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 $X = (X_1, \dots, X_k)$, one selects some putative conditional distributions $Q_1, \dots, Q_k$. Each Q_i is regarded as a possible (or putative) conditional distribution for X_i given $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_k)$. The Q_i are compatible if there is a joint distribution P for X with conditionals $Q_1, \dots, Q_k$. 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 $Q_1, \dots, Q_k$ is requested to belong to some given class $\mathcal{P}_0$ of distributions. Two choices for $\mathcal{P}_0$ are considered, that is, $\mathcal{P}_0 = \{\text{exchangeable laws}\}$ and $\mathcal{P}_0 = \{\text{laws with identical univariate marginals}\}$.

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

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

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

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

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

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

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

* 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].