國立政治大學統計學系

博士論文

離散條件機率分配之相容性研究

On compatibility of discrete conditional distributions

指導教授：姚怡慶 博士

研究生：陳世傑

中華民國104年7月

謝　辭

與其說跟姚老師學統計，倒不如說跟姚老師學數學。與其說跟姚老師學數學，倒不如說跟姚老師學做事情的態度。姚老師做事情就好像在做數學題目一樣，有條有理。我的個性比較散漫，再加上我大學和研究所，都不是念數學系或統計系，所以當我開始和姚老師學習撰寫論文時，就相當吃力。經過姚老師多年來的耐心指導，他改變了我，也才有這篇論文的產生。

有幸請到鄭宗記老師、翁久幸老師、洪英超老師、陳宏老師和程毅豪老師來當我的口試委員，他們所提供的寶貴意見，讓本篇論文有更佳的呈現。

學弟泰期和宇翔，可說是我的貴人。他們在程式上幫了我很大的忙，沒有他們，此篇論文無法完成。

撰寫論文期間，幸虧有內人月霞和兒子昶廷幫忙打字，才能使本篇論文如期完成。

承蒙劉惠美老師的鼓勵，我才能進入統計學的研究領域。

我來政大唸博士班，一待就是十年，過程雖然辛苦，但終於磨成一劍。

國立政治大學
National Chengchi University

Abstract

For two discrete random variables X_1 and X_2 taking values in $\{1, \ldots, I\}$ and $\{1, \ldots, J\}$, respectively, a putative conditional model for the joint distribution of X_1 and X_2 consists of two $I \times J$ matrices representing the conditional distributions of X_1 given X_2 and of X_2 given X_1. We say that two conditional distributions (matrices) A and B are compatible if there exists a joint distribution of X_1 and X_2 whose two conditional distributions are exactly A and B. We present new versions of necessary and sufficient conditions for compatibility of discrete conditional distributions via a graphical representation. Moreover, we show that there is a unique joint distribution for two given compatible conditional distributions if and only if the corresponding graph is connected. Markov chain characterizations are also presented.

Keywords: compatibility of conditional distributions, graph theory, connectedness, spanning tree, Gibbs sampler, MCMC.

National Chengchi University

Contents

國立政治大學

National Chengchi University

1 Introduction

The problem of characterizing a joint distribution by conditional distributions has been extensively studied in the last few decades. Applications may be found in classical construction of joint distribution and elicitation of Bayesian multiparameter prior distribution. We first consider the simple case of two random variables (X, Y) whose joint distribution is to be determined, given the two conditional distributions $P_{X|Y}$ and $P_{Y|X}$. For example, Toffoli et al. (2006) investigated the relationship between adverse reaction to drug treatment and genotype in the gene region UGT1A1 $*$ 28 for metastatic colorectal cancer patients. Drug treatment with irinotecan has been found to be effective for metastatic colorectal cancer patients. However, irinotecan includes toxicity and the individual reaction to the toxicity is known to be different. Genotype in the gene region UGT1A1 $*$ 28 is known to be associated with adverse reaction. Let X be the genotype with three kinds: TA7/TA7, TA7/TA6 and TA6/TA6. Let Y be the adverse reaction with four levels: severe, moderate, mild and nil. Two conditional models are commonly used by clinicians: the diagnostic conditional model $P_{X|Y}$ and the treatment conditional model $P_{Y|X}$. To obtain full information about the probability model, it is necessary to construct a joint distribution from $P_{X|Y}$ and $P_{Y|X}$. Since in practice, $P_{X|Y}$ and $P_{Y|X}$ are either estimated or hypothesized, there may not exist a joint distribution which has the given $P_{X|Y}$ and $P_{Y|X}$ as its conditional distributions. If there is a joint distribution which has $P_{X|Y}$ and $P_{Y|X}$ as its conditional distributions, then $P_{X|Y}$ and $P_{Y|X}$ are said to be compatible. Therefore, it is desired (i) to determine whether the given $P_{X|Y}$ and $P_{Y|X}$ are compatible, and (ii) to find all such (compatible) joint distributions when the given conditional distributions are compatible.

The compatibility problem is most easily visualized in the finite discrete case. A convenient introduction to the topic may be found in Arnold and Press (1989). They proposed "ratio matrix" of two discrete conditional distributions to check the compatibility. They showed that two discrete condi-

國立政治大學 National Chengchi University

tional distributions are compatible if and only if the ratio matrix is of rank one. However, this version can be used only for the discrete conditional distributions without zero elements. If the discrete conditional distributions contain zero elements, the method can not be applied. Song, Li, Chen, Jiang and Kuo (2010) proposed “positive extension ratio matrix” of two discrete conditional distributions to solve the problem. “Positive extension” means inserting positive values for the undefined elements in a ratio matrix. They showed that two discrete conditional distributions are compatible if and only if there exists a positive extension ratio matrix of rank one. They also found that there is a unique joint distribution for two given compatible conditional distributions if and only if the corresponding ratio matrix is irreducible. However, the positive extension ratio matrix may not be unique and may be difficult to find. Moreover, this approach can not be naturally extended to the higher-dimensional case.

A discrete conditional distribution can be considered as linear constraints on the (unknown) joint distribution. Thus, the linear algebra approach is a natural way to deal with the compatibility problem. A given set of discrete conditional distributions is compatible if and only if there is a joint distribution satisfying conditional distribution constraints. Arnold, Castillo and Sarabia (2002) proposed three methods for solving constrained linear equations to obtain the compatible joint distributions. The first method is to find a joint distribution satisfying the given set of discrete conditional distributions directly. The second method is to find the marginal distributions of X and Y which combined with $P_{X|Y}$ and $P_{Y|X}$ will determine the joint distribution. The third method is to find a marginal distribution of X, knowing that a compatible marginal distribution of X, combined with $P_{Y|X}$, will determine the joint distribution. (Both the second and third methods are based on Theorem 2.2.1.) Alternatively, Tian, Tan, Ng and Tang (2009) used the Euclidean distance measure to transform the compatibility problem into a quadratic optimization problem with unit cube constraints. This method is related to the third method of Arnold, Castillo and Sarabia (2002), and requires numerically solving a quadratic optimization problem. While the above linear algebra approach can be naturally extended to the higher-dimensional case, it can be very time consuming for three or more random variables. Therefore, the linear algebra approach is mostly applied to bivariate distributions in which each random variable takes values in a small set.

A discrete conditional distribution can also be treated as a stochastic matrix. Arnold, Castillo and Sarabia (1999) applied Markov chain to check the compatibility of discrete conditional distributions. This approach needs not only to solve a set of stationary marginal distributions, but also to determine whether the set of stationary marginal distributions matches the given set of conditional distributions.

Ip and Wang (2009) proposed using the canonical representation to deal with the compatibility problem. The method does not require solving constrained linear equations, but requires no zero elements. Wang and Kuo (2010) proposed an odds-oriented approach to deal with the case involving zero elements. This approach is computationally demanding. The algorithm proposed by Kuo and Wang (2011) shares some similar features with our approach in Chapter 3, although it did not use the graphical representation and no underlying theory was provided.

In this thesis we reformulate the compatibility problem in terms of a graphical representation where each vertex corresponds to a possible sample point and an edge connects two vertices if and only if the probability ratio of the two corresponding possible sample points is specified through one of the given conditional distributions. A sequence of connected vertices is called a path. A path determines the probability ratio of any two vertices in the path. However, two vertices may be connected by more than one path. We show that a given set of conditional distributions is compatible if and only if the probability ratio of any two vertices in a path does not depend on the chosen path, and that there is a unique joint distribution for a given set of compatible conditional distributions if and only if the corresponding graph is connected.

In addition, we use the spanning tree to determine all the compatible joint distributions when a given set of conditional distributions is compatible.

The rest of the thesis is organized as follows. In Chapter 2, we review the ratio matrix approach for checking compatibility and uniqueness of two discrete conditional distributions. In Chapter 3, we first use bivariate discrete conditional distributions to illustrate the graphical representation approach (cf. Yao, Chen and Wang (2014)). Then, we extend the graphical representation approach to the higher-dimensional case. We also discuss the relationship between the ratio matrix approach and the graphical representation approach. In Chapter 4, we discuss Markov chain characterizations, which help to understand the connection of compatibility with Gibbs sampler. Gibbs sampler is a Markov chain Monte Carlo sampling algorithm for generating joint distribution via individual conditional distributions. For example, the bivariate Gibbs sampler generates X sample from $P_{X|Y}$ and Y sample from $P_{Y|X}$, resulting in two sequences $\{X^{(t)}, Y^{(t)}\}_{t=1}^{\infty}$ and $\{Y^{(t)}, X^{(t)}\}_{t=1}^{\infty}, t = 1, 2, \ldots$. If $P_{X|Y}$ and $P_{Y|X}$ are compatible, they have the same limiting distribution. If $P_{X|Y}$ and $P_{Y|X}$ are incompatible, Liu (1996) observed that the two limiting distributions are different and also derived some results. However, he only considered the two-dimensional case. We will extend his discussions to the higher-dimensional case.

國立政治大學 National Chengchi University

2 Compatible conditional distributions

2.1 Compatibility

Consider two discrete random variables X and Y taking values in $\{x_1, \ldots, x_I\}$ and $\{y_1, \ldots, y_J\}$, respectively. Let $A = (A_{ij})$ and $B = (B_{ij})$ be two $I \times J$ matrices with nonnegative elements such that each column sum of A equals 1 and each row sum of B equals 1.

Definition 2.1.1. A and B are said to be compatible if there exists a joint distribution of X and Y such that $P\left(X = x_i | Y = y_j\right) = A_{ij}$ and $P(Y = y_j | X = x_i) = B_{ij}$ for $i = 1, \ldots, I, j = 1, \ldots, J$.

Example 2.1.1. Consider two conditional distribution matrices:

$$A = \begin{pmatrix} 1/3 & 3/5 \\ 2/3 & 2/5 \end{pmatrix}, \qquad B = \begin{pmatrix} 1/4 & 3/4 \\ 1/3 & 2/3 \end{pmatrix}.$$

Let

$$p = \begin{pmatrix} a & c \\ b & d \end{pmatrix},$$

where

$$a + b + c + d = 1,$$
$$a, b, c, d \geq 0.$$

Suppose that the joint distribution p has A and B as its two conditional distributions, i.e., $p_{X|Y}\left(X = x_i | Y = y_j\right) = A_{ij}$ and $p_{Y|X}\left(Y = y_j | X = x_i\right) = B_{ij}, i, j = 1, 2$. Then a, b, c, d satisfy the following constraints:

$$\frac{a}{b} = \frac{A_{11}}{A_{21}} = \frac{1/3}{2/3} = \frac{1}{2},$$

$$\frac{c}{d} = \frac{A_{12}}{A_{22}} = \frac{3/5}{2/5} = \frac{3}{2},$$

$$\frac{a}{c} = \frac{B_{11}}{B_{12}} = \frac{1/4}{3/4} = \frac{1}{3},$$

國立政治大學
National Chengchi University

$$\frac{b}{d} = \frac{B_{21}}{B_{22}} = \frac{1/3}{2/3} = \frac{1}{2}.$$

However, there are no solutions to the above linear equations. It follows that there does not exist a joint distribution whose two conditional distributions are exactly A and B. Therefore, A and B are incompatible.

Example 2.1.2. Consider two conditional distribution matrices:

$$A = \begin{pmatrix} 1/3 & 3/8 \\ 2/3 & 5/8 \end{pmatrix}, \qquad B = \begin{pmatrix} 1/4 & 3/4 \\ 2/7 & 5/7 \end{pmatrix}.$$

A and B are compatible, since we can find a joint distribution

$$p = \begin{pmatrix} 1/11 & 3/11 \\ 2/11 & 5/11 \end{pmatrix}$$

such that $p_{X|Y}\left(X = x_i | Y = y_j\right) = A_{ij}$ and $p_{Y|X}\left(Y = y_j | X = x_i\right) = B_{ij}, i, j = 1, 2$.

If A and B contain zero elements, there is a necessary condition for the compatibility of A and B.

Definition 2.1.2. The set $N^A = \{(i, j)\colon\ A_{ij} > 0\}$ is called the incidence set of matrix A, the set of locations of non-zero elements in matrix A.

Clearly $N^A = N^B$ (i.e., A and B share a common incidence set) is a necessary condition for compatibility. We assume $N^A = N^B$ and denote the common incidence set by $N = N^A = N^B = \{(i, j) | A_{ij} > 0, B_{ij} > 0\}$. Note that $N^A = N^B$ is trivially satisfied if A and B contain only positive elements.

Example 2.1.3. Consider two conditional distribution matrices:

$$A = \begin{pmatrix} 1/4 & 0 \\ 3/4 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1/3 & 2/3 \\ 1 & 0 \end{pmatrix}.$$

Since A and B do not share a common incidence set, A and B are incompatible.

Example 2.1.4. Consider two conditional distribution matrices:

$$A = \begin{pmatrix} 2/5 & 0 & 0 \\ 0 & 1 & 1 \\ 3/5 & 0 & 0 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \end{pmatrix}$$

A and B are compatible, since $N^A = N^B$ and we can find infinitely many compatible joint distributions

$$P = \begin{pmatrix} 2a & 0 & 0 \\ 0 & b & b \\ 3a & 0 & 0 \end{pmatrix}$$

with $a, b > 0$ and $5a + 2b = 1$.

2.2 Review of the ratio matrix approach for compatibility between two conditional distributions

Suppose that A and B are compatible. Then there exists a joint distribution p such that $p(x_i, y_i) = p_X(x_i) B_{ij} = p_Y(y_j) A_{ij}$ for all i, j. With this simple observation, Arnold and Press (1989) obtained the following theorem.

Theorem 2.2.1. Assume $N^A = N^B$. Then A and B are compatible if and only if there exist two probability vectors $\tau = (\tau_1, \ldots, \tau_I)$ and $\eta = (\eta_1, \ldots, \eta_J)$ such that $\tau_i B_{ij} = \eta_j A_{ij}$ for all i, j.

The vectors τ and η can be readily interpreted as the X- and Y-marginal distributions, respectively. From Theorem 2.2.1, Arnold and Press (1989) derived the following theorem .

Theorem 2.2.2. Assume $N^A = N^B = N$. Then A and B are compatible if and only if there exist two vectors $u = (u_1, \ldots, u_I)$ and $\upsilon = (\upsilon_1, \ldots, \upsilon_J)$ such that

$$C_{ij} = A_{ij}/B_{ij} = u_i \upsilon_j, \quad (i, j) \in N.$$

Definition 2.2.1. Let

$$C_{ij} = \begin{cases} A_{ij}/B_{ij} & \text{if} \quad A_{ij}, B_{ij} > 0 \\ * & \text{if} \quad A_{ij} = B_{ij} = 0, \end{cases}$$

where the symbol $*$ denotes an undefined element. Then the matrix $C = (C_{ij})$ is called the ratio matrix of A and B.

When A and B contain only positive elements, Arnold and Press (1989) applied the ratio matrix of A and B to obtain the following necessary and sufficient condition for the compatibility of A and B.

Theorem 2.2.3. Suppose that A and B contain only positive elements. Then the following statements are equivalent.

(i) A and B are compatible.

(ii) The ratio matrix C is of rank one.

(iii) The compatible joint distribution is unique.

Example 2.2.1. Consider two conditional distribution matrices:

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

The corresponding ratio matrix is

$$C = \begin{pmatrix} 5/6 & 5/7 & 5/7 \\ 7/6 & 1 & 1 \\ 8/6 & 8/7 & 8/7 \end{pmatrix}.$$

Since the ratio matrix C is of rank one, A and B are compatible and the compatible joint distribution is unique. Let p be the compatible joint distribution. We can use a convenient method to find this p. We begin with a 3×3 positive matrix $Q = (Q_{ij})$, where Q_{ij} represents the probability ratio of p_{ij}

國立政治大學 National Chengchi University

to p_{11}. Note that $Q_{11} = p_{11} : p_{11} = 1$. The remaining elements can be derived from A and B. If we start from the first column of A, then we have

$$\frac{Q_{21}}{Q_{11}} = \frac{p_{21}}{p_{11}} = \frac{A_{21}}{A_{11}} = \frac{1/3}{1/6} = \frac{2}{1},$$
$$\frac{Q_{31}}{Q_{11}} = \frac{p_{31}}{p_{11}} = \frac{A_{31}}{A_{11}} = \frac{1/2}{1/6} = \frac{3}{1},$$

and yield $Q_{21} = 2, Q_{31} = 3$.

Next, from first row of B, we have

$$\frac{Q_{12}}{Q_{11}} = \frac{p_{12}}{p_{11}} = \frac{B_{12}}{B_{11}} = \frac{3/5}{1/5} = \frac{3}{1},$$
$$\frac{Q_{13}}{Q_{11}} = \frac{p_{13}}{p_{11}} = \frac{B_{13}}{B_{11}} = \frac{1/5}{1/5} = \frac{1}{1},$$

and yield $Q_{12} = 3, Q_{13} = 1$.

Similarly, from the second and third rows of B, yield $Q_{22} = 2, Q_{23} = 3, Q_{32} = 2, Q_{33} = 3$. Then, we yield the matrix Q:

$$Q = \begin{pmatrix} 1 & 3 & 1 \\ 2 & 2 & 3 \\ 3 & 2 & 3 \end{pmatrix}.$$

Finally, normalize Q to form the compatible joint distribution

$$p = \frac{1}{20} \begin{pmatrix} 1 & 3 & 1 \\ 2 & 2 & 3 \\ 3 & 2 & 3 \end{pmatrix}.$$

If the conditional distribution matrices contain zero elements, Theorem 2.2.3 can not be applied. Song, Li, Chen, Jiang and Kuo (2010) proposed "positive extension" of ratio matrix C to deal with the problem. They extended the ratio matrix by properly assigning positive numbers to all undefined elements.

Definition 2.2.2. A matrix $\bar{C} = (\bar{C}_{ij})$ is called a positive extension of ratio matrix C if all $\bar{C}_{ij} > 0$ and $\bar{C}_{ij} = C_{ij}$ when $C_{ij} > 0$.

國立政治大學 National Chengchi University

If the ratio matrix C contains only positive elements, then the positive extension $\bar{C} = C$. Otherwise, the positive extension $\bar{C}$ is not unique.

Theorem 2.2.4. Assume $N^A = N^B$. Then A and B are compatible if and only if there exists a rank one positive extension $\bar{C}$ of ratio matrix C.

Example 2.2.2. Consider two conditional distribution matrices:

$$A = \begin{pmatrix} 1/3 & 3/4 & 0 \\ 0 & 1/4 & 3/5 \\ 2/3 & 0 & 2/5 \end{pmatrix}, \qquad B = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/4 & 3/4 \\ 1/3 & 0 & 2/3 \end{pmatrix}.$$

The corresponding ratio matrix is

$$C = \begin{pmatrix} 2/3 & 3/2 & * \\ * & 1 & 4/5 \\ 2 & * & 3/5 \end{pmatrix}.$$

In order to make the first and second rows proportional in matrix C, we have to assign $C_{13} = 6/5$ and $C_{21} = 4/9$, and the matrix turns out to be

$$\begin{pmatrix} 2/3 & 3/2 & 6/5 \\ 4/9 & 1 & 4/5 \\ 2 & * & 3/5 \end{pmatrix}.$$

However, no matter what the value of C_{32} is, the last two rows will never be proportional, i.e., there does not exist a rank one positive extension ratio matrix. Therefore, A and B are incompatible.

Definition 2.2.3. A ratio matrix C is said to be reducible if, after interchanging some rows and/or columns, it can be rearranged as

$$\begin{pmatrix} T_1 & * \\ * & T_2 \end{pmatrix},$$

where entries off the diagonal block matrices T_1 and T_2 are all $*$. The matrix C is irreducible if it is not reducible.

國立政治大學 National Chengchi University

If the ratio matrix C is irreducible, Song, Li, Chen, Jiang and Kuo (2010) obtained the following theorem.

Theorem 2.2.5. Assume that $N^A = N^B$ and the ratio matrix C is irreducible. Then the following statements are equivalent.

(i) A and B are compatible.

(ii) C has a unique rank one positive extension $\bar{C}$.

(iii) The compatible joint distribution is unique.

Example 2.2.3. Consider two conditional distribution matrices:

$$A = \begin{pmatrix} 1/3 & 3/4 & 0 \\ 0 & 1/4 & 3/5 \\ 2/3 & 0 & 2/5 \end{pmatrix}, \qquad B = \begin{pmatrix} 1/4 & 3/4 & 0 \\ 0 & 1/4 & 3/4 \\ 1/2 & 0 & 1/2 \end{pmatrix}.$$

The corresponding ratio matrix is

$$C = \begin{pmatrix} 4/3 & 1 & * \\ * & 1 & 4/5 \\ 4/3 & * & 4/5 \end{pmatrix}.$$

The ratio matrix C is obviously irreducible. We can set $C_{13} = 4/5$, $C_{21} = 4/3$ and $C_{32} = 1$, the extension ratio matrix turns out to be

$$\begin{pmatrix} 4/3 & 1 & 4/5 \\ 4/3 & 1 & 4/5 \\ 4/3 & 1 & 4/5 \end{pmatrix}.$$

Since the rank one positive extension ratio matrix is unique, the compatible joint distribution is unique and easily found to be

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

國立政治大學 National Chengchi University

When the ratio matrix C is reducible, Song, Li, Chen, Jiang and Kuo (2010) used the following lemma to rearrange a reducible ratio matrix C as an irreducible block diagonal matrix .

Lemma 2.2.1. For any ratio matrix C, by interchanging some rows and/or columns, it can be rearranged as an irreducible block diagonal matrix, denoted by

$$T\,(C) = \begin{pmatrix} T_1 & * & * & * \\ * & T_2 & * & * \\ * & * & \ddots & * \\ * & * & * & T_k \end{pmatrix},$$

where the diagonal block matrices $T_1, \ldots, T_k\,(k \geq 1)$ are irreducible and elements off these diagonal block matrices are all $*$. When $k = 1$, C itself is irreducible. For simplicity, let $T(C) = \text{Diag}\,(T_1, \ldots, T_k)$ for $k \geq 1$.

Song, Li, Chen, Jiang and Kuo (2010) applied irreducible block diagonal matrix to present another version of necessary and sufficient conditions for compatibility in the following theorem.

Theorem 2.2.6. Assume $N^A = N^B$. Let $T(C) = \text{Diag}(T_1, \ldots, T_k)$ be an irreducible block diagonal matrix of ratio matrix C. Then A and B are compatible if and only if each T_i has a unique rank one positive extension $\bar{T}_i, i = 1, \ldots, k$.

Example 2.2.4. Consider two conditional distribution matrices:

$$A = \begin{pmatrix} 3/4 & 0 & 0 & 1 \\ 0 & 1 & 2/3 & 0 \\ 0 & 0 & 1/3 & 0 \\ 1/4 & 0 & 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 3/5 & 0 & 0 & 2/5 \\ 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

The corresponding ratio matrix is

$$C = \begin{pmatrix} 5/4 & * & * & 5/2 \\ * & 3 & 1 & * \\ * & * & 1/3 & * \\ 1/4 & * & * & * \end{pmatrix}.$$

The reducible ratio matrix C can be rearranged as an irreducible block diagonal matrix

$$T\,(C) = \begin{pmatrix} 5/4 & 5/2 & * & * \\ 1/4 & * & * & * \\ * & * & 3 & 1 \\ * & * & * & 1/3 \end{pmatrix}.$$

Let $T(C) = \mathrm{Diag}(T_1, T_2)$, where

$$T_1 = \begin{pmatrix} 5/4 & 5/2 \\ 1/4 & * \end{pmatrix} \text{ and } T_2 = \begin{pmatrix} 3 & 1 \\ * & 1/3 \end{pmatrix}.$$

Two positive extensions of T_1 and T_2 are

$$\bar{T}_1 = \begin{pmatrix} 5/4 & 5/2 \\ 1/4 & 1/2 \end{pmatrix} \quad \text{and} \quad \bar{T}_2 = \begin{pmatrix} 3 & 1 \\ 1 & 1/3 \end{pmatrix}.$$

Since both $\bar{T}_1$ and $\bar{T}_2$ are of rank one, A and B are compatible.

Theorem 2.2.6 provides an effective method to check compatibility when the ratio matrix C is reducible. However, Song, Li, Chen, Jiang and Kuo did not characterize all compatible joint distributions when the ratio matrix C is reducible. In the next section, we propose a graphical representation approach to deal with this case where the ratio matrix C is reducible.

The ratio matrix approach cannot be naturally extended to the higher –dimensional case. We will use graphical representation approach to solve the problem.

國立政治大學
National Chengchi University

3 Graphical representation approach

3.1 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). For each edge $\{u, v\}$, there is a specified ratio $r(u, v)$ where $r(u, v)$ represents the probability ratio of vertex u to vertex v. 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 $p(u) : p(v) = r(u, v)$ for all $\{u, v\} \in E$, and (ii) to find all (compatible) probability distributions satisfying R when R is compatible.

In this chapter, instead of X and Y, we use X_1 and X_2 to denote two random variables taking values in $\{1, \ldots, I\}$ and $\{1, \ldots, J\}$, respectively. Let $\xi = \{p_{1|2}, p_{2|1}\}$ be a given set of conditional distributions for two random variables $\underline{X} = (X_1, X_2)$. Let $A = (A_{ij}) = (p_{1|2}\{X_1 = i|X_2 = j\})$ and $B = (B_{ij}) = (p_{2|1}\{X_2 = j|X_1 = i\})$. Let $N^A = \{(i, j)|A_{ij} > 0\}$ and $N^B = \{(i, j)|B_{ij} > 0\}$. We assume $N^A = N^B$ and let $N = N^A = N^B = \{(i, j)|A_{ij} > 0, B_{ij} > 0\}$.

We now define the graphical representation (V_ξ, E_ξ, R_ξ) for ξ. Let $V_\xi = N$ be the vertex set. Write $\underline{x} = (x_1, x_2) \in N$ and $\underline{x}' = (x_1', x_2') \in N$. An edge $\{\underline{x}, \underline{x}'\}$ connects two vertices $\underline{x}, \underline{x}' \in V_\xi$ if and only if either $x_1 = x_1'$ or $x_2 = x_2'$. The ratio associated with this edge is given by

$$r(\underline{x}, \underline{x}') = p_{1|2}(x_1|x_2) : p_{1|2}(x_1'|x_2') \text{ if } x_2 = x_2'$$

or

$$r(\underline{x}, \underline{x}') = p_{2|1}(x_2|x_1) : p_{2|1}(x_2'|x_1') \text{ if } x_1 = x_1'.$$

The resulting graphical representation is denoted by (V_ξ, E_ξ, R_ξ). Since each edge connects two vertices $\underline{x}, \underline{x}' \in V_\xi$ through only one conditional distribution in $\xi = \{p_{1|2}, p_{2|1}\}$, the ratio associated with each edge is uniquely defined. The following theorem shows that compatibility of ξ is equivalent to compatibility of R_ξ.

Theorem 3.1.1. Assume that ξ satisfies $N^A = N^B$. Then a joint distribution with support V_ξ satisfies ξ if and only if it satisfies R_ξ. Consequently, ξ is compatible if and only if R_ξ is compatible.

Proof. To prove the "only if" part, suppose that a joint distribution p (with support V_ξ) satisfies ξ, i.e., under p the conditional distribution of X_S given X_T agrees with $p_{S|T}$ for $(S,T) = (1,2)$ and $(2,1)$. Consider an (arbitrary) edge $\{\underline{x}, \underline{x}'\} \in E_\xi$ with an associated ratio given by $r(\underline{x}, \underline{x}') = p_{S|T}(x_S|x_{\mathrm{T}}) : p_{S|T}(x'_S|x'_T)$, where $(S,T) = (1,2)$ or $(2,1)$. By definition, we have $x_T = x'_T$. Then p satisfies the associated ratio specification since

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

from which it follows that p satisfies R_ξ.

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

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

Since this holds for all $\underline{x} \in V_\xi \setminus \{\underline{x}^0\}$ with $x_T = x^0_T$, under p the conditional distribution of X_S given $X_T = x^0_T$ agrees with $p_{S|T}$. As $\underline{x}^0$ (and x^0_T) is arbitrary, $p_{S|T}$ is indeed the conditional distribution of X_S given X_T under p. This shows that p satisfies ξ. 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 $\xi = \{p_{1|2}, p_{2|1}\}$. Let

$A = (A_{ij})$ and $B = (B_{ij})$, 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 3.1.1. 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. 3.1.1, 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 $r((1, 1), (1, 3)) = B_{11} : B_{13} = 1 : 3$ can be derived from $r((1, 1), (1, 2)) = 1 : 2$ and $r((1, 2), (1, 3)) = 2 : 3$.

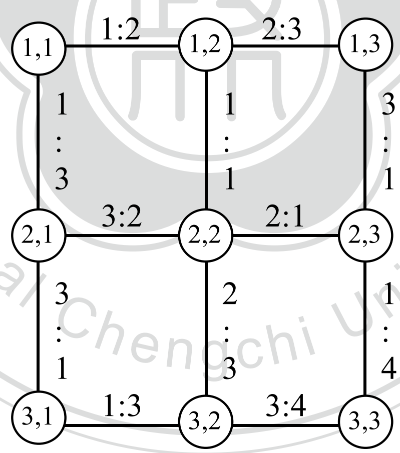

Fig. 3.1.1. Graphical representation for Example 3.1.1.

Remark 3.1.1. For a path $\upsilon_0\upsilon_1 \ldots \upsilon_l$, we have

$$\prod_{i=0}^{l-1} r\left(\upsilon_{i+1}, \upsilon_i\right) = r\left(\upsilon_l, \upsilon_0\right).$$

For example, consider the path $(1,\ 1) \to (1,\ 2) \to (1,\ 3)$ in Fig 3.1.1, $r((1, 2), (1, 1)) \times r((1, 3), (1, 2)) = 2/1 \times 3/2 = 3 : 1 = B_{13} : B_{11} = r((1, 3), (1, 1))$.

Definition 3.1.1. 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 $\{u, v\} \in E \backslash E'$, there exists an E'-path $v_0 v_1 \ldots v_k$ with $v_0 = u, v_k = v$ and $\{v_l, v_{l+1}\} \in E', l = 0, 1, \ldots, k-1$, such that

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

and (iii) for $\{u, v\} \in E' \backslash E$, 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. 3.1.1 is a simplified, but equivalent version of the graphical representation for the given matrices A and B in Example 3.1.1. Another simplified, but equivalent version of the graphical representation is given in Fig. 3.1.1-1. Lemma 3.1.1 states a simple result on equivalent graphical representations, whose proof is straightforward and omitted.

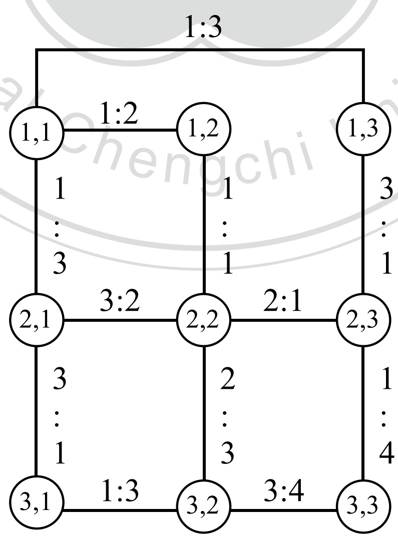

Fig. 3.1.1-1. Another equivalent graphical representation for Example 3.1.1.

Lemma 3.1.1. 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 3.1.2. 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. 3.1.2, 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.

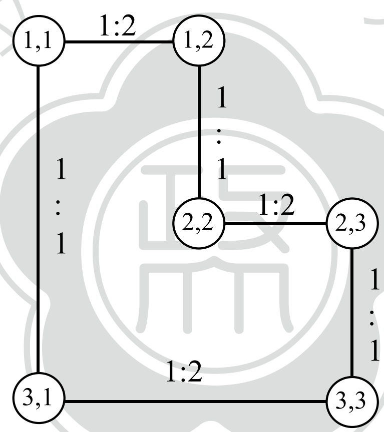

Fig. 3.1.2. Graphical representation for Example 3.1.2.

Definition 3.1.2. A connected graph with no cycles is called a tree. A tree containing every vertex of a graph is called a spanning tree of the graph.

Example 3.1.3.

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

a (simplified but equivalent) graphical representation is given in Fig. 3.1.3, where the four edges $\{(1,j),(3,j)\}, j = 1,2,3$, and $\{(1,1),(1,3)\}$, and the

National Chengchi University

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.

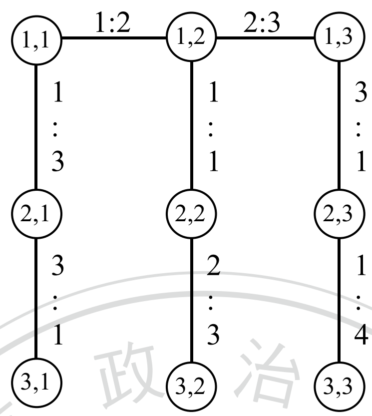

Fig. 3.1.3. Graphical representation for Example 3.1.3.

3.2 Compatibility of a ratio set $\boldsymbol{R}$ and characterization of probability distributions satisfying $\boldsymbol{R}$

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 $(V, E) = \bigcup_{i=1}^{k} (V_i, E_i)$, where each (V_i, E_i) is a connected subgraph and where the symbol $\cup$ denotes disjoint union (implying that $V_i \cap V_j = \emptyset$ and $E_i \cap E_j = \emptyset$ for $i \neq j$).

Theorem 3.2.1. For a graphical representation (V, E, R) where (V, E) is connected, the following statements are equivalent.

(i) R is compatible.

(ii) For any two paths $v_0 v_1 \ldots v_l$ and $w_0 w_1 \ldots w_m$ with $v_0 = w_0, v_l = w_m$,

$$\prod_{i=0}^{l-1} r(v_{i+1}, v_i) = \prod_{i=0}^{m-1} r(w_{i+1}, w_i). \tag{3.1}$$

(iii) For every cycle $v_0 v_1 \ldots v_l v_0$ (i.e., a path whose initial and terminal

國立政治大學 National Chengchi University

vertices are the same), we have

$$\prod_{i=0}^{l} r\left(v_{i+1}, v_i\right) = 1, \tag{3.2}$$

where $v_{l+1} := v_0$.

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 $v_0 v_1 \ldots v_l$ and $w_0 w_1 \ldots w_m$ with $v_0 = w_0, v_l = w_m$, we have

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

$$\frac{p\left(v_l\right)}{p\left(v_0\right)} = \frac{p\left(w_m\right)}{p\left(w_0\right)} = \prod_{i=0}^{m-1} \frac{p\left(w_{i+1}\right)}{p\left(w_i\right)} = \prod_{i=0}^{m-1} r\left(w_{i+1}, \mathrm{w}_i\right).$$

This shows that (3.1) holds.

To show that (ii) implies (i), fix a vertex $v_0 \in V$. Since the graph is connected, for every $v \in V \backslash \{v_0\}$, there exists a path $v_0 v_1 \ldots v_l$ connecting v_0 and $v = v_l$. Define

$$q\left(v\right) := \prod_{i=0}^{l-1} r\left(v_{i+1}, v_i\right).$$

By condition (ii), the definition of $q(v)$ does not depend on the chosen path. Letting $q(v_0) := 1$, define

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

which is a positive probability distribution on V satisfying R. So R is compatible. The proof is complete.

Remark 3.2.1. As a simple application of Theorem 3.2.1, consider Example 3.1.2 for which the graphical representation in Fig. 3.1.2 has only one cycle. There are two methods to check the compatibility for Example 3.1.2. One method is to check condition (ii) in Theorem 3.2.1. Suppose that initial vertex is (1, 1) and terminal vertex is (3, 3). There are two paths from vertex (1, 1) to vertex (3, 3). For the path (1, 1) $\to$ (1, 2) $\to$ (2, 2) $\to$ (2, 3) $\to$ (3, 3), the ratio $r((3, 3), (1, 1)) = 2/1 \times 1/1 \times 2/1 \times 1/1 = 4:1$. For the other path (1, 1) $\to$ (3, 1) $\to$ (3, 3), the ratio $r((3,3),(1,1)) = 1/1 \times 2/1 = 2:1$. Since the ratio $r((3,3),(1,1))$ depends on the chosen path, A and B are incompatible. Another method is to check condition (iii) in Theorem 3.2.1. For the cycle (1, 1) $\to$ (3, 1) $\to$ (3, 3) $\to$ (2, 3) $\to$ (2, 2) $\to$ (1, 2) $\to$ (1, 1), the left hand side of (3.2) for this cycle equals $1/1 \times 2/1 \times 1/1 \times 1/2 \times 1/1 \times 1/2 = 1/2 \neq 1$, implying incompatibility.

Remark 3.2.2. It follows from Theorem 3.2.1 and its proof that when (V, E) is connected and R is compatible, there is a unique probability distribution satisfying R.

Remark 3.2.3. 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 3.2.1 is trivially satisfied. By Remark 3.2.2, there is a unique probability distribution on V satisfying R. Example 3.1.3 is such a case, so it is compatible and has a unique compatible joint distribution. In a graphical representation (V, E, R) where (V, E) is connected, for every spanning tree of the graph (V, E), 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 3.1.1 and note that Example 3.1.3 is derived from Example 3.1.1 with the second and third rows of matrix B removed. As a result, the graphical representation for Example 3.1.3 as shown in Fig. 3.1.3 is a spanning tree of the graphical representation for Example 3.1.1 as shown in Fig. 3.1.1. The unique compatible joint distribution for Example 3.1.3 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 3.1.1, implying compatibility.

If a graph (V, E) is not connected, we have the following theorem.

Theorem 3.2.2. Consider a graphical representation $(V, E, R) = \bigcup_{i=1}^{k}(V_i, E_i, R_i)$ 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, $i = 1, \ldots, k$.

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

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

It is easily verified that $p_{\lambda_1,\ldots,\lambda_k}$ is a probability distribution on V satisfying R, implying that R is compatible.

Remark 3.2.4. In Theorem 3.2.2, 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. Theorem 3.2.1). For any positive numbers $\lambda_1, \ldots, \lambda_k$ with $\sum_{i=1}^{k} \lambda_i = 1$, define

$$p_{\lambda_1,\ldots,\lambda_k}(\upsilon) := \sum_{i=1}^{k} \lambda_i p_i(\upsilon) 1_{V_i}(\upsilon), \upsilon \in V. \tag{3.3}$$

which is a positive probability distribution on V satisfying R (cf. the proof of Theorem 3.2.2). On the other hand, let p be a positive probability distribution on V satisfying R. Define $p_i^*(\upsilon) := p(\upsilon)/\sum_{\upsilon \epsilon V_i} p(\upsilon), \upsilon \in V_i$, which is a positive probability distribution on V_i satisfying R_i. By uniqueness, we have $p_i^* = p_i$.

Letting $\lambda_i = \sum_{\upsilon \epsilon V_i} p(\upsilon)$, it follows that

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

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 3.2.3.

Theorem 3.2.3. Consider a graphical representation $(V, E, R) = \bigcup_{i=1}^{k}(V_i, E_i, R_i)$ 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 (3.3).

By Theorem 3.2.3, we can easily check the compatibility and find all compatible joint distributions for Example 2.2.4.

Example 3.2.1. (Example 2.2.4 continued) For

$$A = \begin{pmatrix} 3/4 & 0 & 0 & 1 \\ 0 & 1 & 2/3 & 0 \\ 0 & 0 & 1/3 & 0 \\ 1/4 & 0 & 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 3/5 & 0 & 0 & 2/5 \\ 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

a graphical representation(V, E, R) is given in Fig.3.2.1. There are two components in the figure. Let $(V_1, E_1,\ R_1)$ be the first component containing vertex (1, 1), (1, 4), (4, 1) and $(V_2, E_2,\ R_2)$ be the second component containing vertex (2, 2), (2, 3), (3, 3). Since (V_1, E_1) and (V_2, E_2) are trees, the corresponding ratio sets R_1 and R_2 are compatible. Therefore, R is compatible. The unique distribution p_1 with support V_1 satisfying R_1 is given by

$$p_1 = \begin{pmatrix} 1/2 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1/6 & 0 & 0 & 0 \end{pmatrix}.$$

The unique distribution p_2 with support V_2 satisfying R_2 is given by

$$p_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/4 & 1/2 & 0 \\ 0 & 0 & 1/4 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The set of positive probability distributions on V satisfying R is

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

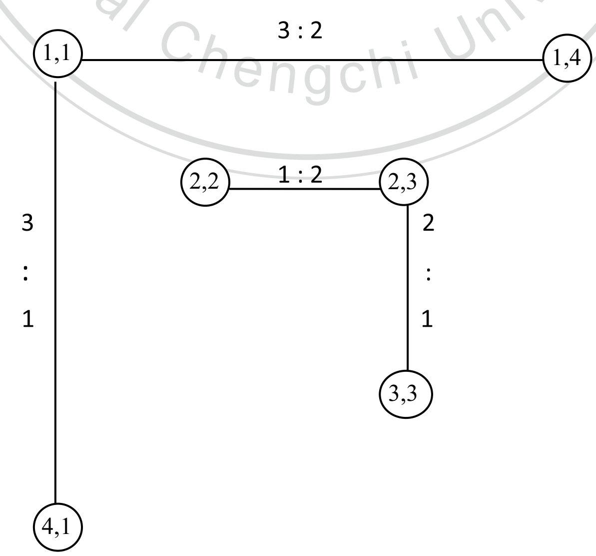

Fig. 3.2.1. Graphical representation for Example 3.2.1.

國立政治大學 National Chengchi University

The graphical representation approach can be readily extended to higher-dimensional case. To illustrate, consider, in the following example, three random variables X_1, X_2 and X_3 all taking values in {1, 2}, and let $\xi = \{p_{1|23}, p_{2|31}$ and $p_{3|12}\}$. Let $A = (A_{ijk})$, $B = (B_{ijk})$ and $C = (C_{ijk})$, where

$$\begin{aligned} A_{ijk} &= p_{1|23}(X_1 = i | X_2 = j, X_3 = k), \\ B_{ijk} &= p_{2|31}(X_2 = j | X_3 = k, X_1 = i), \\ C_{ijk} &= p_{3|12}(X_3 = k | X_1 = i, X_2 = j), i, j, k = 1, 2. \end{aligned}$$

Example 3.2.2. For

$A =$

	$X_1 = 1$	$X_1 = 2$
$X_2 = 1, X_3 = 1$	0.1	0.9
$X_2 = 1, X_3 = 2$	0.9	0.1
$X_2 = 2, X_3 = 1$	0.2	0.8
$X_2 = 2, X_3 = 2$	0.8	0.2

,

$B =$

	$X_2 = 1$	$X_2 = 2$
$X_3 = 1, X_1 = 1$	0.3	0.7
$X_3 = 1, X_1 = 2$	0.7	0.3
$X_3 = 2, X_1 = 1$	0.4	0.6
$X_3 = 2, X_1 = 2$	0.6	0.4

,

$C =$

	$X_3 = 1$	$X_3 = 2$
$X_1 = 1, X_2 = 1$	0.4	0.6
$X_1 = 1, X_2 = 2$	0.6	0.4
$X_1 = 2, X_2 = 1$	0.5	0.5
$X_1 = 2, X_2 = 2$	0.5	0.5

,

a graphical representation is given in Fig. 3.2.2. For the cycle (1, 1, 1) $\rightarrow$ (1, 1, 2) $\rightarrow$ (2, 1, 2) $\rightarrow$ (2, 1, 1) $\rightarrow$ (1, 1, 1), the left hand side of (3.2) for this cycle equals $6/4 \times 1/9 \times 5/5 \times 1/9 = 1/54 \neq 1$, implying incompatibility.

國立政治大學 National Chengchi University

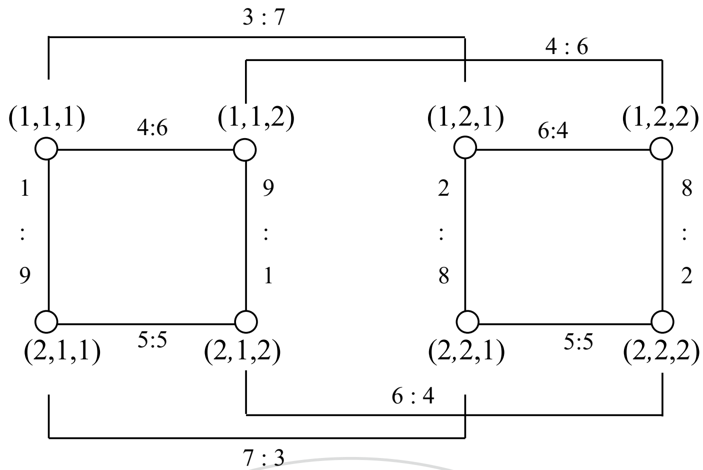

Fig. 3.2.2. Graphical representation for Example 3.2.2.

3.3 The relation between the ratio matrix approach and graphical representation approach.

Restricting attention to the case of two random variables, we have the following result.

Theorem 3.3.1. Suppose that A and B contain only positive elements. Then the following statements are equivalent:

(i) The ratio matrix C is of rank one.

(ii) For every cycle $\upsilon_0\upsilon_1\ldots\upsilon_l\upsilon_0$, we have

$$\prod_{i=0}^{l} r(\upsilon_{i+1}, \upsilon_i) = 1,$$

where $\upsilon_{l+1}\colon = \upsilon_0$.

Proof. To show (i) implies (ii), consider a cycle $(i_0, j_0) \to (i_1, j_0) \to (i_1, j_1) \to (i_2, j_1) \to \ldots \to (i_l, j_l) \to (i_{l+1}, j_l) \to (i_{l+1}, j_{l+1})$ where $i_{l+1} = i_0$ and $j_{l+1} = j_0$. Then the left hand side of (3.2) equals

$$\begin{aligned} &= \frac{A_{i_1,j_0}}{A_{i_0,j_0}} \times \frac{B_{i_1,j_1}}{B_{i_1,j_0}} \times \frac{A_{i_2,j_1}}{A_{i_1,j_1}} \times \ldots \times \frac{A_{i_{l+1},j_l}}{A_{i_l,j_l}} \times \frac{B_{i_{l+1},j_{l+1}}}{B_{i_{l+1},j_l}}, \\ &= \frac{A_{i_1,j_0}}{B_{i_1,j_0}} \times \frac{B_{i_1,j_1}}{A_{i_1,j_1}} \times \frac{A_{i_2,j_1}}{B_{i_2,j_1}} \times \ldots \times \frac{A_{i_{l+1},j_l}}{B_{i_{l+1},j_l}} \times \frac{B_{i_{l+1},j_{l+1}}}{A_{i_0,j_0}}, \end{aligned}$$

國立政治大學 National Chengchi University

$$= \frac{C_{i_1,j_0}}{C_{i_1,j_1}} \times \frac{C_{i_2,j_1}}{C_{i_2,j_2}} \times \ldots \times \frac{C_{i_l,j_{l-1}}}{C_{i_l,j_l}} \times \frac{C_{i_{l+1},j_l}}{C_{i_{l+1},j_{l+1}}}.$$

Since the ratio matrix C is of rank one, there exist two vectors $\tau = (\tau_1, \ldots, \tau_I)$ and $\eta = (\eta_1, \ldots, \eta_J)$ such that $C_{ij} = \tau_i \eta_j$ for all i, j. So

$$\begin{aligned}
&\frac{C_{i_1,j_0}}{C_{i_1,j_1}} \times \frac{C_{i_2,j_1}}{C_{i_2,j_2}} \times \ldots \times \frac{C_{i_l,j_{l-1}}}{C_{i_l,j_l}} \times \frac{C_{i_{l+1},j_l}}{C_{i_{l+1},j_{l+1}}} \\
&= \frac{\tau_{i_1}\eta_{j_0}}{\tau_{i_1}\eta_{j_1}} \times \frac{\tau_{i_2}\eta_{j_1}}{\tau_{i_2}\eta_{j_2}} \times \ldots \times \frac{\tau_{i_l}\eta_{j_{l-1}}}{\tau_{i_l}\eta_{j_l}} \times \frac{\tau_{i_{l+1}}\eta_{j_l}}{\tau_{i_{l+1}}\eta_{j_{l+1}}}, \\
&= \frac{\tau_{i_0}\tau_{i_1}\ldots\tau_{i_l}}{\tau_{i_0}\tau_{i_1}\ldots\tau_{i_l}} \times \frac{\eta_{j_0}\eta_{j_1}\ldots\eta_{j_l}}{\eta_{j_{l+1}}\eta_{j_1}\ldots\eta_{j_l}}, \\
&= 1.
\end{aligned}$$

To show (ii) implies (i), consider a cycle $(i_0, j_0) \to (i_1, j_0) \to (i_1, j_1) \to (i_2, j_1) \to \ldots \to (i_l, j_l) \to (i_{l+1}, j_l) \to (i_{l+1}, j_{l+1})$ where $i_{l+1} = i_0$ and $j_{l+1} = j_0$. Then by (3.2),

$$\frac{A_{i_1,j_0}}{A_{i_0,j_0}} \times \frac{B_{i_1,j_1}}{B_{i_1,j_0}} \times \frac{A_{i_2,j_1}}{A_{i_1,j_1}} \times \ldots \times \frac{A_{i_{l+1},j_l}}{A_{i_l,j_l}} \times \frac{B_{i_{l+1},j_{l+1}}}{B_{i_{l+1},j_l}} = 1,$$

which implies that

$$\begin{aligned}
1 &= \frac{A_{i_1,j_0}}{B_{i_1,j_0}} \times \frac{B_{i_1,j_1}}{A_{i_1,j_1}} \times \frac{A_{i_2,j_1}}{B_{i_2,j_1}} \times \ldots \times \frac{A_{i_{l+1},j_l}}{B_{i_{l+1},j_l}} \times \frac{B_{i_{l+1},j_{l+1}}}{A_{i_0,j_0}} \\
&= \frac{C_{i_1,j_0}}{C_{i_1,j_1}} \times \frac{C_{i_2,j_1}}{C_{i_2,j_2}} \times \ldots \times \frac{C_{i_l,j_{l-1}}}{C_{i_l,j_l}} \times \frac{C_{i_{l+1},j_l}}{C_{i_{l+1},j_{l+1}}}.
\end{aligned}$$

Therefore, for every cycle $(i_a, j_b) \to (i_{a+1}, j_b) \to (i_{a+1}, j_{b+1}) \to (i_a, j_{b+1}) \to (i_a, j_b)$, we have

$$\frac{C_{i_{a+1},j_b}}{C_{i_{a+1},j_{b+1}}} \times \frac{C_{i_a,j_{b+1}}}{C_{i_a,j_b}} = 1,$$

$$\frac{C_{i_a,j_b}}{C_{i_{a+1},j_b}} = \frac{C_{i_a,j_{b+1}}}{C_{i_{a+1},j_{b+1}}}.$$

So the ratio matrix C is of rank one. The proof is complete.

Example 3.3.1. (Example 2.1.2 continued) For

$$A = \begin{pmatrix} 1/3 & 3/8 \\ 2/3 & 5/8 \end{pmatrix} \qquad B = \begin{pmatrix} 1/4 & 3/4 \\ 2/7 & 5/7 \end{pmatrix},$$

a graphical representation is given in Fig. 3.3.1.

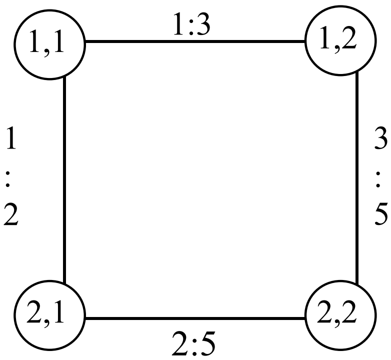

Fig. 3.3.1. Graphical representation for Example 3.3.1.

The corresponding ratio matrix is

$$C = \begin{pmatrix} 4/3 & 4/8 \\ 7/3 & 7/8 \end{pmatrix}.$$

Then, we have

$$\frac{4/3}{7/3} = \frac{4/8}{7/8},$$

$$\frac{\frac{\frac{1/3}{1/4}}{\frac{2/3}{2/7}}}{} \times \frac{\frac{\frac{3/4}{3/8}}{\frac{5/7}{5/8}}}{} = 1,$$

$$\frac{3/4}{1/4} \times \frac{5/8}{3/8} \times \frac{2/7}{5/7} \times \frac{1/3}{2/3} = 1,$$

$$\frac{3}{1} \times \frac{5}{3} \times \frac{2}{5} \times \frac{1}{2} = 1,$$

$r\left((1,2),(1,1)\right) \times r\left((2,2),(1,2)\right) \times r\left((2,1),(2,2)\right) \times r\left((1,1),(2,1)\right) = 1.$

Theorem 3.3.2. The following statements are equivalent:

(i) The ratio matrix C is irreducible.

(ii) The graph (V, E) is connected.

Proof. To show (i) implies (ii), suppose that the graph (V, E) is not connected and it can be decomposed into some $k > 1$ components (disjoint

國立政治大學 National Chengchi University

connected subgraphs), written $(V,E) = \bigcup_{i=1}^{k} (V_i, E_i)$, where each (V_i, E_i) is a connected subgraph and where the symbol $\cup$ denotes disjoint union. Obviously, if $a \neq b$, then, for two vertices $(i_a, j_a) \in V_a$ and $(i_b, j_b) \in V_b$, we have $i_a \neq i_b$ and $j_a \neq j_b$. So if a vertex (i_a, j_a) is in V_a, then all vertices on the same row or same column are in V_a. Let α_i = the set of row indices in $V_i \subset \{1, \ldots, I\}$ and β_i = the set of the column indices in $V_i \subset \{1, \ldots, J\}$, $i = 1, \ldots, k$. Then

$$\alpha_a \cap \alpha_b = \emptyset \text{ and } \beta_a \cap \beta_b = \emptyset \text{ for } a \neq b,$$

$$\bigcup_{i=1}^{k} \alpha_i = \{1, \ldots, I\} \text{ and } \bigcup_{i=1}^{k} \beta_i = \{1, \ldots, J\}.$$

Therefore, by interchanging some rows and/or columns, the ratio matrix C can be rearranged as a block diagonal matrix

$$\begin{pmatrix} T_1 & * & * & * \\ * & T_2 & * & * \\ * & * & \ddots & * \\ * & * & * & T_k \end{pmatrix},$$

where T_i corresponds to rows in α_i and columns in β_i.

By definition 2.2.3, the ratio matrix C is not irreducible. So (i) implies (ii).

To show (ii) implies (i). Suppose that the ratio matrix C is not irreducible. By Lemma 2.2.1, after interchanging some rows and/or columns, the ratio matrix C can be rearranged as an irreducible block diagonal matrix

$$\begin{pmatrix} T_1 & * & * & * \\ * & T_2 & * & * \\ * & * & \ddots & * \\ * & * & * & T_k \end{pmatrix},$$

where the diagonal block matrices $T_1, \ldots, T_k$ are irreducible and elements off these diagonal block matrices are all $*$. Let V_i denote the vertices in T_i.

Clearly, $V_1, \ldots, V_k$ are not connected. So the graph (V, E) is not connected and (ii) implies (i). The proof is complete.

Since the ratio matrix C is irreducible if and only if the graph $(V,\ E)$ is connected, we can combine Theorem 2.2.5 with Theorem 3.2.1 into the following theorem.

Theorem 3.3.3. Suppose that the graph (V, E) is connected. Then the following statements are equivalent.

(i) C has a unique rank one positive extension $\bar{C}$.

(ii) For every cycle $\upsilon_0\upsilon_1 \ldots \upsilon_l\upsilon_0$, we have

$$\prod_{i=0}^{l} r(\upsilon_{i+1}, \upsilon_i) = 1,$$

where $\upsilon_{l+1} := \upsilon_0$.

國立政治大學 National Chengchi University

4 Markov chain characterizations

4.1 Compatibility by the Gibbs sampler

Suppose that X and Y are two random variables taking values in $\{x_1, \ldots, x_I\}$ and $\{y_1, \ldots, y_J\}$, respectively. Consider two conditional probability matrices $A = (A_{ij}) = (P\{X = x_i|Y = y_j\})$ and $B = (B_{ij}) = (P\{Y = y_j|X = x_i\})$. Arnold, Castillo and Sarabia (1999) treated the matrix A' (transpose of A) as a transition matrix from Y to X and the matrix B as a transition matrix from X to Y, and then applied the Gibbs sampler to obtain stationary distributions. We describe the method as follows. For ease of discussion, we assume $A_{ij} > 0, B_{ij} > 0$ for all i, j.

We begin with an initial $X^{(1)}$. Conditioning on $X^{(1)}$, draw a $Y^{(1)}$ from B. Next, conditioning on $Y^{(1)}$, draw a $X^{(2)}$ from A'. So we have the following transitions:

$$X^{(1)} \xrightarrow{B} Y^{(1)} \xrightarrow{A'} X^{(2)} \xrightarrow{B} Y^{(2)} \xrightarrow{A'} X^{(3)} \xrightarrow{B} Y^{(3)} \to \ldots$$

This is a Markov chain, but not homogeneous. We then combine two transitions into a single one, so that we have the following two homogeneous chains:

$$X^{(1)} \to X^{(2)} \to X^{(3)} \ldots$$
$$Y^{(1)} \to Y^{(2)} \to Y^{(3)} \ldots$$

The transition matrix of the first chain is BA', and the transition matrix of the second chain is $A'B$. Each chain determines a stationary distribution, say $\tau = (\tau_i)$ and $\eta = (\eta_j)$ where $\tau_i = P(X = x_i)$ and $\eta_j = P(Y = y_j)$. That is, τ and η are solutions of the following systems:

$$\tau BA' = \tau, \tag{4.1}$$

$$\eta A'B = \eta. \tag{4.2}$$

Note that both transition matrices BA' and $A'B$ are irreducible, so that the respective stationary distributions τ and η are unique.

τ and B together determine a joint distribution $f(x_i, y_j) = \tau_i B_{ij}$, and

η and A together determine a joint distribution $g(x_i, y_j) = \eta_j A_{ij}$.

Let $f(x_i, +) = \sum_{y_j} f(x_i, y_j)$ and $f(+, y_j) = \sum_{x_i} f(x_i, y_j)$, so that $f(x_i, +)$ and $f(+, y_j)$ are the marginal distributions of f. Arnold, Castillo and Sarabia (1999) obtained the following theorem.

Theorem 4.1.1.

(i) Whether A and B are compatible or not, both joint distributions f and g have the same marginal distributions. That is, $f(x_i, +) = g(x_i, +)$ and $f(+, y_j) = g(+, y_j)$ for all i, j.

(ii) A and B are compatible if and only if the stationary distributions τ and η of the respective transition matrices BA' and $A'B$ satisfy $\tau_i B_{ij} = \eta_j A_{ij}$ for all i, j, i.e., $f(x_i, y_j) = g(x_i, y_j)$ for all i, j.

Proof:

(i) Note that

$$f(+, y_j) = \sum_{x_i} f(x_i, y_j) = \sum_i \tau_i B_{ij} = (\tau B)_j \, .$$

So the row vector τB corresponds to the Y-marginal distribution of f.

Similarly, $\eta A'$ corresponds to the X-marginal distribution of g.

Multiplying B to equation (4.1) yields

$$(\tau B) A' B = (\tau B),$$

which together with (4.2) implies

$$\tau B = \eta.$$

So the Y-marginal distribution of $f = \tau B = \eta =$ the Y-marginal distribution of g.

國立政治大學 National Chengchi University

Multiplying A' to equation (4.2) yields

$$(\eta A')BA' = (\eta A').$$

From equation (4.1), we have

$$\eta A' = \tau.$$

So the X-marginal distribution of $g = \eta A' = \tau =$ the X-marginal distribution of f. This proves that both joint distributions have the same marginal distributions.

(ii) Suppose that A and B are compatible, implying that there exists a joint distribution $h(x_i, y_j)$ such that

$$\frac{h(x_i, y_j)}{h(+, y_j)} = A_{ij} \text{ and } \frac{h(x_i, y_j)}{h(x_i, +)} = B_{ij}.$$

Let

$$h_X = (h(x_1, +), \ldots, h(x_I, +)) \text{ and } h_Y = (h(+, y_1), \ldots, h(+, y_J)),$$

which correspond to the X- and Y-marginal distributions of h.

So

$$h_X B = h_Y, \tag{4.3}$$
$$h_Y A' = h_X. \tag{4.4}$$

Multiplying A' to equation (4.3) and B to equation (4.4) yields

$$h_X BA' = h_Y A' = h_X,$$
$$h_Y A' B = h_X B = h_Y.$$

From (4.1) and (4.2), we have

$$\tau = h_X \text{ and } \eta = h_Y.$$

It follows that $f(x_i, y_j) = g(x_i, y_j) = h(x_i, y_j)$ for all i, j.

Conversely, suppose that $f(x_i, y_j) = g(x_i, y_j)$ for all i, j. Since A is the conditional distribution of X given Y under g and B is the conditional distribution of Y given X under f, it follows that $f = g$ has A and B as its two conditional distributions. This proves that A and B are compatible.

Example 4.1.1. Consider two conditional distribution matrices:

$$A = \begin{bmatrix} 0.4 & 0.2 \\ 0.6 & 0.8 \end{bmatrix}, \qquad B = \begin{bmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{bmatrix}.$$

Then, we have

$$BA' = \begin{bmatrix} 0.38 & 0.62 \\ 0.26 & 0.74 \end{bmatrix}, \qquad A'B = \begin{bmatrix} 0.54 & 0.46 \\ 0.42 & 0.58 \end{bmatrix}.$$

Solving

$$\tau BA' = \tau \text{ and } \eta A'B = \eta,$$

yields

$$\tau = (0.29546, 0.70454),$$
$$\eta = (0.47727, 0.52273).$$

τ and B together determine a joint distribution

$$(f(x_i, y_j)) = (\tau_i B_{ij}) = \begin{bmatrix} 0.26591 & 0.02955 \\ 0.21136 & 0.49318 \end{bmatrix},$$

while η and A together determine a joint distribution

$$(g(x_i, y_j)) = (\eta_j A_{ij}) = \begin{bmatrix} 0.19091 & 0.10455 \\ 0.28636 & 0.41818 \end{bmatrix}.$$

The two joint distributions are different, so A and B are incompatible. However, they have the same marginal distributions, τ and η.

國立政治大學 National Chengchi University

Arnold, Castillo, Sarabia (1999) only considered Markov chain characterization involving two random variables. We now consider the three-dimensional case where X, Y and Z are discrete random variables with I, J and K possible values, respectively. Three conditional distributions are given by

$$\begin{aligned} A_{ijk} &= P(X = x_i | Y = y_j, Z = z_k), \\ B_{ijk} &= P(Y = y_j | X = x_i, Z = z_k), \\ C_{ijk} &= P(Z = z_k | X = x_i, Y = y_j). \end{aligned}$$

Again for ease of discussion, we assume A_{ijk}, B_{ijk} and C_{ijk} are all positive.

We generate a Markov chain $X^{(1)}, Y^{(1)}, Z^{(1)}, X^{(2)}, Y^{(2)}, Z^{(2)}, \ldots$ as follows.

We start with $(X^{(1)}, Y^{(1)})$. Then generate $Z^{(1)}$ using C together with $(X^{(1)}, Y^{(1)})$. Thus we move from $(X^{(1)}, Y^{(1)})$ to $(Y^{(1)}, Z^{(1)})$. Next, we generate $X^{(2)}$ using A together with $(Y^{(1)}, Z^{(1)})$, resulting in a movement from $(Y^{(1)}, Z^{(1)})$ to $(Z^{(1)}, X^{(2)})$. Note that in each transition, one of the two components remains the same. So we have the following transitions

$$(X^{(1)}, Y^{(1)}) \to (Y^{(1)}, Z^{(1)}) \to (Z^{(1)}, X^{(2)}) \to (X^{(2)}, Y^{(2)}) \to (Y^{(2)}, Z^{(2)}) \to \ldots$$

This is a Markov chain, but not homogeneous. We then combine three transitions into a single one, so that we have three homogeneous chains,

$$\begin{aligned} &(X^{(1)}, Y^{(1)}) \to (X^{(2)}, Y^{(2)}) \to (X^{(3)}, Y^{(3)}) \to \ldots \\ &(Y^{(1)}, Z^{(1)}) \to (Y^{(2)}, Z^{(2)}) \to (Y^{(3)}, Z^{(3)}) \to \ldots \\ &(Z^{(1)}, X^{(2)}) \to (Z^{(2)}, X^{(3)}) \to (Z^{(3)}, X^{(4)}) \to \ldots \end{aligned}$$

Let $\bar{A}$ be the transition matrix from (Y, Z) to (Z, X):

$$\bar{A}((j,k),(h,i)) = P(Z = z_h, X = x_i | Y = y_j, Z = z_k) = \begin{cases} A_{ijk} & \text{if } h = k \\ 0 & \text{if } h \neq k, \end{cases}$$

$\bar{B}$ the transition matrix from (Z, X) to (X, Y):

$$\bar{B}((k,i),(h,j)) = P(X = x_h, Y = y_j | Z = z_k, X = x_i) = \begin{cases} B_{ijk} & \text{if } h = i \\ 0 & \text{if } h \neq i, \end{cases}$$

and $\bar{C}$ the transition matrix from (X, Y) to (Y, Z):

$$\bar{C}((i,j),(h,k)) = P(Y = y_h, Z = z_k | X = x_i, Y = y_j) = \begin{cases} C_{ijk} & \text{if } h = j \\ 0 & \text{if } h \neq j. \end{cases}$$

The transition matrix of the first chain is $\bar{C}\bar{A}\bar{B}$, the transition matrix of the second chain is $\bar{A}\bar{B}\bar{C}$ and the transition matrix of the third chain is $\bar{B}\bar{C}\bar{A}$. Each chain has a unique stationary distribution, say $\tau = (\tau(i,j))$ of dimension IJ, $\eta = (\eta(j,k))$ of dimension JK and $\theta = (\theta(k,i))$ of dimension KI. That is, τ, η and θ satisfy

$$\tau\bar{C}\bar{A}\bar{B} = \tau, \tag{4.5}$$

$$\eta\bar{A}\bar{B}\bar{C} = \eta, \tag{4.6}$$

$$\theta\bar{B}\bar{C}\bar{A} = \theta. \tag{4.7}$$

國立政治大學 National Chengchi University

τ and C together determine a joint distribution, $f(x_i, y_j, z_k) = \tau(i,j)C_{ijk}$, η and A together determine a joint distribution, $g(x_i, y_j, z_k) = \eta(j,k)A_{ijk}$, and θ and B together determine a joint distribution, $h(x_i, y_j, z_k) = \theta(k,i)B_{ijk}$.

We have the following result.

Theorem 4.1.2

(i) The (Y, Z)-distribution under f is the same as that under g, the (X, Z)-distribution under g is the same as that under h, and the (X, Y)-distribution under h is the same as that under f. That is,

$$f(+, y_j, z_k) = g(+, y_j, z_k) \text{ for all } j, k,$$
$$g(x_i, +, z_k) = h(x_i, +, z_k) \text{ for all } i, k,$$
$$h(x_i, y_j, +) = f(x_i, y_j, +) \text{ for all } i, j.$$

Consequently, f, g and h have the same X-, Y- and Z-marginal distributions.

(ii) A, B and C are compatible if and only if the stationary distributions τ, η and θ of the respective transition matrices $\bar{C}\bar{A}\bar{B}$, $\bar{A}\bar{B}\bar{C}$ and

$\bar{B}\bar{C}\bar{A}$ satisfy $\tau(i,j)C_{ijk} = \eta(j,k)A_{ijk} = \theta(k,i)B_{ijk}$ for all i,j,k, i.e., $f(x_i,y_j,z_k) = g(x_i,y_j,z_k) = h(x_i,y_j,z_k)$ for all i,j,k.

Proof:

(i) The distribution of (YZ) under f is

$$\begin{aligned} f(+,y_j,z_k) &= \sum_i \tau(i,j)C_{ijk} \\ &= \sum_{i,h} \tau\,(i,h)\bar{C}(\,(i,h)\,,(j,k)) \\ &= \text{the } (j,k) \text{ component of } \tau\bar{C}. \end{aligned}$$

So $\tau\bar{C}$ corresponds to the (Y,Z)-distribution under f. Similarly, $\eta\bar{A}$ and $\theta\bar{B}$ correspond respectively to the (Z,X) and (X,Y) distribution under g and h. Multiplying $\bar{C}$ to equation (4.5) yields

$$(\tau\bar{C})\bar{A}\bar{B}\bar{C} = (\tau\bar{C}),$$

which together with (4.6) implies

$$\tau\bar{C} = \eta.$$

So the (Y,Z)-distribution under $f = \tau\bar{C} = \eta =$ the (Y,Z)-distribution under g. Multiplying $\bar{A}$ to equation (4.6) yields

$$(\eta\bar{A})\bar{B}\bar{C}\bar{A} = (\eta\bar{A}).$$

From equation (4.7), we have

$$\eta\bar{A} = \theta.$$

So the (Z,X)-distribution under $g = \eta\bar{A} = \theta =$ the (Z,X)-distribution under h. Multiplying $\bar{B}$ to equation (4.7) yields

$$(\theta\bar{B})\bar{C}\bar{A}\bar{B} = (\theta\bar{B}).$$

From equation (4.5), we have

$$\theta\bar{B} = \tau.$$

So the (X, Y)-distribution under $h = \theta\bar{B} = \tau =$ the (X, Y)-distribution under f.

We have shown that the (Y, Z)-distribution under f is the same as that under g, the (X, Z)-distribution under g is the same as that under h, and the (X, Y)-distribution under h is the same as that under f. That is,

$$\begin{aligned} f(+, y_j, z_k) &= g(+, y_j, z_k) \text{ for all } j, k, \\ g(x_i, +, z_k) &= h(x_i, +, z_k) \text{ for all } i, k, \\ h(x_i, y_j, +) &= f(x_i, y_j, +) \text{ for all } i, j. \end{aligned}$$

Consequently, f, g and h have the same X-, Y- and Z-marginal distributions.

(ii) Suppose that A, B and C are compatible, implying that there exists a joint distribution $d(x_i, y_j, z_k)$ such that

$$\begin{aligned} A_{ijk} &= d(x_i, y_j, z_k)/d(+, y_j, z_k), \\ B_{ijk} &= d(x_i, y_j, z_k)/d(x_i, +, z_k), \\ C_{ijk} &= d(x_i, y_j, z_k)/d(x_i, y_j, +). \end{aligned}$$

Let

$$\begin{aligned} d_{X,Y} &= (d(x_1, y_1, +), \ldots, d(x_I, y_J, +)), \\ d_{Y,Z} &= (d(+, y_1, z_1), \ldots, d(+, y_J, z_K)), \\ d_{Z,X} &= (d(x_1, +, z_1), \ldots, d(x_I, +, z_K)). \end{aligned}$$

So

$$d_{X,Y}\bar{C} = d_{Y,Z}, \tag{4.8}$$

$$d_{Y,Z}\bar{A} = d_{Z,X}, \tag{4.9}$$
$$d_{Z,X}\bar{B} = d_{X,Y}. \tag{4.10}$$

Multiplying $\bar{A}\bar{B}$ to equation (4.8), $\bar{B}\bar{C}$ to equation (4.9) and $\bar{C}\bar{A}$ to equation (4.10) yields

$$d_{X,Y}\bar{C}\bar{A}\bar{B} = d_{X,Y},$$
$$d_{Y,Z}\bar{A}\bar{B}\bar{C} = d_{Y,Z},$$
$$d_{Z,X}\bar{B}\bar{C}\bar{A} = d_{Z,X}.$$

From (4.5), (4.6) and (4.7), we have

$$\tau = d_{X,Y},$$
$$\eta = d_{Y,Z},$$
$$\theta = d_{Z,X}.$$

It follows that $f(x_i, y_j, z_k) = g(x_i, y_j, z_k) = h(x_i, y_j, z_k) = d(x_i, y_j, z_k)$ for all i, j, k.

Conversely, suppose $f(x_i, y_j, z_k) = g(x_i, y_j, z_k) = h(x_i, y_j, z_k)$ for all i, j, k. Since A is the conditional distribution of X given (Y, Z) under g, B is the conditional distribution of Y given (Z, X) under h and C is the conditional distribution of Z given (X, Y) under f, it follows that $f = g = h$ has A, B and C as its three conditional distributions. This proves that A, B and C are compatible.

Example 4.1.2. (Example 3.2.2 continued)

Consider three random variables X, Y and Z with possible values (x_1, x_2), (y_1, y_2) and (z_1, z_2), and three matrices A, B and C.

$A =$

	x_1,	x_2
y_1, z_1	0.1	0.9
y_1, z_2	0.9	0.1
y_2, z_1	0.2	0.8
y_2, z_2	0.8	0.2

國立政治大學 National Chengchi University

$$B = \begin{array}{c|cc} & y_1, & y_2 \\ \hline z_1, x_1 & 0.3 & 0.7 \\ z_1, x_2 & 0.7 & 0.3 \\ z_2, x_1 & 0.4 & 0.6 \\ z_2, x_2 & 0.6 & 0.4 \end{array}$$

$$C = \begin{array}{c|cc} & z_1, & z_2 \\ \hline x_1, y_1 & 0.4 & 0.6 \\ x_1, y_2 & 0.6 & 0.4 \\ x_2, y_1 & 0.5 & 0.5 \\ x_2, y_2 & 0.5 & 0.5 \end{array}$$

Suppose that our generation sequence is $X^{(1)}, Y^{(1)}, Z^{(1)}, X^{(2)}, Y^{(2)}, Z^{(2)} \dots$ Then $\bar{C}$ is the following transition matrix from (X, Y) to (Y, Z):

	y_1, z_1	y_1, z_2	y_2, z_1	y_2, z_2
x_1, y_1	0.4	0.6	0	0
x_1, y_2	0	0	0.6	0.4
x_2, y_1	0.5	0.5	0	0
x_2, y_2	0	0	0.5	0.5

$\bar{A}$ is the following transition matrix from (Y, Z) to (Z, X):

	z_1, x_1	z_1, x_2	z_2, x_1	z_2, x_2
y_1, z_1	0.1	0.9	0	0
y_1, z_2	0	0	0.9	0.1
y_2, z_1	0.2	0.8	0	0
y_2, z_2	0	0	0.8	0.2

$\bar{B}$ is the following transition matrix from (Z, X) to (X, Y):

	x_1, y_1	x_1, y_2	x_2, y_1	x_2, y_2
z_1, x_1	0.3	0.7	0	0
z_1, x_2	0	0	0.7	0.3
z_2, x_1	0.4	0.6	0	0
z_2, x_2	0	0	0.6	0.4

國立政治大學 National Chengchi University

Then $\bar{C}\bar{A}\bar{B}$ is the following transition matrix from (X, Y) to (X, Y)

	x_1, y_1	x_1, y_2	x_2, y_1	x_2, y_2
x_1, y_1	0.228	0.352	0.288	0.132
x_1, y_2	0.164	0.276	0.384	0.176
x_2, y_1	0.195	0.305	0.345	0.155
x_2, y_2	0.190	0.310	0.340	0.160

$\bar{A}\bar{B}\bar{C}$ is the following transition matrix from (Y, Z) to (Y, Z)

	y_1, z_1	y_1, z_2	y_2, z_1	y_2, z_2
y_1, z_1	0.327	0.333	0.177	0.163
y_1, z_2	0.174	0.246	0.344	0.236
y_2, z_1	0.304	0.316	0.204	0.176
y_2, z_2	0.188	0.252	0.328	0.232

$\bar{B}\bar{C}\bar{A}$ is the following transition matrix from (Z, X) to (Z, X)

	$z_1, x_1,$	z_1, x_2	z_2, x_1	z_2, x_2
z_1, x_1	0.096	0.444	0.386	0.074
z_1, x_2	0.065	0.435	0.435	0.065
z_2, x_1	0.088	0.432	0.408	0.072
z_2, x_2	0.070	0.430	0.430	0.070

Suppose that τ, η and θ satisfy the following systems:

$$\tau\bar{C}\bar{A}\bar{B} = \tau,$$
$$\eta\bar{A}\bar{B}\bar{C} = \eta,$$
$$\theta\bar{B}\bar{C}\bar{A} = \theta.$$

We find

$$\tau = (0.1910322, 0.3058966, 0.3452520, 0.1578192),$$
$$\eta = (0.2490389, 0.2872453, 0.2624476, 0.2012682),$$
$$\theta = (0.0773934, 0.4340930, 0.4195354, 0.0689782).$$

From τ and C, we can determine a joint distribution $(f(x_i, y_j, z_k)) = (\tau(i,j)C_{ijk})$

$$
\begin{aligned}
&= (f(x_1, y_1, z_1), f(x_1, y_1, z_2), f(x_1, y_2, z_1), f(x_1, y_2, z_2), \\
&\quad f(x_2, y_1, z_1), f(x_2, y_1, z_2), f(x_2, y_2, z_1), f(x_2, y_2, z_2)) \\
&= (0.07641288, 0.1146193, 0.183538, 0.1223586, \\
&\quad 0.172626, 0.172626, 0.0789096, 0.0789096).
\end{aligned}
$$

Then

the X-marginal distribution of

$f = (f(x_1, +, +),\ f(x_2, +, +)) = (0.4969288,\ 0.5030712)$,

the Y-marginal distribution of

$f = (f(+, y_1, +),\ f(+, y_2, +)) = (0.5362842,\ 0.4637158)$,

the Z-marginal distribution of

$f = (f(+, +, z_1),\ f(+, +, z_2)) = (0.5114865,\ 0.4885135)$.

From η and A, we can determine a joint distribution $(g(x_i, y_j, z_k)) = (\eta(j,k)A_{ijk})$

$$
\begin{aligned}
&= (g(x_1, y_1, z_1), g(x_1, y_1, z_2), g(x_1, y_2, z_1), g(x_1, y_2, z_2), \\
&\quad g(x_2, y_1, z_1), g(x_2, y_1, z_2), g(x_2, y_2, z_1), g(x_2, y_2, z_2)) \\
&= (0.02490389, 0.2585208, 0.05248952, 0.1610146, \\
&\quad 0.224135, 0.02872453, 0.2099581, 0.04025364).
\end{aligned}
$$

Then

the X-marginal distribution of

$g = (g(x_1, +, +),\ g(x_2, +, +)) = (0.4969288,\ 0.5030712)$,

the Y-marginal distribution of

$g = (g(+, y_1, +),\ g(+, y_2, +)) = (0.5362842,\ 0.4637158)$,

the Z-marginal distribution of

$g = (g(+, +, z_1),\ g(+, +, z_2)) = (0.5114865,\ 0.4885135)$.

From θ and B, we can determine a joint distribution $(h(x_i, y_j, z_k)) = (\theta(k,i)B_{ijk})$

$$
\begin{aligned}
&= \left(h\left(x_1,y_1,z_1\right), h\left(x_1,y_1,z_2\right), h\left(x_1,y_2,z_1\right), h\left(x_1,y_2,z_2\right),\right. \\
&\quad \left. h\left(x_2,y_1,z_1\right), h\left(x_2,y_1,z_2\right), h\left(x_2,y_2,z_1\right), h\left(x_2,y_2,z_2\right)\right) \\
&= (0.02321802, 0.1678142, 0.05417538, 0.2517212, \\
&\quad 0.3038651, 0.04138691, 0.1302279, 0.02759127).
\end{aligned}
$$

Then

the X-marginal distribution of

$h = (h(x_1, +, +),\ h(x_2+, +)) = (0.4969288,\ 0.5030712)$,

the Y-marginal distribution of

$h = (h(+, y_1, +),\ h(+, y_2, +)) = (0.5362842,\ 0.4637158)$,

the Z-marginal distribution of

$h = (h(+, +, z_1),\ h(+, +, z_2)) = (0.5114864,\ 0.4885136)$.

So that

$$
\begin{aligned}
f(x_i,+,+) &= g(x_i,+,+) = h(x_i,+,+) \text{ for all } i, \\
f(+,y_j,+) &= g(+,y_j,+) = h(+,y_j,+) \text{ for all } j, \\
f(+,+,z_k) &= g(+,+,z_k) = h(+,+,z_k) \text{ for all } k.
\end{aligned}
$$

All these three joint distributions are different, so A, B and C are incompatible. However, they have the same marginal distributions.

In fact, we can consider an alternative Markov chain $X^{(1)}, Z^{(1)}, Y^{(1)}, X^{(2)}$, $Z^{(2)}, Y^{(2)} \ldots$ Specifically, start with $(X^{(1)}, Z^{(1)})$, then we generate $Y^{(1)}$ using B. Thus we move from $(X^{(1)}, Z^{(1)})$ to $(Z^{(1)}, Y^{(1)})$. Next, we generate $X^{(2)}$ using A together with $(Z^{(1)}, Y^{(1)})$, resulting in a movement from $(Z^{(1)}, Y^{(1)})$, to $(Y^{(1)}, X^{(2)})$. Note that in each transition, one of the two components remains the same. So we have the following transitions

$$
(X^{(1)}, Z^{(1)}) \to (Z^{(1)}, Y^{(1)}) \to (Y^{(1)}, X^{(2)}) \to (X^{(2)}, Z^{(2)}) \to (Z^{(2)}, Y^{(2)}) \to \ldots
$$

國立政治大學 National Chengchi University

This is a Markov chain, but not homogeneous. We then combine three transitions into a single one, so that we have three homogeneous chains,

$$(X^{(1)}, Z^{(1)}) \to (X^{(2)}, Z^{(2)}) \to (X^{(3)}, Z^{(3)}) \to \dots$$
$$(Z^{(1)}, Y^{(1)}) \to (Z^{(2)}, Y^{(2)}) \to (Z^{(3)}, Y^{(3)}) \to \dots$$
$$(Y^{(1)}, X^{(2)}) \to (Y^{(2)}, X^{(3)}) \to (Y^{(3)}, X^{(4)}) \to \dots$$

Let $\tilde{A}$ be the following transition matrix from (Z, Y) to (Y, X):

$$\tilde{A}((k,j),(h,i)) = P(Y = y_h, X = x_i | Z = z_k, Y = y_j) = \begin{cases} A_{ijk} & \text{if } h = j \\ 0 & \text{if } h \neq j. \end{cases}$$

$\tilde{B}$ be the following transition matrix from (X, Z) to (Z, Y):

$$\tilde{B}((i,k),(h,j)) = P(Z = z_h, Y = y_j | X = x_i, Z = z_k) = \begin{cases} B_{ijk} & \text{if } h = k \\ 0 & \text{if } h \neq k. \end{cases}$$

$\tilde{C}$ be the following transition matrix from (Y, X) to (X, Z):

$$\tilde{C}((j,i),(h,k)) = P(X = x_h, Z = z_k | Y = y_j, X = x_i) = \begin{cases} C_{ijk} & \text{if } h = i \\ 0 & \text{if } h \neq i. \end{cases}$$

The transition matrix of the first chain is $\tilde{B}\tilde{A}\tilde{C}$, the transition matrix of the second chain is $\tilde{A}\tilde{C}\tilde{B}$ and the transition matrix of the third chain is $\tilde{C}\tilde{B}\tilde{A}$. Each chain has a unique stationary distribution, say $\tilde{\tau} = (\tilde{\tau}(i,k))$ of dimension IK, $\tilde{\eta} = (\tilde{\eta}(k,j))$ of dimension KJ and $\tilde{\theta} = (\tilde{\theta}(j,i))$ of dimension JI. That is, $\tilde{\tau}, \tilde{\eta}$ and $\tilde{\theta}$ are solutions of the following systems:

$$\tilde{\tau}\tilde{B}\tilde{A}\tilde{C} = \tilde{\tau}, \tag{4.11}$$
$$\tilde{\eta}\tilde{A}\tilde{C}\tilde{B} = \tilde{\eta}, \tag{4.12}$$
$$\tilde{\theta}\tilde{C}\tilde{B}\tilde{A} = \tilde{\theta}, \tag{4.13}$$

$\tilde{\tau}$ and B together determine a joint distribution, $\tilde{f}(x_i, y_j, z_k) = \tilde{\tau}(i,k)B_{ijk}$, $\tilde{\eta}$ and A together determine a joint distribution, $\tilde{g}(x_i, y_j, z_k) = \tilde{\eta}(k,j)A_{ijk}$, and $\tilde{\theta}$ and C together determine a joint distribution, $\tilde{h}(x_i, y_j, z_k) = \tilde{\theta}(j,i)C_{ijk}$.

Following the proof of Theorem 4.1.2, we obtain the following theorem.

Theorem 4.1.3

(i) The (Y, Z)-distribution under $\tilde{f}$ is the same as that under $\tilde{g}$, the (X, Y)-distribution under $\tilde{g}$ is the same as that under $\tilde{h}$, and the (X, Z)-distribution under $\tilde{h}$ is the same as that under $\tilde{f}$. That is,

$$\tilde{f}(+, y_j, z_k) = \tilde{g}(+, y_j, z_k) \text{ for all } j, k,$$
$$\tilde{g}(x_i, y_j, +) = \tilde{h}(x_i, y_j, +) \text{ for all } i, j,$$
$$\tilde{h}(x_i, +, z_k) = \tilde{f}(x_i, +, z_k) \text{ for all } i, k.$$

Consequently, $\tilde{f}, \tilde{g}$ and $\tilde{h}$ have the same X-, Y- and Z-marginal distributions.

(ii) A, B and C are compatible if and only if the stationary distributions $\tilde{\tau}, \tilde{\eta}$ and $\tilde{\theta}$ of respective transition matrices $\tilde{B}\tilde{A}\tilde{C}$, $\tilde{A}\tilde{C}\tilde{B}$ and $\tilde{C}\tilde{B}\tilde{A}$ satisfy $\tilde{\tau}(i,k)B_{ijk} = \tilde{\eta}(k,j)A_{ijk} = \tilde{\theta}(j,i)C_{ijk}$ for all i, j, k. That is $\tilde{f}(x_i, y_j, z_k) = \tilde{g}(x_i, y_j, z_k) = \tilde{h}(x_i, y_j, z_k)$ for all i, j, k.

Example 4.1.3. (Example 4.1.2 continued)

$A =$

	x_1,	x_2
z_1, y_1	0.1	0.9
z_1, y_2	0.2	0.8
z_2, y_1	0.9	0.1
z_2, y_2	0.8	0.2

$B =$

	y_1,	y_2
x_1, z_1	0.3	0.7
x_1, z_2	0.4	0.6
x_2, z_1	0.7	0.3
x_2, z_2	0.6	0.4

國立政治大學 National Chengchi University

$C =$

	$z_1,$	z_2
y_1, x_1	0.4	0.6
y_1, x_2	0.5	0.5
y_2, x_1	0.6	0.4
y_2, x_2	0.5	0.5

Suppose that our generation sequence is $X^{(1)}$, $Z^{(1)}, Y^{(1)}, X^{(2)}, Z^{(2)}, Y^{(2)} \ldots$ Then $\tilde{C}$ is the following transition matrix from (Y, X) to (X, Z):

	x_1, z_1	x_1, z_2	x_2, z_1	x_2, z_2
y_1, x_1	0.4	0.6	0	0
y_1, x_2	0	0	0.5	0.5
y_2, x_1	0.6	0.4	0	0
y_2, x_2	0	0	0.5	0.5

$\tilde{A}$ is the following transition matrix from (Z, Y) to (Y, X):

	y_1, x_1	y_1, x_2	y_2, x_1	y_2, x_2
z_1, y_1	0.1	0.9	0	0
z_1, y_2	0	0	0.2	0.8
z_2, y_1	0.9	0.1	0	0
z_2, y_2	0	0	0.8	0.2

$\tilde{B}$ is the following transition matrix from (X, Z) to (Z, Y):

	z_1, y_1	z_1, y_2	z_2, y_1	z_2, y_2
x_1, z_1	0.3	0.7	0	0
x_1, z_2	0	0	0.4	0.6
x_2, z_1	0.7	0.3	0	0
x_2, z_2	0	0	0.6	0.4

國立政治大學 National Chengchi University

Then $\tilde{B}\tilde{A}\tilde{C}$ is the following transition matrix from (X, Z) to (X, Z)

	x_1, z_1	x_1, z_2	x_2, z_1	x_2, z_2
x_1, z_1	0.096	0.074	0.415	0.415
x_1, z_2	0.432	0.408	0.080	0.080
x_2, z_1	0.064	0.066	0.435	0.435
x_2, z_2	0.408	0.452	0.070	0.070

$\tilde{A}\tilde{C}\tilde{B}$ is the following transition matrix from (Z, Y) to (Z, Y)

	z_1, y_1	z_1, y_2	z_2, y_1	z_2, y_2
z_1, y_1	0.327	0.163	0.294	0.216
z_1, y_2	0.316	0.204	0.272	0.208
z_2, y_1	0.143	0.267	0.246	0.344
z_2, y_2	0.214	0.366	0.188	0.232

$\tilde{C}\tilde{B}\tilde{A}$ is the following transition matrix from (Y, X) to (Y, X)

	$y_1, x_1,$	y_1, x_2	y_2, x_1	y_2, x_2
y_1, x_1	0.228	0.132	0.344	0.296
y_1, x_2	0.305	0.345	0.190	0.160
y_2, x_1	0.162	0.178	0.276	0.384
y_2, x_2	0.305	0.345	0.190	0.160

Suppose that $\tilde{\tau}, \tilde{\eta}$ and $\tilde{\theta}$ satisfy the following systems:

$$\tilde{\tau}\tilde{B}\tilde{A}\tilde{C} = \tilde{\tau},$$

$$\tilde{\eta}\tilde{A}\tilde{C}\tilde{B} = \tilde{\eta},$$

$$\tilde{\theta}\tilde{C}\tilde{B}\tilde{A} = \tilde{\theta}.$$

We find

$\tilde{\tau} = (0.25, 0.25, 0.25, 0.25),$

$\tilde{\eta} = (0.25, 0.25, 0.25, 0.25),$

$\tilde{\theta} = (0.25, 0.25, 0.25, 0.25).$

國立政治大學 National Chengchi University

From $\tilde{\tau}$ and B, we can determine a joint distribution $(\tilde{f}(x_i, y_j, z_k)) = (\tilde{\tau}(i,k)B_{ijk})$

$$\begin{aligned} &= (\tilde{f}\left(x_1, y_1, z_1\right), \tilde{f}\left(x_1, y_1, z_2\right), \tilde{f}\left(x_1, y_2, z_1\right), \tilde{f}\left(x_1, y_2, z_2\right) \\ &\quad \tilde{f}\left(x_2, y_1, z_1\right), \tilde{f}\left(x_2, y_1, z_2\right), \tilde{f}\left(x_2, y_2, z_1\right), \tilde{f}(x_2, y_2, z_2)) \\ &= (0.075, 0.100, 0.175, 0.150, 0.175, 0.150, 0.075, 0.100). \end{aligned}$$

Then

the X-marginal distribution of $\tilde{f} = (\tilde{f}(x_1, +, +), \tilde{f}(x_2, +, +)) = (0.5, 0.5)$,

the Y-marginal distribution of $\tilde{f} = (\tilde{f}(+, y_1, +), \tilde{f}(+, y_2, +)) = (0.5, 0.5)$,

the Z-marginal distribution of $\tilde{f} = (\tilde{f}(+, +, z_1), \tilde{f}(+, +, z_2)) = (0.5, 0.5)$.

From $\tilde{\eta}$ and A, we can determine a joint distribution $(\tilde{g}(x_i, y_j, z_k)) = (\tilde{\eta}(k,j)A_{ijk})$

$$\begin{aligned} &= (\tilde{g}\left(x_1, y_1, z_1\right), \tilde{g}\left(x_1, y_1, z_2\right), \tilde{g}\left(x_1, y_2, z_1\right), \tilde{g}\left(x_1, y_2, z_2\right) \\ &\quad \tilde{g}\left(x_2, y_1, z_1\right), \tilde{g}\left(x_2, y_1, z_2\right), \tilde{g}\left(x_2, y_2, z_1\right), \tilde{g}(x_2, y_2, z_2)) \\ &= (0.025, 0.225, 0.050, 0.200, 0.225, 0.025, 0.200, 0.050). \end{aligned}$$

Then

the X-marginal distribution of $\tilde{g} = (\tilde{g}(x_1, +, +), \tilde{g}(x_2, +, +)) = (0.5, 0.5)$,

the Y-marginal distribution of $\tilde{g} = (\tilde{g}(+, y_1, +), \tilde{g}(+, y_2, +)) = (0.5, 0.5)$,

the Z-marginal distribution of $\tilde{g} = (\tilde{g}(+, +, z_1), \tilde{g}(+, +, z_2)) = (0.5, 0.5)$.

From $\tilde{\theta}$ and C, we can determine a joint distribution $(\tilde{h}(x_i, y_j, z_k)) = (\tilde{\theta}(j,i)B_{ijk})$

$$\begin{aligned} &= (\tilde{h}\left(x_1, y_1, z_1\right), \tilde{h}\left(x_1, y_1, z_2\right), \tilde{h}\left(x_1, y_2, z_1\right), \tilde{h}\left(x_1, y_2, z_2\right) \\ &\quad \tilde{h}\left(x_2, y_1, z_1\right), \tilde{h}\left(x_2, y_1, z_2\right), \tilde{h}\left(x_2, y_2, z_1\right), \tilde{h}(x_2, y_2, z_2)) \\ &= (0.100, 0.150, 0.150, 0.100, 0.125, 0.125, 0.125, 0.125). \end{aligned}$$

Then

the X-marginal distribution of $\tilde{h} = (\tilde{h}(x_1, +, +), \tilde{h}(x_2 +, +)) = (0.5, 0.5)$,

the Y-marginal distribution of $\tilde{h} = (\tilde{h}(+, y_1, +), \tilde{h}(+, y_2, +)) = (0.5, 0.5)$,

國立政治大學 National Chengchi University

the Z-marginal distribution of $\tilde{h} = (\tilde{h}(+,+,z_1), \tilde{h}(+,+,z_2)) = (0.5, 0.5)$. So that

$$\begin{aligned}
\tilde{f}(x_i,+,+) = \tilde{g}(x_i,+,+) = \tilde{h}(x_i,+,+) \quad &\text{for all} \quad i, \\
\tilde{f}(+,y_j,+) = \tilde{g}(+,y_j,+) = \tilde{h}(+,y_j,+) \quad &\text{for all} \quad j, \\
\tilde{f}(+,+,z_k) = \tilde{g}(+,+,z_k) = \tilde{h}(+,+,z_k) \quad &\text{for all} \quad k.
\end{aligned}$$

Although all these three joint distributions are different, they have the same marginal distributions.

4.2 Simulations

We applied the Gibbs sampler to generate simulations for Example 4.1.1. The results are given in Table 4.2.1 where the second column is the empirical distribution of $(X^{(i)}, Y^{(i)})$, $i = 1, \ldots, n$ while the third column is the empirical distribution of $(Y^{(i)}, X^{(i)})$, $i = 1, \ldots, n$.

Table 4.2.1: Empirical distributions for the Gibbs sampler in Example 4.1.1

Sample size n	Sampling sequence $X^{(1)} \to Y^{(1)} \to X^{(2)} \to Y^{(2)} \ldots$	Sampling sequence $Y^{(1)} \to X^{(1)} \to Y^{(2)} \to X^{(2)} \ldots$
1000	$\begin{pmatrix} 0.26600 & 0.03800 \\ 0.20600 & 0.49000 \end{pmatrix}$	$\begin{pmatrix} 0.21400 & 0.10400 \\ 0.27000 & 0.41200 \end{pmatrix}$
10,000	$\begin{pmatrix} 0.27860 & 0.02680 \\ 0.20380 & 0.49080 \end{pmatrix}$	$\begin{pmatrix} 0.18140 & 0.10700 \\ 0.29920 & 0.41240 \end{pmatrix}$
100,000	$\begin{pmatrix} 0.26238 & 0.02904 \\ 0.21116 & 0.49742 \end{pmatrix}$	$\begin{pmatrix} 0.18876 & 0.10678 \\ 0.28556 & 0.41890 \end{pmatrix}$
1,000,000	$\begin{pmatrix} 0.26662 & 0.02928 \\ 0.21114 & 0.49295 \end{pmatrix}$	$\begin{pmatrix} 0.19212 & 0.10445 \\ 0.28667 & 0.41675 \end{pmatrix}$
Stationary distribution	$\begin{pmatrix} 0.26591 & 0.02955 \\ 0.21136 & 0.49318 \end{pmatrix}$	$\begin{pmatrix} 0.19091 & 0.10455 \\ 0.28636 & 0.41818 \end{pmatrix}$

From Table 4.2.1, we find that the Gibbs sampler has two different empirical joint distributions, one based on $(X^{(i)}, Y^{i)})$, $i = 1, 2, \ldots, n$, and the other based on $(Y^{(i)},\ X^{(i)})$, $i = 1, 2, \ldots, n$. Each empirical joint distribution is very close to its stationary joint distribution when the sample size n is large.

We also applied the Gibbs sampler to generate simulations for Example 4.1.2. The results are given in Tables 4.2.2−4.2.4. Table 4.2.2 is the empirical distribution of $(X^{(i)}, Y^{(i)}, Z^{(i)})$, $i = 1, \ldots, n$, Table 4.2.3 is the empirical distribution of $(Y^{(i)}, Z^{(i)}, X^{(i)})$, $i = 1, \ldots, n$, and Table 4.2.4 is the empirical distribution of $(Z^{(i)}, X^{(i)}, Y^{(i)})$, $i = 1, \ldots, n$.

Table 4.2.2: Empirical distribution of $(X^{(i)}, Y^{(i)}, Z^{(i)})$ for the Gibbs sampler in Example 4.1.2

Sample size n	Sampling sequence $X^{(1)} \to Y^{(1)} \to Z^{(1)} \to X^{(2)} \to Y^{(2)} \to Z^{(2)} \ldots$
1000	$\begin{pmatrix} 0.07200 & 0.09400 & 0.18200 & 0.09200 \\ 0.20800 & 0.18200 & 0.09600 & 0.07400 \end{pmatrix}$
10,000	$\begin{pmatrix} 0.07740 & 0.11740 & 0.18880 & 0.12640 \\ 0.16720 & 0.17140 & 0.08060 & 0.07080 \end{pmatrix}$
100,000	$\begin{pmatrix} 0.07602 & 0.11520 & 0.18406 & 0.12286 \\ 0.17150 & 0.17018 & 0.07986 & 0.08032 \end{pmatrix}$
1,000,000	$\begin{pmatrix} 0.07636 & 0.11502 & 0.18346 & 0.12226 \\ 0.17279 & 0.17265 & 0.07890 & 0.07856 \end{pmatrix}$
Stationary distribution	$\begin{pmatrix} 0.07641 & 0.11462 & 0.18354 & 0.12236 \\ 0.17263 & 0.17263 & 0.07891 & 0.07891 \end{pmatrix}$

國立政治大學 National Chengchi University

Table 4.2.3: Empirical distribution of $(Y^{(i)}, Z^{(i)}, X^{(i)})$ for the Gibbs sampler in Example 4.1.2

Sample size n	Sampling sequence $Y^{(1)} \to Z^{(1)} \to X^{(1)} \to Y^{(2)} \to Z^{(2)} \to X^{(2)} \dots$
1000	$\begin{pmatrix} 0.02800 & 0.25200 & 0.05600 & 0.12400 \\ 0.23600 & 0.03600 & 0.22400 & 0.04400 \end{pmatrix}$
10,000	$\begin{pmatrix} 0.02540 & 0.25440 & 0.05820 & 0.16200 \\ 0.22300 & 0.02500 & 0.21240 & 0.03960 \end{pmatrix}$
100,000	$\begin{pmatrix} 0.02570 & 0.25268 & 0.05308 & 0.16290 \\ 0.22450 & 0.02902 & 0.21228 & 0.03984 \end{pmatrix}$
1,000,000	$\begin{pmatrix} 0.02489 & 0.25862 & 0.05267 & 0.16052 \\ 0.22425 & 0.02902 & 0.21015 & 0.03988 \end{pmatrix}$
Stationary distribution	$\begin{pmatrix} 0.02490 & 0.25852 & 0.05249 & 0.16101 \\ 0.22414 & 0.02872 & 0.20996 & 0.04025 \end{pmatrix}$

Table 4.2.4: Empirical distribution of $(Z^{(i)}, X^{(i)}, Y^{(i)})$ for the Gibbs sampler in Example 4.1.2

Sample size n	Sampling sequence $Z^{(1)} \to X^{(1)} \to Y^{(1)} \to Z^{(2)} \to X^{(2)} \to Y^{(2)} \dots$
1000	$\begin{pmatrix} 0.01800 & 0.14400 & 0.07000 & 0.23400 \\ 0.33400 & 0.03000 & 0.14400 & 0.02600 \end{pmatrix}$
10,000	$\begin{pmatrix} 0.02280 & 0.17720 & 0.05940 & 0.25300 \\ 0.28620 & 0.04260 & 0.13120 & 0.02760 \end{pmatrix}$
100,000	$\begin{pmatrix} 0.02232 & 0.16818 & 0.05386 & 0.25374 \\ 0.30480 & 0.04170 & 0.12954 & 0.02586 \end{pmatrix}$
1,000,000	$\begin{pmatrix} 0.02340 & 0.16777 & 0.05399 & 0.25179 \\ 0.30357 & 0.04152 & 0.13074 & 0.02722 \end{pmatrix}$
Stationary distribution	$\begin{pmatrix} 0.02322 & 0.16781 & 0.05418 & 0.25172 \\ 0.30387 & 0.04139 & 0.13023 & 0.02759 \end{pmatrix}$

From Tables 4.2.2−4.2.4, we find that the Gibbs sampler has three different empirical joint distributions, one based on $(X^{(i)}, Y^{(i)}, Z^{(i)})$, $i = 1, 2, \dots, n$, another based on $(Y^{(i)}, Z^{(i)}, X^{(i)})$, $i = 1, \ldots, n$, and the third based on $(Z^{(i)}, X^{(i)}, Y^{(i)})$, $i = 1, 2, \ldots, n$. Each empirical joint distribution is very close to its stationary joint distribution when the sample size n is large. The result is consistent with Theorem 4.1.2.

Finally, we apply the Gibbs sampler to generate simulations for Example 4.1.3. The results are given in Tables 4.2.5−4.2.7. Table 4.2.5 is the empirical distribution of $(X^{(i)}, Z^{(i)}, Y^{(i)})$, $i = 1, \ldots, n$, Table 4.2.6 is the empirical distribution of $(Z^{(i)}, Y^{(i)}, X^{(i)})$, $i = 1, \ldots, n$, and Table 4.2.7 is the empirical distribution of $(Y^{(i)}, X^{(i)}, Z^{(i)})$, $i = 1, \ldots, n$.

Table 4.2.5: Empirical distribution of $(X^{(i)}, Z^{(i)}, Y^{(i)})$ for the Gibbs sampler in Example 4.1.3

Sample size n	Sampling sequence $X^{(1)} \to Z^{(1)} \to Y^{(1)} \to X^{(2)} \to Z^{(2)} \to Y^{(2)} \ldots$
1000	$\begin{pmatrix} 0.08000 & 0.11000 & 0.17000 & 0.12600 \\ 0.20200 & 0.14600 & 0.08400 & 0.08200 \end{pmatrix}$
10,000	$\begin{pmatrix} 0.07280 & 0.10340 & 0.17660 & 0.15100 \\ 0.17300 & 0.14640 & 0.07160 & 0.10520 \end{pmatrix}$
100,000	$\begin{pmatrix} 0.07400 & 0.10106 & 0.17426 & 0.14968 \\ 0.17366 & 0.14996 & 0.07692 & 0.10005 \end{pmatrix}$
1,000,000	$\begin{pmatrix} 0.07515 & 0.09987 & 0.17536 & 0.15041 \\ 0.17523 & 0.14979 & 0.07442 & 0.09976 \end{pmatrix}$
Stationary distribution	$\begin{pmatrix} 0.07500 & 0.10000 & 0.17500 & 0.15000 \\ 0.17500 & 0.15000 & 0.07500 & 0.10000 \end{pmatrix}$

國立政治大學 National Chengchi University

Table 4.2.6: Empirical distribution of $(Z^{(i)}, Y^{(i)}, X^{(i)})$ for the Gibbs sampler in Example 4.1.3

Sample size n	Sampling sequence $Z^{(1)} \to Y^{(1)} \to X^{(1)} \to Z^{(2)} \to Y^{(2)} \to X^{(2)} \dots$
1000	$\begin{pmatrix} 0.03000 & 0.21600 & 0.05200 & 0.19200 \\ 0.23600 & 0.02000 & 0.19000 & 0.06400 \end{pmatrix}$
10,000	$\begin{pmatrix} 0.02100 & 0.23200 & 0.05200 & 0.19340 \\ 0.22480 & 0.02560 & 0.20080 & 0.05040 \end{pmatrix}$
100,000	$\begin{pmatrix} 0.02494 & 0.22412 & 0.05098 & 0.20044 \\ 0.22496 & 0.02638 & 0.19998 & 0.04820 \end{pmatrix}$
1,000,000	$\begin{pmatrix} 0.02537 & 0.22567 & 0.04998 & 0.19914 \\ 0.22503 & 0.02507 & 0.19967 & 0.05004 \end{pmatrix}$
Stationary distribution	$\begin{pmatrix} 0.02500 & 0.22500 & 0.05000 & 0.20000 \\ 0.22500 & 0.02500 & 0.20000 & 0.05000 \end{pmatrix}$

Table 4.2.7: Empirical distribution of $(Y^{(i)}, X^{(i)}, Z^{(i)})$ for the Gibbs sampler in Example 4.1.3

Sample size n	Sampling sequence $Y^{(1)} \to X^{(1)} \to Z^{(1)} \to Y^{(2)} \to X^{(2)} \to Z^{(2)} \dots$
1000	$\begin{pmatrix} 0.01100 & 0.13400 & 0.12000 & 0.10000 \\ 0.14400 & 0.13200 & 0.14600 & 0.11200 \end{pmatrix}$
10,000	$\begin{pmatrix} 0.09900 & 0.15080 & 0.15140 & 0.10560 \\ 0.12160 & 0.12260 & 0.12640 & 0.12260 \end{pmatrix}$
100,000	$\begin{pmatrix} 0.10006 & 0.15218 & 0.15052 & 0.10040 \\ 0.12500 & 0.12386 & 0.12300 & 0.12498 \end{pmatrix}$
1,000,000	$\begin{pmatrix} 0.09994 & 0.15049 & 0.14981 & 0.10042 \\ 0.12487 & 0.12390 & 0.12494 & 0.12563 \end{pmatrix}$
Stationary distribution	$\begin{pmatrix} 0.10000 & 0.15000 & 0.15000 & 0.10000 \\ 0.12500 & 0.12500 & 0.12500 & 0.12500 \end{pmatrix}$

From Tables 4.2.5−4.2.7, we find that the Gibbs sampler has three different empirical joint distributions, one based on $(X^{(i)}, Z^{(i)}, Y^{(i)})$, $i = 1, 2, \dots, n$,

國立政治大學 National Chengchi University

another based on $(Z^{(i)}, Y^{(i)}, X^{(i)})$, $i = 1, \ldots, n$, and the third based on $(Y^{(i)}, X^{(i)}, Z^{(i)})$, $i = 1, 2, \ldots, n$. Each empirical joint distribution is very close to its stationary joint distribution when the sample size n is large. The result is consistent with Theorem 4.1.3.

國立政治大學
National Chengchi University

5 Conclusions

Although the ratio matrix approach can deal with the compatibility of discrete conditional distributions, it can be only applied to two-dimensional case. Our graphical representation approach, using basic ideas in graph theory, can be extended to higher-dimensional case. This approach can not only check the compatibility but also find the set of all compatible joint distributions when the given conditional distributions are compatible. It works for general n-variate cases and allows for zero elements. Moreover, when the graph is connected, we can use a spanning tree to check the compatibility and find the unique probability distribution if the given conditional distributions are compatible.

In the present paper, we restrict attention to the case where each random variable takes values in a finite set and the given conditional distributions are full. If a random variable takes values in an infinite set, e.g., Poisson variate, the compatibility problem should be extended to the infinite setting. However, in the literature little has been done for the infinite setting. So our graphical representation approach can not be readily extended to this setting.

Consider random variables $X_1, \ldots, X_n$, a conditional distribution $p_{S|T}$, where $S \neq \emptyset$, $S \cap T = \emptyset$ and $S \cup T = \{1, \ldots, n\}$, is called a full conditional distribution because all variables are involved. For instance, if $n = 3$, then $p_{12|3}$ is a full conditional distribution but $p_{1|2}$ is not. Since specifying a full conditional distribution $p_{S|T}$ 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 $\{1, \ldots, n\}$, the given conditional distributions can be equivalently described in terms of probability ratios between vertices. However, this is not so for general conditional disibutions as considered by Gelmen and Speed (1993). For example, if $n = 3$ and the given conditional distributions are $p_{1|2}, p_{2|3}$ and $p_{3|1}$, then we can not find any probability ratios between vertices. 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 by Arnold, Castillo and Sarabia (2002) and Chen, Ip and Wang (2011). It will be of great interest to formulate and solve this problem in terms of a graphical representation.

We also present the relation of compatibility with Gibbs sampler in higher-dimensional case. We transfer a nonhomogeneous Markov chain into all kinds of homogeneous Markov chains by combining successive transitions into a single one. We prove that a given set of conditional distributions is compatible if and only if all homogeneous Markov chains give rise to the same joint distribution. This result can be extended to general n-variate cases.

國立政治大學 National Chengchi University

References

[1] Arnold, B. C., Castillo, E., and Sarabia, J. M. (1999). *Conditional specification of statistical models.* Springer, New York.

[2] Arnold, B. C., Castillo, E., and Sarabia, J. M. (2001). Conditionally specified distributions: an introduction (with discussions). *Statistical Science,* 16, 249-274.

[3] Arnold, B. C., Castillo, E., and Sarabia, J. M. (2002). Exact and near compatibility of discrete distributions. *Computational Statistics and Data Analysis*, 40, 231-252.

[4] Arnold, B. C., Castillo, E., and Sarabia, J. M. (2004). Compatibility of partial or complete conditional probability specifications. *Journal of Statistical Planning and Inference,* 123, 133-159.

[5] Arnold, B. C. and Gokhale, D. V. (1998). Distributions most nearly compatible with given families of conditional distributions. *Test* 7, 377-390.

[6] Arnold, B. C. and Press, S. J. (1989). Campatible conditional distributions. *Journal of the American Statistical Association*, 84, 152-156.

[7] Besag, J., (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B* 36, 192-236.

[8] Gelman, A. and Speed, T. P. (1993). Characterizing a joint probability distribution by conditionals. *Journal of the Royal Statistical Society. Series B* 55, 185-188.

[9] Gourieroux, C. and Monfort, A. (1979). On the characterization of a joint probability distribution by conditional distributions. *Journal of Econometrics,* 10, 115-118.

[10] Hobert, J. P. and Casella, G. (1998) Functional compatibility, markov chains, and Gibbs sampling with improper posteriors. *Journal of Computational and Graphical Statistics,* 7, 42-60.

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

[12] Kuo, K-L, Wang, Y. J. (2011) A simple algorithm for checking compatibility among discrete conditional distributions. *Computational Statistics and Data Analysis,* 5, 2457-2462.

[13] Liu, J. S. (1996) Discussion of "Statistical inference and Monte Carlo algorithms" by Casella, G. *Test* 5, 305-310.

[14] Slavkovic, A. B., Sullivant, S., (2006) The space of compatible full conditionals is a unimodular toric variety. *Journal of Symbolic Computation,* 41, 196-209.

[15] Song, C. C., Li, L. A., Chen, C. H., Jiang, T. J. and Kuo, K. L. (2010). Compatibilty of finie discrete conditional distributions. *Statistica Sinica,* 20, 423-440.

[16] Tian, G. L., Tan, M., Ng, K. W. and Tang, M. L. (2009). A unified method for checking compatibility and uniqueness for finite discrete conditional distributions. *Communications in Statistics-Theory and Models,* 38, 115-129.

[17] Toffoli, E., Cecchin, E., Corona, G., Russo, A., Buonadonna, A., D'Andrea, M., Pasetto, L., Pessa, S., Errante, D., De Pangher, V., Giusto, M., Medici, M., Gaion, F., Sandri, P., Galligioni, E., Bonura, S., Boccalon, M., Biason, P., Frustaci, S. (2006). The role of UGT1A1*28 polymorphism in the pharmacodynamics and pharmacokinetics of irinotecan in patients with metastatic colorectal cancer. *Jounal of Clinical Oncology* 24, 3061-3068.

國 立 政 治 大 學
National Chengchi University

[18] Wang, Y. J., Kuo, K-L. (2010) Compatibility of discrete conditional distributions with structural zeros. *Journal of Multivariate Analysis,* 101, 191-199.

[19] Yao, Y. C., Chen, S. C., Wang, S. H. (2014). On compatibility of discrete full conditional distributions: A graphical representation approach. *Journal of Multivariate Analysis,* 124, 1-9.

國立政治大學
National Chengchi University