Journal of Multivariate Analysis 125 (2014) 190–203

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Compatibility results for conditional distributions

Patrizia Berti[a], Emanuela Dreassi[b], Pietro Rigo[c,*]

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

[b] Dipartimento di Statistica, Informatica, Applicazioni "G. Parenti", Universita' di Firenze, viale Morgagni 59, 50134 Firenze, Italy

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

ARTICLE INFO

Article history:
Received 10 December 2012
Available online 13 January 2014

AMS 2010 subject classifications:
62F15
62H05
62G99
60A05

Keywords:
Bayesian inference
Compatibility of conditional distributions
Exchangeability
Gibbs sampling
Markov random fields
Multiple imputation

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. The law P with conditionals $Q_1, \ldots, Q_k$ is requested to belong to some given class $\mathscr{P}_0$ of distributions. Two choices for $\mathscr{P}_0$ are considered, that is, $\mathscr{P}_0 = \{\text{exchangeable laws}\}$ and $\mathscr{P}_0 = \{\text{laws with identical univariate marginals}\}$.

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 $\mathscr{P}$ the set of all probability distributions for the process

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

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

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

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

$$Q_H = P_H, \quad \text{a.s. with respect to } P, \text{ for all } H \in \mathscr{H}. \tag{1}$$

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

* Corresponding author.
E-mail addresses: patrizia.berti@unimore.it (P. Berti), dreassi@disia.unifi.it (E. Dreassi), pietro.rigo@unipv.it (P. Rigo).

0047-259X/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmva.2013.12.009

An obvious version of the previous definition is the following. Fix $\mathcal{P}_0 \subset \mathcal{P}$. 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 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

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

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 (*Gibbs Sampling, Multiple Imputation, Markov Random Fields*)**.** Let $I = \{1, \ldots, k\}$ and $H_i = \{i\}$. 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 all $i \in I$. Usually, the P_{H_i} are obtained from a given $P \in \mathcal{P}$. But sometimes P is not assessed. Rather, one selects a collection $\{Q_{H_i} : i \in I\}$ of putative conditional distributions and use Q_{H_i} in the place of P_{H_i}. Formally, this procedure makes sense only if $\{Q_{H_i} : i \in I\}$ is compatible. Essentially the same situation occurs in missing data imputation and spatial data modeling. Again, P is not explicitly assessed and $X = (X_1, \ldots, X_k)$ is modeled by some collection $\{Q_{H_i} : i \in I\}$ of putative conditional distributions. As a (remarkable) particular case, in Markov random fields, each Q_{H_i} depends only on $(X_j : j \in N_i)$, where N_i denotes the set of neighbors of i. See [5,6,11,13,16,15] and references therein.

We point out that Gibbs sampling, multiple imputation and spatial data modeling are different statistical issues, but they share the structure of the putative conditional distributions $\{Q_{H_i} : i \in I\}$. From the point of view of compatibility, hence, they can be unified.

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 (proper) prior μ, or equivalently, whether Q_Y and Q_Θ are compatible.

Compatibility results, if usable, have significant practical implications. In fact, in frameworks such as Examples 1 and 2 (Example 3 is a little more problematic), the statistical procedures based on $\{Q_H : H \in \mathcal{H}\}$ request compatibility. If $\{Q_H : H \in \mathcal{H}\}$ fails to be compatible, such procedures are questionable, or perhaps they do not make sense at all. In any case, a preliminary test of compatibility is fundamental.

Example 1 has been largely investigated (see e.g. [10]) while Example 3 reduces to Example 2 with $k = 2$ by taking X_1 and X_2 as random vectors of suitable dimensions. Thus, in this paper, we focus on the framework of Example 2.

In the sequel, we let

$$I = \{1, \ldots, k\} \quad \text{and} \quad X = (X_1, \ldots, X_k)$$

for some $k \geq 2$. We also let $H_i = \{i\}$ and we write

$$Q_i = Q_{\{i\}} \quad \text{for } i = 1, \ldots, k.$$

Accordingly, Q_i is to be regarded as the (putative) conditional distribution of X_i given $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k)$.

Three different types of compatibility results are provided. Most of them hold for arbitrary k, even if they take a nicer form for small 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 [3].

In Section 4, as in most real problems, $Q_1, \ldots, Q_k$ have densities with respect to reference measures. Under this assumption, compatibility is characterized in Theorem 10. Such a result improves and extends to any k a well known criterion which holds for $k = 2$. In particular, no positivity assumption on the conditional densities is requested. See [2,1,5,8,12–14,17]. See also Example 9 and the remarks after Theorem 10.

In Section 5, $\mathcal{P}_0$-compatibility is investigated under two different choices for $\mathcal{P}_0$. We let $\mathcal{P}_0 = \mathcal{E}$ and $\mathcal{P}_0 = \mathcal{I}$ where

$$\mathcal{E} = \{P \in \mathcal{P} : X \text{ exchangeable under } P\} \quad \text{and}$$
$$\mathcal{I} = \{P \in \mathcal{P} : X_1, \ldots, X_k \text{ identically distributed under } P\}.$$

Note that $\mathcal{E} \subset \mathcal{I}$. Among other things it is shown that, if $Q_1 = \cdots = Q_k$ and Q_1 meets a certain invariance condition, then $Q_1, \ldots, Q_k$ are $\mathcal{E}$-compatible if and only if they are compatible (Theorem 12). Moreover, if $k = 2$ and X_1, X_2 take values in a countable set $\mathcal{X}$, a necessary and sufficient condition for $\mathcal{I}$-compatibility is provided (Theorem 17). The latter condition becomes quite simple and practically useful when $\mathcal{X}$ is finite. In this case, if the (finitely many) values of Q_1 and Q_2 are uploaded into a computer, one obtains an on-line, definitive answer on whether Q_1 and Q_2 are $\mathcal{I}$-compatible or not.

Finally, some examples are given, mainly in Section 5. Suppose that, according to Q_i, the conditional distribution of X_i given $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k)$ is

$$N\left(\alpha\,\frac{\sum_{j\neq i} X_j}{k-1},\ 1\right) \quad \text{for some } \alpha \in \mathbb{R} \text{ and all } i = 1, \ldots, k.$$

Then, those values of α which make $Q_1, \ldots, Q_k$ compatible can be exactly determined. If $k = 3$, for instance, it turns out that Q_1, Q_2, Q_3 are compatible if and only if $\alpha \in (-2, 1)$. In addition, Q_1, Q_2, Q_3 are actually $\mathcal{E}$-compatible for $\alpha \in (-2, 1)$. As another example, suppose $k = 2$ and Q_1 is the kernel corresponding to the symmetric random walk on the integers. According to Q_1, thus, X_1 takes values $j - 1$ and $j + 1$ with equal probability $1/2$ conditionally on $X_2 = j$. Then, there is no putative conditional distribution Q_2 which is $\mathcal{I}$-compatible with such Q_1.

2. Notation and basic definitions

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 ith 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 notations will be used. If $i \in I$, $x \in \Omega_i$ 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 $\mathcal{B}(S)$ be the Borel σ-field on S. 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) $E_P\,\{I_B(Y_i)\,P_i(Y_i, A)\} = P(X_i \in A,\ Y_i \in B)$ for $A \in \mathcal{B}(\Omega_i)$ and $B \in \mathcal{B}(\mathcal{Y}_i)$.

Such a 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, or a *kernel*, is a function Q_i with the same domain as P_i, satisfying conditions (i)–(ii) but not necessarily (iii). In the sequel,

$$Q_1, \ldots, Q_k \quad \text{are given kernels.}$$

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

Some general compatibility criterions have been obtained in [3]. While quite abstract, such criterions simplify when adapted to the framework of this paper. All results in this section are actually proved by applying Theorem 6 of [3] to the present setting.

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

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

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

To state our first result, we let

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

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

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

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

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

for all $U_2, \ldots, U_k \in \mathcal{L}$.

Proof. In the notation of [3], 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 $\mathcal{B}$ such that

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

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

$$\int_{\Omega} U(v)\, \mu_i(\omega)(dv) = \int_{\Omega_i} U(x,\ Y_i(\omega))\, Q_i\big(Y_i(\omega),\ dx\big) = E\big(U \mid Y_i = Y_i(\omega)\big).$$

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

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

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

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

for suitable $U_2, \ldots, U_k \in \mathcal{L}$. Thus, under (2), compatibility of $Q_1, \ldots, Q_k$ follows from Theorem 6-(a) of [3]. This proves the "if" part. Conversely, suppose $Q_1, \ldots, Q_k$ are compatible. Take $U_2, \ldots, U_k \in \mathcal{L}$ and define h according to (3). By compatibility, there is $P \in \mathcal{P}$ such that $E\big(U_i \mid Y_i = Y_i(\cdot)\big)$ and $E\big(U_i \mid Y_1 = Y_1(\cdot)\big)$ are both conditional expectations under P for all i. It follows that

$$\begin{aligned}\sup_{\omega\in\Omega} h(\omega) \ge E_P(h) &= \sum_{i=2}^{k} E_P\Big\{E\big(U_i \mid Y_i = Y_i(\cdot)\big) - E\big(U_i \mid Y_1 = Y_1(\cdot)\big)\Big\}\\ &= \sum_{i=2}^{k} \big\{E_P(U_i) - E_P(U_i)\big\} = 0.\end{aligned}$$

Hence, condition (2) holds. □

Under the assumptions of Theorem 4, the sup in condition (2) is attained. Thus, condition (2) is equivalent to: for all $U_2, \ldots, U_k \in \mathcal{L}$, there is $\omega \in \Omega$ such that

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

For instance, 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

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

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

$$\begin{aligned}&\text{for all } U,\ V \in \mathcal{L},\ \text{there is } (x, y, z) \in \Omega \text{ such that}\\ &\quad E\big(U \mid X_1 = x, X_3 = z\big) + E\big(V \mid X_1 = x, X_2 = y\big) \ge E\big(U + V \mid X_2 = y, X_3 = z\big).\end{aligned}$$

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*

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

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

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

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

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 [3], 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 [3] instead of Theorem 6-(a). □

3.2. Examples

The possible applications of Theorems 4 and 5 depend on the choice of $\mathcal{L}$. We just give two examples for $k = 2$. Recall that $Y_1 = X_2$ and $Y_2 = X_1$ when $k = 2$.

Example 6 (*Putative Conditional Moments*)**.** Suppose Ω_1 and Ω_2 are compact 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 $j \geq 1$. Here, X_2^j and X_1^j are the jth powers of X_2 and X_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

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

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

$$\begin{aligned} h(x, y) &= E\big(U \mid X_1 = x\big) - E\big(U \mid X_2 = y\big) \\ &= \sum_{0 \leq r, s \leq n} c(r, s)\, \left\{ x^r\, E\big(X_2^s \mid X_1 = x\big) - y^s\, E\big(X_1^r \mid X_2 = y\big) \right\}. \end{aligned}$$

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 U on Ω of the type

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

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

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

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 Theorem 4. Define in fact

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

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 all choices of the constants $c(a, b)$.

One drawback of Theorem 4 is that condition (2) is to be checked for infinitely many choices of elements of $\mathcal{L}$. For instance, in Example 7 with Ω_1 and Ω_2 finite, one has to verify whether $\max h \geq 0$ for every choice of the constants $c(a, b)$. This fact reduces the practical scope of Theorem 4. The same is true for Theorem 5. However, Theorems 4 and 5 are of possible theoretical interest. Furthermore, since they give necessary and sufficient conditions, they can be useful to quickly prove non compatibility. As a trivial example, suppose $\Omega_1 = \Omega_2 = \{1, 2, 3\}$ and

$$Q_1 = \begin{pmatrix} 1/2 & * & * \\ 2/3 & * & * \\ 2/5 & 1/5 & 2/5 \end{pmatrix}, \qquad Q_2 = \begin{pmatrix} 1/7 & 2/7 & 4/7 \\ * & * & 1/3 \\ * & * & 1/4 \end{pmatrix}.$$

Such Q_1 and Q_2 are not compatible, whatever the $*$ are specified. This can be seen by restricting to $c(a, b) \in \{0, 1\}$ for all $a, b \in \{1, 2, 3\}$. For instance, one obtains $h(x, y) < 0$ for all $x, y \in \{1, 2, 3\}$ in case $c(1, 1) = c(1, 2) = c(2, 3) = c(3, 3) = 1$ and $c(a, b) = 0$ otherwise.

4. The dominated case

In Theorems 4 and 5, $Q_1, \ldots, Q_k$ are arbitrary kernels. In almost all applications, however, each Q_i has a density with respect to some reference measure λ_i. In this case, simpler results are available.

For each $i \in I$, let λ_i denote a σ-finite measure on $\mathcal{B}(\Omega_i)$. For instance, some Ω_i could be countable with λ_i the counting measure and some other Ω_j could be an interval with λ_j the Lebesgue measure. In this section, it is assumed 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 well known result which holds for $k = 2$; see e.g. [2,1] and references therein. Let

$$\lambda = \lambda_1 \times \cdots \times \lambda_k$$

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*

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

and

$$\int_\Omega I_{\{v>0\}}(y) f_2(y \mid x)\, u(x)\, \lambda(dx, dy) = \int_{\Omega_1} u\, d\lambda_1 = \int_{\{v>0\}} 1/v\, d\lambda_2 = 1. \tag{5}$$

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.

To our knowledge, no version of Theorem 8 includes condition (5). But some form of (5) is necessary to characterize compatibility. In fact, some of the existing versions of Theorem 8, as they stand, can give rise to misunderstandings.

Example 9. According to Theorems 3.1 and 4.1 of [2] and Theorem 1 of [1], 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 $\int_{\Omega_1} u\, d\lambda_1 < \infty$. Actually, such conditions suffice for compatibility of f_1 and f_2, but they are not necessary (even if they are asked λ-a.e. only). For instance, take $\Omega_1 = \Omega_2 = [0, 1]$, $\lambda_1 = \lambda_2 =$ Lebesgue measure, and

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

where $S = [1/2,\ 1] \times [1/2,\ 1]$. Let f be the uniform density on S, 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,

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

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

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

on $\mathcal{B}(\mathcal{Y}_i)$. Recall that, according to Section 2, X_i is the ith 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 $i < k$,

$$f_i(X_i \mid Y_i) = f_k(X_k \mid Y_k)\, u_i(Y_i)\, u_k(Y_k), \tag{6}$$
λ-*a.e. on the set* $\{u_i(Y_i) > 0,\ u_k(Y_k) > 0\}$,

and

$$\int_{\Omega} I_{\{u_i>0\}}(Y_i) f_k(X_k \mid Y_k)\, u_k(Y_k)\, d\lambda = \int_{\mathcal{Y}_k} u_k\, d\lambda_k^* = \int_{\{u_i>0\}} 1/u_i\, d\lambda_i^* = 1. \tag{7}$$

Moreover:

(a) *If* $f_1, \dots, f_k$ *are compatible and* $P \in \mathcal{P}$ *has conditional distributions* $Q_1, \dots, Q_k$, *then* $P \ll \lambda$. *If, in addition,* $f_i > 0$ *for all* $i \in I$, *then* $P \sim \lambda$.
(b) *If conditions* (6)–(7) *hold for some* $u_1, \dots, u_k$, *then* $f = f_k(X_k \mid Y_k)\, u_k(Y_k)$ *is a density with respect to* λ *and* $f_1, \dots, f_k$ *are the conditional densities induced by* f.

The proof of Theorem 10 is postponed to a final Appendix while some examples are given in Section 5. Here, we make a few brief remarks.

For $k = 2$, Theorem 10 implies Theorem 8 (with $u = u_2$ and $v = u_1$). For $k = 3$, if (x, y, z) denotes a point of $\Omega_1 \times \Omega_2 \times \Omega_3 = \Omega$, condition (6) 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). Similarly, for condition (7). In general, to investigate compatibility of $f_1, \dots, f_k$, one has to handle $2\,(k-1)$ constraints. Such constraints reduce to $k - 1$ provided $f_i > 0$ for all $i \in I$ and $\int_{\mathcal{Y}_k} u_k\, d\lambda_k^* = 1$. In fact, the following result is available.

Corollary 11. *Suppose condition* (4) *holds with* $f_i > 0$ *for all* $i \in I$. *Then,* $f_1, \dots, f_k$ *are compatible if and only if there are strictly positive Borel functions* $u_1, \dots, u_k$ *satisfying condition* (6) *as well as* $\int_{\mathcal{Y}_k} u_k\, d\lambda_k^* = 1$.

Proof. Suppose $f_1, \dots, f_k$ are compatible. Since $f_i > 0$ for all $i \in I$, points (a)–(b) of Theorem 10 imply $u_i(Y_i) > 0$, λ-a.e., for all $i \in I$. Thus, $u_1, \dots, u_k$ can be taken to be strictly positive. Conversely, if $u_i > 0$ for all $i \in I$, condition (7) follows from condition (6) and $\int_{\mathcal{Y}_k} u_k\, d\lambda_k^* = 1$. □

Theorem 10 is inspired to Theorem 8, which in turn underlies most results in compatibility theory. Furthermore, at least for low values of k, Theorem 10 is useful in real problems. Despite these facts, no explicit version of Theorem 10 has been stated so far. To our knowledge, the existing results focus on particular cases only and/or request some positivity condition on $f_1, \dots, f_k$. See [2,1,5,8,12–14,17].

A last note is that Theorem 10 provides information on $\mathcal{P}_0$-compatibility as well. This is apparent if $\mathcal{P}_0 = \{P \in \mathcal{P} : P \ll \lambda\}$ or $\mathcal{P}_0 = \{P \in \mathcal{P} : P \sim \lambda\}$, but Theorem 10 may be instrumental also for $\mathcal{P}_0 = \{P \in \mathcal{P} : X \text{ exchangeable under } P\}$; see Section 5.1.

5. $\mathcal{P}_0$-compatibility

In this section, $\mathcal{P}_0$-compatibility is investigated under two different choices for $\mathcal{P}_0$. We let

$$\Omega_1 = \cdots = \Omega_k = \mathcal{X} \quad \text{for some } \mathcal{X} \in \mathcal{B}(\mathbb{R}).$$

As a consequence, $\Omega = \mathcal{X}^k$ and $\mathcal{Y}_i = \mathcal{X}^{k-1}$ for all $i \in I$.

5.1. Exchangeability

For each $n \geq 1$, let Π_n denote the set of all permutations of $\mathcal{X}^n$, that is, those mappings $\pi : \mathcal{X}^n \to \mathcal{X}^n$ of the form

$$\pi(x_1, \dots, x_n) = (x_{\pi_1}, \dots, x_{\pi_n}) \quad \text{for all } (x_1, \dots, x_n) \in \mathcal{X}^n,$$

where $(\pi_1, \dots, \pi_n)$ is a fixed permutation of $(1, \dots, n)$. The random vector $X = (X_1, \dots, X_k)$ is *exchangeable* if X is distributed as $\pi(X)$ for all $\pi \in \Pi_k$. Let

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

Exchangeability plays a role in various frameworks where compatibility issues arise. In Bayesian statistics, observations are usually i.i.d. conditionally on some random parameter, so that they are actually exchangeable. Or else, in some problems of spatial statistics, the joint distribution of the random variables associated to the sites is invariant under permutations of the sites; see e.g. [7,9]. Accordingly, in this subsection, we let $\mathcal{P}_0 = \mathcal{E}$ and we investigate $\mathcal{E}$-compatibility.

If $Q_1, \dots, Q_k$ are the conditional distributions of some $P \in \mathcal{E}$, there is an invariant kernel Q such that $Q_1 = \cdots = Q_k = Q$, P-a.s. Here, invariance of Q means that

$$Q\big(\pi(y),\ \cdot\big) = Q(y, \cdot) \quad \text{for all } y \in \mathcal{X}^{k-1} \text{ and } \pi \in \Pi_{k-1}. \tag{8}$$

Thus, it makes sense to assume

$$Q_1 = \cdots = Q_k = Q \tag{9}$$

for some kernel Q satisfying (8). But conditions (8)–(9) are not enough, even for compatibility alone. As an example, take $k = 2$, $\mathcal{X} = \mathbb{R}$ and $Q_1 = Q_2 = Q$, where $Q(y, \cdot) = N(y, 1)$ for all $y \in \mathbb{R}$. Then, conditions (8)–(9) are trivially true but Q_1 and Q_2 fail to be compatible; see forthcoming Example 15.

Based on the previous remarks, a natural question is whether $Q_1, \dots, Q_k$ are $\mathcal{E}$-compatible provided they are compatible and conditions (8)–(9) hold. For some time, we conjectured a negative answer. Instead, the answer is yes.

Theorem 12. *Suppose conditions (8)–(9) hold. Then, $Q_1, \dots, Q_k$ are $\mathcal{E}$-compatible if and only if they are compatible.*

Proof. Suppose $Q_1, \dots, Q_k$ are compatible and fix $P \in \mathcal{P}$ with conditionals $Q_1, \dots, Q_k$. It suffices to prove that, for all $i \in I$ and $\pi \in \Pi_k$,

$$Q \text{ is a version of the conditional distribution of } X_i \text{ given } Y_i \text{ under } P \circ \pi^{-1}. \tag{10}$$

In fact, suppose (10) holds and define

$$P^* = \frac{1}{k!} \sum_{\pi \in \Pi_k} P \circ \pi^{-1}.$$

By definition, $P^* \in \mathcal{E}$. Fix $i \in I$. For each $\pi \in \Pi_k$, let

$$\mu^*(\cdot) = P^*(Y_i \in \cdot) \quad \text{and} \quad \mu_\pi(\cdot) = P \circ \pi^{-1}\big(Y_i \in \cdot\big)$$

be the marginal distributions of Y_i under P^* and $P \circ \pi^{-1}$. By (10),

$$\begin{aligned} \int_B Q(y, A)\, \mu^*(dy) &= \frac{1}{k!} \sum_{\pi \in \Pi_k} \int_B Q(y, A)\, \mu_\pi(dy) \\ &= \frac{1}{k!} \sum_{\pi \in \Pi_k} P \circ \pi^{-1}\big(X_i \in A,\ Y_i \in B\big) = P^*\big(X_i \in A,\ Y_i \in B\big) \end{aligned}$$

for all $A \in \mathcal{B}(\mathcal{X})$ and $B \in \mathcal{B}(\mathcal{X}^{k-1})$. Hence, Q is a version of the conditional distribution of X_i given Y_i under P^*.

It remains to prove condition (10). Since $P \circ \pi^{-1}$ is the distribution of $\pi(X)$ under P, it suffices to show that, for all $i \in I$ and $\psi \in \Pi_{k-1}$,

$$Q \text{ is a version of the conditional distribution of } X_i \text{ given } \psi(Y_i) \text{ under } P.$$

Fix $i \in I$, $\psi \in \Pi_{k-1}$, and define

$$\mu(\cdot) = P\big(\psi(Y_i) \in \cdot\big) \quad \text{and} \quad \nu(\cdot) = P\big(Y_i \in \cdot\big)$$

to be the marginal distributions of $\psi(Y_i)$ and Y_i under P. Then,

$$\mu \circ \psi(B) = \mu\big(\psi(B)\big) = P\big(\psi(Y_i) \in \psi(B)\big) = P\big(Y_i \in B\big) = \nu(B)$$

for all $B \in \mathcal{B}(\mathcal{X}^{k-1})$. Thus, $\mu \circ \psi = \nu$. Together with (8), this fact implies

$$\begin{aligned} \int_B Q(y, A)\, \mu(dy) &= \int_B Q\big(\psi^{-1}(y),\ A\big)\, \mu(dy) = \int_{\psi^{-1}(B)} Q(y, A)\, \mu \circ \psi(dy) \\ &= \int_{\psi^{-1}(B)} Q(y, A)\, \nu(dy) = P\big(X_i \in A,\ Y_i \in \psi^{-1}(B)\big) = P\big(X_i \in A,\ \psi(Y_i) \in B\big) \end{aligned}$$

for all $A \in \mathcal{B}(\mathcal{X})$ and $B \in \mathcal{B}(\mathcal{X}^{k-1})$. Hence, Q is a version of the conditional distribution of X_i given $\psi(Y_i)$ under P. This concludes the proof. □

In view of Theorem 12, $\mathcal{E}$-compatibility reduces to compatibility as far as conditions (8)–(9) are satisfied. In turn, in many real problems, compatibility can be tested via Theorem 10. This provides a usable strategy for checking $\mathcal{E}$-compatibility. Moreover, under some conditions, Theorem 12 gives a necessary condition for compatibility as well.

Corollary 13. *Suppose condition (4) holds with*

$$f_1 = \cdots = f_k \quad \textit{and} \quad \lambda_1 = \cdots = \lambda_k.$$

Suppose also that

$$f_1 > 0 \quad \textit{and} \quad f_1\big(\cdot \mid \pi(y)\big) = f_1(\cdot \mid y) \quad \textit{for all } y \in \mathcal{X}^{k-1} \textit{ and } \pi \in \Pi_{k-1}.$$

Then, $f_1, \ldots, f_k$ are compatible if and only if they are $\mathcal{E}$-compatible, if and only if there is a strictly positive Borel function g on $\mathcal{X}^k$ such that

$$g = g \circ \pi \quad \textit{for all } \pi \in \Pi_k, \quad g \textit{ is a density with respect to } \lambda,$$
$$f_1(x \mid y) = \frac{g(x,y)}{\int_{\mathcal{X}} g(u,y)\,\lambda_1(du)} \quad \textit{for } \lambda\textit{-almost all } (x,y) \in \mathcal{X}^k.$$

Proof. Since conditions (8)–(9) hold, it suffices to see that f_1 can be represented as asserted whenever $f_1, \ldots, f_k$ are compatible. Suppose $f_1, \ldots, f_k$ are compatible. Then, $f_1, \ldots, f_k$ are actually $\mathcal{E}$-compatible. Fix $P \in \mathcal{E}$ with conditional densities $f_1, \ldots, f_k$. Since $f_1 > 0$, Theorem 10-(a) yields $P \sim \lambda$. Let g be a density of P with respect to λ. Since $P \sim \lambda$, $P \in \mathcal{E}$, and $\lambda = \lambda_1^k$ is invariant under permutations, up to modifying g on a λ-null set, it can be assumed $g > 0$ and $g = g \circ \pi$ for all $\pi \in \Pi_k$. Further, $f_1(x \mid y) = \left\{\int_{\mathcal{X}} g(u,y)\,\lambda_1(du)\right\}^{-1} g(x,y)$ for λ-almost all $(x,y) \in \mathcal{X}^k$. □

To exploit Corollary 13, the following remark is useful.

Remark 14. Let $\lambda_1 = \cdots = \lambda_k$ and let φ and h be real Borel functions on $\mathcal{X}^k$ and $\mathcal{X}^{k-1}$, respectively. If $\varphi = \varphi \circ \pi$ for all $\pi \in \Pi_k$ and

$$h(y) = \varphi(x,y) \quad \text{for } \lambda\text{-almost all } (x,y) \in \mathcal{X}^k,$$

then h is constant, λ_1^{k-1}-a.e. We omit the proof of this fact.

Example 15 (*Normal Distributions Depending on the Sample Mean*). Let $\mathcal{X} = \mathbb{R}$ and

$$Q_1(y,\cdot) = \cdots = Q_k(y,\cdot) = N\big(\alpha\,\overline{y},\ 1\big)$$

where $\alpha \in \mathbb{R}$, $y = (y_1, \ldots, y_{k-1}) \in \mathbb{R}^{k-1}$ and $\overline{y} = (1/(k-1)) \sum_{i=1}^{k-1} y_i$. We aim to identify those values of α which make $Q_1, \ldots, Q_k$ compatible. Let $f_i = f_1$ and $\lambda_i = \lambda_1$ for all $i \in I$, where f_1 is a normal density with mean $\alpha\,\overline{y}$ and unit variance while λ_1 is Lebesgue measure. We first assume $k = 2$. Write

$$f_1(x \mid y) = (2\,\pi)^{-1/2} \exp\big\{-(1/2)(x - \alpha\,y)^2\big\} = \frac{\varphi(x,y)}{h(y)} \quad \text{where}$$
$$\varphi(x,y) = (2\,\pi)^{-1/2} \exp\big\{-(1/2)(x^2 + y^2) + \alpha\,x\,y\big\}, \qquad h(y) = \exp\big\{(1/2)\,y^2(\alpha^2 - 1)\big\}.$$

If $|\alpha| < 1$, then $0 < \int_{\mathbb{R}} h(y)\,dy < \infty$. Letting

$$g(x,y) = \frac{\varphi(x,y)}{\int_{\mathbb{R}} h(y)\,dy},$$

Corollary 13 implies that Q_1 and Q_2 are compatible. Next, suppose $|\alpha| \geq 1$. If Q_1 and Q_2 are compatible, Corollary 13 yields

$$\frac{\int_{\mathbb{R}} g(u,y)\,du}{h(y)} = \frac{g(x,y)}{\varphi(x,y)}$$

for a suitable density function g and λ-almost all $(x,y) \in \mathbb{R}^2$. Since the right-hand member is invariant under permutations of (x,y) while the left-hand member depends on y only, Remark 14 implies $\int_{\mathbb{R}} g(u,y)\,du = c\,h(y)$ for some constant $c > 0$ and λ_1-almost all y. But since $|\alpha| \geq 1$, one obtains

$$\int_{\mathbb{R}^2} g\,d\lambda = \int_{\mathbb{R}} \int_{\mathbb{R}} g(u,y)\,du\,dy = \int_{\mathbb{R}} c\ \exp\big\{(1/2)\,y^2(\alpha^2 - 1)\big\}\,dy = \infty,$$

contrary to the assumption that g is a density with respect to λ. To sum up, Q_1 and Q_2 are compatible if and only if $|\alpha| < 1$.

The previous argument actually works for any k. In fact, f_1 can be factorized as

$$f_1(x \mid y) = (2\,\pi)^{-1/2} \exp\big\{-(1/2)(x - \alpha\,\overline{y})^2\big\} = \frac{\varphi(x,y)}{h(y)}$$

where φ is invariant under permutations of $(x,y) \in \mathbb{R}^k$ and h depends on $y \in \mathbb{R}^{k-1}$ only. Then, $Q_1, \ldots, Q_k$ are compatible exactly for those values of α such that $\int_{\mathbb{R}^{k-1}} h(y)\,dy < \infty$. For $k = 3$, for instance, Q_1, Q_2, Q_3 are compatible if and only if $4 - \alpha^2 > |2\,\alpha + \alpha^2|$, that is, $\alpha \in (-2, 1)$.

A last note is in order before leaving this Subsection. For $k = 2$, condition (8) is trivially true. Furthermore, if $Q_1 = Q_2 = Q$, compatibility of Q_1 and Q_2 amounts to reversibility of the kernel Q. We recall that, for $k = 2$ and $\Omega_1 = \Omega_2 = \mathcal{X}$, a kernel Q is *reversible* if there is a probability measure μ on $\mathcal{B}(\mathcal{X})$ such that

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

The conditional distributions of an (exchangeable) law $P \in \mathcal{E}$ are actually reversible; see e.g. Theorem 3.2 of [4].

Theorem 16. *Suppose $k = 2$ and $Q_1 = Q_2 = Q$ for some kernel Q. Then, Q_1 and Q_2 are compatible if and only if they are $\mathcal{E}$-compatible, if and only if Q is a reversible kernel.*

Proof. By Theorem 12, it suffices to prove that Q_1 and Q_2 are $\mathcal{E}$-compatible if and only if Q is reversible. Suppose Q is reversible. 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, $P \in \mathcal{E}$. Also, Q is a conditional distribution, under P, for X_1 given X_2 as well as for X_2 given X_1. Conversely, suppose Q_1 and Q_2 are $\mathcal{E}$-compatible. Letting $\mu(\cdot) = P(X_1 \in \cdot)$, where $P \in \mathcal{E}$ has conditionals Q_1 and Q_2, it is straightforward to see that Q meets condition (11). $\square$

5.2. Identical marginal distributions

If X is exchangeable, X_i is distributed as X_1 for all $i \in I$, but not conversely. In a number of frameworks, when modeling the joint distribution P of X via a set of putative conditional distributions, one is actually looking for some P which makes $X_1, \ldots, X_k$ identically distributed. Thus, it makes sense to study $\mathcal{I}$-compatibility, where

$$\mathcal{I} = \{P \in \mathcal{P} : X_1, \ldots, X_k \text{ identically distributed under } P\}.$$

If only $Q_1, \ldots, Q_k$ are assigned, as in this paper, to investigate $\mathcal{I}$-compatibility for $k > 2$ looks quite difficult (to us). But for $k = 2$ and $\mathcal{X}$ countable, a useful result can be obtained.

Let $k = 2$. By adapting the proof of Theorem 16, it is not hard to prove that Q_1 and Q_2 are $\mathcal{I}$-compatible if and only if there is a probability measure μ on $\mathcal{B}(\mathcal{X})$ such that

$$\int_A Q_2(x, B)\,\mu(dx) = \int_B Q_1(x, A)\,\mu(dx) \quad \text{for all } A,\ B \in \mathcal{B}(\mathcal{X}).$$

In fact, under such condition, there is $P \in \mathcal{I}$ satisfying: (i) P has conditionals Q_1 and Q_2; (ii) both X_1 and X_2 have marginal distribution μ under P.

Suppose that $\mathcal{X}$ is countable and Q is a kernel on $\mathcal{X}$. As usual, we will write $Q(a, b)$ instead of $Q(a, \{b\})$ whenever $a,\ b \in \mathcal{X}$. We also need the following (well known) definition. Given $a,\ b \in \mathcal{X}$, *a path connecting a and b* is a finite sequence $x_0, x_1, \ldots, x_n \in \mathcal{X}$ such that $x_0 = a$, $x_n = b$ and $Q(x_{i-1},\ x_i) > 0$ for all i. Also, Q is *irreducible* if any pair of points in $\mathcal{X}$ are connected by a path.

We are now able to state our last result.

Theorem 17. *Suppose $k = 2$, $\mathcal{X}$ countable and Q_1 irreducible. Fix $a \in \mathcal{X}$. Then, Q_1 and Q_2 are $\mathcal{I}$-compatible if and only if*

$$\prod_{i=1}^{n} Q_1(x_{i-1},\ x_i) = \prod_{i=1}^{n} Q_2(x_i,\ x_{i-1}) \tag{12}$$

whenever $x_0, x_1, \ldots, x_n \in \mathcal{X}$ and $x_n = x_0$,

$$Q_1(x, y) > 0 \Longleftrightarrow Q_2(y, x) > 0 \tag{13}$$

for all $x,\ y \in \mathcal{X}$, and

$$\sum_{x\in\mathcal{X}} \prod_{i=1}^{n} \frac{Q_1\big(b^x_{i-1},\ b^x_i\big)}{Q_2\big(b^x_i,\ b^x_{i-1}\big)} < \infty \tag{14}$$

whenever $b^x_0, \ldots, b^x_n$ is a path connecting a and x. (Hence, $b^x_0 = a$, $b^x_n = x$ and $Q_1\big(b^x_{i-1},\ b^x_i\big) > 0$ for all i).

The proof of Theorem 17 is deferred to the Appendix.

Theorem 17 notably simplifies in some special cases. Firstly, if there is a point $a \in \mathcal{X}$ such that $Q_1(a, x) > 0$ for all $x \in \mathcal{X}$, one can take $n = 1$, $b_0^x = a$ and $b_1^x = x$ in condition (14). Thus, such condition reduces to

$$\sum_{x\in\mathcal{X}} \frac{Q_1(a,x)}{Q_2(x,a)} < \infty.$$

More importantly, condition (14) can be dropped at all when $\mathcal{X}$ is finite.

Corollary 18. *If $k = 2$, $\mathcal{X}$ is finite and Q_1 irreducible, then Q_1 and Q_2 are $\mathcal{I}$-compatible if and only if conditions (12)–(13) hold.*

Corollary 18 provides a simple and effective criterion for $\mathcal{I}$-compatibility. Condition (13), in fact, is trivially seen to be true or false. Suppose it is true. Then, to check (12), one can restrict to those sequences $x_0, x_1, \ldots, x_n \in \mathcal{X}$ such that $x_n = x_0$ and $Q_1(x_{i-1}, x_i) > 0$ for all i. Moreover, as it is easily verified by induction, it can be assumed $x_i \neq x_j$ for all $0 \le i < j < n$. Thus, when $\mathcal{X}$ is finite and Q_1 irreducible, $\mathcal{I}$-compatibility can be tested via a *finite number* of straightforward conditions. If the values of Q_1 and Q_2 are uploaded into a computer, one obtains an on-line, definitive answer on whether Q_1 and Q_2 are $\mathcal{I}$-compatible or not.

To be concrete, we give a numerical example.

Example 19. With $\mathcal{X} = \{1, 2, 3, 4\}$, let

$$Q_1 = \begin{pmatrix} 1/10 & 0 & 3/10 & 3/5 \\ 0 & 2/11 & 4/11 & 5/11 \\ 4/15 & 1/5 & 8/15 & 0 \\ 1/4 & 3/10 & 0 & 9/20 \end{pmatrix} \quad \text{and} \quad Q_2 = \begin{pmatrix} 1/10 & 0 & 2/5 & 1/2 \\ 0 & 2/11 & 3/11 & 6/11 \\ 1/5 & 4/15 & 8/15 & 0 \\ 3/10 & 1/4 & 0 & 9/20 \end{pmatrix}.$$

Such Q_1 and Q_2 are $\mathcal{I}$-compatible, and this can be proved as follows. First note that Q_1 is irreducible and condition (13) is trivially true. By Corollary 18, thus, it suffices to check condition (12). Let $x_0, x_1, \ldots, x_n \in \mathcal{X}$ be such that

$$x_n = x_0, \qquad x_i \neq x_j \quad \text{for } 0 \le i < j < n, \qquad \prod_{i=1}^{n} Q_1(x_{i-1}, x_i) > 0. \tag{15}$$

To fix ideas, let $x_0 = 1$. It must be $1 \le n \le 4$. Since $Q_1(1, 1) = Q_2(1, 1)$, condition (12) holds for $n = 1$ and $x_0 = 1$. For $n = 3$, no path satisfies (15) and $x_0 = 1$. For $n = 2$ and $n = 4$, the paths satisfying (15) and $x_0 = 1$ are

$$x_0 = 1,\ x_1 = 3,\ x_2 = 1; \qquad x_0 = 1,\ x_1 = 4,\ x_2 = 1;$$
$$x_0 = 1,\ x_1 = 3,\ x_2 = 2,\ x_3 = 4,\ x_4 = 1; \qquad x_0 = 1,\ x_1 = 4,\ x_2 = 2,\ x_3 = 3,\ x_4 = 1.$$

All such paths meet condition (12). Similarly, (12) is immediately seen to be true for $x_0 > 1$. Therefore, Q_1 and Q_2 are $\mathcal{I}$-compatible.

We finally give an example with an infinite state space $\mathcal{X}$.

Example 20 (*Random Walk on the Integers*)**.** Let $\mathcal{X} = \mathbb{Z}$ be the integers and let Q be the kernel of the symmetric random walk on $\mathbb{Z}$, that is, $Q(x, y) = 1/2$ if $y \in \{x-1,\ x+1\}$ and $Q(x, y) = 0$ if $y \notin \{x-1,\ x+1\}$. A first (obvious) question is whether Q is compatible with itself. More precisely, letting $Q_1 = Q_2 = Q$, the question is whether Q_1 and Q_2 are compatible. Since Q is clearly not reversible, the answer is no because of Theorem 16. The second possible question is the following. Let $Q_1 = Q$. Is there a kernel Q_2 on $\mathbb{Z}$ such that Q_1 and Q_2 are $\mathcal{I}$-compatible? Fix a kernel Q_2. Since $Q_1 = Q$ is irreducible, Theorem 17 applies. Thus, if Q_1 and Q_2 are $\mathcal{I}$-compatible, condition (13) implies $Q_2(x, y) > 0$ if $y \in \{x-1,\ x+1\}$ and $Q_2(x, y) = 0$ if $y \notin \{x-1,\ x+1\}$. Let $\alpha(x) = Q_2(x, x+1)$. For each $x \in \mathbb{Z}$, condition (12) yields

$$1/4 = Q_1(x, x+1)Q_1(x+1, x) = Q_2(x+1, x)Q_2(x, x+1) = \{1 - \alpha(x+1)\}\alpha(x).$$

Therefore,

$$\alpha(x+1) = 1 - \frac{1}{4\,\alpha(x)}. \tag{16}$$

To fix ideas, suppose $\alpha(0) \ge 1/2$. Then, condition (16) implies $\alpha(x) \ge 1/2$ for all $x \ge 1$, so that

$$\frac{Q_1(0,1)Q_1(1,2)\cdots Q_1(x-1,x)}{Q_2(1,0)Q_2(2,1)\cdots Q_2(x,x-1)} = \frac{(1/2)^x}{(1-\alpha(1))(1-\alpha(2))\cdots(1-\alpha(x))} \ge 1$$

for all $x \ge 1$. Hence, condition (14) fails (just let $a = 0$, $n = x$ and $b_i^x = i$). Similarly, condition (14) fails if $\alpha(0) < 1/2$. By Theorem 17, thus, no kernel Q_2 is $\mathcal{I}$-compatible with $Q_1 = Q$.

Appendix

We have to prove Theorems 10 and 17. We begin with point (a) of Theorem 10.

Lemma 21. *Suppose (4) holds and $P \in \mathscr{P}$ 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 distribution of Y_k under P. Fix $A \in \mathscr{B}(\Omega)$ such that $\lambda(A) = 0$ and define

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

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 distribution of X_1 under P. If $A \in \mathscr{B}(\Omega_1)$ 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 distribution of (X_1, X_2) under P. For μ_1-almost all $x \in \Omega_1$, one obtains

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

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

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

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 $\mu \ll \lambda_1 \times \cdots \times \lambda_{k-1} = \lambda_k^*$. 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 $A \in \mathscr{B}(\Omega_i)$ and $\lambda_i(A) > 0$. Based on this fact, $P \sim \lambda$ can be proved exactly as above. $\square$

Proof of Theorem 10. Point (a) has been proved in Lemma 21. Recall also that

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

Suppose $f_1, \ldots, f_k$ are compatible and fix $P \in \mathscr{P}$ with conditional distributions $Q_1, \ldots, Q_k$. By point (a), 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

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

and note that

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

Let $H_i = \{u_i(Y_i) > 0\}$. 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.,

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

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

Therefore, conditions (6)–(7) hold. Conversely, suppose (6)–(7) hold for some functions $u_1, \ldots, u_k$. Define again $H_i = \{u_i(Y_i) > 0\}$. By (7),

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

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 (7) 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 (6), 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,

$$f = \frac{f_i(X_i \mid Y_i)}{u_i(Y_i)} I_{H_i}, \quad \lambda\text{-a.e. for all } i < k. \tag{17}$$

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 (17) 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 (17) implies $f/\phi_i(Y_i) = f_i(X_i \mid Y_i)$, λ-a.e. on H_i. Since point (b) is obvious, this concludes the proof. $\square$

We finally turn to Theorem 17.

Proof of Theorem 17. Assume conditions (12)–(14). Let $x_0, x_1, \ldots, x_r$ and $y_0, y_1, \ldots, y_s$ be any two paths connecting a and x. Take a further path $z_0, z_1, \ldots, z_t$ connecting x and a. On noting that $x_0 = y_0 = z_t = a$ and $x_r = y_s = z_0 = x$, condition (12) yields

$$\begin{aligned}\prod_{i=1}^r Q_1(x_{i-1}, x_i) \prod_{i=1}^t Q_1(z_{i-1}, z_i) \prod_{i=1}^s Q_2(y_i, y_{i-1}) &= \prod_{i=1}^r Q_2(x_i, x_{i-1}) \prod_{i=1}^t Q_2(z_i, z_{i-1}) \prod_{i=1}^s Q_2(y_i, y_{i-1}) \\ &= \prod_{i=1}^r Q_2(x_i, x_{i-1}) \prod_{i=1}^t Q_1(z_{i-1}, z_i) \prod_{i=1}^s Q_1(y_{i-1}, y_i).\end{aligned}$$

By condition (13) and the definition of path, all factors are strictly positive. Hence,

$$\prod_{i=1}^r \frac{Q_1(x_{i-1}, x_i)}{Q_2(x_i, x_{i-1})} = \prod_{i=1}^s \frac{Q_1(y_{i-1}, y_i)}{Q_2(y_i, y_{i-1})}.$$

Next, define

$$\nu\{x\} = \prod_{i=1}^r \frac{Q_1(x_{i-1}, x_i)}{Q_2(x_i, x_{i-1})}.$$

By what already proved, the definition of $\nu\{x\}$ does not depend on the path connecting a and x. Hence, ν is a (well defined) measure on the power set of $\mathcal{X}$, and $\nu(\mathcal{X}) = \sum_{x\in\mathcal{X}} \nu\{x\} < \infty$ because of (14). Define $\mu = \nu/\nu(\mathcal{X})$. To conclude the proof of the "if" part, it suffices to see that

$$\mu\{x\}\, Q_1(x, y) = \mu\{y\}\, Q_2(y, x) \quad \text{for all } x, y \in \mathcal{X}. \tag{18}$$

In view of (13), to check condition (18) it can be assumed $Q_1(x, y) > 0$. In this case, the very definition of μ yields

$$\mu\{x\}\,\frac{Q_1(x, y)}{Q_2(y, x)} = \mu\{y\}.$$

Conversely, suppose Q_1 and Q_2 are $\mathcal{I}$-compatible. Take a probability μ satisfying condition (18). Summing over $x \in \mathcal{X}$, one obtains

$$\mu\{y\} = \sum_{x\in\mathcal{X}} \mu\{y\}\, Q_2(y, x) = \sum_{x\in\mathcal{X}} \mu\{x\}\, Q_1(x, y) \quad \text{for all } y \in \mathcal{X}.$$

Thus, μ is an invariant probability for the irreducible kernel Q_1, and this fact implies $\mu\{x\} > 0$ for all $x \in \mathcal{X}$. Therefore, condition (13) follows from (18) and $\mu\{x\} > 0$ for all $x \in \mathcal{X}$. Next, let $x_0, x_1, \ldots, x_n \in \mathcal{X}$ with $x_n = x_0$. If $Q_1(x_{i-1}, x_i) = 0$ for some i, condition (13) yields

$$\prod_{i=1}^{n} Q_1(x_{i-1}, x_i) = 0 = \prod_{i=1}^{n} Q_2(x_i, x_{i-1}).$$

If $Q_1(x_{i-1}, x_i) > 0$ for all i, one obtains

$$\prod_{i=1}^{n} \frac{Q_1(x_{i-1}, x_i)}{Q_2(x_i, x_{i-1})} = \prod_{i=1}^{n} \frac{\mu\{x_i\}}{\mu\{x_{i-1}\}} = \frac{\mu\{x_n\}}{\mu\{x_0\}} = \frac{\mu\{x_0\}}{\mu\{x_0\}} = 1.$$

Thus, condition (12) holds. Finally, as to (14), it suffices to note that

$$\sum_{x\in\mathcal{X}} \prod_{i=1}^{n} \frac{Q_1\big(b_{i-1}^x, b_i^x\big)}{Q_2\big(b_i^x, b_{i-1}^x\big)} = \sum_{x\in\mathcal{X}} \prod_{i=1}^{n} \frac{\mu\{b_i^x\}}{\mu\{b_{i-1}^x\}} = \sum_{x\in\mathcal{X}} \frac{\mu\{x\}}{\mu\{a\}} = \frac{1}{\mu\{a\}}$$

whenever $b_0^x, \ldots, b_n^x$ is a path connecting a and x. □

References

[1] B.C. Arnold, E. Castillo, J.M. Sarabia, Conditionally specified distributions: an introduction (with discussion), Statist. Sci. 16 (2001) 249–274.
[2] B.C. Arnold, S.J. Press, Compatible conditional distributions, J. Amer. Statist. Assoc. 84 (1989) 152–156.
[3] P. Berti, E. Dreassi, P. Rigo, A consistency theorem for regular conditional distributions, Stochastics 85 (2013) 500–509.
[4] P. Berti, L. Pratelli, P. Rigo, Limit theorems for empirical processes based on dependent data, Electron. J. Probab. 17 (2012) 1–18.
[5] J. Besag, Spatial interaction and the statistical analysis of lattice systems (with discussion), J. R. Stat. Soc. Ser. B Stat. Methodol. 36 (1974) 192–236.
[6] J. Besag, On statistical analysis of dirty pictures (with discussion), J. R. Stat. Soc. Ser. B Stat. Methodol. 48 (1996) 259–302.
[7] J. Besag, J. York, A. Mollie, Bayesian image restoration, with two applications in spatial statistics (with discussion), Ann. Inst. Statist. Math. 43 (1991) 1–59.
[8] H.Y. Chen, Compatibility of conditionally specified models, Statist. Probab. Lett. 80 (2010) 670–677.
[9] D. Clayton, J. Kaldor, Empirical Bayes estimates of age-standardized relative risks for use in disease mapping, Biometrics 43 (1987) 671–681.
[10] H.O. Georgii, Gibbs Measures and Phase Transitions, second ed., in: de Gruyter Studies in Mathematics, vol. 9, 2011.
[11] J.P. Hobert, G. Casella, Functional compatibility, Markov chains, and Gibbs sampling with improper posteriors, J. Comput. Graph. Statist. 7 (1998) 42–60.
[12] E.H. Ip, Y.J. Wang, Canonical representation of conditionally specified multivariate discrete distributions, J. Multivariate Anal. 100 (2009) 1282–1290.
[13] M.S. Kaiser, N. Cressie, The construction of multivariate distributions from Markov random fields, J. Multivariate Anal. 73 (2000) 199–220.
[14] C-C. Song, L-A. Li, C-H. Chen, T.J. Jiang, K-L. Kuo, Compatibility of finite discrete conditional distributions, Statist. Sinica 20 (2010) 423–440.
[15] S. van Buuren, Multiple imputation of discrete and continuous data by fully conditional specification, Stat. Methods Med. Res. 16 (2007) 219–242.
[16] S. van Buuren, J.P.L. Brand, C.G.M. Groothuis-Oudshoorn, D.B. Rubin, Fully conditional specification in multivariate imputation, J. Stat. Comput. Simul. 76 (2006) 1049–1064.
[17] Y.J. Wang, E.H. Ip, Conditionally specified continuous distributions, Biometrika 95 (2008) 735–746.