A NOTE ON COMPATIBILITY OF CONDITIONAL AUTOREGRESSIVE MODELS

EMANUELA DREASSI AND PIETRO RIGO

Abstract. Suppose that, to assess the joint distribution of a random vector $(X_1, \ldots, X_n)$, one selects the kernels $Q_1, \ldots, Q_n$ with Q_i to be regarded as a possible conditional distribution for X_i given $(X_j : j \neq i)$; $Q_1, \ldots, Q_n$ are compatible if there exists a joint distribution for $(X_1, \ldots, X_n)$ with conditionals $Q_1, \ldots, Q_n$. Similarly, $Q_1, \ldots, Q_n$ are improperly compatible if they can be obtained, according to the usual rule, with an improper distribution in place of a probability distribution. In this paper, compatibility and improper compatibility of $Q_1, \ldots, Q_n$ are characterized under some assumptions on their functional form. The characterization applies, in particular, if each Q_i belongs to a one parameter exponential family. Special attention is paid to Gaussian conditional autoregressive models.

1. Introduction

Let $I = \{1, \ldots, n\}$. For each $i \in I$, let $\mathcal{X}_i$ be a Polish space (complete separable metric space) and X_i an $\mathcal{X}_i$-valued random variable.

Sometimes, every X_i is requested to have an *assigned* conditional distribution given $(X_j : j \neq i)$. The main reason is to assess the joint distribution of the vector $(X_1, \ldots, X_n)$ by specifying some of its conditionals. Quoting from [18, page 171]: "It is frequently difficult or impossible in complex situations to specify a model through formulation of a joint distribution for a complete set of response variables. Even in the Gaussian case, where it may be possible to write such a joint distribution, the relative merits of conditional specification versus simultaneous specification of a statistical model may lead one to prefer the conditional approach". Indeed, such a conditional approach is standard practice in various fields, including spatial statistics, statistical mechanics, Bayesian image analysis, multiple data imputation and Gibbs sampling. See e.g. [1]-[3], [7]-[10], [12], [16]-[17], [21].

To handle situations of this type, it is convenient to let $\mathcal{X} = \prod_{i \in I} \mathcal{X}_i$ and to take $X_1, \ldots, X_n$ to be the canonical projections on $\mathcal{X}$, namely

$$X_i(x) = x_i \quad \text{for } i \in I \text{ and } x \in \mathcal{X}$$

where $x = (x_1, \ldots, x_n)$ and x_i denotes the i-th coordinate of x. Also, for each $i \in I$, we let

$$\mathcal{X}_{-i} = \prod_{j \neq i} \mathcal{X}_j$$

and we fix a function Q_i on $\mathcal{X}_{-i}$ satisfying

2010 *Mathematics Subject Classification.* 62F15, 62H05, 62E99, 62G99, 60A05.

Key words and phrases. Bayesian inference, CAR models, Compatibility of conditional distributions, Gibbs sampling, Improper distribution, Markov random fields.

(i) $Q_i(y)$ is a probability measure on $\mathcal{B}(\mathcal{X}_i)$ for fixed $y \in \mathcal{X}_{-i}$,
(ii) $y \mapsto Q_i(y)(B)$ is measurable for fixed $B \in \mathcal{B}(\mathcal{X}_i)$.

Here and in the sequel, "measurable" stands for "Borel measurable" and $\mathcal{B}(S)$ denotes the Borel σ-field on S for any topological space S.

Depending on the framework, $Q_1, \ldots, Q_n$ are given various names, such as *probability kernels* (or merely kernels), *putative conditional distributions*, *full conditional distributions* and *conditional autoregressive* (CAR) models; see e.g. [1], [3], [6], [7], [9], [10], [12], [16], [17].

Let $\mathcal{P}$ be the set of all probability measures on $\mathcal{B}(\mathcal{X})$. Each Q_i should be regarded as a conditional distribution of X_i given $(X_j : j \neq i)$. But of course, since Q_i is only subjected to (i)-(ii), it may be that no $P \in \mathcal{P}$ admits $Q_1, \ldots, Q_n$ as conditional distributions. In this case, $Q_1, \ldots, Q_n$ are not compatible. Instead, $Q_1, \ldots, Q_n$ are *compatible* if they are the conditionals of some $P \in \mathcal{P}$. A seminal paper on compatibility is Besag's [7]. Some other references, without any claim to be exhaustive, are [1]-[2], [5]-[12], [14]-[16], [18]-[22].

An *improper* distribution is an infinite measure on $\mathcal{B}(\mathcal{X})$. Even if $Q_1, \ldots, Q_n$ are not compatible, it may be that they can be obtained, according to the usual rule, with an improper distribution in place of a probability distribution. To give a formal definition, we need some notation. For each $i \in I$, let λ_i be a σ-finite measure on $\mathcal{B}(\mathcal{X}_i)$ and let $\lambda = \lambda_1 \times \ldots \times \lambda_n$ denote the corresponding product measure on $\mathcal{B}(\mathcal{X})$. Suppose that $Q_i(y)$ has a density $f_i(\cdot \mid y)$ with respect to λ_i, namely,

$$Q_i(y)(dz) = f_i(z \mid y)\, \lambda_i(dz) \tag{1}$$

for all $i \in I$ and $y \in \mathcal{X}_{-i}$. Further, define

$$x_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \quad \text{for all } x \in \mathcal{X} \text{ and } i \in I.$$

Under (1), we say that $Q_1, \ldots, Q_n$ are *improperly compatible* if

$$f_i(x_i \mid x_{-i}) = \frac{f(x)}{\int_{\mathcal{X}_i} f(x_1, \ldots, x_{i-1}, z, x_{i+1}, \ldots, x_n)\, \lambda_i(dz)} \tag{2}$$
$$\text{for all } i \in I \text{ and } \lambda\text{-almost all } x \in \mathcal{X},$$

where f is a strictly positive measurable function on $\mathcal{X}$ and the integral in the denominator is finite.

Let Q be the measure on $\mathcal{B}(\mathcal{X})$ with density f with respect to λ. If $\int_{\mathcal{X}} f\, d\lambda < \infty$, then Q can be normalized to be a probability measure. In this case, $Q_1, \ldots, Q_n$ are even compatible (and not only improperly compatible). Instead, $Q_1, \ldots, Q_n$ are improperly compatible but not compatible whenever $\int_{\mathcal{X}} f\, d\lambda = \infty$. Even in this case, however, they can be *formally* obtained by the usual rule starting from the improper distribution Q.

Since $f > 0$ everywhere, definition (2) makes sense only if $f_i > 0$ for all $i \in I$, which is certainly a restrictive condition. As remarked by an anonymous referee, such definition could be generalized by introducing suitable conventions (such as $0/0 = 0$) and allowing for $\lambda(f = 0) > 0$. As it stands, however, definition (2) suffices for our purposes and we do not try to generalize it in this paper.

Improper compatibility is meaningless in probability theory. According to the latter, $Q_1, \ldots, Q_n$ are either compatible or not compatible, without intermediate cases. Nevertheless, the improper case is statistically meaningful. It is very popular, for instance, in Bayesian statistics; see e.g. [4], [13] and references therein. In that framework, the prior distribution of a random parameter is often improper, mainly if such prior is requested to be "non informative" (in some sense). However, though the literature on improper distributions is virtually endless, improper compatibility has not been paid much attention to date. Instead, it often occurs in applications.

Example 1. **(Gaussian CAR models).** For each $i \in I$, let $\mathcal{X}_i = \mathbb{R}$ and

$$Q_i(y) = N\Big(\sum_{j \neq i} w_{ij} y_j \;,\; \sigma_i^2\Big)$$

where $\sigma_i^2 > 0$, w_{ij} is any real number and $y = (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n) \in \mathbb{R}^{n-1}$. Such $Q_1, \ldots, Q_n$ are usually called *Gaussian CAR models* or *Gaussian auto-models*; see e.g. [3], [7], [9], [17]. An important special case, introduced in [9, Section 3], is that of Gaussian *intrinsic* CAR models. The latter are precisely those Gaussian CAR models which are improperly compatible but not compatible. Roughly speaking, intrinsic CAR models can be seen as a limiting form of compatible CAR models. They are frequently used in real problems, mainly to model spatial random effects; see e.g. [3, Section 4.3] and [17]. A well known example is

$$Q_i(y) = N\Big(\frac{\sum_{j \in N_i} y_j}{\operatorname{card}(N_i)} \;,\; \frac{c}{\operatorname{card}(N_i)}\Big)$$

where $c > 0$ is any constant and N_i a non-empty subset of I (to be regarded as the collection of "neighbors" of i).

In this paper, improper compatibility is explicitly taken into account. Our main results (Theorem 5 and Corollary 6) complete, unify and slightly extend the material in [7, Section 4]; see also [18] and [20]. A joint treatment of compatibility and improper compatibility, for random variables with values in quite general spaces (such as Polish spaces), is provided. In addition, a class of kernels larger than the exponential family is covered. Special attention is finally paid to Gaussian CAR models.

2. Preliminaries

From now on, condition (1) is assumed to hold, namely, $Q_i(y)$ admits a density $f_i(\cdot \mid y)$ with respect to λ_i. Since $\mathcal{B}(\mathcal{X}_i)$ is countably generated, $f_i(z \mid y)$ can be taken to be jointly measurable in z and y (i.e., $(z, y) \mapsto f_i(z \mid y)$ is measurable).

Accordingly we fix $f_1, \ldots, f_n$, jointly measurable and satisfying (1), and we say that $f_1, \ldots, f_n$ are compatible or improperly compatible to mean that $Q_1, \ldots, Q_n$ are compatible or improperly compatible. Furthermore,

$$\lambda = \lambda_1 \times \ldots \times \lambda_n \quad \text{and} \quad \lambda_{-i} = \lambda_1 \times \ldots \times \lambda_{i-1} \times \lambda_{i+1} \times \ldots \times \lambda_n$$

denote the product measures on $\mathcal{B}(\mathcal{X})$ and $\mathcal{B}(\mathcal{X}_{-i})$, respectively.

In the next result, u_i is a non-negative function on $\mathcal{X}_{-i}$ and

$$H_i = \big\{x \in \mathcal{X} : u_i(x_{-i}) > 0,\, u_n(x_{-n}) > 0\big\}.$$

Theorem 2. (Theorem 10 of [6]**).** *$f_1, \ldots, f_n$ are compatible if and only if there are measurable functions $u_i : \mathcal{X}_{-i} \to [0, \infty)$, $i \in I$, such that*

$$f_i(x_i \mid x_{-i}) = f_n(x_n \mid x_{-n})\, u_i(x_{-i})\, u_n(x_{-n}) \quad \textit{for } \lambda\textit{-almost all } x \in H_i \tag{3}$$

and

$$\int_{H_i} f_n(x_n \mid x_{-n})\, u_n(x_{-n})\, \lambda(dx) = \int_{\{u_i>0\}} 1/u_i \, d\lambda_{-i} = \int_{\mathcal{X}_{-n}} u_n \, d\lambda_{-n}$$

for all $i < n$, *where* $0 < \int_{\mathcal{X}_{-n}} u_n \, d\lambda_{-n} < \infty$.

We have reported Theorem 2 for completeness, but we actually need the following fact only.

Corollary 3. (Corollary 11 of [6]**).** *Suppose* $f_i > 0$ *for all* $i \in I$. *Then,* $f_1, \ldots, f_n$ *are compatible if and only if condition* (3) *holds for some strictly positive measurable functions* u_i *on* $\mathcal{X}_{-i}$, $i \in I$, *such that* $\int_{\mathcal{X}_{-n}} u_n \, d\lambda_{-n} < \infty$.

A simple modification of Corollary 3 allows to characterize improper compatibility. As can be expected, it suffices to drop the integrability condition on u_n.

Theorem 4. *Suppose* $f_i > 0$ *for all* $i \in I$. *Then,* $f_1, \ldots, f_n$ *are improperly compatible if and only if condition* (3) *holds for some (strictly positive, measurable) functions* $u_1, \ldots, u_n$.

The proof of Theorem 4, as well as of all other results in this paper, is postponed to a final appendix.

A last remark is in order. Suppose $f_1, \ldots, f_n$ are improperly compatible, fix $u_1, \ldots, u_n$ satisfying condition (3), and define $c = \int_{\mathcal{X}_{-n}} u_n \, d\lambda_{-n}$. Then, $f_1, \ldots, f_n$ are compatible if and only if $c < \infty$, and in that case they are the conditional densities of $f(x) = (1/c)\, f_n(x_n \mid x_{-n})\, u_n(x_{-n})$, $x \in \mathcal{X}$; see Corollary 3 and forthcoming Lemma 7. In particular, u_n/c is the marginal density of $(X_1, \ldots, X_{n-1})$ whenever $c < \infty$. As an example, take $n = 2$, $\mathcal{X}_1 = \mathcal{X}_2 = \mathbb{R}$ and $Q_1(y) = Q_2(y) = N(y, 1)$. Then, Q_1 and Q_2 are improperly compatible but not compatible. In fact, with $\lambda_1 = \lambda_2 =$ Lebesgue measure, one obtains $f_1 = f_2$. Hence, (3) holds with $u_1 = u_2 = 1$ but clearly $c = \int_{-\infty}^{\infty} u_2(y)\, dy = \infty$.

3. A class of CAR models

In this section, the abbreviation a.e. is always meant with respect to the corresponding product measure. More exactly, suppose $\mathcal{X}^* = \prod_{j \in J} \mathcal{X}_j$ for some $J \subset I$ and $\lambda^* = \prod_{j \in J} \lambda_j$ is the corresponding product measure. Then, a statement on $\mathcal{X}^*$ is said to hold a.e., or for almost all $x \in \mathcal{X}^*$, provided it holds λ^*-a.e. Moreover, a point $y \in \mathcal{X}_{-i}$ is written as $y = (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$.

We assume f_i to be of the form

$$f_i(z \mid y) = g_i(z)\, h_i(y) \prod_{j \neq i} q_{ij}(z, y_j) \tag{4}$$

for all $i \in I$, $z \in \mathcal{X}_i$ and $y \in \mathcal{X}_{-i}$, where

$$g_i : \mathcal{X}_i \to (0, \infty), \quad h_i : \mathcal{X}_{-i} \to (0, \infty), \quad q_{ij} : \mathcal{X}_i \times \mathcal{X}_j \to (0, \infty)$$

are strictly positive measurable functions.

We also need that the ratio q_{ij}/q_{ji} is either a.e. constant or can not be factorized as

$$\frac{q_{ij}(s,t)}{q_{ji}(t,s)} = v_i(s)\, v_j(t) \quad \text{for almost all } (s,t) \in \mathcal{X}_i \times \mathcal{X}_j, \tag{5}$$

where $v_i : \mathcal{X}_i \to (0,\infty)$ and $v_j : \mathcal{X}_j \to (0,\infty)$ are strictly positive measurable functions. Precisely, we say that q_{ij}/q_{ji} is *irreducible* if condition (5) implies that q_{ij}/q_{ji} is a.e. constant.

We begin with the following result.

Theorem 5. *Suppose that f_i satisfies condition* (4) *for all $i \in I$ and the ratios q_{ij}/q_{ji} are irreducible for all $i \neq j$. Then, $f_1, \ldots, f_n$ are improperly compatible if and only if*

$$q_{ij}(s,t) = c_{ij}\, q_{ji}(t,s), \tag{6}$$

for all $i \neq j$, some constant $c_{ij} > 0$, and almost all $(s,t) \in \mathcal{X}_i \times \mathcal{X}_j$. Moreover, $f_1, \ldots, f_n$ are compatible if and only if condition (6) *holds and $\int_{\mathcal{X}_{-n}} u_n\, d\lambda_{-n} < \infty$, where*

$$u_n(y) = \frac{1}{h_n(y)} \prod_{j\neq n} g_j(y_j) \prod_{j\neq n} \prod_{k\neq j,n} \sqrt{q_{jk}(y_j, y_k)} \qquad \text{for all } y \in \mathcal{X}_{-n}.$$

A meaningful case of Theorem 5 is when f_i belongs to a certain type of exponential family. In fact, condition (4) holds whenever

$$f_i(z \mid y) = g_i(z)\, h_i(y)\, \exp\big\{a_i(y) b_i(z)\big\} \tag{7}$$

where $b_i : \mathcal{X}_i \to \mathbb{R}$ is any measurable function and $a_i : \mathcal{X}_{-i} \to \mathbb{R}$ can be written as

$$a_i(y) = \alpha_i + \sum_{j\neq i} \beta_{ij}\, b_j(y_j) \tag{8}$$

for some real numbers α_i and β_{ij}.

Corollary 6. *For each $i \in I$, suppose that f_i satisfies conditions* (7)-(8) *and b_i is not a.e. constant. Then, $f_1, \ldots, f_n$ are improperly compatible if and only if $\beta_{ij} = \beta_{ji}$ for all $i \neq j$. Moreover, $f_1, \ldots, f_n$ are compatible if and only if $\beta_{ij} = \beta_{ji}$ for all $i \neq j$ and $\int_{\mathcal{X}_{-n}} u_n\, d\lambda_{-n} < \infty$, where*

$$u_n(y) = \frac{1}{h_n(y)} \prod_{j\neq n} g_j(y_j)\ \exp\Big\{\sum_{j\neq n} \alpha_j b_j(y_j) + \frac{1}{2}\sum_{j\neq n}\sum_{k\neq j,n} \beta_{jk} b_j(y_j) b_k(y_k)\Big\}$$

for all $y \in \mathcal{X}_{-n}$.

Corollary 6 is possibly the most important case of Theorem 5, but there are also situations where the latter applies while the former does not. An example is

$$f_i(z \mid y) = h_i(y)\, \exp\Big(-\sum_{j\neq i} \beta_{ij} |z - y_j|\Big)$$

where $\mathcal{X}_i = \mathbb{R}$, $\lambda_i =$ Lebesgue measure and the β_{ij} are non-negative coefficients.

A *neighborhood system* is a collection $\mathcal{N} = \big\{N_i : i \in I\big\}$, where N_i is a subset of I to be regarded as the class of "neighbors" of i. As usual, it is assumed that $i \notin N_i$ and $i \in N_j \Leftrightarrow j \in N_i$. Recall also that a *clique* is a subset $C \subset I$ such that $\operatorname{card}(C) = 1$ or C consists of sites that are all neighbors of each other.

In *Markov Random Fields* (also known as *Undirected Graphical Models*) a neighborhood system $\mathcal{N}$ is *given* and $f_1, \ldots, f_n$ are required to be consistent with $\mathcal{N}$; see e.g. [3, page 79]. Consistency means that $f_i(x_i \mid x_{-i})$ does not depend on x_j for every $j \notin N_i \cup \{i\}$. Thus, in Markov Random Fields, $f_1, \ldots, f_n$ can be taken to satisfy condition (4) when no clique has cardinality greater than 2.

Theorem 5 and Corollary 6 apply regardless of whether $\mathcal{N}$ is given or not. Under conditions (7)-(8), however, $f_1, \ldots, f_n$ are actually consistent with the neighborhood system $N_i = \{j : j \neq i,\, \beta_{ij} \neq 0\}$. In fact,

$$h_i(x_{-i}) = \Big(\int_{\mathcal{X}_i} g_i(z)\, \exp\big\{a_i(x_{-i}) b_i(z)\big\}\, \lambda_i(dz) \Big)^{-1} \quad \text{for all } x \in \mathcal{X},$$

so that $f_i(x_i \mid x_{-i})$ does not depend on x_j if $\beta_{ij} = 0$.

Some form of condition (8) is quite usual when $f_1, \ldots, f_n$ are from the exponential family; see e.g. [7], [8], [10], [16], [17]. Moreover, under some assumptions, (8) becomes a necessary condition for compatibility. In fact, (8) holds whenever $f_1, \ldots, f_n$ are compatible and consistent with a neighborhood system $\mathcal{N}$ such that no clique has cardinality greater than 2; see [7], [18] and [20].

We next turn to Gaussian CAR models, as defined in Example 1. Accordingly, for each $i \in I$, we let $\mathcal{X}_i = \mathbb{R}$, $\lambda_i =$ Lebesgue measure, and

$$f_i(z \mid y) = (2\, \pi \sigma_i^2)^{-1/2}\, \exp\Big\{ -\frac{(z - \sum_{j\neq i} w_{ij} y_j)^2}{2\, \sigma_i^2} \Big\}$$

where $z \in \mathbb{R}$, $y \in \mathbb{R}^{n-1}$, w_{ij} and σ_i^2 are real numbers and $\sigma_i^2 > 0$. Such f_i can be written as in (7) with

$$a_i(y) = \sum_{j\neq i} \frac{w_{ij}}{\sigma_i^2}\, y_j \quad \text{and} \quad b_i(z) = z.$$

By Corollary 6, it follows that $f_1, \ldots, f_n$ are improperly compatible if and only if

$$w_{ij}\, \sigma_j^2 = w_{ji}\, \sigma_i^2 \quad \text{for all } i \neq j. \tag{9}$$

Furthermore, under (9), $f_1, \ldots, f_n$ are compatible if and only if

$$\int_{\mathbb{R}^{n-1}} \exp\big\{-(1/2)\, y' A y\big\}\, dy < \infty, \tag{10}$$

where $y \in \mathbb{R}^{n-1}$ denotes here a column vector and A is the symmetric matrix

$$a_{ii} = \frac{1 - w_{ni} w_{in}}{\sigma_i^2}, \quad a_{ij} = -\frac{w_{ij} + w_{in} w_{nj}}{\sigma_i^2} \quad \text{for } i < j.$$

In fact, λ_{-n} is Lebesgue measure on $\mathbb{R}^{n-1}$ and the function u_n of Corollary 6 reduces to $u_n(y) = \exp\big\{-(1/2)\, y' A y\big\}$ (up to a multiplicative constant) because of condition (9).

Some remarks are in order.

An in-depth discussion of Gaussian CAR models, including compatibility and the interpretation of w_{ij} and σ_i^2, is in [3, Section 4.3]. Our treatment of Gaussian CAR models is actually connected to this reference.

Condition (9) appears in [9, page 734] to make symmetric a certain precision matrix. We came across (9) for a different purpose (improper compatibility) independently of [9].

Usually, the weights w_{ij} are non-negative. In principle, however, there might be (extreme) situations where $w_{ij} < 0$ makes sense. In medical statistics, for instance, a decrease of the epidemic in site i could be obtained at the expense of site j, implying competition between i and j.

Condition (10) is equivalent to positive definiteness of the matrix A. Accordingly, Gaussian CAR models are compatible if and only if A is positive definite and

$w_{ij}\,\sigma_j^2 = w_{ji}\,\sigma_i^2$ for $i \neq j$. Though implicit in the existing literature, to our knowledge, this result has not been explicitly stated so far. A few sufficient conditions are available, but the latter can be quickly deduced from such a result. For instance, by [9, page 734], Gaussian CAR models are compatible provided the graph induced by $\mathcal{N}$ is connected, $w_{ij}\,\sigma_j^2 = w_{ji}\,\sigma_i^2$ and $w_{ij} \geq 0$ for all $i \neq j$, and

$$\sum_{j \neq i} w_{ij} \leq 1 \quad \text{for all } i \in I \text{ with strict inequality for at least one } i.$$

In fact, under such conditions, A is easily seen to be positive definite.

Finally, fix a neighborhood system $\mathcal{N}$ and define

$$w_{ij} = 1_{N_i}(j)\,\frac{r_{ij}}{\sum_{k \in N_i} r_{ik}} \quad \text{and} \quad \sigma_i^2 = \frac{c}{\sum_{k \in N_i} r_{ik}}$$

where 1_{N_i} is the indicator function of N_i, the r_{ij} are strictly positive numbers satisfying $r_{ij} = r_{ji}$, and c is a strictly positive constant. Then, $f_1, \ldots, f_n$ are consistent with $\mathcal{N}$ and condition (9) holds. As an example, suppose that a covariate c_i is associated to each site i. Then, one could take $r_{ij} = \varphi(c_i, c_j)$ where φ is a symmetric strictly positive function. In particular, if φ is identically constant, w_{ij} reduces to

$$w_{ij} = \frac{1_{N_i}(j)}{\operatorname{card}(N_i)}.$$

The latter is the intrinsic CAR model already mentioned in Example 1. Such a model is quite intuitive and frequently used in applications. However, it is also subjected to some criticism. Specifically, it is not appropriate for those situations where neighborhood data are only weakly correlated; see [9] and [17].

Appendix

Proof of Theorem 4. Suppose that $f_1, \ldots, f_n$ are improperly compatible, namely, condition (2) holds for some strictly positive measurable function f. For $y \in \mathcal{X}_{-n}$, define $u_n(y) = \int_{\mathcal{X}_n} f(y,z)\,\lambda_n(dz)$ if the integral is finite and $u_n(y) = 1$ otherwise. Similarly, for $i < n$ and $y \in \mathcal{X}_{-i}$, let

$$u_i(y) = \Big(\int_{\mathcal{X}_i} f(y_1, \ldots, y_{i-1}, z, y_{i+1}, \ldots, y_n)\,\lambda_i(dz) \Big)^{-1}$$

if the integral is finite and $u_i(y) = 1$ otherwise. Exploiting such $u_1, \ldots, u_n$, it is straightforward to verify condition (3).

Conversely, suppose (3) holds for some strictly positive measurable functions $u_1, \ldots, u_n$. Define

$$f(x) = f_n(x_n \mid x_{-n})\,u_n(x_{-n}) \quad \text{for all } x \in \mathcal{X}.$$

For $y \in \mathcal{X}_{-n}$, since $f_n(\cdot \mid y)$ is a probability density with respect to λ_n,

$$\int_{\mathcal{X}_n} f(y,z)\,\lambda_n(dz) = u_n(y) \int_{\mathcal{X}_n} f_n(z \mid y)\,\lambda_n(dz) = u_n(y).$$

Hence,

$$\frac{f(x)}{\int_{\mathcal{X}_n} f(x_{-n}, z)\,\lambda_n(dz)} = \frac{f_n(x_n \mid x_{-n})\,u_n(x_{-n})}{u_n(x_{-n})} = f_n(x_n \mid x_{-n}) \quad \text{for all } x \in \mathcal{X}.$$

Given $i < n$, because of (3), there is a set $G \in \mathcal{B}(\mathcal{X})$ such that

$$\lambda(G^c) = 0 \quad \text{and} \quad f(x) = f_n(x_n \mid x_{-n})\, u_n(x_{-n}) = \frac{f_i(x_i \mid x_{-i})}{u_i(x_{-i})} \quad \text{for all } x \in G.$$

Let

$$F_i = \big\{x \in \mathcal{X} : (x_1, \ldots, x_{i-1}, z, x_{i+1}, \ldots, x_n) \in G \text{ for almost all } z \in \mathcal{X}_i\big\}.$$

Since $\lambda(G^c) = 0$, then $\lambda(F_i^c) = 0$. Furthermore,

$$f_i(x_i \mid x_{-i}) = u_i(x_{-i})\, f(x) = \frac{u_i(x_{-i})\, f(x)}{\int_{\mathcal{X}_i} f_i(z \mid x_{-i})\, \lambda_i(dz)}$$

$$= \frac{f(x)}{\int_{\mathcal{X}_i} f(x_1, \ldots, x_{i-1}, z, x_{i+1}, \ldots, x_n)\, \lambda_i(dz)} \quad \text{for all } x \in F_i \cap G.$$

Hence, $f_1, \ldots, f_n$ are improperly compatible. □

Proof of Theorem 5. First note that

$$\frac{f_i(x_i \mid x_{-i})}{f_j(x_j \mid x_{-j})} = \frac{g_i(x_i)\, h_i(x_{-i})}{g_j(x_j)\, h_j(x_{-j})}\, \frac{\prod_{k \neq i,j} q_{ik}(x_i, x_k)}{\prod_{k \neq i,j} q_{jk}(x_j, x_k)}\, \frac{q_{ij}(x_i, x_j)}{q_{ji}(x_j, x_i)} \tag{11}$$

whenever $i \neq j$ and $x \in \mathcal{X}$.

Suppose $f_1, \ldots, f_n$ are improperly compatible and fix $i \neq j$. To make the notation easier, we assume $n > 2$, $i = 1$ and $j = 2$, but exactly the same argument applies if $n = 2$ or if $(i, j) \neq (1, 2)$. By (11), one can write

$$\frac{f_1(x_1 \mid x_{-1})}{f_2(x_2 \mid x_{-2})} = \psi_1(x_{-1})\, \psi_2(x_{-2})\, \frac{q_{1,2}(x_1, x_2)}{q_{2,1}(x_2, x_1)} \quad \text{for all } x \in \mathcal{X},$$

where ψ_1 and ψ_2 are suitable functions on $\mathcal{X}_{-1}$ and $\mathcal{X}_{-2}$, respectively. By Theorem 4, condition (3) holds for some strictly positive measurable functions u_i on $\mathcal{X}_{-i}$, $i \in I$. Hence, there is a set $G \in \mathcal{B}(\mathcal{X})$ such that $\lambda(G^c) = 0$ and

$$\frac{f_1(x_1 \mid x_{-1})}{f_2(x_2 \mid x_{-2})} = \frac{u_1(x_{-1})}{u_2(x_{-2})} \quad \text{for all } x \in G.$$

Therefore,

$$\frac{q_{1,2}(x_1, x_2)}{q_{2,1}(x_2, x_1)} = \varphi_1(x_{-1})\, \varphi_2(x_{-2})$$

for all $x \in G$ and suitable functions φ_1 on $\mathcal{X}_{-1}$ and φ_2 on $\mathcal{X}_{-2}$. Since $\lambda(G^c) = 0$, there is a point $x^* \in \prod_{j=3}^n \mathcal{X}_j$ such that $(s, t, x^*) \in G$ for almost all $(s, t) \in \mathcal{X}_1 \times \mathcal{X}_2$. Fix one such x^* and define $v_1(s) = \varphi_2(s, x^*)$ for all $s \in \mathcal{X}_1$ and $v_2(t) = \varphi_1(t, x^*)$ for all $t \in \mathcal{X}_2$. If $(s, t) \in \mathcal{X}_1 \times \mathcal{X}_2$ and $(s, t, x^*) \in G$, then

$$\frac{q_{1,2}(s, t)}{q_{2,1}(t, s)} = \varphi_1(t, x^*)\, \varphi_2(s, x^*) = v_2(t)\, v_1(s).$$

Hence, $q_{1,2}(s, t) = v_1(s)\, v_2(t)\, q_{2,1}(t, s)$ for almost all $(s, t) \in \mathcal{X}_1 \times \mathcal{X}_2$. By irreducibility of $q_{1,2}/q_{2,1}$, there is a constant $c_{1,2} > 0$ such that $q_{1,2}(s, t) = c_{1,2}\, q_{2,1}(t, s)$ for almost all $(s, t) \in \mathcal{X}_1 \times \mathcal{X}_2$. Thus, condition (6) holds.

Conversely, suppose condition (6) holds. Fix $i < n$. By (6), there are constants $c_{ij} > 0$ and a set $G \in \mathcal{B}(\mathcal{X})$ such that $\lambda(G^c) = 0$ and

$$q_{ij}(x_i, x_j) = c_{ij}\, q_{ji}(x_j, x_i) \quad \text{whenever } j \neq i \text{ and } x \in G.$$

Define $c_i = c_{in} \sqrt{\prod_{j \neq i,n} c_{ij}}$ and

$$u_i(y) = c_i\, h_i(y) \left\{ \prod_{j \neq i} g_j(y_j) \prod_{j \neq i,n} q_{nj}(y_n, y_j) \prod_{j \neq i,n} \prod_{k \neq i,j,n} \sqrt{q_{jk}(y_j, y_k)} \right\}^{-1}$$

where $y \in \mathcal{X}_{-i}$. Also, define u_n as in the statement of the theorem, i.e.

$$u_n(y) = \frac{1}{h_n(y)} \prod_{j \neq n} g_j(y_j) \prod_{j \neq n} \prod_{k \neq j,n} \sqrt{q_{jk}(y_j, y_k)} \quad \text{for all } y \in \mathcal{X}_{-n}.$$

Then, for every $x \in \mathcal{X}$, a direct calculation shows that

$$u_i(x_{-i})\, u_n(x_{-n}) = c_i \frac{g_i(x_i)\, h_i(x_{-i})}{g_n(x_n)\, h_n(x_{-n})} \frac{\sqrt{\prod_{j \neq i,n} q_{ij}(x_i, x_j)\, q_{ji}(x_j, x_i)}}{\prod_{j \neq i,n} q_{nj}(x_n, x_j)}.$$

Further, for every $x \in G$, such an equation can be rewritten as

$$u_i(x_{-i})\, u_n(x_{-n}) = \frac{g_i(x_i)\, h_i(x_{-i})}{g_n(x_n)\, h_n(x_{-n})} \frac{\prod_{j \neq i,n} q_{ij}(x_i, x_j)}{\prod_{j \neq i,n} q_{nj}(x_n, x_j)} \frac{q_{in}(x_i, x_n)}{q_{ni}(x_n, x_i)} = \frac{f_i(x_i \mid x_{-i})}{f_n(x_n \mid x_{-n})}.$$

Thus, condition (3) holds and $f_1, \ldots, f_n$ are improperly compatible by Theorem 4. In addition, if $\int_{\mathcal{X}_{-n}} u_n\, d\lambda_{-n} < \infty$, then $f_1, \ldots, f_n$ are compatible by Corollary 3.

Since compatibility implies improper compatibility, it remains only to show that $\int_{\mathcal{X}_{-n}} u_n\, d\lambda_{-n} < \infty$ whenever $f_1, \ldots, f_n$ are compatible. Suppose that $f_1, \ldots, f_n$ are actually compatible. By Corollary 3, there are strictly positive measurable functions $u_1^*, \ldots, u_n^*$ such that

$$\int_{\mathcal{X}_{-n}} u_n^*\, d\lambda_{-n} < \infty \quad \text{and} \quad u_i^*(x_{-i})\, u_n^*(x_{-n}) = \frac{f_i(x_i \mid x_{-i})}{f_n(x_n \mid x_{-n})}$$

for all $i < n$ and almost all $x \in \mathcal{X}$. Thus,

$$u_i(x_{-i})\, u_n(x_{-n}) = u_i^*(x_{-i})\, u_n^*(x_{-n}) \quad \text{for all } i < n \text{ and almost all } x \in \mathcal{X}$$

where u_i is defined as above. Such an equation can be rewritten as

$$\phi_1(x_{-1}) = \ldots = \phi_n(x_{-n}) \quad \text{for almost all } x \in \mathcal{X},$$

where $\phi_n = u_n / u_n^*$ and $\phi_i = u_i^* / u_i$ for $i < n$. Applying Lemma 7, one obtains $u_n = c\, u_n^*$ a.e. for some constant $c > 0$. Hence,

$$\int_{\mathcal{X}_{-n}} u_n\, d\lambda_{-n} = c \int_{\mathcal{X}_{-n}} u_n^*\, d\lambda_{-n} < \infty.$$

$\square$

Proof of Corollary 6. First note that f_i satisfies condition (4) with $q_{ij}(s,t) = \exp\big\{\beta_{ij}\, b_i(s)\, b_j(t)\big\}$ where $(s,t) \in \mathcal{X}_i \times \mathcal{X}_j$. We have to prove that the ratios q_{ij}/q_{ji} are irreducible. Fix $i \neq j$ and suppose that

$$\exp\Big\{(\beta_{ij} - \beta_{ji})\, b_i(s)\, b_j(t)\Big\} = \frac{q_{ij}(s,t)}{q_{ji}(t,s)} = v_i(s)\, v_j(t) \quad \text{for almost all } (s,t) \in \mathcal{X}_i \times \mathcal{X}_j$$

and some strictly positive measurable functions v_i on $\mathcal{X}_i$ and v_j on $\mathcal{X}_j$. If $\beta_{ij} \neq \beta_{ji}$, one obtains

$$b_i(s)\, b_j(t) = \psi_i(s) + \psi_j(t) \quad \text{for almost all } (s,t) \in \mathcal{X}_i \times \mathcal{X}_j$$

where ψ_i and ψ_j are suitable functions on $\mathcal{X}_i$ and $\mathcal{X}_j$, respectively. By Lemma 7, it follows that b_i is a.e. constant or b_j is a.e. constant, contrary to the assumptions. Therefore, it must be $\beta_{ij} = \beta_{ji}$, and in that case q_{ij}/q_{ji} is trivially constant. Thus, q_{ij}/q_{ji} is irreducible. The corollary is now an obvious consequence of Theorem 5 and the fact that q_{ij}/q_{ji} is a.e. constant if and only if $\beta_{ij} = \beta_{ji}$.

□

Lemma 7. *Let $\phi_i : \mathcal{X}_{-i} \to \mathbb{R}$ and $\psi_i : \mathcal{X}_i \to \mathbb{R}$ be measurable functions, $i \in I$.*

- *If $\phi_1(x_{-1}) = \ldots = \phi_n(x_{-n})$ for almost all $x \in \mathcal{X}$, there is a constant c such that $\phi_n = c$ a.e.*
- *If $i \neq j$ and $\psi_i(s) + \psi_j(t) = b_i(s)\, b_j(t)$ for almost all $(s,t) \in \mathcal{X}_i \times \mathcal{X}_j$, then b_i is a.e. constant or b_j is a.e. constant.*

Proof. Suppose $\phi_1(x_{-1}) = \ldots = \phi_n(x_{-n})$ for all $x \in G$, where $G \in \mathcal{B}(\mathcal{X})$ and $\lambda(G^c) = 0$. For $n = 2$, this reduces to $\phi_1(t) = \phi_2(s)$ for all $(s,t) \in G$. Fix $t_0 \in \mathcal{X}_2$ and define $H = \big\{s \in \mathcal{X}_1 : (s, t_0) \in G\big\}$. Then, $\phi_2(s) = \phi_1(t_0)$ for all $s \in H$. Further, since $\lambda(G^c) = 0$, the point t_0 can be taken such that $\lambda_1(H^c) = 0$. Thus, the result is (trivially) true for $n = 2$. The general case follows by induction on n.

Next, suppose $\psi_i(s) + \psi_j(t) = b_i(s)\, b_j(t)$ for all $(s,t) \in B$, where $B \in \mathcal{B}(\mathcal{X}_i \times \mathcal{X}_j)$ and $\lambda_i \times \lambda_j(B^c) = 0$. Given $s_0 \in \mathcal{X}_i$ and $t_0 \in \mathcal{X}_j$, such an equation can be rewritten as

$$m_i(s) + m_j(t) = \big\{b_i(s) - b_i(s_0)\big\}\, \big\{b_j(t) - b_j(t_0)\big\} \quad \text{for all } (s,t) \in B$$

where m_i and m_j are suitable functions on $\mathcal{X}_i$ and $\mathcal{X}_j$. Since $\lambda_i \times \lambda_j(B^c) = 0$, the points s_0 and t_0 can be taken such that $(s, t_0) \in B$ for almost all $s \in \mathcal{X}_i$ and $(s_0, t) \in B$ for almost all $t \in \mathcal{X}_j$. Thus, for $s = s_0$, one obtains $m_j(t) = -m_i(s_0)$ for almost all $t \in \mathcal{X}_j$. Similarly, letting $t = t_0$ yields $m_i(s) = -m_j(t_0)$ for almost all $s \in \mathcal{X}_i$. Hence, the map

$$(s,t) \mapsto \big\{b_i(s) - b_i(s_0)\big\}\, \big\{b_j(t) - b_j(t_0)\big\}$$

is a.e. constant. In turn, this implies that b_i is a.e. constant or b_j is a.e. constant.

□

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, *Statist. Science*, 16, 249-274.

[3] Banerjee S., Carlin B.P., Gelfand A.E. (2015) *Hierarchical modeling and analysis for spatial data (Second Edition)*, Crc Press, Chapman & Hall, Boca Raton.

[4] Bernardo J.M., Smith A.F.M. (2009) *Bayesian theory*, Wiley, Chichester.

[5] Berti P., Dreassi E., Rigo P. (2013) A consistency theorem for regular conditional distributions, *Stochastics*, 85, 500-509.

[6] Berti P., Dreassi E., Rigo P. (2014) Compatibility results for conditional distributions, *J. Multivar. Analysis*, 125, 190-203.

[7] Besag J. (1974) Spatial interaction and the statistical analysis of lattice systems, *J. Royal Statist. Soc., Ser B*, 36, 192-236.

[8] Besag J., York J., Mollie A. (1991) Bayesian image restoration, with two applications in spatial statistics (with discussion), *Ann. Inst. Statist. Math.*, 43, 1-59.

[9] Besag J., Kooperberg C. (1995) On conditional and intrinsic autoregressions, *Biometrika*, 82, 733-746.

[10] Besag J., Green P., Higdon D., Mengersen K. (1995) Bayesian computation and stochastic systems, *Statist. Science*, 10, 3-66.

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

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

[13] Hartigan J.A. (1983) *Bayes theory*, Springer-Verlag, New York Inc.

[14] Hobert J.P., Casella G. (1998) Functional compatibility, Markov Chains, and Gibbs sampling with improper posteriors, *J. Comput. Graphic Statis.*, 7, 42-60.

[15] Ip E.H., Wang Y.J. (2009) Canonical representation of conditionally specified multivariate discrete distributions, *J. Multivar. Analysis*, 100, 1282-1290.

[16] Kaiser M.S., Cressie N. (2000) The construction of multivariate distributions from Markov random fields, *J. Multivar. Analysis*, 73, 199-220.

[17] Lee D. (2011) A comparison of conditional autoregressive models used in Bayesian disease mapping, *Spatial and Spatio-temporal Epidemiology*, 2, 79-89.

[18] Lee J., Kaiser M.S., Cressie N. (2001) Multiway dependence in exponential family conditional distributions, *J. Multivar. Analysis*, 79, 171-190.

[19] 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.

[20] Tansey W., Madrid Padilla O.H., Suggala A.S., Ravikumar P. (2015) Vector-space Markov random fields via exponential families, *Proceedings of the* 32^{nd} *International Conference on Machine Learning*, Lille.

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

[22] Wang Y.J., Ip E.H. (2008) Conditionally specified continuous distributions, *Biometrika*, 95, 735-746.

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

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

Pietro Rigo (corresponding author), Dipartimento di Matematica "F. Casorati", Universita' di Pavia, via Ferrata 1, 27100 Pavia, Italy

E-mail address: pietro.rigo@unipv.it