Dipartimento di Statistica
"Giuseppe Parenti"

WORKING PAPER 2012/08

Compatibility results for conditional distributions

Patrizia Berti,
Emanuela Dreassi, Pietro Rigo

Università degli Studi
di Firenze

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

COMPATIBILITY RESULTS FOR CONDITIONAL DISTRIBUTIONS

PATRIZIA BERTI, EMANUELA DREASSI, AND PIETRO RIGO

Abstract. In various frameworks, to assess the joint distribution of a k-dimensional random vector $X = (X_1, \ldots, X_k)$, one selects some putative conditional distributions $Q_1, \ldots, Q_k$. Each Q_i is regarded as a possible (or putative) conditional distribution for X_i given $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k)$. The Q_i are compatible if there is a joint distribution P for X with conditionals $Q_1, \ldots, Q_k$. Three types of compatibility results are given in this paper. First, the X_i are assumed to take values in compact subsets of $\mathbb{R}$. Second, the Q_i are supposed to have densities with respect to reference measures. Third, a stronger form of compatibility is investigated. Indeed, the law P with conditionals $Q_1, \ldots, Q_k$ is requested to be exchangeable.

1. Introduction

Let I be a countable index set and, for each $i \in I$, let X_i be a real random variable. Denote by $\mathcal{P}$ the set of all probability distributions for the process

$$X = (X_i : i \in I).$$

Also, for each $P \in \mathcal{P}$ and $H \subset I$ (with $H \neq \emptyset$ and $H \neq I$), denote by P_H the conditional distribution of

$$(X_i : i \in H) \quad \text{given} \quad (X_i : i \in I \setminus H) \quad \text{under } P.$$

The P_H are determined by P (up to P-null sets). In fact, to get P_H, the obvious strategy is to first select $P \in \mathcal{P}$ and then calculate P_H. Sometimes, however, this procedure is reverted. Let $\mathcal{H}$ be a class of subsets of I (all different from $\emptyset$ and I). One first selects a collection $\{Q_H : H \in \mathcal{H}\}$ of *putative* conditional distributions, and then looks for some $P \in \mathcal{P}$ inducing the Q_H as conditional distributions, in the sense that

$$(1) \qquad Q_H = P_H, \quad \text{a.s. with respect to } P, \text{ for all } H \in \mathcal{H}.$$

But such a P can fail to exist. Accordingly, a set $\{Q_H : H \in \mathcal{H}\}$ of putative conditional distributions is said to be *compatible*, or *consistent*, if there exists $P \in \mathcal{P}$ satisfying condition (1). (See Section 2 for formal definitions).

A natural version of the previous definition is the following. Fix a proper subset $\mathcal{P}_0 \subset \mathcal{P}$. For instance, $\mathcal{P}_0$ could be the set of those $P \in \mathcal{P}$ which make X exchangeable, or else which are absolutely continuous with respect to some reference measure. A natural question is whether there is $P \in \mathcal{P}_0$ with given conditional

2010 *Mathematics Subject Classification.* 62F15, 62H05, 62G99, 60A05, 60A10.

Key words and phrases. Compatibility of conditional distributions, Exchangeability, Gibbs measure, Gibbs sampling, Multiple imputation, Posterior distribution.

distributions $\{Q_H : H \in \mathcal{H}\}$. Thus, a set $\{Q_H : H \in \mathcal{H}\}$ of putative conditional distributions is $\mathcal{P}_0$-*compatible* if condition (1) holds for some $P \in \mathcal{P}_0$.

To better frame the problem, we next give three examples where compatibility issues arise.

Example 1. (Gibbs measures). Think of I as a lattice and of X_i as the spin at site $i \in I$. The equilibrium distribution of a finite system of statistical physics is generally believed to be the Boltzmann-Gibbs distribution. Thus, when I is finite, one can let

$$P(dx) = a \exp\Big\{ - b \sum_{H \subset I} U_H(x) \Big\} \lambda(dx)$$

where $a, b > 0$ are constants and λ is a suitable reference measure. Roughly speaking, $U_H(x)$ quantifies the energy contribution of the subsystem $(X_i : i \in H)$ at point x. This simple scheme breaks down when I is countably infinite. However, for each finite $H \subset I$, the Boltzmann-Gibbs distribution can still be attached to $(X_i : i \in H)$ conditionally on $(X_i : i \in I \setminus H)$. If Q_H denotes such Boltzmann-Gibbs distribution, we thus obtain a family $\{Q_H : H \text{ finite}\}$ of putative conditional distributions. But a law $P \in \mathcal{P}$ having the Q_H as conditional distributions can fail to exist. So, it is crucial to decide whether $\{Q_H : H \text{ finite}\}$ is compatible. See [10].

Example 2. (Complex data systems and Gibbs sampling). A joint modeling of a k-dimensional random vector $X = (X_1, \ldots, X_k)$ is often very hard. A conditional specification, which should capture the various features of X separately, may be more convenient. Well known examples are missing data imputation and spatial data modeling. In these cases, X is modeled by some collection $\{Q_H : H \in \mathcal{H}\}$ of putative conditional distributions. But of course this makes sense only if $\{Q_H : H \in \mathcal{H}\}$ is compatible. A similar example is the Gibbs sampler. Let $H_i = \{i\}$. For the Gibbs sampler to apply, one needs

$$P_{H_i}(\cdot) = P\big(X_i \in \cdot \mid X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k\big)$$

for $i = 1, \ldots, k$. The P_{H_i} are usually obtained from a given $P \in \mathcal{P}$. But sometimes P is not assessed. Rather, one selects a collection $\{Q_{H_i} : i = 1, \ldots, k\}$ of putative conditional distributions and use Q_{H_i} in the place of P_{H_i}. Again, this makes sense only if $\{Q_{H_1}, \ldots, Q_{H_k}\}$ is compatible. See [6], [7], [13], [15], [18], [19] and references therein.

Example 3. (Bayesian inference). Let $X = (X_1, \ldots, X_n, \ldots, X_m)$. Think of $Y = (X_1, \ldots, X_n)$ as the data and of $\Theta = (X_{n+1}, \ldots, X_m)$ as a random parameter. As usual, a *prior* is a marginal distribution for Θ, a *statistical model* a conditional distribution for Y given Θ, and a *posterior* a conditional distribution for Θ given Y. The statistical model, say Q_Y, is supposed to be assigned. Then, the standard Bayes scheme is to select a prior μ, to obtain the joint distribution of (Y, Θ), and to calculate (or to approximate) the posterior. To assess μ is typically very arduous. Sometimes, it may be convenient to avoid the choice of μ and to assign directly a putative conditional distribution Q_Θ, to be viewed as the posterior.

The alternative Bayes scheme sketched above is not unusual. Suppose Q_Θ is the formal posterior of an improper prior, or it is obtained by some empirical Bayes method, or else it is a fiducial distribution. In all these cases, Q_Θ is assessed without explicitly selecting any (proper) prior. Such a Q_Θ may look reasonable or not (there are indeed different opinions). But a basic question is whether Q_Θ is the actual posterior of Q_Y and some prior μ, or equivalently, whether Q_Y and Q_Θ are compatible.

Incidentally, the alternative Bayes scheme agrees with the subjective view of probability and has been investigated in a coherence framework; see [3], [12], [14], [16] and references therein. However, in a coherence framework, the compatibility of Q_Y and Q_Θ is studied in a finitely additive setting.

Other significant compatibility examples are in [8], [11], [17], [20].

This paper includes three different types of compatibility results. We always focus on finite I, say $I = \{1, \ldots, k\}$, and we let $H_i = \{i\}$ for $i = 1, \ldots, k$. Most results hold for arbitrary $k \geq 2$, even if they take a nicer form for low values of k.

In Section 3, each X_i (or each X_i but one) takes values in a compact subset of the real line. Then, necessary and sufficient conditions for compatibility are obtained as a consequence of a general result in [5].

In Section 4, as in most real problems, the Q_{H_i} have densities with respect to reference measures. Under this assumption, compatibility is characterized by Theorem 10. The latter result extends to any $k \geq 2$ a well known criterion which holds for $k = 2$. See [1], [2] and Remark 9.

Finally, $\mathcal{P}_0$-compatibility is concerned in Section 5. Various conditions for $\mathcal{P}_0$-compatibility are provided in case $\mathcal{P}_0 = \{P \in \mathcal{P} : X \text{ exchangeable under } P\}$.

2. Notation and basic definitions

In the rest of this paper, we let

$$I = \{1, \ldots, k\} \quad \text{and} \quad H_i = \{i\} \text{ for } i = 1, \ldots, k.$$

With reference to such a case, we next make precise some definitions informally given in Section 1.

Since we are only concerned with distributions (both conditional and unconditional) the X_i can be taken to be coordinate random variables. Thus, for each i, we fix a Borel set $\Omega_i \subset \mathbb{R}$ to be regarded as the collection of "admissible" values for X_i (possibly, $\Omega_i = \mathbb{R}$). We define $\Omega = \prod_{j=1}^k \Omega_j$ and we take X_i to be the i-th coordinate map on Ω. We define also

$$Y_i = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k) \quad \text{and} \quad \mathcal{Y}_i = \prod_{j \neq i} \Omega_j.$$

The following notation will be often used. Let $i \in I$, $x \in \Omega_i$ and $y \in \mathcal{Y}_i$. Then, (x, y) denotes that point $\omega \in \Omega$ such that $X_i(\omega) = x$ and $Y_i(\omega) = y$.

For any topological space S, we let $\mathcal{B}(S)$ denote the Borel σ-field on S. Also, if μ and ν are measures on the same σ-field, $\mu \ll \nu$ means that $\mu(A) = 0$ whenever A is measurable and $\nu(A) = 0$, and $\mu \sim \nu$ stands for $\mu \ll \nu$ and $\nu \ll \mu$.

A probability distribution for

$$X = (X_1, \ldots, X_k)$$

is a probability measure on $\mathcal{B}(\Omega)$. Let $\mathcal{P}$ denote the set of all such probability measures. Fix $P \in \mathcal{P}$ and $i \in I$. The conditional distribution of X_i given Y_i, under P, is a function P_i of the pair (y, A), where $y \in \mathcal{Y}_i$ and $A \in \mathcal{B}(\Omega_i)$, satisfying

(i) $A \mapsto P_i(y, A)$ is a probability measure for fixed y;

(ii) $y \mapsto P_i(y, A)$ is a Borel measurable function for fixed A;

(iii) $E_P\Big\{I_B(Y_i)\, P_i(Y_i,\, A)\Big\} = P(X_i \in A,\, Y_i \in B)$ for $A \in \mathcal{B}(\Omega_i)$ and $B \in \mathcal{B}(\mathcal{Y}_i)$.

Such P_i is P-essentially unique. Clearly, $P_i(y, A)$ should be regarded as the conditional probability of $\{X_i \in A\}$ given that $Y_i = y$ under P.

A putative conditional distribution is a function Q_i, with the same domain as P_i, satisfying conditions (i)-(ii) but not necessarily (iii). In the sequel,

$$Q_1, \ldots, Q_k \quad \text{are putative conditional distributions.}$$

We say that $Q_1, \ldots, Q_k$ are compatible if there is $P \in \mathcal{P}$ such that

$$Q_i(y, \cdot) = P_i(y, \cdot)$$

for all $i \in I$ and P-almost all $y \in \mathcal{Y}_i$. In addition, given $\mathcal{P}_0 \subset \mathcal{P}$, we say that $Q_1, \ldots, Q_k$ are $\mathcal{P}_0$-compatible if such a condition holds for some $P \in \mathcal{P}_0$.

3. Compactly supported distributions

3.1. **Two compatibility results.** Let $\mathcal{L}$ be a set of real bounded Borel functions on Ω which is both a linear space and a determining class. By a determining class we mean that, given any $P \in \mathcal{P}$ and $Q \in \mathcal{P}$,

$$E_P(f) = E_Q(f) \quad \text{for all } f \in \mathcal{L} \quad \Longleftrightarrow \quad P = Q.$$

For instance, $\mathcal{L}$ could be the set of real bounded continuous functions on Ω.

For $f \in \mathcal{L}$ and $i \in I$, write

$$E\big(f \mid Y_i = y\big) = \int_{\Omega_i} f(x, y)\, Q_i(y, dx) \quad \text{for all } y \in \mathcal{Y}_i.$$

Our first result follows from applying to the present framework a compatibility criterion stated in [5]. See also [14].

Theorem 4. *Suppose that, for all $f \in \mathcal{L}$ and $i \in I$,*

$$\Omega_i \text{ is compact and } y \mapsto E\big(f \mid Y_i = y\big) \text{ is a continuous function.}$$

Then, $Q_1, \ldots, Q_k$ are compatible if and only if

$$\sup_{\omega \in \Omega} \sum_{i=2}^{k} \Big\{E\big(f_i \mid Y_i = Y_i(\omega)\big) - E\big(f_i \mid Y_1 = Y_1(\omega)\big)\Big\} \geq 0 \tag{2}$$

for all $f_2, \ldots, f_k \in \mathcal{L}$.

Proof. In the notation of [5], define $\mathcal{B} = \mathcal{B}(\Omega)$ and $\mathcal{A}_i = \sigma(Y_i)$. Also, for each $\omega \in \Omega$ and $i \in I$, take $\mu_i(\omega)$ to be the only probability on $\mathcal{B}$ satisfying

$$\mu_i(\omega)\big(X_i \in A,\, Y_i \in B\big) = I_B\big(Y_i(\omega)\big)\, Q_i\big(Y_i(\omega),\, A\big)$$

where $A \in \mathcal{B}(\Omega_i)$ and $B \in \mathcal{B}(\mathcal{Y}_i)$. Then, for each bounded Borel function $f : \Omega \to \mathbb{R}$, one obtains

$$\int_{\Omega} f(z)\, \mu(\omega)(dz) = \int_{\Omega_i} f(x,\, Y_i(\omega))\, Q_i\big(Y_i(\omega),\, dx\big) = E\big(f \mid Y_i = Y_i(\omega)\big).$$

Next, let $\mathcal{H}$ be the linear space generated by all functions

$$\omega \mapsto E\big(f \mid Y_i = Y_i(\omega)\big) - E\big(f \mid Y_1 = Y_1(\omega)\big)$$

for $f \in \mathcal{L}$ and $i = 2, \ldots, k$. Since $\mathcal{L}$ is a linear space, each $h \in \mathcal{H}$ can be written as

$$h(\omega) = \sum_{i=2}^{k} \Big\{ E\big(f_i \mid Y_i = Y_i(\omega)\big) - E\big(f_i \mid Y_1 = Y_1(\omega)\big) \Big\} \tag{3}$$

for suitable $f_2, \ldots, f_k \in \mathcal{L}$. Thus, under (2), compatibility of $Q_1, \ldots, Q_k$ follows from Theorem 6-(a) of [5]. This proves the "if" part. Conversely, suppose $Q_1, \ldots, Q_k$ are compatible. Take $f_2, \ldots, f_k \in \mathcal{L}$ and define h according to (3). By compatibility, there is $P \in \mathcal{P}$ such that $E\big(f_i \mid Y_i = Y_i(\cdot)\big)$ and $E\big(f_i \mid Y_1 = Y_1(\cdot)\big)$ are both conditional expectations under P for all i. It follows that

$$\sup_{\omega \in \Omega} h(\omega) \geq E_P(h)$$

$$= \sum_{i=2}^{k} E_P \Big\{ E\big(f_i \mid Y_i = Y_i(\cdot)\big) - E\big(f_i \mid Y_1 = Y_1(\cdot)\big) \Big\}$$

$$= \sum_{i=2}^{k} \big\{ E_P(f_i) - E_P(f_i) \big\} = 0.$$

Hence, condition (2) holds. □

A few brief remarks are in order.

First, under the assumptions of Theorem 4, the sup in condition (2) is attained. Thus, condition (2) is equivalent to: for all $f_2, \ldots, f_k \in \mathcal{L}$, there is $\omega \in \Omega$ such that

$$\sum_{i=2}^{k} E\big(f_i \mid Y_i = Y_i(\omega)\big) \geq \sum_{i=2}^{k} E\big(f_i \mid Y_1 = Y_1(\omega)\big).$$

Second, let $k = 2$ and let (x, y) denote a point of $\Omega_1 \times \Omega_2 = \Omega$. Since $Y_2 = X_1$ and $Y_1 = X_2$, condition (2) reduces to

$$\text{for each } f \in \mathcal{L}, \text{ there is } (x, y) \in \Omega \text{ such that}$$
$$E\big(f \mid X_1 = x\big) \geq E\big(f \mid X_2 = y\big).$$

Similarly, if $k = 3$ and (x, y, z) denotes a point of $\Omega_1 \times \Omega_2 \times \Omega_3 = \Omega$, condition (2) can be written as

$$\text{for each } f,\, g \in \mathcal{L}, \text{ there is } (x, y, z) \in \Omega \text{ such that}$$
$$E\big(f \mid X_1 = x, X_3 = z\big) + E\big(g \mid X_1 = x, X_2 = y\big) \geq E\big(f + g \mid X_2 = y, X_3 = z\big).$$

Third, for Theorem 4 to apply, each Ω_i has to be compact. This is certainly a strong restriction, which rules out various interesting applications. However, the compactness assumption can be weakened at the price of replacing (2) with a more involved condition. We give an explicit statement for $k = 2$ only.

Theorem 5. *Suppose $k = 2$, Ω_1 is compact, and*

$$x \mapsto E\big(f \mid X_1 = x\big) \quad \text{and} \quad x \mapsto \int_{\Omega_2} E\big(f \mid X_2 = y\big)\, Q_2(x, dy)$$

are continuous functions on Ω_1 for all $f \in \mathcal{L}$. Then, Q_1 and Q_2 are compatible if and only if

$$\sup_{x \in \Omega_1} \Big\{ E\big(f \mid X_1 = x\big) - \int_{\Omega_2} E\big(f \mid X_2 = y\big)\, Q_2(x, dy) \Big\} \geq 0$$

for all $f \in \mathcal{L}$.

Proof. We just give a sketch of the proof. The "only if" part can be proved as in Theorem 4. As to the "if" part, in the notation of [5], take $j = 2$ and $\phi = Y_2 = X_1$. Define also $\mathcal{A}_i$, μ_i and $\mathcal{B}$ as in the proof of Theorem 4. Now, proceed as in such a proof but apply Theorem 6-(b) of [5] instead of Theorem 6-(a). □

3.2. **Examples.** The possible applications of Theorems 4-5 depend on the choice of $\mathcal{L}$. We just give two examples for $k = 2$.

Example 6. (Putative conditional moments). Suppose Ω_1 and Ω_2 are compact intervals and

$$x \mapsto E\big(X_2^j \mid X_1 = x\big) \quad \text{and} \quad y \mapsto E\big(X_1^j \mid X_2 = y\big)$$

are continuous functions for all $j \geq 1$. Then, $\mathcal{L}$ can be taken to be the class of polynomials on Ω. Practically, this amounts to testing compatibility of Q_1 and Q_2 via conditional moments. Let

$$f(x, y) = \sum_{0 \leq r,s \leq n} c(r, s)\, x^r\, y^s$$

where $(x, y) \in \Omega$, $n \geq 1$ and the $c(r, s)$ are real coefficients. Define

$$h(x, y) = E\big(f \mid X_1 = x\big) - E\big(f \mid X_2 = y\big)$$
$$= \sum_{0 \leq r,s \leq n} c(r, s) \Big\{ x^r\, E\big(X_2^s \mid X_1 = x\big) - y^s\, E\big(X_1^r \mid X_2 = y\big) \Big\}.$$

By Theorem 4, Q_1 and Q_2 are compatible if and only if $\sup h \geq 0$ for every $n \geq 1$ and every choice of the constants $c(r, s)$.

Example 7. (Discrete random variables). Suppose Ω_1 is finite and Ω_2 countably infinite. Let $I(a, b)$ denote the indicator of the point $(a, b) \in \Omega$. Take $\mathcal{L}$ to be the class of all functions f on Ω of the type

$$f = \sum_{a \in \Omega_1} \sum_{b \in B} c(a, b)\, I(a, b)$$

where $B \subset \Omega_2$ is a finite subset and the $c(a, b)$ are real constants. Writing $Q_i(u, v)$ instead of $Q_i(u, \{v\})$, one obtains

$$h(x) = E\big(f \mid X_1 = x\big) - \int_{\Omega_2} E\big(f \mid X_2 = y\big)\, Q_2(x, dy)$$
$$= \sum_{b \in B} c(x, b)\, Q_2(x, b) - \sum_{a \in \Omega_1} \sum_{b \in B} c(a, b)\, Q_1(b, a)\, Q_2(x, b)$$

for all $x \in \Omega_1$. By Theorem 5, Q_1 and Q_2 are compatible if and only if $\max h \geq 0$ for all finite $B \subset \Omega_2$ and all choices of the constants $c(a, b)$. Suppose now that Ω_1 and Ω_2 are both finite. Then, $\mathcal{L}$ can be taken as above with $B = \Omega_2$ and Theorem 5 can be replaced by the simpler Theorem 4. Define in fact

$$h(x, y) = E\big(f \mid X_1 = x\big) - E\big(f \mid X_2 = y\big) = \sum_{b \in \Omega_2} c(x, b)\, Q_2(x, b) - \sum_{a \in \Omega_1} c(a, y)\, Q_1(y, a)$$

for all $(x, y) \in \Omega$. By Theorem 4, Q_1 and Q_2 are compatible if and only if $\max h \geq 0$ for every choice of the constants $c(a, b)$.

4. The absolutely continuous case

In Theorems 4 and 5, $Q_1, \ldots, Q_k$ are not requested to have densities with respect to reference measures. When this happens, however, stronger results are available.

For each $i \in I$, let λ_i denote a σ-finite measure on $\mathcal{B}(\Omega_i)$. For instance, Ω_i could be countable and λ_i the counting measure. Or else, Ω_i could be an interval and λ_i the Lebesgue measure. In almost all applications, it happens that

$$Q_i(y, A) = \int_A f_i(x \mid y)\, \lambda_i(dx) \tag{4}$$

for all $i \in I$, $y \in \mathcal{Y}_i$ and $A \in \mathcal{B}(\Omega_i)$. Here, f_i is a putative conditional density, that is, $(x, y) \mapsto f_i(x \mid y)$ is a non-negative Borel function on Ω satisfying

$$\int_{\Omega_i} f_i(x \mid y)\, \lambda_i(dx) = 1 \quad \text{for each } y \in \mathcal{Y}_i.$$

Under (4), we will say indifferently that $f_1, \ldots, f_k$ are compatible or that $Q_1, \ldots, Q_k$ are compatible.

We first report a classical result which holds for $k = 2$; see e.g. [1]-[2] and references therein. Let

$$\lambda = \lambda_1 \times \ldots \times \lambda_k$$

denote the product measure on $\mathcal{B}(\Omega)$.

Theorem 8. *Suppose $k = 2$ and condition (4) holds. Then, f_1 and f_2 are compatible if and only if there are two Borel functions $u : \Omega_1 \to [0, \infty)$ and $v : \Omega_2 \to [0, \infty)$ such that*

$$f_1(x \mid y) = f_2(y \mid x)\, u(x)\, v(y), \tag{5}$$
$$\lambda\text{-a.e. on the set } \{(x, y) : u(x) > 0,\ v(y) > 0\},$$

and

$$\int_\Omega I_{\{v>0\}}(y)\, f_2(y \mid x)\, u(x)\, \lambda(dx, dy) = \int_{\Omega_1} u\, d\lambda_1 = \int_{\{v>0\}} 1/v\, d\lambda_2 = 1. \tag{6}$$

Our next goal is extending Theorem 8 from $k = 2$ to an arbitrary $k \geq 2$. Before doing this, however, a remark is in order.

Remark 9. The informal idea of Theorem 8 is that the ratio f_1/f_2, where it is defined, factorizes in the product of a function of x alone times a function of y alone. Such an idea is realized by condition (5). Instead, as far as we know, no version of Theorem 8 includes condition (6). But some form of (6) seems to be unavoidable to characterize compatibility. For instance, according to Theorem 1 of [2], f_1 and f_2 are compatible if and only if

$$\{f_1 > 0\} = \{f_2 > 0\} = N \quad \text{(say) and}$$
$$\frac{f_1(x \mid y)}{f_2(y \mid x)} = u(x)\, v(y) \quad \text{for } (x, y) \in N$$

for some u, v such that $\int_{\Omega_1} u\, d\lambda_1 < \infty$. But, as it stands, such result does not work. In fact, the requested conditions suffice for compatibility of f_1 and f_2, but they are not necessary (even if they are asked λ-a.e. only). As a trivial example, take $\Omega_1 = \Omega_2 = [0,1]$, $\lambda_1 = \lambda_2 =$ Lebesgue measure, and

$$f_1(x \mid y) = I_{[0,\,1/2)}(y) + 2\, I_{[1/2,\,1]}(x)\, I_{[1/2,\,1]}(y),$$
$$f_2(y \mid x) = I_{[0,\,1/2)}(x) + 2\, I_{[1/2,\,1]}(y)\, I_{[1/2,\,1]}(x).$$

Let f be the uniform density on $S := [1/2,\,1] \times [1/2,\,1]$, that is, $f(x,y) = 4\, I_S(x,y)$. Then, f_1 and f_2 are compatible, for they agree on S with the conditional densities induced by f. Nevertheless,

$$\lambda(f_1 = 0,\, f_2 > 0) = \lambda(f_1 > 0,\, f_2 = 0) = 1/4.$$

In the next result, λ_i^* denotes the product measure

$$\lambda_i^* = \lambda_1 \times \ldots \times \lambda_{i-1} \times \lambda_{i+1} \times \ldots \times \lambda_k$$

on $\mathcal{B}(\mathcal{Y}_i)$. Recall that, according to Section 2, X_i is the i-th coordinate map on $\Omega = \prod_{j=1}^k \Omega_j$ and $Y_i = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k)$.

Theorem 10. *Suppose condition* (4) *holds. Then, $f_1, \ldots, f_k$ are compatible if and only if there are Borel functions*

$$u_i : \mathcal{Y}_i \to [0,\infty), \quad i \in I,$$

such that, for each $i < k$,

$$f_i(X_i \mid Y_i) = f_k(X_k \mid Y_k)\, u_i(Y_i)\, u_k(Y_k), \quad \lambda\text{-a.e. on the set } \{u_i(Y_i) > 0,\, u_k(Y_k) > 0\}, \tag{7}$$

and

$$\int_\Omega I_{\{u_i>0\}}(Y_i)\, f_k(X_k \mid Y_k)\, u_k(Y_k)\, d\lambda = \int_{\mathcal{Y}_k} u_k\, d\lambda_k^* = \int_{\{u_i>0\}} 1/u_i\, d\lambda_i^* = 1. \tag{8}$$

Moreover,

- (i) *If $f_1, \ldots, f_k$ are compatible and $P \in \mathcal{P}$ has conditional distributions $Q_1, \ldots, Q_k$, then $P \ll \lambda$. If, in addition, $f_i > 0$ for all $i \in I$, then $P \sim \lambda$.*
- (ii) *If conditions* (7)-(8) *hold for some $u_1, \ldots, u_k$, then $f = f_k(X_k \mid Y_k)\, u_k(Y_k)$ is a density with respect to λ and $f_1, \ldots, f_k$ are the conditional densities induced by f.*

In a sense, Theorem 10 is in the folklore of the existing literature on compatibility. In fact, the spirit of Theorem 10 is the same as that of Theorem 8, even if the conditions become less manageable as k increases. Overall, for low values of k, Theorem 10 is useful in real problems, mainly in connection with Gibbs sampling, missing data imputation and spatial data modeling. Nevertheless, to our knowledge, no explicit version of Theorem 10 has been stated so far.

To illustrate a few particular cases, suppose $k = 2$ and (x,y) denotes a point of $\Omega_1 \times \Omega_2 = \Omega$. Then, $f_1(X_1 \mid Y_1) = f_1(x \mid y)$ and $f_2(X_2 \mid Y_2) = f_2(y \mid x)$ so that Theorem 10 reduces to Theorem 8 with $u = u_2$ and $v = u_1$. Similarly, if $k = 3$ and (x,y,z) denotes a point of $\Omega_1 \times \Omega_2 \times \Omega_3 = \Omega$, condition (7) can be written as

$$f_1(x \mid y,z) = f_3(z \mid x,y)\, u_1(y,z)\, u_3(x,y) \quad \text{if } u_1(y,z) > 0 \text{ and } u_3(x,y) > 0,$$
$$f_2(y \mid x,z) = f_3(z \mid x,y)\, u_2(x,z)\, u_3(x,y) \quad \text{if } u_2(x,z) > 0 \text{ and } u_3(x,y) > 0,$$

for λ-almost all (x,y,z).

Note also that, if $u_i > 0$ for all i, condition (8) reduces to $\int_{\mathcal{Y}_k} u_k \, d\lambda_k^* = 1$.

We finally prove Theorem 10. We begin with the following lemma.

Lemma 11. *Suppose* (4) *holds,* $Q_1, \ldots, Q_k$ *are compatible and* $P \in \mathcal{P}$ *has conditional distributions* $Q_1, \ldots, Q_k$*. Then* $P \ll \lambda$*, and* $P \sim \lambda$ *if* $f_i > 0$ *for all* $i \in I$*.*

Proof. We first prove $P \ll \lambda$. Let $\mu(\cdot) = P\big(Y_k \in \cdot\big)$ be the marginal of Y_k under P. Fix $A \in \mathcal{B}(\Omega)$ such that $\lambda(A) = 0$ and define

$$A_y = \{x \in \Omega_k : (x, y) \in A\} \text{ for } y \in \mathcal{Y}_k \text{ and } B = \{y \in \mathcal{Y}_k : \lambda_k(A_y) = 0\}.$$

Since

$$\int_{\mathcal{Y}_k} \lambda_k(A_y) \, \lambda_k^*(dy) = \int_{\mathcal{Y}_k} \int_{\Omega_k} I_A(x, y) \, \lambda_k(dx) \, \lambda_k^*(dy) = \lambda(A) = 0,$$

then $\lambda_k^*(B^c) = 0$. Thus, if $\mu \ll \lambda_k^*$, condition (4) yields

$$P(A) = \int_{\mathcal{Y}_k} Q_k(y, A_y) \, \mu(dy) = \int_B Q_k(y, A_y) \, \mu(dy) = 0.$$

Therefore, to get $P \ll \lambda$, it suffices to show that $\mu \ll \lambda_k^*$. Let μ_1 be the marginal of X_1 under P. If $A \in \mathcal{B}(\Omega_1)$ and $\lambda_1(A) = 0$, condition (4) implies

$$\mu_1(A) = P(X_1 \in A) = E_P\big\{Q_1(Y_1, A)\big\} = 0.$$

Hence, $\mu_1 \ll \lambda_1$. Next, let $\mu_{1,2}$ be the marginal of (X_1, X_2) under P. For μ_1-almost all $x \in \Omega_1$, one obtains

$$P\big(X_2 \in A \mid X_1 = x\big) = E_P\Big\{Q_2\big((x, X_3, \ldots, X_k), A\big) \mid X_1 = x\Big\} \quad \text{for each } A \in \mathcal{B}(\Omega_2).$$

Hence, for μ_1-almost all $x \in \Omega_1$,

$$P\big(X_2 \in A \mid X_1 = x\big) = 0 \quad \text{provided } A \in \mathcal{B}(\Omega_2) \text{ and } \lambda_2(A) = 0.$$

Since $\mu_1 \ll \lambda_1$, the above condition implies $\mu_{1,2} \ll \lambda_1 \times \lambda_2$. Proceeding in this way, one finally obtains $\mu \ll \lambda_1 \times \ldots \times \lambda_{k-1} = \lambda_k^*$. This proves $P \ll \lambda$. Next, suppose $f_i > 0$ for all $i \in I$. Then $Q_i(y, A) > 0$, for all $i \in I$ and $y \in \mathcal{Y}_i$, provided $A \in \mathcal{B}(\Omega_i)$ and $\lambda_i(A) > 0$. Basing on this fact, $P \sim \lambda$ can be proved exactly as above. □

Proof of Theorem 10. Write $H_i = \{u_i(Y_i) > 0\}$ and recall

$$\int_{\Omega_i} f_i(x \mid y) \, \lambda_i(dx) = 1 \quad \text{for all } i \in I \text{ and } y \in \mathcal{Y}_i.$$

Note also that point (i) follows from Lemma 11.

Suppose $f_1, \ldots, f_k$ are compatible and fix $P \in \mathcal{P}$ having $Q_1, \ldots, Q_k$ as conditional distributions. By point (i), P has a density f with respect to λ. Let

$$\phi_i(y) = \int_{\Omega_i} f(x, y) \, \lambda_i(dx), \quad y \in \mathcal{Y}_i,$$

be the marginal of f with respect to λ_i^*. Define also

$$u_i = I_{\{0 < \phi_i < \infty\}} \, (1/\phi_i) \text{ for } i < k, \quad u_k = I_{\{\phi_k < \infty\}} \, \phi_k,$$

and note that

$$\{0 < \phi_i < \infty\} = \{u_i > 0\} \quad \text{and} \quad \lambda_i^*(\phi_i = \infty) = 0 \quad \text{for all } i \in I.$$

Given $i < k$, since $f_1, \ldots, f_k$ are the conditional densities induced by f, one trivially obtains

$$f_i(X_i \mid Y_i) = \frac{f}{\phi_i(Y_i)} = \frac{f}{\phi_k(Y_k)}\, u_i(Y_i)\, \phi_k(Y_k) = f_k(X_k \mid Y_k)\, u_i(Y_i)\, u_k(Y_k),$$

λ-a.e. on the set $H_i \cap H_k$. Further, since $f = f_k(X_k \mid Y_k)\, u_k(Y_k)$, λ-a.e.,

$$\int_{\mathcal{Y}_k} u_k\, d\lambda_k^* = \int_{\mathcal{Y}_k} \phi_k\, d\lambda_k^* = 1, \quad \int_{\{u_i>0\}} 1/u_i\, d\lambda_i^* = \int_{\mathcal{Y}_i} \phi_i\, d\lambda_i^* = 1,$$

$$\int_\Omega I_{H_i}\, f_k(X_k \mid Y_k)\, u_k(Y_k)\, d\lambda = \int_\Omega I_{H_i}\, f\, d\lambda = P\big(0 < \phi_i(Y_i) < \infty\big) = 1.$$

Therefore, conditions (7)-(8) hold. Conversely, suppose (7)-(8) hold. By (8),

$$\int_\Omega f_k(X_k \mid Y_k)\, u_k(Y_k)\, d\lambda = \int_{\mathcal{Y}_k} \int_{\Omega_k} f_k(x \mid y)\, \lambda_k(dx)\, u_k(y)\, \lambda_k^*(dy) = \int_{\mathcal{Y}_k} u_k\, d\lambda_k^* = 1.$$

Thus, $f := f_k(X_k \mid Y_k)\, u_k(Y_k)$ is a density with respect to λ. By definition, $f = 0$ on H_k^c. If $i < k$, condition (8) yields

$$\int_{H_i^c} f\, d\lambda = 1 - \int_{H_i} f\, d\lambda = 1 - 1 = 0.$$

Hence $f = 0$, λ-a.e., on $\cup_{i=1}^k H_i^c$. By (7), it follows that

$$f = f\, I_{H_i}\, I_{H_k} = \frac{f_i(X_i \mid Y_i)}{u_i(Y_i)}\, I_{H_i}\, I_{H_k}, \quad \lambda\text{-a.e. for all } i < k.$$

Moreover,

$$\begin{aligned}
\int_{H_k^c} I_{H_i} \frac{f_i(X_i \mid Y_i)}{u_i(Y_i)}\, d\lambda &= \int_\Omega I_{H_i} \frac{f_i(X_i \mid Y_i)}{u_i(Y_i)}\, d\lambda - \int_{H_k} I_{H_i} \frac{f_i(X_i \mid Y_i)}{u_i(Y_i)}\, d\lambda \\
&= \int_{\{u_i>0\}} \int_{\Omega_i} f_i(x \mid y)\, \lambda_i(dx)\, \frac{1}{u_i(y)}\, \lambda_i^*(dy) - \int_\Omega f\, d\lambda \\
&= \int_{\{u_i>0\}} 1/u_i\, d\lambda_i^* - 1 = 0.
\end{aligned}$$

Thus,

$$f = \frac{f_i(X_i \mid Y_i)}{u_i(Y_i)}\, I_{H_i}, \quad \lambda\text{-a.e. for all } i < k. \tag{9}$$

Next, define the marginal ϕ_i of f as above. Then, it suffices to prove that

$$\frac{f}{\phi_i(Y_i)} = f_i(X_i \mid Y_i), \quad \lambda\text{-a.e. on the set } \{0 < \phi_i(Y_i) < \infty\}, \text{ for all } i \in I.$$

Since $\phi_k = u_k$, such condition holds for $i = k$. If $i < k$, condition (9) yields

$$\phi_i(Y_i) = \int_{\Omega_i} \frac{f_i(x \mid Y_i)}{u_i(Y_i)}\, I_{H_i}\, \lambda_i(dx) = \frac{I_{H_i}}{u_i(Y_i)}.$$

Thus, $\{0 < \phi_i(Y_i) < \infty\} = H_i$, and condition (9) implies $f/\phi_i(Y_i) = f_i(X_i \mid Y_i)$, λ-a.e. on H_i. Since point (ii) is obvious, this concludes the proof. □

5. Compatibility under an exchangeable law

We now turn to $\mathcal{P}_0$-compatibility. Various choices of $\mathcal{P}_0$ could be of interest. Two of them are $\mathcal{P}_0 = \{P \in \mathcal{P} : P \ll \lambda\}$ or $\mathcal{P}_0 = \{P \in \mathcal{P} : P \sim \lambda\}$ but they are already covered by Theorem 10. Another option is

$$\mathcal{P}_0 = \{P \in \mathcal{P} : X \text{ is exchangeable under } P\}.$$

Recall that $X = (X_1, \ldots, X_k)$ is exchangeable in case $(X_{j_1}, \ldots, X_{j_k})$ is distributed as $(X_1, \ldots, X_k)$ for all permutations $(j_1, \ldots, j_k)$ of $(1, \ldots, k)$.

The latter choice of $\mathcal{P}_0$ looks intriguing (to us). Indeed, exchangeability plays a role in various frameworks where compatibility issues arise, such as Bayesian and/or spatial statistics.

In this section, we just let $\mathcal{P}_0 = \{P \in \mathcal{P} : X \text{ exchangeable under } P\}$. Then, it makes sense to assume

$$\Omega_1 = \ldots = \Omega_k = \mathcal{X} \quad \text{and} \quad \lambda_1 = \ldots = \lambda_k = \gamma \tag{10}$$

where $\mathcal{X} \in \mathcal{B}(\mathbb{R})$ and γ is a σ-finite measure on $\mathcal{B}(\mathcal{X})$. Note that condition (10) implies $\Omega = \mathcal{X}^k$, $\lambda = \gamma^k$, $\mathcal{Y}_i = \mathcal{X}^{k-1}$ and $\lambda_i^* = \gamma^{k-1}$ for all $i \in I$.

If $Q_1, \ldots, Q_k$ are the conditional distributions of $P \in \mathcal{P}_0$, then $Q_1 = \ldots = Q_k$, P-a.s.. Thus, we also assume

$$Q_1 = \ldots = Q_k.$$

But such condition is not enough, even for compatibility alone. For instance, if $k = 2$, $\mathcal{X} = \mathbb{R}$ and $Q_1(x, \cdot) = Q_2(x, \cdot) = N(x, 1)$ for all $x \in \mathbb{R}$, then Q_1 and Q_2 are not compatible (just apply Theorem 8).

Basing on the previous remarks, a question is whether $Q_1, \ldots, Q_k$ are $\mathcal{P}_0$-compatible provided they are compatible and $Q_1 = \ldots = Q_k$. For some time, we conjectured a negative answer. Instead, the answer is yes for $k = 2$. To prove this fact, a definition is to be recalled.

Let Q be a putative conditional distribution, with $k = 2$ and $\Omega_1 = \Omega_2 = \mathcal{X}$. Say that Q is a *reversible kernel* if there is a probability measure μ on $\mathcal{B}(\mathcal{X})$ such that

$$\int_A Q(x, B)\, \mu(dx) = \int_B Q(x, A)\, \mu(dx) \quad \text{for all } A,\, B \in \mathcal{B}(\mathcal{X}). \tag{11}$$

If $P \in \mathcal{P}_0$ has conditional distributions Q_1 and Q_2, then $Q_1 = Q_2 = Q$, P-a.s., for some reversible kernel Q; see e.g. Theorem 3.2 of [4]. Neglecting the a.s., the converse becomes true as well.

Theorem 12. *Suppose $k = 2$, $\Omega_1 = \Omega_2 = \mathcal{X}$ and $Q_1 = Q_2$. The following statements are equivalent:*

(a) *Q_1 and Q_2 are $\mathcal{P}_0$-compatible;*
(b) *Q_1 and Q_2 are compatible;*
(c) *Q_1 is a reversible kernel.*

Proof. Write $Q_1 = Q_2 = Q$ and note that "(a) $\Rightarrow$ (b)" is trivial.

"(b) $\Rightarrow$ (c)" Fix $P \in \mathcal{P}$ with conditionals Q_1 and Q_2. Let $\mu_1(\cdot) = P(X_1 \in \cdot)$ and $\mu_2(\cdot) = P(X_2 \in \cdot)$ be the marginal distributions of X_1 and X_2 under P. Since $Q_1 = Q_2 = Q$,

$$\int_A Q(x, B)\, \mu_1(dx) = \int_A Q_2(x, B)\, \mu_1(dx) = P\big(X_1 \in A,\, X_2 \in B\big) = \int_B Q(x, A)\, \mu_2(dx)$$

for all $A,\, B \in \mathcal{B}(\mathcal{X})$. Hence, condition (11) holds with $\mu = (\mu_1 + \mu_2)/2$, that is, Q is a reversible kernel.

"(c) $\Rightarrow$ (a)" Fix a probability measure μ on $\mathcal{B}(\mathcal{X})$ satisfying (11) and define

$$P(A) = \int_{\mathcal{X}} \int_{\mathcal{X}} I_A(x,y)\, Q(x,dy)\, \mu(dx) \quad \text{for } A \in \mathcal{B}(\mathcal{X}^2).$$

Since Q is reversible,

$$P\big(X_1 \in A,\, X_2 \in B\big) = \int_A Q(x,B)\, \mu(dx) = \int_B Q(x,A)\, \mu(dx) = P\big(X_1 \in B,\, X_2 \in A\big)$$

for all $A,\, B \in \mathcal{B}(\mathcal{X})$. Hence, $P \in \mathcal{P}_0$. Also, Q is a conditional distribution, under P, for X_1 given X_2 as well as for X_2 given X_1. □

Reversible kernels admit sometimes simple characterizations.

Example 13. Let $\mathcal{X}$ be countable. Write $Q(x,y)$ instead of $Q(x,\{y\})$ and suppose Q irreducible (in the sense of Markov chains). There is a non zero measure μ on $\mathcal{B}(\mathcal{X})$ satisfying (11) if and only if

$$Q(x,y) > 0 \;\Leftrightarrow Q(y,x) > 0 \quad \text{and} \quad \prod_{i=1}^n Q(x_{i-1},x_i) = \prod_{i=1}^n Q(x_i,x_{i-1})$$

whenever $x,y,x_0,x_1,\ldots,x_n \in \mathcal{X}$ and $x_n = x_0$. See e.g. page 303 of [9]. However, μ needs not be a probability measure and some extra condition is needed in order that $\mu(\mathcal{X}) < \infty$. As an extreme example, suppose there is $a \in \mathcal{X}$ satisfying $Q(x,a) > 0$ for all $x \in \mathcal{X}$. Then, $\mu(\mathcal{X}) < \infty$ (so that Q is reversible) if and only if $\sum_x Q(a,x)/Q(x,a) < \infty$.

We finally turn to $k \geq 2$. For arbitrary $Q_1,\ldots,Q_k$, Theorem 12 does not admit nice extensions to $k \geq 2$. Hence, $Q_1,\ldots,Q_k$ are assumed to have densities. Next result is quite expected but may be useful in real problems. In fact, it provides simple (and easily checkable) conditions for $\mathcal{P}_0$-compatibility.

Theorem 14. *Suppose conditions* (4) *and* (10) *hold. Then,* $Q_1,\ldots,Q_k$ *are* $\mathcal{P}_0$*-compatible provided* $f_1 = \ldots = f_k$ *and*

$$f_1(x \mid y) = g(x,y)\, h(y) \quad \text{for all } x \in \mathcal{X} \text{ and } y \in \mathcal{X}^{k-1},$$

where g *and* h *are Borel functions (on* $\mathcal{X}^k$ *and* $\mathcal{X}^{k-1}$*, respectively) satisfying*

$$h > 0, \quad \int_{\mathcal{X}^{k-1}} 1/h\, d\gamma^{k-1} = 1, \quad g = g \circ \pi \text{ for all permutations } \pi \text{ of } \mathcal{X}^k.$$

Proof. Since g is invariant under permutations, conditions (7)-(8) trivially hold with $u_k = 1/h$ and $u_i = h$ for $i < k$. Define

$$f = f_k(X_k \mid Y_k)\, u_k(Y_k) = g(X_k, Y_k)$$

and $P(A) = \int_A f\, d\lambda$ for all $A \in \mathcal{B}(\Omega)$. By point (ii) of Theorem 10, $Q_1,\ldots,Q_k$ are the conditional distributions induced by P. Also $P \in \mathcal{P}_0$, for both $f = g$ and $\lambda = \gamma^k$ are invariant under permutations. □

References

[1] Arnold B.C., Press S.J. (1989) Compatible conditional distributions, *J. Amer. Statist. Assoc.*, 84, 152-156.

[2] Arnold B.C., Castillo E., Sarabia J.M. (2001) Conditionally Specified distributions: An introduction (with discussion), *Statist. Science*, 16, 249-274.

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

[4] Berti P., Pratelli L., Rigo P. (2012) Limit theorems for empirical processes based on dependent data, *Electron. J. Probab.*, 17, 1-18.

[5] Berti P., Dreassi E., Rigo P. (2012) A consistency theorem for regular conditional distributions, *Stochastics*, to appear, currently available at: http://economia.unipv.it/pagp/pagine_personali/prigo/notaio.pdf

[6] Besag J. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion), *J. Royal Statist. Soc., Ser B*, 36, 192236.

[7] Besag J. (1996) On statistical analysis of dirty pictures (with discussion), *J. Royal Statist. Soc., Ser B*, 48, 259302.

[8] Chen H.Y. (2010) Compatibility of conditionally specified models, *Statist. Probab. Letters*, 80, 670-677.

[9] Durrett R. (1996) *Probability: Theory and Examples* (Second Edition), Duxbury Press.

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

[11] Gourieroux C., Montfort A. (1979) On the characteristerization of a joint probability distribution by conditional distributions, *J. Econometrics*, 10, 115118.

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

[13] Hobert J.P., Casella G. (1998) Functional compatibility, Markov Chains, and Gibbs sampling with improper posteriors, *J. Comput. Graphic Statis.*, 7, 4260.

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

[15] Raghunathan T.E., Lepkowski J.M., van Hoewyk J., Solenberger P. (2001) A multivariate technique for multiply imputing missing values using a sequence of regression models, *Survey Methodology*, 27, 8595.

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

[17] Song C-C., Li L-A., Chen C-H., Jiang T.J., Kuo K-L. (2010) Compatibility of finite discrete conditional distributions, *Statist. Sinica*, 20, 423-440.

[18] van Buuren S., Brand H.P.L., Groothuis-Oudshoorn C.G.M., Rubin D.B. (2006) Fully conditional specification in multivariate imputation, *H. Statist. Comput. Simul.*, 76, 1049-1064.

[19] van Buuren S. (2007) Multiple imputation of discrete and continuous data by fully conditional specification, *Statis. Meth. Medical Res.*, 16, 219242.

[20] Wang Y.J., Ip E.H. (2008) Conditionally specified continuous distributions, *Biometrika*, 95, 735746.

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 Matematica "F. Casorati", Universita' di Pavia, via S. Ferrata 1, 27100 Pavia, Italy
E-mail address: pietro.rigo@unipv.it