma(Y).$$

Since P_θ is perfect, μ is a perfect probability measure on $\mathcal{D}$. Each atom F of $\mathcal{D}$ is of the form $F = A \cap \{Y = y\}$ for some y, and

$$\mu(F) = \frac{P_\theta(A \cap \{Y = y\})}{P_\theta(A)} = 0.$$

In view of (j), the set of $\mathcal{D}$-atoms has the cardinality of the continuum, so that

$$\text{card}(L(C)) \geq \text{card}\{y \in \mathcal{Y} : \{Y = y\} \cap A \neq \emptyset\} = \text{card}\{\mathcal{D}\text{-atoms}\} = \text{card}(\mathbb{R}).$$

This is a contradiction, since $\text{card}(L(C)) < \text{card}(\mathbb{R})$. Hence, it must be $P_\theta(A) = 0$ for all θ. To conclude the proof, just note that

$$\{x : (x, \theta) \in C\} \subset \bigcup_{y \in L(C)} \{x \in \mathcal{X} : Y(x) = y\} \quad \text{for all } \theta.$$

It follows that

$$P(C) = \int P_\theta\{x : (x, \theta) \in C\} \, \nu(\mathrm{d}\theta) = 0.$$

Hence, (3) holds and this concludes the proof.   □

Theorem 6 improves (Berti and Rigo 1996, Theorem 3.1) where the probability m involved in equation (2) is only finitely additive.

In the subjective framework, Theorem 6 has a nice interpretation in terms of sufficiency; see Berti and Rigo (1996). In fact, think of Y as a statistic. Also, given a posterior Q, say that Y is *sufficient for* Q if $Q_{x_1} = Q_{x_2}$ whenever $Y(x_1) = Y(x_2)$. Then, Theorem 6 essentially states that, for any prior ν and any statistic Y, there is a posterior Q which makes Y sufficient provided only that $P_\theta(Y = y) = 0$ for all θ and y. This seems in line with both the substantial meaning of sufficiency and the

subjective view of probability. Indeed, the assessment of $\mathcal{Q}$ can be split into two steps: First, the inferrer selects a partition of $\mathcal{X}$, by grouping those samples which, according to him/her, have the same inferential content. This step precisely amounts to the choice of a sufficient statistic Y. Subsequently, a probability law on $\mathcal{F}$ is attached to every element in the partition. If no such element has positive probability under the statistical model, Theorem 6 implies that at least a posterior $\mathcal{Q}$ is consistent with this procedure.

In addition to sufficiency, another intriguing point is whether improper priors can be recovered when posteriors are regarded as strategies. This issue is actually connected to compatibility. Thus, improper priors are postponed to Example 9.

4 Further consequences

In principle, in every framework where conditional probability plays a role, things are quite different according to whether conditional probability is meant as an r.c.d. or as a strategy. In Bayesian statistics, for instance, Theorem 6 would not be available if a posterior would be regarded as an r.c.d. This section is in the spirit of the previous one; namely, the different behaviors of r.c.d.'s and strategies are compared in some special situations. Needless to say that many other analogous examples could be given.

Example 7 (Exchangeability) Let $(\Omega, \mathcal{A}) = (\mathcal{X}^\infty, \mathcal{E}^\infty)$ where $(\mathcal{X}, \mathcal{E})$ is a standard space. To each $n \in \mathbb{N}$ and each permutation $(\pi_1, \ldots, \pi_n)$ of $(1, \ldots, n)$, we can associate a function $f : \Omega \to \Omega$ defined by

$$f(x_1, x_2, \ldots) = (x_{\pi_1}, \ldots, x_{\pi_n}, x_{n+1}, x_{n+2}, \ldots) \qquad \text{for all } (x_1, x_2, \ldots) \in \Omega.$$

Let F denote the class of all such functions, for all $n \in \mathbb{N}$ and all permutations of $(1, \ldots, n)$. A probability measure P on $\mathcal{A}$ is exchangeable if $P \circ f^{-1} = P$ for all $f \in F$. The symmetric sub-σ-field is

$$\mathcal{G} = \{A \in \mathcal{A} : f^{-1}(A) = A \text{ for all } f \in F\}.$$

Note that the $\mathcal{G}$-atoms can be written as

$$H(\omega) = \{f(\omega) : f \in F\} \qquad \text{for all } \omega \in \Omega.$$

Suppose P exchangeable and $P(\Delta) = 0$, where $\Delta = \{(x, x, \ldots) : x \in \mathcal{X}\}$ is the diagonal. Since $(\mathcal{X}, \mathcal{E})$ is a standard space, there is an r.c.d. Q^* for P given $\mathcal{G}$. Since P is exchangeable, by de Finetti's theorem, $Q^*(\omega)$ is an i.i.d. probability measure on $\mathcal{A} = \mathcal{E}^\infty$ for almost all $\omega \in \Omega$. Now, an i.i.d. probability measure vanishes on singletons unless it is degenerate. Since $P(\Delta) = 0$ and $H(\omega)$ is countable, it follows that

$$Q^*(\omega, H(\omega)) = 0 \qquad \text{for almost all } \omega \in \Omega.$$

On the other hand, because of Theorem 5, P also admits a $\mathcal{G}$-strategy Q. By definition, Q satisfies

$$Q(\omega, H(\omega)) = 1 \qquad \text{for all } \omega \in \Omega.$$

Therefore, $Q(\omega)$ and $Q^*(\omega)$ are even singular for almost all $\omega \in \Omega$. Another curious fact is that $Q^*(\omega)$ can be shown to be $\{0, 1\}$-valued on $\mathcal{G}$, despite $Q^*(\omega, H(\omega)) = 0$, for almost all $\omega \in \Omega$; see Berti and Rigo (2008).

Example 8 (Compatibility) Let $(\Omega, \mathcal{A})$ be a measurable space, $\mathcal{G}_i \subset \mathcal{A}$ a sub-σ-field and $Q_i = \{Q_i(\omega) : \omega \in \Omega\}$ a collection of probability measures on $\mathcal{A}$, where $i = 1, 2$. Generally speaking, Q_1 and Q_2 are compatible if there is a probability measure P on $\mathcal{A}$ which admits Q_1 and Q_2 as conditional probabilities given $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively; see, e.g., Berti et al. (2014) and references therein. Once again, this general idea can be realized differently according to the selected notion of conditional probability.

If conditional probabilities are meant as r.c.d.'s, a necessary condition for compatibility is $\sigma(Q_1) \subset \mathcal{G}_1$ and $\sigma(Q_2) \subset \mathcal{G}_2$. In that case, Q_1 and Q_2 are compatible if there is a probability measure P on $\mathcal{A}$ such that

$$P(A \cap B) = \int_B Q_i(\omega, A)\, P(\mathrm{d}\omega) \qquad \text{whenever } i = 1, 2, \ B \in \mathcal{G}_i \text{ and } A \in \mathcal{A}. \quad (4)$$

If conditional probabilities are meant as strategies, the necessary condition for compatibility turns into

$$Q_i(\omega) = Q_i(\upsilon) \text{ if } H_i(\omega) = H_i(\upsilon) \quad \text{and} \quad Q_i(\omega) = \delta_\omega \text{ on } \mathcal{G}_i$$

for $i = 1, 2$ and all $\omega, \upsilon \in \Omega$, where $H_i(\omega)$ is the $\mathcal{G}_i$-atom including ω. Under such condition, Q_1 and Q_2 are compatible whenever

$$\int Q_1(\omega, A)\, \widehat{P}_1(\mathrm{d}\omega) = \int Q_2(\omega, A)\, \widehat{P}_2(\mathrm{d}\omega), \qquad A \in \mathcal{A}, \quad (5)$$

for some probability measures $\widehat{P}_1$ on $\sigma(Q_1)$ and $\widehat{P}_2$ on $\sigma(Q_2)$.

Condition (5) looks intriguing and possibly easier than (4) to work with.

A weaker version of (5) is obtained allowing $\widehat{P}_1$ and $\widehat{P}_2$ to be finitely additive probabilities. In that case, compatibility essentially reduces to a notion of consistency, introduced in Lane and Sudderth (1983) for Bayesian statistical inference; see also Heath and Sudderth (1978).

Example 9 (Improper priors) We adopt the notation and the assumptions of Sect. 3. In addition, we assume the model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ dominated, namely $P_\theta(\mathrm{d}x) = f(x, \theta)\,\lambda(\mathrm{d}x)$ for all $\theta \in \Theta$, where λ is a σ-finite measure on $\mathcal{E}$ and f a nonnegative measurable function on $(\Omega, \mathcal{A}) = (\mathcal{X} \times \Theta, \mathcal{E} \otimes \mathcal{F})$. An *improper prior* is a σ-finite measure γ on $\mathcal{F}$ such that $\gamma(\Theta) = \infty$. Let

$$\psi(x) = \int f(x, \theta)\,\gamma(\mathrm{d}\theta) \qquad \text{for all } x \in \mathcal{X}.$$

A standard practice is to fix an improper prior γ and to let

$$Q_x(\mathrm{d}\theta) = \frac{f(x, \theta)}{\psi(x)} \gamma(\mathrm{d}\theta) \quad \text{whenever } \psi(x) \in (0, \infty). \tag{6}$$

Notice that no prior probability on $\mathcal{F}$ has been selected. In the sequel, we assume $\psi(x) \in (0, \infty)$ for all $x \in \mathcal{X}$, and we let $\mathcal{Q} = \{Q_x : x \in \mathcal{X}\}$ with Q_x given by (6).

Define

$$\mathcal{G}_1 = \{A \times \Theta : A \in \mathcal{E}\} \quad \text{and} \quad \mathcal{G}_2 = \{\mathcal{X} \times B : B \in \mathcal{F}\}.$$

For $C \in \mathcal{A}$ and $\omega = (x, \theta) \in \Omega$, define also

$$T_1(\omega)(C) = Q_x\{t \in \Theta : (x, t) \in C\} \quad \text{and} \quad T_2(\omega)(C) = P_\theta\{z \in \mathcal{X} : (z, \theta) \in C\}.$$

Then, $\mathcal{G}_1$ and $\mathcal{G}_2$ are sub-σ fields of $\mathcal{A}$ while $T_1(\omega)$ and $T_2(\omega)$ are probability measures on $\mathcal{A}$. We say that $\mathcal{P}$ and $\mathcal{Q}$ are compatible to mean that $\mathcal{T}_1$ and $\mathcal{T}_2$ are compatible, where $\mathcal{T}_i = \{T_i(\omega) : \omega \in \Omega\}$ for $i = 1, 2$.

From the point of view of probability theory, using $\mathcal{Q}$ as a posterior makes sense only if $\mathcal{P}$ and $\mathcal{Q}$ are compatible; see also (Berti et al. 2014, Example 3). However, measurability of f implies $\sigma(\mathcal{T}_i) \subset \mathcal{G}_i$ for $i = 1, 2$. Therefore, as regards compatibility of $\mathcal{P}$ and $\mathcal{Q}$, using r.c.d.'s or strategies is equivalent. The situation is slightly different if the assumption $0 < \psi < \infty$ is dropped.

On the other hand, in most real problems, $\mathcal{P}$ and $\mathcal{Q}$ fail to be compatible. To get compatibility and thus to make improper priors admissible, finitely additive probabilities are to be involved; see Heath and Sudderth (1978), Heath and Sudderth (1989) and Lane and Sudderth (1983).

References

Berti, P., Rigo, P.: On the existence of inferences which are consistent with a given model. Ann. Stat. **24**, 1235–1249 (1996)

Berti, P., Rigo, P.: Sufficient conditions for the existence of disintegrations. J. Theor. Probab. **12**, 75–86 (1999)

Berti, P., Rigo, P.: On coherent conditional probabilities and disintegrations. Ann. Math. Artif. Intell. **35**, 71–82 (2002)

Berti, P., Rigo, P.: 0–1 laws for regular conditional distributions. Ann. Probab. **35**, 649–662 (2007)

Berti, P., Rigo, P.: A conditional 0–1 law for the symmetric sigma-field. J. Theor. Probab. **21**, 517–526 (2008)

Berti, P., Dreassi, E., Rigo, P.: Compatibility results for conditional distributions. J. Multivar. Anal. **125**, 190–203 (2014)

Berti, P., Rigo, P.: Finitely additive mixtures of probability measures, *submitted*, (2018) currently available at: http://www-dimat.unipv.it/~rigo/bprv.pdf

Blackwell, D., Dubins, L.E.: On existence and non-existence of proper, regular, conditional distributions. Ann. Probab. **3**, 741–752 (1975)

Dubins, L.E.: Finitely additive conditional probabilities, conglomerability and disintegrations. Ann. Probab. **3**, 88–99 (1975)

Dubins, L.E., Prikry, K.: On the existence of disintegrations. In: Azema, J., Emery, M., Meyer, P.-A., Yor, M (eds.) Seminaire de Probab. XXIX, Lect. Notes in Math., Springer, 1613, 248–259 (1995)

Heath, D., Sudderth, W.D.: On finitely additive priors, coherence and extended admissibility. Ann. Stat. **6**, 333–345 (1978)

Heath, D., Sudderth, W.D.: Coherent inference from improper priors and from finitely additive priors. Ann. Stat. **17**, 907–919 (1989)

Jirina, M.: Conditional probabilities on σ-algebras with countable basis. Czech. Math. J. 4(79), 372–380. English translation in: Selected Transl. Math. Stat. and Prob., Am. Math. Soc., 2, 79–86 (1962) (1954)

Lane, D.A., Sudderth, W.D.: Coherent and continuous inference. Ann. Stat. **11**, 114–120 (1983)

Maitra, A., Ramakrishnan, S.: Factorization of measures and normal conditional distributions. Proc. Am. Math. Soc. **103**, 1259–1267 (1988)

Prikry, K., Sudderth, W.D.: Singularity with respect to strategic measures, Illinois. J. Math. **43**, 139–153 (1982)