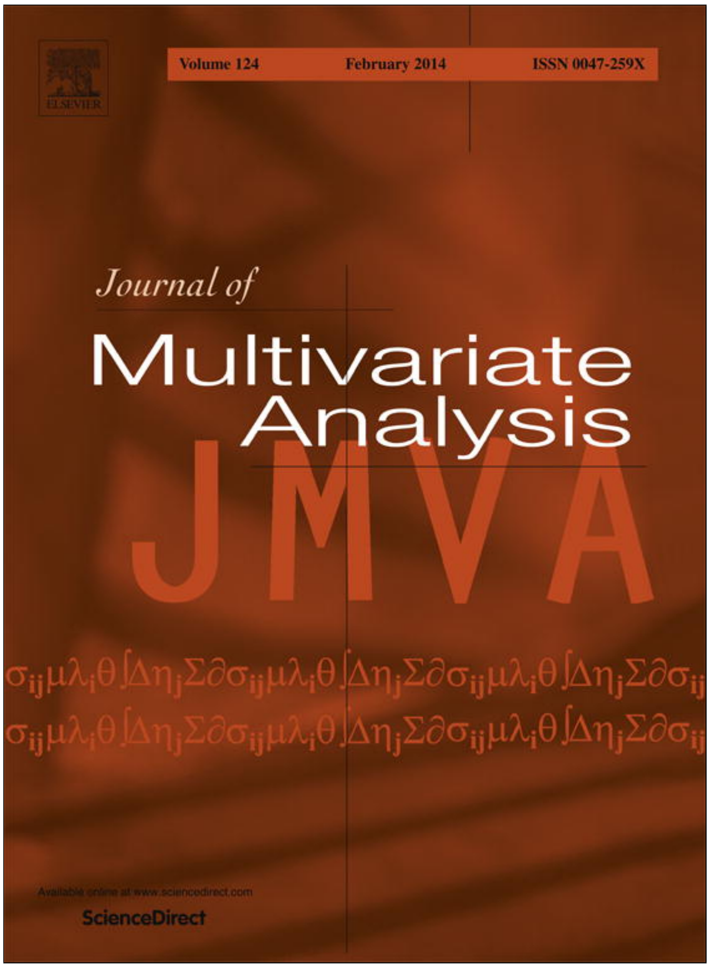
Volume 124
February 2014
ISSN 0047-259X
Journal of
Multivariate
Analysis
JMVA
ScienceDirect

Journal of Multivariate Analysis 124 (2014) 1–9

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

On compatibility of discrete full conditional distributions: A graphical representation approach

Yi-Ching Yao [a,b,*], Shih-chieh Chen [b], Shao-Hsuan Wang [c]

[a] *Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan, ROC*
[b] *Department of Statistics, National Chengchi University, Taipei 116, Taiwan, ROC*
[c] *Department of Mathematics, National Taiwan University, Taipei 106, Taiwan, ROC*

ARTICLE INFO

Article history:
Received 8 December 2011
Available online 26 October 2013

AMS subject classifications:
primary 62E10
secondary 62E15

Keywords:
Connected graph
Full conditional
Graph theory
Spanning tree

ABSTRACT

To deal with the compatibility issue of full conditional distributions of a (discrete) random vector, a graphical representation is introduced where a vertex corresponds to a configuration of the random vector and an edge connects two vertices if and only if the ratio of the probabilities of the two corresponding configurations is specified through one of the given full conditional distributions. Compatibility of the given full conditional distributions is equivalent to compatibility of the set of all specified probability ratios (called the ratio set) in the graphical representation. Characterizations of compatibility of the ratio set are presented. When the ratio set is compatible, the family of all probability distributions satisfying the specified probability ratios is shown to be the set of convex combinations of k probability distributions where k is the number of components of the underlying graph.

1. Introduction

The problem of characterizing a joint probability distribution by conditionals has been extensively studied in the last few decades. Consider a set of n real-valued random variables $\underline{X} = (X_1, \ldots, X_n)$ whose joint distribution is to be determined. Given some conditional distributions of the form $p_{S|T}(\underline{x}_S|\underline{x}_T)$ (the conditional probability mass/density function of $(X_i)_{i\in S}$ at $\underline{x}_S$ given $(X_i)_{i\in T} = \underline{x}_T$), where S ($\neq\emptyset$) and T are disjoint subsets of $\{1, \ldots, n\}$, it is desired (i) to determine whether these conditionals are compatible in the sense that they are the conditional distributions of some joint distribution, and (ii) to find all such (compatible) joint distributions when the given conditionals are compatible. See [1,2,6,10,11] for general results and comprehensive discussions. See also [4,13–18] for more recent developments. (It should be noted that most of the above papers consider the case where the given conditionals of the form $p_{S|T}$ are full in the sense that $S \cup T = \{1, \ldots, n\}$.) In case that $X_1, \ldots, X_n$ refer to certain observations at n locations in a region, the problem falls in the area of spatial statistics. In particular, when $X_1, \ldots, X_n$ form a Markov random field, the famous Hammersley–Clifford theorem characterizes the joint distribution via the Gibbs measure, cf. [7].

In the present paper, we restrict attention to the case where each X_i takes values in a finite set and the given conditionals are full. Note that specifying a full conditional $p_{S|T}$ amounts to specifying the probability ratio $p(\underline{x})/p(\underline{x}')$ for all $\underline{x} = (\underline{x}_S, \underline{x}_T)$ and $\underline{x}' = (\underline{x}'_S, \underline{x}'_T)$ with $\underline{x}_T = \underline{x}'_T$ where $\underline{x}_S$ denotes $\underline{x}$ restricted to the subset S of $\{1, \ldots, n\}$. With this simple observation, in the next section, we reformulate the problem more generally in terms of a graphical representation where each vertex corresponds to a configuration of $\underline{X}$ and an edge connects two vertices if and only if the ratio of the probabilities of the two corresponding configurations is specified. (It should be remarked that the graphical representation introduced here is different

* Corresponding author at: Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan, ROC.
E-mail address: yao@stat.sinica.edu.tw (Y.-C. Yao).

0047-259X/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmva.2013.10.007

from the graph for a Markov random field where a vertex (usually called a *site*) corresponds to a random variable in the Markov random field and an edge connects two sites (called *neighbors*) if the two corresponding random variables have *local interactions*; see e.g. [7] for the precise definitions of technical terms in Markov random fields.) In Section 3, results on compatibility of the specified probability ratios in a graphical representation are presented. Furthermore, all compatible joint distributions are characterized when the specified probability ratios are compatible, in which case there is a unique compatible joint distribution if and only if the underlying graph is connected. Section 4 presents a simple algorithm for checking compatibility of the specified probability ratios in a graphical representation. Section 5 contains concluding remarks.

2. Graphical representation

Consider a graph with vertex set V and edge set E where an edge connecting vertices $u, v \in V$ is denoted by $\{u, v\}$ (so that E is identified with a subset of the collection of all 2-element subsets of V). See e.g. [8] for an introduction to graph theory. For each edge $\{u, v\}$, there is a specified ratio $r(u, v) : r(v, u)$ where $r(u, v)$ and $r(v, u)$ are positive numbers. Let $R = R(E)$ denote the collection of all the specified ratios (to be called the ratio set), and we refer to (V, E, R) as a graphical representation. It is desired (i) to determine whether R is compatible in the sense that there is a (positive) probability distribution $(p(v))_{v\in V}$ on the vertex set V such that $p(u) : p(v) = r(u, v) : r(v, u)$ for all $\{u, v\} \in E$, and (ii) to find all (compatible) probability distributions satisfying R when R is compatible.

Let $\mathscr{C} = \{p_{S|T}\}$ be a given set of full conditionals for n discrete random variables $\underline{X} = (X_1, \ldots, X_n)$ taking values in $\mathscr{X} = \mathscr{X}_1 \times \cdots \times \mathscr{X}_n$. If $p_{S|T}$ and $p_{S'|T'} \in \mathscr{C}$ are conditionals of some joint distribution p and if $p_{S'|T'}(\underline{x}^0_{S'}|\underline{x}^0_{T'}) = 0$ for some $\underline{x}^0 \in \mathscr{X}$, then $p(\underline{x}^0)$ is necessarily zero, which in turn implies $p_{S|T}(\underline{x}^0_S|\underline{x}^0_T) = 0$. Consequently, a necessary condition for $\mathscr{C}$ to be compatible is that if $\underline{x}^0 \in \mathscr{X}$ is such that $p_{S'|T'}(\underline{x}^0_{S'}|\underline{x}^0_{T'}) = 0$ for some $p_{S'|T'} \in \mathscr{C}$, then $p_{S|T}(\underline{x}^0_S|\underline{x}^0_T) = 0$ for all $p_{S|T} \in \mathscr{C}$. We will refer to this necessary condition as condition (C1). Letting

$$O_{\mathscr{C}} := \{\underline{x} \in \mathscr{X} : p_{S|T}(\underline{x}_S|\underline{x}_T) = 0 \text{ for some } p_{S|T} \in \mathscr{C}\},$$

condition (C1) is equivalent to

$$O_{\mathscr{C}} = \{\underline{x} \in \mathscr{X} : p_{S|T}(\underline{x}_S|\underline{x}_T) = 0 \text{ for all } p_{S|T} \in \mathscr{C}\}.$$

Note that condition (C1) can also be stated as: "For each $p_{S|T} \in \mathscr{C}$, the set $O(p_{S|T}) := \{\underline{x} \in \mathscr{X} : p_{S|T}(\underline{x}_S|\underline{x}_T) = 0\}$ is the same for all $p_{S|T} \in \mathscr{C}$". Theorem 1.1(i) of [1] gives this necessary condition for $n = 2$ and $\mathscr{C} = \{p_{\{1\}|\{2\}}, p_{\{2\}|\{1\}}\}$. In what follows, we will always assume that $\mathscr{C}$ satisfies condition (C1). Define $V_{\mathscr{C}} = \mathscr{X} \setminus O_{\mathscr{C}}$. Note that if all $p_{S|T} \in \mathscr{C}$ are conditionals of some joint distribution p (implying that $\mathscr{C}$ is compatible), then p has $V_{\mathscr{C}}$ as its support.

We now define the graphical representation for $\mathscr{C}$. Let $V_{\mathscr{C}}$ be the vertex set. An edge $\{\underline{x}, \underline{x}'\}$ connects two vertices $\underline{x}, \underline{x}' \in V_{\mathscr{C}}$ if and only if $\underline{x}_T = \underline{x}'_T$ for some $p_{S|T} \in \mathscr{C}$; the ratio associated with this edge is given by

$$r(\underline{x}, \underline{x}') : r(\underline{x}', \underline{x}) = p_{S|T}(\underline{x}_S|\underline{x}_T) : p_{S|T}(\underline{x}'_S|\underline{x}'_T).$$

The resulting graphical representation is denoted by $(V_{\mathscr{C}}, E_{\mathscr{C}}, R_{\mathscr{C}})$. It should be noted that an edge may connect two vertices $\underline{x}$ and $\underline{x}' \in V_{\mathscr{C}}$ through two or more conditionals in $\mathscr{C}$, resulting in possibly different ratios associated with this edge. For example, consider $n = 3$ and $\mathscr{C} = \{p_{\{1,2\}|\{3\}}, p_{\{1,3\}|\{2\}}\}$. Then an edge connects $\underline{x} = (x_1, x_2, x_3)$ and $\underline{x}' = (x'_1, x_2, x_3)$ (with $x_1 \neq x'_1$) through either of the two conditionals in $\mathscr{C}$, resulting in two ratios given by

$$p_{\{1,2\}|\{3\}}((x_1, x_2)|x_3) : p_{\{1,2\}|\{3\}}((x'_1, x_2)|x_3) \quad \text{and} \quad p_{\{1,3\}|\{2\}}((x_1, x_3)|x_2) : p_{\{1,3\}|\{2\}}((x'_1, x_3)|x_2).$$

(Such cases cannot happen if S is a singleton for every $p_{S|T} \in \mathscr{C}$.) Clearly in order for $\mathscr{C}$ to be compatible, the ratios associated with an edge are necessarily equal if the edge is formed through two or more conditionals in $\mathscr{C}$. Below we will also assume that $\mathscr{C}$ satisfies this (second) necessary condition (to be referred to as condition (C2)), so that exactly one ratio is associated with each edge. Note that condition (C2) is trivially satisfied if S is a singleton for every $p_{S|T} \in \mathscr{C}$. The following theorem shows that compatibility of $\mathscr{C}$ is equivalent to compatibility of $R_{\mathscr{C}}$.

Theorem 1. *Assume that $\mathscr{C}$ satisfies conditions* (C1) *and* (C2). *Then a joint distribution with support $V_{\mathscr{C}}$ satisfies $\mathscr{C}$ if and only if it satisfies $R_{\mathscr{C}}$. Consequently, $\mathscr{C}$ is compatible if and only if $R_{\mathscr{C}}$ is compatible.*

Proof. To prove the "only if" part, suppose that a joint distribution p (with support $V_{\mathscr{C}}$) satisfies $\mathscr{C}$, i.e. under p the conditional distribution of $\underline{X}_S$ given $\underline{X}_T$ agrees with $p_{S|T}$ for every $p_{S|T} \in \mathscr{C}$. Consider an (arbitrary) edge $\{\underline{x}, \underline{x}'\} \in E_{\mathscr{C}}$ with an associated ratio given by $r(\underline{x}, \underline{x}') : r(\underline{x}', \underline{x}) = p_{S|T}(\underline{x}_S|\underline{x}_T) : p_{S|T}(\underline{x}'_S|\underline{x}'_T)$ (for some $p_{S|T} \in \mathscr{C}$). By definition, we have $\underline{x}_T = \underline{x}'_T$. Then p satisfies the associated ratio specification since

$$\frac{p(\underline{x})}{p(\underline{x}')} = \frac{p_{S|T}(\underline{x}_S|\underline{x}_T)}{p_{S|T}(\underline{x}'_S|\underline{x}'_T)} = \frac{r(\underline{x}, \underline{x}')}{r(\underline{x}', \underline{x})},$$

from which it follows that p satisfies $R_{\mathscr{C}}$.

To prove the "if" part of the theorem, suppose that a joint distribution p (with support $V_{\mathscr{C}}$) satisfies $R_{\mathscr{C}}$. Consider a conditional $p_{S|T} \in \mathscr{C}$. Fix an (arbitrary) $\underline{x}^0 \in V_{\mathscr{C}}$. For every $\underline{x} \in V_{\mathscr{C}}$ with $\underline{x}_T = \underline{x}^0_T$, we have (by definition) $\{\underline{x}^0, \underline{x}\} \in E_{\mathscr{C}}$ and $r(\underline{x}^0, \underline{x}) : r(\underline{x}, \underline{x}^0) = p_{S|T}(\underline{x}^0_S|\underline{x}^0_T) : p_{S|T}(\underline{x}_S|\underline{x}_T)$. It follows that

$$\frac{p(\underline{x})}{p(\underline{x}^0)} = \frac{r(\underline{x}, \underline{x}^0)}{r(\underline{x}^0, \underline{x})} = \frac{p_{S|T}(\underline{x}_S|\underline{x}_T)}{p_{S|T}(\underline{x}^0_S|\underline{x}^0_T)} = \frac{p_{S|T}(\underline{x}_S|\underline{x}^0_T)}{p_{S|T}(\underline{x}^0_S|\underline{x}^0_T)}.$$

Since this holds for all $\underline{x} \in V_{\mathscr{C}}$ with $\underline{x}_T = \underline{x}_T^0$, under p the conditional distribution of $\underline{X}_S$ given $\underline{X}_T = \underline{x}_T^0$ agrees with $p_{S|T}$. As $\underline{x}^0$ (and $\underline{x}_T^0$) is arbitrary, $p_{S|T}$ is indeed the conditional distribution of $\underline{X}_S$ given $\underline{X}_T$ under p. This shows that p satisfies $\mathscr{C}$. The proof of the theorem is complete. □

To illustrate, consider, in the following examples, two random variables X_1 and X_2 both taking values in $\{1, 2, 3\}$, and let $\mathscr{C} = \{p_{\{1\}|\{2\}}, p_{\{2\}|\{1\}}\}$. Let $A = (a_{ij})$, $B = (b_{ij})$ where $a_{ij} := p_{\{1\}|\{2\}}(i|j) = p(X_1 = i|X_2 = j)$, $b_{ij} := p_{\{2\}|\{1\}}(j|i) = p(X_2 = j|X_1 = i)$, $i, j = 1, 2, 3$.

Example 1 (*Cf. Example 2.3 on page 24 of [1]*)**.** For

$$A = \begin{pmatrix} 1/5 & 2/7 & 3/8 \\ 3/5 & 2/7 & 1/8 \\ 1/5 & 3/7 & 1/2 \end{pmatrix}, \qquad B = \begin{pmatrix} 1/6 & 1/3 & 1/2 \\ 1/2 & 1/3 & 1/6 \\ 1/8 & 3/8 & 1/2 \end{pmatrix},$$

a graphical representation is given in Fig. 1, where a vertex labeled (i, j) corresponds to a configuration (i, j) of (X_1, X_2), and ratios attached to vertical edges are derived from A while ratios attached to horizontal edges are derived from B. Note that the six edges $\{(i, 1), (i, 3)\}$, $\{(1, j), (3, j)\}$, $i, j = 1, 2, 3$ and the associated ratios are not shown. These six ratios can be derived from those shown in the figure. For example, the ratio $r((1, 1), (1, 3)) : r((1, 3), (1, 1)) = b_{11} : b_{13} = 1 : 3$ can be derived from $r((1, 1), (1, 2)) : r((1, 2), (1, 1)) = 1 : 2$ and $r((1, 2), (1, 3)) : r((1, 3), (1, 2)) = 2 : 3$.

Definition 1. Two graphical representations (V, E, R) and (V, E', R') with the same vertex set V are said to be equivalent if (i) R and R' agree on $E \cap E'$, (ii) for $\{u, v\} \in E \setminus E'$, there exists an E'-path $v_0 v_1 \cdots v_k$ with $v_0 = u$, $v_k = v$ and $\{v_\ell, v_{\ell+1}\} \in E'$, $\ell = 0, 1, \ldots, k - 1$, such that

$$\frac{r(v, u)}{r(u, v)} = \prod_{\ell=0}^{k-1} \frac{r'(v_{\ell+1}, v_\ell)}{r'(v_\ell, v_{\ell+1})},$$

and (iii) for $\{u, v\} \in E' \setminus E$, a condition similar to (ii) is satisfied with the roles of (E, R) and (E', R') interchanged.

In words, two graphical representations (V, E, R) and (V, E', R') are equivalent if the two ratio sets R and R' agree on all common edges and a ratio in only one of R and R' can be derived from ratios in the other set. Strictly speaking, Fig. 1 is a simplified, but equivalent version of the graphical representation for the given matrices A and B in Example 1. Lemma 1 states a simple result on equivalent graphical representations, whose proof is straightforward and omitted.

Lemma 1. *Suppose (V, E, R) and (V, E', R') are equivalent. Then a positive probability distribution p on V satisfies R if and only if p satisfies R'. Consequently, R is compatible if and only if R' is compatible.*

Example 2 (*Cf. Example 2.2 on page 23 of [1]*)**.** For

$$A = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}, \qquad B = \begin{pmatrix} 1/3 & 2/3 & 0 \\ 0 & 1/3 & 2/3 \\ 1/3 & 0 & 2/3 \end{pmatrix},$$

a graphical representation is given in Fig. 2, where there are only six vertices corresponding to the six (possible) configurations of (X_1, X_2). Note that the underlying graph has exactly one cycle.

Example 3. For

$$A = \begin{pmatrix} 1/5 & 2/7 & 3/8 \\ 3/5 & 2/7 & 1/8 \\ 1/5 & 3/7 & 1/2 \end{pmatrix}, \qquad B = \begin{pmatrix} 1/6 & 1/3 & 1/2 \\ ? & ? & ? \\ ? & ? & ? \end{pmatrix},$$

a (simplified but equivalent) graphical representation is given in Fig. 3, where the four edges $\{(1, j), (3, j)\}$, $j = 1, 2, 3$, and $\{(1, 1), (1, 3)\}$, and the associated ratios are not shown. This example admits a graphical representation even though the conditional probabilities $p_{\{2\}|\{1\}}(j|i)$, $j = 1, 2, 3$, $i = 2, 3$, are unavailable. Note that the underlying graph is a tree (a connected graph with no cycles).

3. Compatibility of a ratio set *R* and characterization of probability distributions satisfying *R*

A graph (V, E) is connected if every pair of vertices are connected by a path. If (V, E) is not connected, it can be decomposed into some $k > 1$ components (disjoint connected subgraphs), written $(V, E) = \sqcup_{i=1}^k (V_i, E_i)$, where each (V_i, E_i) is a connected subgraph and where the symbol $\sqcup$ denotes disjoint union (implying that $V_i \cap V_j = \emptyset$ and $E_i \cap E_j = \emptyset$ for $i \neq j$).

Theorem 2. *Consider a graphical representation $(V, E, R) = \sqcup_{i=1}^k (V_i, E_i, R_i)$ where each (V_i, E_i) is a connected subgraph of (V, E) and R_i is R restricted to E_i. Then R is compatible if and only if R_i is compatible, $i = 1, \ldots, k$.*

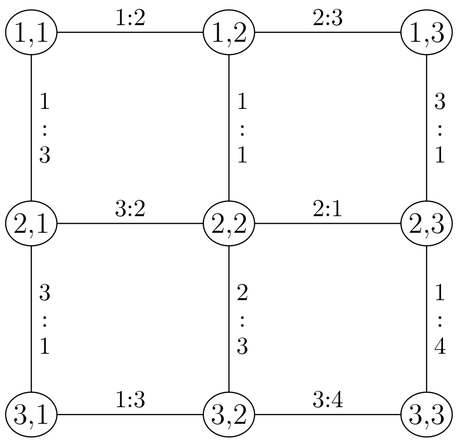

Fig. 1. Graphical representation for Example 1.

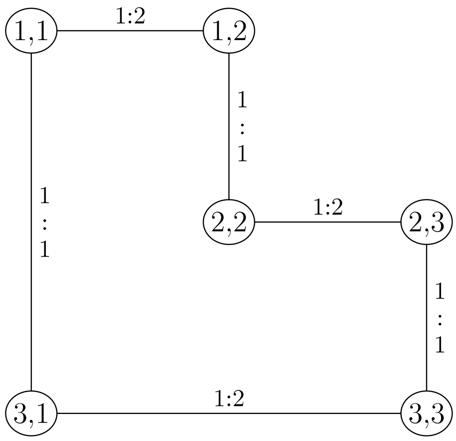

Fig. 2. Graphical representation for Example 2.

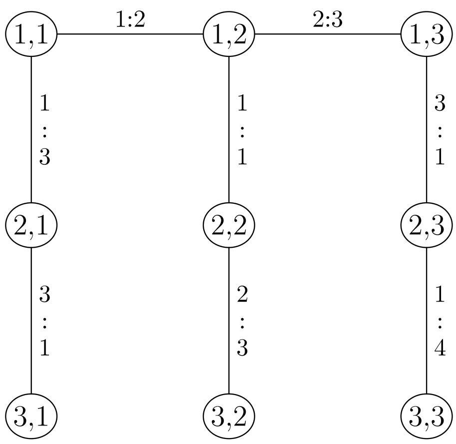

Fig. 3. Graphical representation for Example 3.

Proof. To show the "only if" part, suppose R is compatible. Let p be a (positive) probability distribution on V satisfying R. Let p_i be a probability distribution on V_i defined by $p_i(v) := p(v)/\sum_{v\in V_i} p(v)$, $v \in V_i$. It is easily verified that p_i satisfies R_i, implying that R_i is compatible. To show the "if" part, suppose R_i is compatible for $i = 1, \ldots, k$. Let p_i be a (positive) probability distribution on V_i satisfying R_i, $i = 1, \ldots, k$. For any given positive numbers $\lambda_1, \ldots, \lambda_k$ with $\sum_{i=1}^k \lambda_i = 1$, define $p_{\lambda_1,\ldots,\lambda_k}(v) := \sum_{i=1}^k \lambda_i p_i(v)\mathbf{1}_{V_i}(v)$, $v \in V$. It is easily verified that $p_{\lambda_1,\ldots,\lambda_k}$ is a probability distribution on V satisfying R, implying that R is compatible. $\square$

By Theorem 2, to check compatibility of a ratio set R in a graphical representation $(V, E, R) = \sqcup_{i=1}^k (V_i, E_i, R_i)$, it suffices to check compatibility of each R_i (on a connected subgraph (V_i, E_i)).

Theorem 3. *For a graphical representation* (V, E, R)*, the following statements are equivalent.*

(i) *R is compatible.*
(ii) *For any two paths* $v_0 v_1 \cdots v_\ell$ *and* $w_0 w_1 \cdots w_m$ *with* $v_0 = w_0$, $v_\ell = w_m$,

$$\prod_{i=0}^{\ell-1} \frac{r(v_{i+1}, v_i)}{r(v_i, v_{i+1})} = \prod_{i=0}^{m-1} \frac{r(w_{i+1}, w_i)}{r(w_i, w_{i+1})}. \tag{1}$$

(iii) *For every cycle* $v_0 v_1 \cdots v_\ell v_0$ *(i.e. a path whose initial and terminal vertices are the same), we have*

$$\prod_{i=0}^{\ell} \frac{r(v_{i+1}, v_i)}{r(v_i, v_{i+1})} = 1, \tag{2}$$

where $v_{\ell+1} := v_0$.

Proof. The equivalence of (ii) and (iii) is obvious. To show that (i) implies (ii), suppose R is compatible and let p be a (positive) probability distribution on V satisfying R. For two paths $v_0 v_1 \cdots v_\ell$ and $w_0 w_1 \cdots w_m$ with $v_0 = w_0$, $v_\ell = w_m$, we have

$$\frac{p(v_\ell)}{p(v_0)} = \prod_{i=0}^{\ell-1} \frac{p(v_{i+1})}{p(v_i)} = \prod_{i=0}^{\ell-1} \frac{r(v_{i+1}, v_i)}{r(v_i, v_{i+1})},$$

$$\frac{p(v_\ell)}{p(v_0)} = \frac{p(w_m)}{p(w_0)} = \prod_{i=0}^{m-1} \frac{p(w_{i+1})}{p(w_i)} = \prod_{i=0}^{m-1} \frac{r(w_{i+1}, w_i)}{r(w_i, w_{i+1})},$$

showing that (1) holds.

Suppose (ii) holds. To show R is compatible, by Theorem 2 it suffices to consider the case that (V, E) is connected. Fix a vertex $v_0 \in V$. For every $v \in V \setminus \{v_0\}$, define

$$q(v) := \prod_{i=0}^{\ell-1} \frac{r(v_{i+1}, v_i)}{r(v_i, v_{i+1})},$$

where $v_0 v_1 \cdots v_\ell$ is a path connecting v_0 and $v = v_\ell$. By condition (ii), the definition of $q(v)$ does not depend on the chosen path. Letting $q(v_0) := 1$, define

$$p(v) := q(v) / \sum_{v \in V} q(v), \quad v \in V,$$

which is a positive probability distribution on V and satisfies R. So R is compatible. The proof is complete. □

Remark 1. As a simple application of Theorem 3, consider Example 2 for which the graphical representation in Fig. 2 has only one cycle. The left hand side of (2) for this cycle (clockwise) equals $2 \neq 1$, implying incompatibility.

Remark 2. If (V, E) is a tree (i.e. a connected graph with no cycles), then any ratio set R is compatible since condition (iii) in Theorem 3 is trivially satisfied. Indeed, there is a unique probability distribution on V satisfying R. Example 3 is such a case, so it is compatible. In a graphical representation (V, E, R) where (V, E) is connected, for every spanning tree of the graph (V, E) (i.e. a tree containing every vertex of the graph), there is a unique probability distribution which satisfies R restricted to the spanning tree. Thus, R is compatible if and only if all spanning trees give rise to the same probability distribution. Alternatively, to check compatibility of R, it may be easier to first choose a convenient spanning tree and find the unique probability distribution p which satisfies R restricted to the spanning tree, and then check if this p satisfies R. As an illustration, consider Example 1 and note that Example 3 is derived from Example 1 with the second and third rows of matrix B removed. As a result, the graphical representation for Example 3 as shown in Fig. 3 is a spanning tree of the graphical representation for Example 1 as shown in Fig. 1. The unique compatible joint distribution for Example 3 is easily found to be

$$p = \begin{pmatrix} 1/20 & 1/10 & 3/20 \\ 3/20 & 1/10 & 1/20 \\ 1/20 & 3/20 & 1/5 \end{pmatrix}.$$

For this p, it is readily shown that $p_{\{2\}|\{1\}}$ agrees with B in Example 1, implying compatibility. The next section gives a further and detailed discussion.

Remark 3. In Theorem 2, a graphical representation (V, E, R) is written as a disjoint union of (V_i, E_i, R_i), $i = 1, \ldots, k$ where each (V_i, E_i) is a connected subgraph of (V, E). To show that R is compatible, it suffices to show that each R_i is compatible. Suppose now R is compatible. We want to characterize all positive probability distributions on V satisfying R. Since each R_i is compatible, there is a unique positive probability distribution p_i on V_i satisfying R_i (cf. Remark 2). For any positive numbers $\lambda_1, \ldots, \lambda_k$ with $\sum_{i=1}^k \lambda_i = 1$, define

$$p_{\lambda_1,\ldots,\lambda_k}(v) := \sum_{i=1}^k \lambda_i p_i(v)\mathbf{1}_{V_i}(v), \quad v \in V, \tag{3}$$

which is a positive probability distribution on V satisfying R (cf. the proof of Theorem 2). On the other hand, let p be a positive probability distribution on V satisfying R. Define $p_i^*(v) := p(v)/\sum_{v\in V_i} p(v)$, $v \in V_i$, which is a positive probability distribution on V_i satisfying R_i. By uniqueness, we have $p_i^* = p_i$. Letting $\lambda_i = \sum_{v\in V_i} p(v)$, it follows that

$$p(v) = \sum_{i=1}^k \lambda_i p_i^*(v)\mathbf{1}_{V_i}(v) = \sum_{i=1}^k \lambda_i p_i(v)\mathbf{1}_{V_i}(v) = p_{\lambda_1,\ldots,\lambda_k}(v).$$

Thus the set of all positive probability distributions on V satisfying R is $\{p_{\lambda_1,\ldots,\lambda_k} : \lambda_i > 0, \sum_{i=1}^k \lambda_i = 1\}$, the set of all convex combinations of $p_1, \ldots, p_k$ with positive coefficients. We summarize this result in Theorem 4.

Theorem 4. *Consider a graphical representation $(V, E, R) = \sqcup_{i=1}^k (V_i, E_i, R_i)$ where each (V_i, E_i) is a connected subgraph of (V, E). Suppose R is compatible. Let p_i be the unique probability distribution on V_i satisfying R_i, $i = 1, \ldots, k$. Then the set of all positive probability distributions on V satisfying R is $\{p_{\lambda_1,\ldots,\lambda_k} : \lambda_i > 0, \sum_{i=1}^k \lambda_i = 1\}$ where $p_{\lambda_1,\ldots,\lambda_k}$ is given in (3).*

4. An algorithm

For a graphical representation $(V, E, R) = \sqcup_{i=1}^k (V_i, E_i, R_i)$ where each (V_i, E_i) is a connected subgraph (component) of (V, E) and R_i is R restricted to E_i, we have by Theorem 2 that R is compatible if and only if each R_i is compatible. While condition (ii) (as well as condition (iii)) in Theorem 3 (applied to (V_i, E_i, R_i)) is necessary and sufficient for R_i to be compatible, it is usually cumbersome to verify either (ii) or (iii). A relatively simple way to check compatibility of R consists of the following steps: (1) identify each of the components (V_i, E_i) of (V, E), $i = 1, \ldots, k$; (2) for each i, choose a (convenient) spanning tree of (V_i, E_i); (3) determine the unique probability distribution p_i on V_i which satisfies the ratio specifications associated with the edges of the spanning tree; and (4) check if the probability distribution p_i satisfies R_i. If p_i satisfies R_i, then R_i is compatible and p_i is the unique probability distribution satisfying R_i. If p_i does not satisfy R_i, then R_i (and hence R) is incompatible.

More precisely, first choose a convenient way to number and order all vertices of V, denoted by $v_1 < v_2 < \cdots$. This linear order on V then induces a linear order on E as follows. Let $\min\{v_\ell, v_m\} := v_{\min\{\ell,m\}}$ and $\max\{v_\ell, v_m\} := v_{\max\{\ell,m\}}$. For edges in E, define $\{v_\ell, v_m\} < \{v_{\ell'}, v_{m'}\}$ if and only if $\min\{v_\ell, v_m\} < \min\{v_{\ell'}, v_{m'}\}$ or $\min\{v_\ell, v_m\} = \min\{v_{\ell'}, v_{m'}\}$ and $\max\{v_\ell, v_m\} < \max\{v_{\ell'}, v_{m'}\}$. This defines a linear order on E. The following steps consist of identifying the component of (V, E) that contains the lowest vertex v_1 and also selecting a spanning tree of this component.

Step 1 (initialization): Set $\mathscr{V} \leftarrow \{v_1\}$ and $\mathscr{E} \leftarrow \emptyset$.

Step 2: Find the lowest edge $\{v_\ell, v_m\} \in E \setminus \mathscr{E}$ such that exactly one of v_ℓ and v_m (say v_ℓ) is in $\mathscr{V}$. If no such edge exists, go to Step 4.

Step 3: Set $\mathscr{V} \leftarrow \mathscr{V} \cup \{v_m\}$ and $\mathscr{E} \leftarrow \mathscr{E} \cup \{\{v_\ell, v_m\}\}$. Go to Step 2.

Step 4: Set $V_1 = \mathscr{V}$ and $E_1 = E \cap (V_1 \times V_1)$, which is the component of (V, E) that contains v_1. Note that $\mathscr{E}$ is a spanning tree of the component (V_1, E_1).

Step 5: Determine the unique distribution p_1 satisfying the ratio specifications associated with the edges in $\mathscr{E}$.

Step 6: Check if p_1 satisfies R_1 (which is R restricted to E_1). If so, R_1 is compatible. If not, then R_1 (and hence R) is not compatible.

If R_1 is compatible and V is not exhausted (i.e. $V \neq V_1$), we move on to finding the second component that contains the lowest vertex in $V \setminus V_1$, say v_h, by resetting $\mathscr{V} \leftarrow \{v_h\}$ and $\mathscr{E} \leftarrow \emptyset$ and then repeating Steps 2–6. Repeat this procedure until either an incompatible component is found (corresponding to the case of incompatible R) or V is exhausted and all of the (compatible) components are found (corresponding to the case of compatible R). In the latter case (with R compatible), if (V, E) has k components, the algorithm will find each of the components (V_i, E_i, R_i), $i = 1, \ldots, k$ along with the unique distribution p_i on V_i that satisfies R_i. By Theorem 4, the set of all positive probability distributions on V satisfying R is $\{p_{\lambda_1,\ldots,\lambda_k} : \lambda_i > 0, \sum_{i=1}^k \lambda_i = 1\}$ where $p_{\lambda_1,\ldots,\lambda_k}$ is given in (3).

Let $\mathscr{C}$ be a given set of conditionals for $\underline{X} = (X_1, \ldots, X_n)$ taking values in $\mathscr{X} = \mathscr{X}_1 \times \cdots \times \mathscr{X}_n$. To check compatibility of $\mathscr{C}$, we need to first check if $\mathscr{C}$ satisfies the two necessary conditions (C1) and (C2) stated in Section 2. If either one of the conditions is not satisfied, then $\mathscr{C}$ is incompatible. Provided that $\mathscr{C}$ satisfies both conditions, we then work with a (possibly simplified but equivalent) graphical representation for $\mathscr{C}$, denoted $(V_{\mathscr{C}}, E_{\mathscr{C}}, R_{\mathscr{C}})$. Without loss of generality, assume $\mathscr{X}_i = \{1, 2, \ldots, L_i\}$ ($L_i = |\mathscr{X}_i|$, the cardinality of $\mathscr{X}_i$). To apply the above algorithm to $(V_{\mathscr{C}}, E_{\mathscr{C}}, R_{\mathscr{C}})$, we need to specify a linear order on $V_{\mathscr{C}}$ (which then induces a linear order on $E_{\mathscr{C}}$). A convenient linear order on $V_{\mathscr{C}}$ is lexicographical, i.e. $(x_1, \ldots, x_n) < (x_1', \ldots, x_n')$ if and only if there is an $m \in \{1, \ldots, n\}$ such that $x_m < x_m'$ and $x_\ell = x_\ell'$ for all $\ell < m$. With this linear order on $V_{\mathscr{C}}$, the resulting algorithm determines whether $R_{\mathscr{C}}$ (and hence $\mathscr{C}$) is compatible and also finds, in case that $\mathscr{C}$ is compatible, the set of all probability distributions that satisfy $\mathscr{C}$. As an illustration, we apply the algorithm to the following example.

Example 4. Consider $\underline{X} = (X_1, X_2)$ taking values in $\mathscr{X} = \{1, 2, 3, 4, 5, 6, 7\} \times \{1, 2, 3, 4, 5\}$. Let $\mathscr{C} = \{p_{\{1\}|\{2\}}, p_{\{2\}|\{1\}}\}$ where $A = p_{\{1\}|\{2\}}$ and $B = p_{\{2\}|\{1\}}$ are given by

$$A = \begin{pmatrix} 1/10 & 0 & 3/14 & 0 & 2/5 \\ 0 & 1/6 & 0 & 1/3 & 0 \\ 2/5 & 0 & 5/14 & 0 & 1/5 \\ 0 & 1/3 & 0 & 1/2 & 0 \\ 1/5 & 0 & 2/7 & 0 & 3/10 \\ 0 & 1/2 & 0 & 1/6 & 0 \\ 3/10 & 0 & 1/7 & 0 & 1/10 \end{pmatrix}, \qquad B = \begin{pmatrix} 1/8 & 0 & 3/8 & 0 & 1/2 \\ 0 & 1/3 & 0 & 2/3 & 0 \\ 4/11 & 0 & 5/11 & 0 & 2/11 \\ 0 & 2/5 & 0 & 3/5 & 0 \\ 2/9 & 0 & 4/9 & 0 & 1/3 \\ 0 & 3/4 & 0 & 1/4 & 0 \\ 1/2 & 0 & 1/3 & 0 & 1/6 \end{pmatrix}.$$

Since A and B share a common incidence set, i.e. $\{(i, j) : a_{ij} > 0\} = \{(i, j) : b_{ij} > 0\}$, condition (C1) is satisfied. Also as $|S| = 1$ for each $P_{S|T} \in \mathscr{C}$, condition (C2) is trivially satisfied. A simplified but equivalent graphical representation $(V_{\mathscr{C}}, E_{\mathscr{C}}, R_{\mathscr{C}})$ is shown in Fig. 4 where some of the edges and associated ratios are removed so that $E_{\mathscr{C}}$ consists only of 24 edges. The vertex set $V_{\mathscr{C}}$ is the same as the common incidence set of A and B, which consists of 18 configurations. We label these 18 configurations by lexicographical order as follows:

$$\mathbf{1}(1, 1), \mathbf{2}(1, 3), \mathbf{3}(1, 5), \mathbf{4}(2, 2), \mathbf{5}(2, 4), \mathbf{6}(3, 1), \mathbf{7}(3, 3), \mathbf{8}(3, 5), \mathbf{9}(4, 2), \mathbf{10}(4, 4),$$
$$\mathbf{11}(5, 1), \mathbf{12}(5, 3), \mathbf{13}(5, 5), \mathbf{14}(6, 2), \mathbf{15}(6, 4), \mathbf{16}(7, 1), \mathbf{17}(7, 3), \mathbf{18}(7, 5).$$

The linear order on $E_{\mathscr{C}}$ induced by the lexicographical order on $V_{\mathscr{C}}$ gives $\{(1, 1), (1, 3)\}$ as the lowest edge followed by $\{(1, 1), (3, 1)\}, \{(1, 3), (1, 5)\}, \ldots$. A ratio in the ratio set $R_{\mathscr{C}}$ is determined by either A or B. As an example, the ratio associated with the edge $\{(1, 1), (1, 3)\}$ is

$$\begin{aligned} r((1, 1), (1, 3)) : r((1, 3), (1, 1)) &= p_{\{2\}|\{1\}}(1|1) : p_{\{2\}|\{1\}}(3|1) \\ &= b_{11} : b_{13} = 1 : 3. \end{aligned}$$

We now apply the algorithm to this graphical representation. For ease of notation, we write **1** for vertex $(1, 1)$ and $\{\mathbf{1}, \mathbf{2}\}$ for edge $\{(1, 1), (1, 3)\}$, and similarly for other vertices and edges. To identify the component containing vertex **1** and also select a spanning tree of this component, the algorithm found 11 edges in the order

$$\{\mathbf{1}, \mathbf{2}\}, \{\mathbf{1}, \mathbf{6}\}, \{\mathbf{2}, \mathbf{3}\}, \{\mathbf{2}, \mathbf{7}\}, \{\mathbf{3}, \mathbf{8}\}, \{\mathbf{6}, \mathbf{11}\}, \{\mathbf{7}, \mathbf{12}\}, \{\mathbf{8}, \mathbf{13}\}, \{\mathbf{11}, \mathbf{16}\}, \{\mathbf{12}, \mathbf{17}\}, \{\mathbf{13}, \mathbf{18}\},$$

which form a spanning tree of the component (shown on the left panel of Fig. 4, denoted (V_1, E_1, R_1)) that contains vertex **1**. Here V_1 consists of the 12 vertices that appear in the 11 edges, i.e. $V_1 = \{\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{6}, \mathbf{7}, \mathbf{8}, \mathbf{11}, \mathbf{12}, \mathbf{13}, \mathbf{16}, \mathbf{17}, \mathbf{18}\}$. Also $E_1 = E_{\mathscr{C}} \cap (V_1 \times V_1)$ and R_1 is $R_{\mathscr{C}}$ restricted to E_1. The unique distribution p_1 with support V_1 that satisfies the ratio specifications associated with the 11 edges of the spanning tree is given by

$$p_1 = \begin{pmatrix} 1/34 & 0 & 3/34 & 0 & 2/17 \\ 0 & 0 & 0 & 0 & 0 \\ 2/17 & 0 & 5/34 & 0 & 1/17 \\ 0 & 0 & 0 & 0 & 0 \\ 1/17 & 0 & 2/17 & 0 & 3/34 \\ 0 & 0 & 0 & 0 & 0 \\ 3/34 & 0 & 1/17 & 0 & 1/34 \end{pmatrix}.$$

It can be verified that p_1 satisfies R_1, so that R_1 is compatible. The lowest vertex in $V_{\mathscr{C}} \setminus V_1$ is $\mathbf{4} = (2, 2)$. To identify the component that contains vertex **4** and select a spanning tree, the algorithm found 5 edges in the order

$$\{\mathbf{4}, \mathbf{5}\}, \{\mathbf{4}, \mathbf{9}\}, \{\mathbf{5}, \mathbf{10}\}, \{\mathbf{9}, \mathbf{14}\}, \{\mathbf{10}, \mathbf{15}\},$$

which form a spanning tree of the second component (shown on the right panel of Fig. 4, denoted (V_2, E_2, R_2)). Here V_2 consists of the 6 vertices appearing in the spanning tree, i.e. $V_2 = \{\mathbf{4}, \mathbf{5}, \mathbf{9}, \mathbf{10}, \mathbf{14}, \mathbf{15}\}$. Also $E_2 = E_{\mathscr{C}} \cap (V_2 \times V_2)$ and R_2 is $R_{\mathscr{C}}$ restricted to E_2. The unique distribution p_2 with support V_2 that satisfies the ratio specifications associated with the 5 edges of the spanning tree is given by

$$p_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1/12 & 0 & 1/6 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1/6 & 0 & 1/4 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 1/12 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

It can be verified that p_2 satisfies R_2, so that R_2 is compatible. Since $V_{\mathscr{C}} \setminus (V_1 \cup V_2) = \emptyset$, and since R_1 and R_2 are both compatible, we conclude that $R_{\mathscr{C}}$ (and hence $\mathscr{C}$) is compatible. The set of positive probability distributions on $V_{\mathscr{C}}$ that satisfy $\mathscr{C}$ is

$$\{\lambda_1 p_1 + \lambda_2 p_2 : \lambda_1 > 0, \lambda_2 > 0, \lambda_1 + \lambda_2 = 1\}.$$

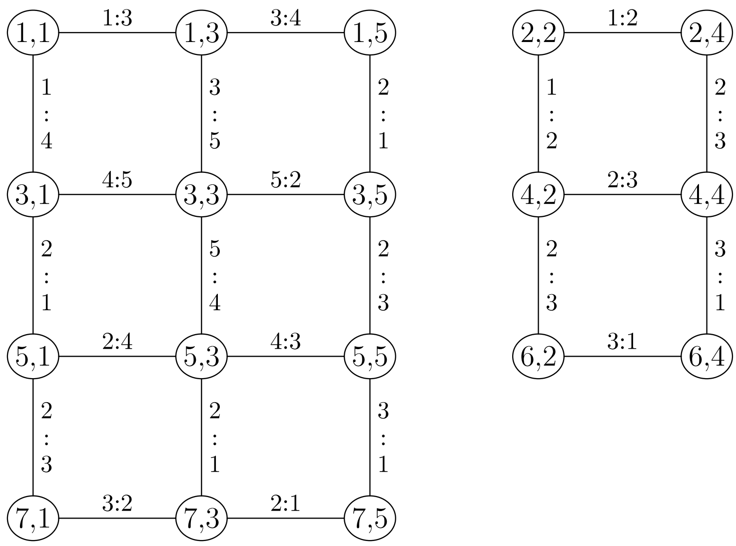

Fig. 4. Graphical representation for Example 4.

5. Concluding remarks

We introduced a graphical representation to deal with compatibility of a given set $\mathscr{C}$ of full conditional distributions for random variables $X_1, \ldots, X_n$. This approach is conceptually useful which provides characterizations of compatibility using basic ideas in graph theory. It also shows why the case of full conditional distributions (as studied by most papers in the literature) is easier to deal with than more general cases as considered in [10]. Based on the graphical representation along with the natural lexicographical ordering, we presented an algorithm to check compatibility of $\mathscr{C}$ as well as to determine, in case that $\mathscr{C}$ is compatible, the set of all probability distributions that satisfy $\mathscr{C}$. It works for general n and allows for general structural zeros although it can be time consuming when n is not small and/or some X_i takes values in a large set. In the literature, the algorithm proposed by Kuo and Wang [14] shares some similar features although it does not use the graphical representation and no underlying theory is given. It does not present the set of all probability distributions satisfying $\mathscr{C}$ when $\mathscr{C}$ is compatible. Ip and Wang [13] proposed using the canonical representation to deal with the compatibility problem which, however, requires that no structural zeros are present. See page 2458 in [14] for a discussion of limitations of other methods in the literature.

The graphical representation approach works when the information contained in given conditionals (such as full conditionals) can be equivalently described in terms of probability ratios between configurations. However, this is not so for general conditionals as considered in [10]. This is a major limitation of the approach.

In practical applications, since specified conditional distributions are typically subject to errors, it is unlikely for them to be exactly compatible. An issue of practical relevance is to find a probability distribution that is "most nearly compatible" with the given conditional distributions, which has been addressed in [1,3,5]. See also [9,12] for related work. It will be of great interest to formulate and solve this problem in terms of a graphical representation.

Acknowledgments

The authors are grateful to the two reviewers for helpful comments that led to an improved presentation. The first author gratefully acknowledges support from the National Science Council of Taiwan.

References

[1] B.C. Arnold, E. Castillo, J.M. Sarabia, Conditional Specification of Statistical Models, Springer, New York, 1999.
[2] B.C. Arnold, E. Castillo, J.M. Sarabia, Conditionally specified distributions: an introduction (with discussions), Statist. Sci. 16 (2001) 249–274.
[3] B.C. Arnold, E. Castillo, J.M. Sarabia, Exact and near compatibility of discrete conditional distributions, Comput. Statist. Data Anal. 40 (2002) 231–252.
[4] B.C. Arnold, E. Castillo, J.M. Sarabia, Compatibility of partial or complete conditional probability specifications, J. Statist. Plann. Infer. 123 (2004) 133–159.
[5] B.C. Arnold, D.V. Gokhale, Distributions most nearly compatible with given families of conditional distributions, Test 7 (1998) 377–390.
[6] B.C. Arnold, S.J. Press, Compatible conditional distributions, J. Amer. Statist. Assoc. 84 (1989) 152–156.
[7] J. Besag, Spatial interaction and the statistical analysis of lattice systems (with discussions), J. R. Stat. Soc. Ser. B Stat. Methodol. 36 (1974) 192–236.
[8] B. Bollobas, Modern Graph Theory, Springer, New York, 1998.
[9] S.-H. Chen, E.H. Ip, Y.J. Wang, Gibbs ensembles for nearly compatible and incompatible conditional models, Comput. Statist. Data Anal. 55 (2011) 1760–1769.
[10] A. Gelman, T.P. Speed, Characterizing a joint probability distribution by conditionals, J. R. Stat. Soc. Ser. B Stat. Methodol. 55 (1993) 185–188.
[11] A. Gelman, T.P. Speed, Corrigendum: characterizing a joint probability distribution by conditionals, J. R. Stat. Soc. Ser. B Stat. Methodol. 61 (1999) 483.
[12] D. Heckerman, D.M. Chickering, C. Meek, R. Rounthwaite, C. Kadie, Dependency networks for inference, collaborative filtering, and data visualization, J. Machine Learning Research 1 (2000) 49–75.

[13] E.H. Ip, Y.J. Wang, Canonical representation of conditionally specified multivariate discrete distributions, J. Multivariate Anal. 100 (2009) 1282–1290.
[14] K.-L. Kuo, Y.J. Wang, A simple algorithm for checking compatibility among discrete conditional distributions, Comput. Statist. Data Anal. 55 (2011) 2457–2462.
[15] A.B. Slavkovic, S. Sullivant, The space of compatible full conditionals is a unimodular toric variety, J. Symbolic Comput. 41 (2006) 196–209.
[16] C.-C. Soong, L.-A. Li, C.-H. Chen, T.J. Jiang, K.-L Kuo, Compatibility of finite discrete conditional distributions, Statistica Sinica 20 (2010) 423–440.
[17] G. Tian, M. Tan, K.W. Ng, M. Tang, A unified method for checking compatibility and uniqueness for finite discrete conditional distributions, Comm. Statist. Theory Methods 38 (2009) 115–129.
[18] Y.J. Wang, K.-L. Kuo, Compatibility of discrete conditional distributions with structural zeros, J. Multivariate Anal. 101 (2010) 191–199.