Spanish Journal of Statistics
Vol. **6** No. 1 2024, Pages 21–38
doi:https://doi.org/10.37830/SJS.2024.1.03

Regular Article

Finding most nearly compatible conditionals under a finite discrete set-up: An overview and recent developments

Indranil Ghosh

Department of Mathematics and Statistics, University of North Carolina, ghoshi@uncw.edu
Received: February 26, 2024. Accepted: July 31, 2024.

Abstract:

In modeling complicated real-life scenarios, one objective is to capture the dependence being observed. Consequently, conditional specification is a worthy alternative to the joint-distribution models. Since its' inception, the use of divergence measures have been instrumental in determining the closeness between two probability distributions, especially when joint distributions are specified by the corresponding conditional distributions. Conditional specification of distributions is a developing area with several applications. This work gives an overview of a variety of divergence measures including, but not limited to, Kullback-Leibler divergence measure, Power-divergence statistic, Hellinger distance along with some newly developed divergence measures and its role in addressing various compatible conditions in search for a most-nearly compatible for a finite discrete case, and also identifying compatibility under conditional and marginal information under some additional information in the form of marginal and/or conditional summary. Finally, we provide some numerical examples to illustrate each of the scenarios.

Keywords: ncompatible conditionals, divergence measures, iterative algorithm, conditional specification, near compatibility

MSC: 62H05, 62E17

1 Introduction

The problem of determining whether two families of conditional distributions are compatible or minimally incompatible has been considered by several authors and the problem is well established in the literature. For an excellent survey on this topic, an interested reader is referred to the scholarly works by Arnold and Press (1989) and Arnold et al. (1999) and the references cited therein. A non-exhaustive list of pertinent references can be cited as follows, for example, in the works by Gelman and Speed (1993), and Arnold and Gokhale (1994, 1998). Arnold et al. (1992) provided a useful survey of distributions being obtained in such a fashion. Several alternative approaches exist in the literature with regard to the problem of determining the possible compatibility of two families of conditional distributions, for example in the works of Arnold and Press (1989); Arnold and Gokhale (1994); Cacoullos and Papageorgiou (1983); Wesolowski (1996). In addition, the problem of determining most nearly compatible distributions, in the absence of compatibility, has been addressed (Arnold and Gokhale, 1998; Arnold et al., 1999, 2001). In this paper, our our main objective is concentrated on cases in which the conditional specifications are incompatible. In addition, we envision a scenario in which case, from our informed expert and/or practitioner who is working in this field has provided a set of additional information in the form of conditional moments/percentiles; marginal moments etc. We want to examine to what extent such amount of additional information is compatible with the given two conditional probability matrices in search for a most nearly compatible (equivalently minimally incompatible) probability distribution. It is safe to say that the problem has been explored by Arnold et al. (2001) in which the authors derived this problem as a set of non-linear equations involving some constraints.

Our search for a compatible P in terms of equations subject to inequality constraints is based on the fact that we really need to find one compatible marginal, say that corresponding to the random variable X, and we consider the fact that when this is combined with B will give us P. However, in this paper, we look at a different objective which is not discussed in Arnold et al. (2001). Here, we explore the applicability of several measures of divergence (alias pseudo-distance measures) in finding a most nearly compatible distributions by incorporating the additional sets of information along with the complete specification of two conditionals. For an excellent survey on the use of divergence measures in various aspects of distribution theory and associated statistical inference, one is suggested to take a look at the book by Pardo (2006).

In particular, we examine the relative performance of these measures of divergence based on at what stage of iterative algorithm in search for a most nearly compatible P, the adopted procedure converges based on a user defined level of precision which is described later. Needless to say, compatible conditional and marginal specifications of distributions are of fundamental importance in modeling scenarios. Moreover in Bayesian prior elicitation contexts, inconsistent conditional specifications are to be expected. In such situations interest will center on most nearly compatible distributions.

The remainder of the paper is organized as follows. In Section 2, we provide some basic preliminaries regarding compatibility of two discrete conditionals. Section 3 deals with various necessary conditions for compatibility. In Section 4, we discuss the role of pseudo-distance measures in identifying a most nearly compatible probability distribution starting from two given conditional probability matrices under a finite discrete set-up. In Section 5, various methods of finding most nearly compatible distributions are discussed. Section 6 provides an overview on the topic of using pseudo-divergence measures in the presence of additional marginal and/or conditional information. Several illustrative examples are provided in Section 7. Finally, some concluding remarks are presented in Section 8.

2 Basic preliminaries

Let A and B be two $(I \times J)$ matrices with non-negative elements such that $\sum_{i=1}^{I} a_{ij} = 1, \forall j = 1, \dots, J$ and $\sum_{j=1}^{J} b_{ij} = 1, \forall i = 1, 2, \dots, I$. Without loss of generality, it can be assumed that $I \leq J$. Matrices A and B are said to form a compatible conditional specification for the distribution of (X, Y) if there exists some $(I \times J)$ matrix P with non-negative entries p_{ij} and with $\sum_{i=1}^{I}\sum_{j=1}^{J} p_{ij} = 1$ such that, for every (i,j), $a_{ij} = \frac{p_{ij}}{p_{.j}}$ and $b_{ij} = \frac{p_{ij}}{p_{i.}}$, where $p_{i.} = \sum_{j=1}^{J} p_{ij}$ and $p_{i.} = \sum_{i=1}^{I} p_{ij}$. If such a matrix P exists, then, if we assume that $p_{ij} = P(X = x_i, Y = y_j)$, $i = 1,2,\cdots,I$, $j = 1,2,\cdots,J$, we will have $a_{ij} = P(X = x_i|Y = y_j)$, $i = 1,2,\cdots,I$, $j = 1,2,\cdots,J$, and $b_{ij} = P(Y = y_j|X = x_i)$, $i = 1,2,\cdots,I$, $j = 1,2,\cdots,J$. Equivalently, A and B are compatible if there exist stochastic vectors $\underline{\tau} = (\tau_1,\tau_2,\cdots,\tau_J)$ and $\underline{\eta} = (\eta_1,\eta_2,\cdots,\eta_I)$ such that

$$a_{ij}\tau_j = b_{ij}\eta_i,$$

for every (i,j). In the case of compatibility, $\underline{\eta}$ and $\underline{\tau}$ can be readily interpreted as the resulting marginal distributions of X and Y, respectively. For any probability vector $\eta = (\eta_1,\eta_2,\ldots,\eta_I)$, $p_{ij} = b_{ij}\eta_i$ is a probability distribution on the IJ cells. So, the conditional probability matrix, denoted by A, and its elements (a_{ij}) will be given by

$$a_{ij} = \frac{p_{ij}}{\sum_{s=1}^{I} p_{sj}} = \frac{b_{ij}\eta_i}{\sum_{s=1}^{I} b_{sj}\eta_s}, \tag{1}$$

for every i,j. If A and B are compatible, then

$$a_{ij}\sum_{s=1}^{I} b_{sj}\eta_s = b_{ij}\eta_i.$$

We then have

$$\tau_j = \sum_{s=1}^{I} b_{ij}\eta_s, \forall j = 1,\ldots,J.$$

In this case, the expressions given in (1) can be rewritten as

$$a_{ij}\sum_{s=1}^{I} b_{sj}\eta_s - b_{ij}\eta_i = 0.$$

3 Compatibility conditions

Conditions for compatibility are listed in the following theorems which are due to Arnold and his co-authors.

Suppose that A and B have identical incidence sets then they are compatible if and only if either of the following two conditions hold.

(a) There exist stochastic vectors $\vec{\tau} = (\tau_1,\tau_2,....,\tau_I)$ and $\vec{\eta} = (\eta_1,\eta_2,....,\eta_J)$ such that $\eta_j a_{ij} = (\tau_i b_{ij}),\forall i,j$. In the case of compatibility, the vectors $\vec{\tau}$ and $\vec{\eta}$ can readily interpreted being proportional to the marginal distributions of X and Y respectively.

(b) There exists vectors $\vec{u}$ and $\vec{v}$ for which $d_{ij} = \frac{a_{ij}}{b_{ij}} = u_i v_j,\forall i,j \in N$.

This suggests the use of log-linear models to fit the matrix D. Indeed, if the log-linear model has all interactions equal to zero, then we have compatibility. Otherwise, A and B are incompatible.
If $N = \{1,2,...I\} \times \{1,2,...J\}$, i.e; if all the entries in A and B are positive, then we have the following theorem given by due to Arnold and Gokhale (1994).

1. A and B are compatible iff they have identical uniform marginal representations(UMRs) (Mosteller, 1968).
2. A and B are compatible iff all cross product ratios of A are identical to those of B.

Note: Some restrictions on the common incidence set of A and B is necessary for the above theorem. For example if we consider

$A = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}$

and $B = \begin{pmatrix} 1/3 & 2/3 & 0 \\ 0 & 1/3 & 2/3 \\ 2/3 & 0 & 1/3 \end{pmatrix}$

It may be verified here here that A and B have equal cross product ratios(there are no positive 2×2 submatrices)and have identical uniform marginal representations but A and B are not compatible. Compatibility of A and B of course does not confirm a unique compatible matrix P. The simplest sufficient condition is positivity, i.e; $(a_{ij} b_{ij}) \geq 0$ and $\forall i, j$.

4 Measures of divergence

In this section, we list several useful divergence measures which will be utilized in this paper for finding the ϵ-compatible distributions under the finite discrete set-up. In addition, we provide some useful relationships among these divergence measures. Some of these results have been independently derived and discussed in Ghosh and Sunoj (2024) and Borzadaran and Amini (2010) in the context of copula-based divergence measures. We begin our discussion with the power divergence statistics as a measure of divergence, for pertinent details, see Cressie and Read (1984). A divergence measure between two probability distributions $\underline{p}$ and $\underline{q}$ (which are of the same dimension) returns a measure of similarity or distance between them. It is non-negative. It measures the divergence between the population distribution $\underline{\pi} = (\pi_1, \pi_2, \ldots, \pi_k)$ and the uniform distribution $\left(\frac{1}{k}, \ldots, \frac{1}{k}\right)$, where a value closer to zero represents a wider divergence from the uniform distribution. A natural generalization, when considered in this way, is to define a measure of divergence between two general distributions. This concept was first considered by Kullback (1959) in his directed divergence measure. It was followed up by Arnold and Gokhale (1994, 1998) while considering minimum incompatibility via the K-L criterion. It is of the form

$$K\left(\underline{p} : \underline{q}\right) = \sum_{i=1}^{k} p_i \log_2 \left(\frac{p_i}{q_i}\right), \tag{2}$$

where $\underline{p}$ and $\underline{q}$ are two discrete probability distributions defined on the $(k-1)$ dimensional simplex

$$\Delta_k = \left\{ \underline{\pi} : \pi_i \geq 0; i = 1, \ldots, k; \sum_{i=1}^{k} \pi_i = 1 \right\}.$$

Here, we adopt the convention that $p_i \log_2 \left(\frac{p_i}{q_i}\right) = 0$ when $p_i = 0$ and for any $0 \leq q_i \leq 1$. A family of power divergence statistics indexed by $\lambda \in \mathbb{R}$ for $\underline{p} = (p_1, p_2, \ldots, p_k)$, $\underline{q} = (q_1, q_2, \ldots, q_k)$ can be defined as

$$I^{\lambda}\left(\underline{p}:\underline{q}\right)=\frac{1}{\lambda(\lambda+1)}\sum_{i=1}^{k}p_i\left[\left(\frac{p_i}{q_i}\right)^{\lambda}-1\right] \tag{3}$$

with the convention $p_i = 0$ whenever $q_i = 0$. Note that (3) generalizes (2) in the same way the Rényi entropy (Rényi, 1961) generalizes the Shannon entropy (Shannon, 1951).

1. Considering the fact that a matrix can be written as an array of column vectors, we define the power divergence statistic for matrices A and B as:

$$\begin{aligned} D_1 &= I^{\lambda}\left(p_{ij}:a_{ij}p_{\cdot j}\right)+I^{\lambda}\left(p_{ij}:b_{ij}p_{i\cdot}\right) \\ &= \frac{1}{\lambda(\lambda+1)}\left[\sum_{i=1}^{I}\sum_{j=1}^{J}p_{ij}\left(\left(\frac{p_{ij}}{a_{ij}p_{\cdot j}}\right)^{\lambda}-1\right)+\sum_{i=1}^{I}\sum_{j=1}^{J}p_{ij}\left(\left(\frac{p_{ij}}{b_{ij}p_{i\cdot}}\right)^{\lambda}-1\right)\right], \end{aligned}$$

   where $\lambda \in \mathbb{R}$ is a parameter. The power divergence statistic is undefined for $\lambda = -1$ or $\lambda = 0$. However, if we define these two cases as continuous limits of D_1 for $\lambda \to -1$ and $\lambda \to 0$, then D_1 is continuous in λ.
   The name power divergence derives from the fact that the statistic D_1 measures the divergence of p_{ij} from $(a_{ij}p_{\cdot j})$ and $(b_{ij}p_{i\cdot})$ through a weighted sum of powers of the terms $\left(\frac{p_{ij}}{a_{ij}p_{\cdot j}}\right)$ and $\left(\frac{p_{ij}}{b_{ij}p_{i\cdot}}\right)$ for all $(i,j) \in N$. We want to minimize D_1 with respect to $\sum\sum_{(i,j)\in N} p_{ij} = 1$.

   Note: On the choice of λ

   In the power divergence statistic, λ is a parameter that can take any real value. A natural question that arises here is: what should be the optimum choice of λ? There are some conflicting recommendations regarding which value of λ results in the optimal test statistic. In all our examples of iterative study discussed in Section 4 later, we find that the rate of convergence is very slow for most values of λ. For example, for $\lambda = 0.2, 0.3$ and 0.5, the iterative procedure for the divergence measure D_λ converges at $n = 20, 27$ and 34, respectively. For negative choices of λ, D_1 is quite big, and moreover the resulting matrix is not a probability matrix. A future work will focus on providing practical guidelines about how to choose λ and also to investigate the sensitivity of solutions in addition to the rate of convergence) when different values of λ's are used in its' permissible range. In the next, we provide a collection of divergence measures which has been utilized to obtain the ϵ-compatible distribution(s) under the finite discrete set-up. For pertinent details, see Ghosh (2011), Ghosh and Balakrishnan (2015), Ghosh and Nadarajah (2017) and the references cited therein.

2. Modified Renyi's divergence measure, see Ghosh (2011)

$$D_2=\frac{1}{(\alpha-1)}\left[\sum_{i=1}^{I}\sum_{j=1}^{J}(a_{ij}p_{.j})^{-1}\log\left(\frac{p_{ij}}{a_{ij}p_{.j}}\right)^{\alpha}+\sum_{i=1}^{I}\sum_{j=1}^{J}(b_{ij}p_{i.})^{-1}\log\left(\frac{p_{ij}}{b_{ij}p_{i.}}\right)^{\alpha}\right] \tag{4}$$

   Note: Nadarajah and Zografos (2003); Zografos and Nadarajah (2005) provided a useful review of Renyi's entropy for different univariate and k-variate random variables.

3. χ^2 measure of divergence
   It is defined as

$$D_3 = \sum_{i=1}^{I}\sum_{j=1}^{J}\left[\left(\frac{p_{ij}}{a_{ij}p_{.j}}\right)^2\right]a_{ij}p_{.j} + \sum_{i=1}^{I}\sum_{j=1}^{J}\left[\left(\frac{p_{ij}}{b_{ij}p_{i.}}\right)^2\right]b_{ij}p_{i.} \tag{5}$$

4. First new measure of divergence (see, Ghosh (2011))

$$D_4 = \sum_{i=1}^{I}\sum_{j=1}^{J}\left[\left(\frac{p_{ij}}{a_{ij}p_{.j} + b_{ij}p_{i.}} - 1\right)^2\right]^{\lambda}, \tag{6}$$

   where $\lambda > 0$ is a constant.

5. Second new measure of divergence (see, Ghosh (2011))

$$D_5 = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\sqrt{p_{ij}} - \sqrt{a_{ij}p_{.j}}\right)^2 + \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\sqrt{p_{ij}} - \sqrt{b_{ij}p_{.j}}\right)^2. \tag{7}$$

Note: It is to be noted that if the two conditional matrices A and B are compatible then each of these measures will be equal to zero.

5 Available methods of obtaining minimally incompatible distributions

In this section, we describe the idea of minimal incompatibility of two given conditional distributions, and then explain some methods of finding minimally incompatible distributions. For pertinent details, see Arnold et al. (1999).

5.1 ϵ-Compatibility

Suppose, we do not insist on precise compatibility, and instead wish to have p_{ij} to be approximately consistent with two given conditional probability matrices A and B. Let W be a weight matrix that represents the relative importance of accuracy in determining the probabilities p_{ij} for each (i, j). For a given weight matrix W which might be uniform, i.e., $w_{ij} = 1, \forall (i, j)$ if all pairs (i, j) were equally important, we may consider the following strategies expressed as non-linear and linear programming problems.

(i) **First method:** Find a matrix P, with $p_{ij} \geq 0 \quad \forall (i, j)$, such that

$$\left| p_{ij} - a_{ij}\sum_{i=1}^{I} p_{ij} \right| \leq \epsilon w_{ij} \quad \forall (i,j) \in N,$$

$$\left| p_{ij} - b_{ij}\sum_{j=1}^{J} p_{ij} \right| \leq \epsilon w_{ij} \quad \forall (i,j) \in N,$$

with the linear constraint $\sum_{i=1}^{I}\sum_{j=1}^{J} p_{ij} = 1$.

(ii) **Second method:** Second method: Seek two probability vectors η and τ such that
$|a_{ij}\eta_j - b_{ij}\tau_i| \le \epsilon w_{ij} \quad \forall (i,j))\,, \quad \sum_j \eta_j = 1, \quad \sum_i \tau_i = 1, \quad$ and $\tau_i \ge 0, \quad \eta_j \ge 0, \forall (i,j) \in N.$

(iii) **Third method:** Find a (marginal) probability vector $\tau \ge 0$, such that and $\tau_i \ge 0, \forall i$.

Clearly, the above methods introduce three different concepts of ϵ-compatibility. If we use Method 1, and if A and B are ϵ-compatible, then the matrix P^* which satisfies Eq. (1) will be said to be most nearly compatible. If we use Method 2 and if A and B are ϵ-compatible, then a reasonable choice for a most nearly compatible matrix P^* will be

$$P^* = \frac{a_{ij}\eta_j^* + b_{ij}\tau_j^*}{2},$$

where η_j^* and τ_j^* satisfy Eq.(2). Finally, if we use Method 3 and if A and B are ϵ-compatible, then a plausible choice for a most nearly compatible P^* will be $P^* = (b_{ij}\tau_i^*)$, where τ_i^* satisfies Eq.(3).

6 Pseudo-divergence measures under additional information

Until now, we have discussed the power divergence statistic as a measure of divergence to obtain minimally incompatible (or equivalently ε-compatible) joint probability distributions from the set of two conditionals. Here, we want to find a procedure from which we would like to get the joint probability distribution from the two conditionals but with some additional information provided on the marginal and conditional probabilities and expectations, i.e., we want to see whether a given set of constraints involving marginal and conditional probabilities and expectations of functions are compatible or minimally incompatible. The finite discrete case (the main focus of the paper) may be viewed as one involving solutions of linear equations in restricted domains. We will consider cases where the given conditional probabilities and expectations are specified. Cases of imprecise specification will be considered later on. So far, in all divergence criteria we minimized the given function based on only one linear constraint: $\sum_{i=1}^{I}\sum_{j=1}^{J} p_{ij} = 1$. Instead, Suppose we are given (by our well-informed expert engaged in this study) the following set of marginal and conditional information (one may call this a set of precise information):

1. $P(\underline{X} \in A_i) = \delta_i$ for specified sets $A_1, A_2, \ldots, A_{n_1}$,
2. $P(\underline{X} \in B_i | \underline{X} \in C_i) = \eta_i$, $i = 1, 2, \ldots, n_2$ for specified sets of $B_1, B_2, \ldots, B_{n_2}$, and $C_1, C_2, \ldots, C_{n_2}$,
3. $E(\epsilon_j(\underline{X})) = \xi_j$, $j = 1, 2, \ldots, n_3$ for specified functions $\epsilon_1, \epsilon_2, \ldots, \epsilon_{n_3}$,
4. $E(\varphi_i(\underline{X}) | \phi_i(\underline{X}) = \lambda_i) = \omega_i$, $i = 1, 2, \ldots, n_4$ for specified functions $\varphi_1, \varphi_2, \ldots, \varphi_{n_4}$ and specified constants $\lambda_1, \lambda_2, \ldots, \lambda_{n_4}$,
5. $P(\nu_i(\underline{X}) \in E_i | \gamma_i(\underline{X}) \in F_i) = \beta_i$, $i = 1, 2, \ldots, n_5$ for specified functions $\nu_1, \nu_2, \ldots, \nu_{n_5}$ and specified sets $E_1, E_2, \ldots, E_{n_5}$ and $F_1, F_2, \ldots, F_{n_5}$.

Note that the above sets of information can be rewritten as follows.

- $P(\underline{X} \in A_i) = \sum_{\underline{X} \in A_i} p(\underline{X}) = \delta_i$,
- $P(\underline{X} \in B_i | \underline{X} \in C_i) = \eta_i$ if and only if $\sum_{\underline{X} \in B_i \cap C_i} p(\underline{X}) - \eta_i \sum_{\underline{X} \in C_i} p(\underline{X}) = 0$,

- $E\left(\epsilon_j\left(\underline{X}\right)\right) = \sum_{\underline{X}} \epsilon_j\left(\underline{X}\right) p\left(\underline{X}\right) = \xi_j$,
- $E\left(\varphi_i\left(\underline{X}\right) \middle| \phi_i\left(\underline{X}\right) = \lambda_i\right) = \omega_i$ if and only if $\sum_{\phi_i(\underline{X})=\lambda_i} \varphi_i\left(\underline{X}\right) p\left(\underline{X}\right) - \omega_i \sum_{\phi_i(\underline{X})=\lambda_i} p\left(\underline{X}\right) = 0$,
- $P\left(\nu_i\left(\underline{X}\right) \in E_i \middle| \gamma_i\left(\underline{X}\right) \in F_i\right) = \beta_i$ if and only if $\sum_{\nu_i(\underline{X})\in E_i \cap \gamma_i(\underline{X})\in F_i} p\left(\underline{X}\right) - \beta_i \left(\sum_{\nu_i(\underline{X})\in F_i} p\left(\underline{X}\right)\right) = 0.$

Thus, if we arrange the values of the joint density $p\left(\underline{X}\right)$ of $\underline{X}$ as a vector of dimension $\Omega = \text{card}(X_1) \times \text{card}(X_2) \times \cdots \times \text{card}(X_k)$, where $\underline{X} = (X_1, X_2, \ldots, X_k)$, then we can write every piece of information given above in the form:

$$M\underline{p} = \underline{\theta}, \tag{8}$$

where the matrix M in Eq.(8) is of order $(r+1) \times \Omega$, assuming r pieces of information are given and rank $r+1 \leq \Omega$. The $\vec{\theta}$ is of order $(r+1) \times 1$. Both M and θ are assumed to be known. The "natural" constraint $\underline{p} \cdot \vec{1} = 1$ is incorporated in Eq.(8) by letting the first row of M consist of all unit elements and the first element of θ equal to unity. The system in Eq.(8) is assumed to be consistent in the sense that there exists a positive probability vector satisfying (1). If $r+1$ is large, it is highly unlikely that $r+1$ pieces of information will be compatible with the given information, in the sense that Eq.(8) has a solution $\underline{p}^*$ with non-negative coordinates adding up to one. In general, it would be more rational to seek approximate equality in Eq.(8) subject to $\underline{p} \geq 0$ and $M\underline{p} = \underline{\theta}$. In other words, we are seeking an almost compatible distribution.

6.1 Power divergence statistic under conditional and marginal information

Our search for a most nearly compatible distribution (equivalently ε compatible) $\underline{p}$ can be viewed as a problem of minimizing $D\left(M\underline{p}, \underline{\theta}\right)$ for a suitable distance measure D subject to the restriction that $\underline{p} \geq 0$ and $M\underline{p} = \underline{\theta}$. One such reasonable distance measure is the power divergence statistic. The determined minimum value of the objective function, in each of the examples, described later, provides a measure of incompatibility of the given information.

In this case, we have $P^{I\times J} = \left(\underline{p}_1, \underline{p}_2, \ldots, \underline{p}_I\right)^{1\times I}$, where $\underline{p}_1 = (p_{11}, p_{12}, \ldots, p_{1J})^{1\times J}$, $\underline{p}_2 = (p_{21}, p_{22}, \ldots, p_{2J})^{1\times J}$, and so on up to $\underline{p}_I = (p_{I1}, p_{I2}, \ldots, p_{IJ})^{1\times J}$, and we have the linear restriction of the form

$$\sum_{u=1}^{I} M_{tu}\underline{p}_u = \theta_t,$$

for $t = 1, 2, \ldots, (r+1)$. The power divergence statistic (PDS) in this case reduces to

$$D_1\left(\underline{p}\right) = \frac{1}{\lambda(\lambda+1)} \sum_{u=1}^{I} \left[\underline{p}_u \left(\left(\frac{\underline{p}_u}{\underline{a}_u p_{\cdot j}}\right)^{\lambda} - 1\right) + \underline{p}_u \left(\left(\frac{\underline{p}_u}{\underline{b}_u p_{i\cdot}}\right)^{\lambda} - 1\right)\right].$$

Now we consider the following Lagrangian function

$$F = D_1\left(\underline{p}\right) + \sum_{t=1}^{r+1} \tau_t \left(\sum_{u=1}^{I} M_{tu} p_u - \theta_t\right),$$

where τ_t, $t = 1, 2, \ldots, (r+1)$ are $(r+1)$ Lagrangian multipliers. To minimize F, we consider simultaneous solution of

$$\frac{\partial F}{\partial \underline{p}_u} = 0. \tag{9}$$

Consequently, the optimal value of $\underline{p}_u$ is

$$\underline{p}_u^* = \frac{\left(\left(\frac{1}{(\underline{a}_u p_{\cdot j})^\lambda} + \frac{1}{(\underline{b}_u p_{i\cdot})^\lambda}\right)^{\frac{1}{\lambda}}\right)^{-1}}{\left(\sum_{u\in N}\left\{\left(\frac{1}{(\underline{a}_u p_{\cdot j})^\lambda} + \frac{1}{(\underline{b}_u p_{i\cdot})^\lambda}\right)^{\frac{1}{\lambda}}\right\}^{-1}\right)^{-1}}.$$

For an iterative study, we consider the following

$$\underline{p}_u^{n+1} = \frac{\left(\frac{1}{(\underline{a}_u p_{\cdot j}^n)^\lambda} + \frac{1}{(\underline{b}_u p_{i\cdot}^n)^\lambda}\right)^{\frac{1}{\lambda}}}{\sum_{u\in N}\left(\frac{1}{\left(\underline{a}_u p_{\cdot j}^n\right)^\lambda} + \frac{1}{(\underline{b}_u p_{i\cdot}^n)^\lambda}\right)^{\frac{1}{\lambda}}},$$

for $n = 0, 1, \ldots$ with the initial choice of $p_{ij}^{(0)} = \frac{1}{IJ}$ for all $(i, j) \in N$. We may use the stopping rule for this iterative algorithm as $\left|\frac{D_1^{(n+1)}}{D_1^{(n)}} - 1\right| \leq 10^{-6}$. In all the examples we considered, our process was found to converge for a wide range of λ.

6.2 Kullback-Leibler divergence criterion under conditional and marginal information

In this case, the K-L divergence statistic is

$$D_2(\underline{p}) = \sum_{u=1}^{I}\left[\underline{a}_u \log\left(\frac{\underline{a}_u p_{\cdot j}}{\underline{p}_u}\right) + \underline{b}_u \log\left(\frac{\underline{b}_u p_{i\cdot}}{\underline{p}_u}\right)\right].$$

Again, we consider the following Lagrangian function

$$F_2 = D_2\left(\underline{p}\right) + \sum_{t=1}^{r+1}\tau_t\left(\sum_{u=1}^{I} M_{tu} p_u - \theta_t\right),$$

where τ_t, $t = 1, 2, \ldots, (r+1)$ are $(r+1)$ Lagrangian multipliers. To minimize F_2, we consider simultaneous solution of

$$\frac{\partial F_2}{\partial \underline{p}_u} = 0,$$

same as in (9). So, the optimal value of $\underline{p}_u$ is

$$\underline{p}_u^* = \frac{\left(\frac{\underline{a}_u + \underline{b}_u}{\frac{1}{p_{i\cdot}} + \frac{1}{p_{\cdot j}}}\right)}{\left(\sum_{u\in N}\frac{\underline{a}_u + \underline{b}_u}{\frac{1}{p_{i\cdot}} + \frac{1}{p_{\cdot j}}}\right)}.$$

For an iterative study, we consider the following

$$\underline{p}_u^{(n+1)} = \frac{\left(\frac{\underline{a}_u + \underline{b}_u}{\frac{1}{p_{i.}^n} + \frac{1}{p_{.j}^n}}\right)}{\left(\sum_{u \in N} \frac{\underline{a}_u + \underline{b}_u}{\frac{1}{p_{i.}^n} + \frac{1}{p_{.j}^n}}\right)}$$

for $n = 0, 1, \ldots$ with the initial choice of $p_{ij}^{(0)} = \frac{1}{IJ}$ for all $(i, j) \in N$. We use the following stopping rule $\left|\frac{D_2^{(n+1)}}{D_2^{(n)}} - 1\right| \leq 10^{-6}$. Here also our iterative algorithm is convergent.

6.3 Modified Renyi's measure of divergence under the marginal and conditional information

Proceeding as before, in this case, the statistic will be

$$D_3 = \frac{1}{(\alpha - 1)}\left[\sum_{i=1}^{I}\sum_{j=1}^{J} (\underline{a}_u p_{.j})^{-1} \log\left(\frac{p_{ij}}{\underline{a}_u p_{.j}}\right)^{\alpha} + \sum_{i=1}^{I}\sum_{j=1}^{J} (\underline{b}_u p_{i.})^{-1} \log\left(\frac{p_{ij}}{\underline{b}_u p_{i.}}\right)^{\alpha}\right]. \tag{10}$$

Next, we consider the following Lagrangian function

$$F_3 = D_3\left(\underline{p}\right) + \sum_{t=1}^{r+1} \tau_t \left(\sum_{u=1}^{I} M_{tu} p_u - \theta_t\right),$$

where τ_t, $t = 1, 2, \ldots, (r + 1)$ are $(r + 1)$ Lagrangian multipliers. Now, to minimize F_3, we consider simultaneous solution of

$$\frac{\partial F_3}{\partial \underline{p}_u} = 0,$$

same as in (9). Consequently, the optimal value of $\underline{p}_u$ is

$$\underline{p}_u^* = \frac{\frac{1}{\underline{a}_u p_{.j}} + \frac{1}{\underline{b}_u p_{i.}}}{\sum\sum_{(i,j)\in N}\left(\frac{1}{\underline{a}_u p_{.j}} + \frac{1}{\underline{b}_u p_{i.}}\right)}.$$

Subsequently, for an iterative study, we consider the following iterative algorithm

$$\underline{p}_u^{(n+1)} = \frac{\frac{1}{\underline{a}_u p_{.j}^{(n)}} + \frac{1}{\underline{b}_u p_{i.}^{(n)}}}{\sum\sum_{(i,j)\in N}\left(\frac{1}{\underline{a}_u p_{.j}^{(n)}} + \frac{1}{\underline{b}_u p_{i.}^{(n)}}\right)}.$$

for $n = 0, 1, \ldots$ with the initial choice of $p_{ij}^{(0)} = \frac{1}{IJ}$ for all $(i, j) \in N$. We use the following stopping rule $\left|\frac{D_3^{(n+1)}}{D_3^{(n)}} - 1\right| \leq 10^{-6}$. Here also our iterative algorithm is convergent based on all the empirical studies that we have made in this regard. A formal mathematical proof is still remains an open problem.

6.4 χ^2 divergence criterion under conditional and marginal information

In this case, our test statistic reduces to

$$D_4 = \sum \sum_{(i,j)\in N} \left[\left(\frac{p_{ij}}{\underline{a}_u p_{.j}} \right)^2 \right] \underline{a}_u p_{.j} + \sum \sum_{(i,j)\in N} \left[\left(\frac{p_{ij}}{\underline{b}_u p_{i.}} \right)^2 \right] \underline{b}_u p_{i.} \qquad (11)$$

Next, we consider the following Lagrangian function

$$F_4 = D_4\left(\underline{p}\right) + \sum_{t=1}^{r+1} \tau_t \left(\sum_{u=1}^{I} M_{tu} p_u - \theta_t \right),$$

where τ_t, $t = 1, 2, \ldots, (r+1)$ are $(r+1)$ Lagrangian multipliers. Now, to minimize F_4, we consider simultaneous solution of

$$\frac{\partial F_4}{\partial \underline{p}_u} = 0,$$

same as in (9). Consequently, the optimal value of $\underline{p}_u$ will be

$$\underline{p}_u^* = \left(\frac{1}{\underline{a}_u p_{.j}} + \frac{1}{\underline{b}_u p_{i.}} \right)^{-1} \left[\sum \sum_{(i,j)\in N} \frac{1}{\underline{a}_u p_{.j}} + \frac{1}{\underline{b}_u p_{i.}} \right]^{-1}$$

Consequently, an iterative algorithm for finding minimally compatible (alias ϵ-compatible) P would be to have

$$\underline{p}_u^{(n+1)} = \left(\frac{1}{\underline{a}_u p_{.j}^{(n)}} + \frac{1}{\underline{b}_u p_{i.}^{(n)}} \right)^{-1} \left[\sum \sum_{(i,j)\in N} \frac{1}{\underline{a}_u p_{.j}^{(n)}} + \frac{1}{\underline{b}_u p_{i.}^{(n)}} \right]^{-1},$$

for $n = 0, 1, \ldots$ with the initial choice of $p_{ij}^{(0)} = \frac{1}{IJ}$ for all $(i,j) \in N$. We use the following stopping rule $\left| \frac{D_4^{(n+1)}}{D_4^{(n)}} - 1 \right| \leq 10^{-6}$. Here also our iterative algorithm is convergent based on all the empirical studies that we have made in this regard. A formal mathematical proof is still remains an open problem.

6.5 Divergence measure D_5 under conditional and marginal information

Here, our test statistic reduces to

$$D_5 = \sum \sum_{(i,j)\in N} \left[\left(\frac{p_{ij}}{\underline{a}_u p_{.j} + \underline{b}_u p_{i.}} - 1 \right)^2 \right]^{\lambda},$$

Next, we consider the following Lagrangian function

$$F_5 = D_5\left(\underline{p}\right) + \sum_{t=1}^{r+1} \tau_t \left(\sum_{u=1}^{I} M_{tu} p_u - \theta_t \right),$$

where τ_t, $t = 1, 2, \ldots, (r+1)$ are $(r+1)$ Lagrangian multipliers. Now, to minimize F_5, we consider simultaneous solution of

$$\frac{\partial F_5}{\partial \underline{p}_u} = 0,$$

same as in (9). Consequently, the optimal value of $\underline{p}_u$ will be

$$\underline{p}_u^* = \frac{(\underline{a}_u p_{.j} + \underline{b}_u p_{i.})^{1-\lambda^{-1}}}{\sum\sum_{(i,j)\in N} (\underline{a}_u p_{.j} + \underline{b}_u p_{i.})^{1-\lambda^{-1}}}$$

Based on the above optimal value, an iterative algorithm could be

$$\underline{p}_u^{(n+1)} = \frac{\left(\underline{a}_u p_{.j}^{(n)} + \underline{b}_u p_{i.}^{(n)}\right)^{1-\lambda^{-1}}}{\sum\sum_{(i,j)\in N} \left(\underline{a}_u p_{.j}^{(n)} + \underline{b}_u p_{i.}^{(n)}\right)^{1-\lambda^{-1}}},$$

for $n = 0, 1, \ldots$ with the initial choice $p_{ij}^{(0)} = \frac{1}{IJ}$ for all $(i,j) \in N$. We use the following stopping rule $\left|\frac{D_5^{(n+1)}}{D_5^{(n)}} - 1\right| \leq 10^{-6}$. Here also our iterative algorithm is convergent based on all the empirical studies that we have made in this regard. A formal mathematical proof is still remains an open problem.

6.6 Divergence measure D_6 under conditional and marginal information

Here, our test statistic reduces to

$$D_6 = \sum\sum_{(i,j)\in N} \left(\sqrt{p_{ij}} - \sqrt{\underline{a}_u p_{.j}}\right)^2 + \sum\sum_{(i,j)\in N} \left(\sqrt{p_{ij}} - \sqrt{\underline{b}_u p_{.j}}\right)^2. \tag{12}$$

Next, we consider the following Lagrangian function

$$F_6 = D_6(\underline{p}) + \sum_{t=1}^{r+1} \tau_t \left(\sum_{u=1}^{I} M_{tu} p_u - \theta_t\right),$$

where τ_t, $t = 1, 2, \ldots, (r+1)$ are $(r+1)$ Lagrangian multipliers. Now, to minimize F_6, we consider simultaneous solution of

$$\frac{\partial F_6}{\partial \underline{p}_u} = 0,$$

same as in (9). Consequently, the optimal value of $\underline{p}_u$ will be

$$\underline{p}_u^* = \frac{(\underline{a}_u p_{.j})^2 + (\underline{b}_u p_{.j})^2}{\sum\sum_{(i,j)\in N} \left\{(\underline{a}_u p_{.j})^2 + (\underline{b}_u p_{.j})^2\right\}}$$

Based on the above optimal value, an iterative algorithm could be

$$\underline{p}_u^{(n+1)} = \frac{\left(\underline{a}_u p_{.j}^{(n)}\right)^2 + \left(\underline{b}_u p_{.j}^{(n)}\right)^2}{\sum\sum_{(i,j)\in N}\left\{\left(\underline{a}_u p_{.j}^{(n)}\right)^2 + \left(\underline{b}_u p_{.j}^{(n)}\right)^2\right\}}$$

for $n = 0, 1, \ldots$ with the initial choice $p_{ij}^{(0)} = \frac{1}{IJ}$ for all $(i, j) \in N$. We use the following stopping rule $\left|\frac{D_6^{(n+1)}}{D_6^{(n)}} - 1\right| \leq 10^{-6}$. Here also our iterative algorithm is convergent based on all the empirical studies that we have made in this regard. A formal mathematical proof is still remains an open problem.

7 Illustrative Examples

In these illustrative examples, we consider conditional probability matrices that are incompatible in nature. These examples, although not taken from a real life scenario, are representative of the fact that given an additional set of precise information, whether the two conditional distributions are compatible or not, and in case they are not, can we find something close to what we call as ε-compatibility. Prominent real life scenarios in which this might be useful are Bayesian networks, model building in classical statistical settings, and elicitation and construction of multiparameter prior distributions in Bayesian scenarios. The dimensions of the matrices A and B are taken to be either 3 or 4 in Examples 1 to 5. The matrix M for each example was easily constructed using `Mathematica` software. The results of the iterative algorithm for the examples are shown in Tables 1 to 3.

- **Example 1.** In this example, we illustrate the above defined method in a simple case. Consider the set (X, Y) of two variables taking values $1, 2, 3, 4$. Let us consider the associated conditional probability matrices, where $I = 4$ and $J = 4$ and

$$A = \begin{pmatrix} 0.27 & 0.4 & 0 & 0.10 \\ 0.18 & 0.20 & 0.50 & 0.40 \\ 0.55 & 0.20 & 0.30 & 0.25 \\ 0 & 0.20 & 0.20 & 0.25 \end{pmatrix},$$

and

$$B = \begin{pmatrix} 0.15 & 0.28 & 0.35 & 0.22 \\ 0.45 & 0 & 0.25 & 0.30 \\ 0.50 & 0.17 & 0.20 & 0.13 \\ 0 & 0.55 & 0.20 & 0.30 \end{pmatrix}.$$

Here, A and B are incompatible since they do not share even a common incidence matrix.

Suppose that we have the following information (from our informed expert) :

  - $E\left(X^2\right) = 7.49$;
  - $P(Y = 3) = 0.38$;
  - $P\left(X^2 = 9 \middle| Y = 2\right) = 0.37$;
  - $P\left(Y^2 = 1 \middle| X = 2\right) = 0.53$.

Here, we have $\underline{p} = (p_{11}, p_{12}, \ldots, p_{44})$. In this case all the above information can be summarized by our M matrix given as follows.

$$M = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 4 & 4 & 4 & 4 & 9 & 9 & 9 & 9 & 16 & 16 & 16 & 16 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0.37 & 0 & 0 & 0 & 0.37 & 0 & 0 & 0 & -0.63 & 0 & 0 & 0 & 0.37 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.47 & 0.53 & 0.53 & 0.53 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Subsequently, $\underline{\theta} = (1, 7.49, 0.38, 0, 0)$. The iterative algorithm results are given in Table 1. In all the examples we considered, the constraints were approximated to a relative absolute error of 10^{-6}. The algorithm was found to converge for a wide range of values of λ.

- **Example 2.** In this example, we consider the set $\{X, Y\}$ of two variables taking values $1, 2, 3$. Let us consider two conditional probability matrices, where $I = 3$ and $J = 3$ and

$$A = \begin{pmatrix} 0.35 & 0.43 & 0 \\ 0 & 0.57 & 0.42 \\ 0.65 & 0 & 0.58 \end{pmatrix},$$

and

$$B = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{3}{4} & 0 & \frac{1}{4} \end{pmatrix}.$$

Here also, one can easily examine that the matrices A and B are incompatible. Suppose that we have the following information:

  - $E(X|Y = 2) = 1.5372$;
  - $P\left(X^2 = 1\middle| Y = 1\right) = 0.4235$;
  - $E\left(X^2\middle| Y^2 = 4\right) = 3.2953$;
  - $P(X < 3|Y > 2) = 0.4367$.

Here, we have $\underline{p} = (p_{11}, p_{12}, \ldots, p_{33})$. Subsequently, in this case, our M matrix is

$$M = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & -0.5372 & 0 & 0 & 0.4728 & 0 & 0 & 1.4728 & 0 \\ 0.5865 & 0 & 0 & -0.4235 & 0 & 0 & -0.4235 & 0 & 0 \\ 0 & -1.2953 & 0 & 0 & 1.6147 & 0 & 0 & 7.6147 & 0 \\ 0 & 0 & 0.5733 & 0 & 0 & 0.5733 & 0 & 0 & -0.4367 \end{pmatrix}.$$

We have $\underline{\theta} = (1, 0, 0, 0, 0)$. The iterative algorithm results are given in Table 1.

- **Example** 3**.** Let us consider two conditional probability matrices, where $I = 3$ and $J = 3$ and

$$A = \begin{pmatrix} \frac{2}{7} & \frac{3}{7} & 0 \\ 0 & \frac{4}{7} & \frac{6}{7} \\ \frac{5}{7} & 0 & \frac{1}{7} \end{pmatrix},$$

and

$$B = \begin{pmatrix} \frac{2}{5} & \frac{3}{5} & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{3}{5} & 0 & \frac{2}{5} \end{pmatrix}.$$

Suppose that we have the following information:

- $P\left(X^2 = 1 \middle| Y = 3\right) = 0$;
- $P\left(X^2 = 9 \middle| Y \geq 1\right) = 0.3956$;
- $E\left(X \left| Y^2 = 4\right.\right) = 1.3726$;
- $P(Y > 2|X < 3) = 0.6849$.

Here, we have $\underline{p} = (p_{11}, p_{12}, \ldots, p_{33})$. Also in this case our M matrix is

$$M = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.3956 & 0 & 0 & -0.3956 & 0 & 0 & 0.6044 & 1 & 1 \\ 0 & -0.3726 & 0 & 0 & 0.6374 & 0 & 0 & 0 & 2.6374 \\ -0.6849 & -0.6849 & -0.6849 & -0.6849 & -0.6849 & -0.6849 & 1 & 1 & 0 \end{pmatrix}.$$

Here, $\underline{\theta} = (1, 0, 0, 0, 0)$. The iterative algorithm results are given in Table 3.

Criterion	Optimal value	Matrix P	No. of iterations
D_1	0.002353209	$\begin{pmatrix} 0.0610 & 0.0557 & 0.0484 & 0.0219 \\ 0.0837 & 0.0103 & 0.0485 & 0.0489 \\ 0.2069 & 0.0377 & 0.1111 & 0.0461 \\ 0.0000 & 0.0815 & 0.1345 & 0.0078 \end{pmatrix}$	8
D_2	0.003245132	$\begin{pmatrix} 0.0583 & 0.0571 & 0.0492 & 0.0227 \\ 0.0837 & 0.0132 & 0.0465 & 0.0489 \\ 0.2062 & 0.0352 & 0.1132 & 0.0460 \\ 0.0000 & 0.0821 & 0.1351 & 0.0084 \end{pmatrix}$	11
D_3	0.002129779	$\begin{pmatrix} 0.0681 & 0.0741 & 0.1590 & 0.0000 \\ 0.0841 & 0.0000 & 0.0419 & 0.0949 \\ 0.0000 & 0.0624 & 0.1004 & 0.1448 \\ 0.0389 & 0.0763 & 0.0547 & 0.0000 \end{pmatrix}$	10
D_4	0.005605187	$\begin{pmatrix} 0.0686 & 0.0711 & 0.1538 & 0.0000 \\ 0.0864 & 0.0000 & 0.0419 & 0.0919 \\ 0.0000 & 0.0635 & 0.1026 & 0.1438 \\ 0.0415 & 0.0781 & 0.0561 & 0.0000 \end{pmatrix}$	12
D_5	0.002219034	$\begin{pmatrix} 0.0682 & 0.0704 & 0.1523 & 0.0000 \\ 0.0868 & 0.0000 & 0.0420 & 0.0901 \\ 0.0000 & 0.0640 & 0.1049 & 0.1418 \\ 0.0426 & 0.0798 & 0.0571 & 0.0000 \end{pmatrix}$	10
D_6	0.001537571	$\begin{pmatrix} 0.0686 & 0.0705 & 0.1528 & 0.0000 \\ 0.0867 & 0.0000 & 0.0420 & 0.0915 \\ 0.0000 & 0.0637 & 0.1037 & 0.1432 \\ 0.0421 & 0.0787 & 0.0565 & 0.0000 \end{pmatrix}$	9

Table 1: Minimal (ϵ) incompatibility results for Example 1.

The small values of divergence in Tables 1 to 3 are quite encouraging. There is no evidence that D_1 decreases/increases with the dimension or the values in A and B. The nature of the results were similar for a wide range of other A, B and for A, B of higher dimensions. A similar approach in the case of continuous probability models still remains an open problem and will be taken up in a future article.

Criterion	Optimal value	Matrix P	No. of iterations
D_1	0.000291763	$\begin{pmatrix} 0.0924 & 0.1638 & 0.0000 \\ 0.0000 & 0.1468 & 0.2605 \\ 0.1030 & 0.0000 & 0.2335 \end{pmatrix}$	6
D_2	0.001796547	$\begin{pmatrix} 0.1113 & 0.1469 & 0.0000 \\ 0.0000 & 0.1734 & 0.1302 \\ 0.2013 & 0.0000 & 0.2365 \end{pmatrix}$	9
D_3	0.001796547	$\begin{pmatrix} 0.1142 & 0.1478 & 0.0000 \\ 0.0000 & 0.1737 & 0.1320 \\ 0.2009 & 0.0000 & 0.2316 \end{pmatrix}$	10
D_4	0.001652207	$\begin{pmatrix} 0.1107 & 0.1472 & 0.0000 \\ 0.0000 & 0.1726 & 0.1298 \\ 0.2013 & 0.0000 & 0.2384 \end{pmatrix}$	11
D_5	0.001079299	$\begin{pmatrix} 0.1104 & 0.1471 & 0.0000 \\ 0.0000 & 0.1723 & 0.1293 \\ 0.2013 & 0.0000 & 0.2396 \end{pmatrix}$	9
D_6	0.000609597	$\begin{pmatrix} 0.1103 & 0.1472 & 0.0000 \\ 0.0000 & 0.1721 & 0.1283 \\ 0.2013 & 0.0000 & 0.2398 \end{pmatrix}$	8

Table 2: Minimal incompatibility results for Example 2.

Criterion	Optimal value	Matrix P	No. of iterations
D_1	0.001992807	$\begin{pmatrix} 0.1052 & 0.1691 & 0.0000 \\ 0.0000 & 0.0587 & 0.07112 \\ 0.3062 & 0.0000 & 0.2895 \end{pmatrix}$	7
D_2	0.001453787	$\begin{pmatrix} 0.0921 & 0.1654 & 0.0000 \\ 0.0000 & 0.0632 & 0.0817 \\ 0.2931 & 0.0000 & 0.3045 \end{pmatrix}$	8
D_3	0.002309232	$\begin{pmatrix} 0.0961 & 0.2339 & 0.0000 \\ 0.0000 & 0.1799 & 0.0691 \\ 0.1194 & 0.0000 & 0.3014 \end{pmatrix}$	8
D_4	0.008158526	$\begin{pmatrix} 0.0904 & 0.2317 & 0.0000 \\ 0.0000 & 0.1721 & 0.0653 \\ 0.1224 & 0.0000 & 0.3182 \end{pmatrix}$	8
D_5	0.004180903	$\begin{pmatrix} 0.0860 & 0.2380 & 0.0000 \\ 0.0000 & 0.1885 & 0.1007 \\ 0.1109 & 0.0000 & 0.2759 \end{pmatrix}$	8
D_6	0.00251268	$\begin{pmatrix} 0.0936 & 0.2246 & 0.0000 \\ 0.0000 & 0.1808 & 0.0755 \\ 0.1317 & 0.0000 & 0.2935 \end{pmatrix}$	8

Table 3: Minimal incompatibility results for Example 3.

7.1 Some observations on the concept of ϵ-compatibility

The advantage of the definition of ϵ- compatibility utilized in this article is that the degree of incompatibility could be determined by standard linear programming techniques which has been advocated by Arnold et al. (2001). However, this simplicity comes at a cost. If the information is found to be, say, $.0058$ compatible it is difficult to interpret the meaning of the quantity $.0058$. It is obvious that 0-compatible means completely compatible and 0.01 compatible is better than 0.023 compatible but no interpretation of 0.01 or 0.02 seems available in the literature.

8 Concluding remarks

The problem of finding most nearly compatible distribution(s) starting from two given conditionals (that are incompatible) is not new in the literature. However, there is a scarcity of scholarly work on this topic when in addition to complete specification of two given conditional probability matrices, our informed expert has some additional information in the form of say, conditional percentiles and/or conditional moments etc., among others. Arnold et al. (2001) has provided a brief overview on the issue of finding minimally incompatible distribution in the presence of additional information. However, the role of various existing as well as comparatively newly defined pseudo-divergence measures in search for a minimally incompatible under the presence of additional information has not been adequately addressed. In this paper, we explore the relative performance (equivalently the applicability) of some of the well-known measures of divergence in finding a most nearly compatible distribution in the presence of additional information. The survey made in this paper is far from complete. Compatibility in higher dimensions, such as, given three conditional matrices, say X given Y and Z; Y given X and Z; and Z given X and Y in the presence of additional information (in terms of marginal/conditional moments, percentiles etc.) will be the subject matter of a separate article.

References

Arnold, B.C., E. Castillo, and J.M. Sarabia (1999). *Conditional Specification of Statistical Models*. New York: Springer Verlag.

Arnold, B.C., E. Castillo, and J.M. Sarabia (2001). Quantification of incompatibility of conditional and marginal information. *Communications in Statistics: Theory and Methods 30*, 381–395.

Arnold, B.C. and D.V. Gokhale (1994). On uniform marginal representations of contingency tables. *Statistics and Probability Letters 21*, 311–316.

Arnold, B.C. and D.V. Gokhale (1998). Distributions most nearly compatible with given families of conditional distributions. the finite discrete case. *Test 7*, 377–390.

Arnold, Barry C, Enrique Castillo, José-Mariá Sarabia, Barry C Arnold, Enrique Castillo, and José-Mariá Sarabia (1992). Conditional specification. In *Conditionally Specified Distributions*, pp. 1–6. Springer.

Arnold, Barry C and S James Press (1989). Compatible conditional distributions. *Journal of the American Statistical Association 84*(405), 152–156.

Borzadaran, GR Mohtashami and M Amini (2010). Information measures via copula functions. *Journal of Statistical Research Iran 7*, 47–60.

Cacoullos, Theophilos and H Papageorgiou (1983). Characterizations of discrete distributions by a conditional distribution and a regression function. *Annals of the Institute of Statistical Mathematics 35*, 95–103.

Cressie, Noel and Timothy RC Read (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society Series B: Statistical Methodology 46*(3), 440–464.

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

Ghosh, Indranil (2011). *Inference for the bivariate and multivariate hidden truncated Pareto (type II) and Pareto (type IV) distribution and some measures of divergence related to incompatibility of probability distribution*. University of California, Riverside.

Ghosh, I. and N. Balakrishnan (2015). Study of incompatibility or near compatibility of bivariate discrete conditional probability distributions through divergence measures. *Journal of Statistical Computation and Simulation 85*, 117–130.

Ghosh, I. and S. Nadarajah (2017). On the construction of a joint distribution given two discrete conditionals. *Studia Scientiarum Mathematicarum Hungarica 54*, 178–204.

Ghosh, Indranil and SM Sunoj (2024). Copula-based mutual information measures and mutual entropy: A brief survey. *Mathematical Methods of Statistics 33*(3), 297–309.

Kullback, S. (1959). *Information Theory and Statistics*. New York: Wiley.

Mosteller, Frederick (1968). Association and estimation in contingency tables. *Journal of the American Statistical Association 63*(321), 1–28.

Nadarajah, Saralees and Kostas Zografos (2003). Formulas for rényi information and related measures for univariate distributions. *Information Sciences 155*(1-2), 119–138.

Pardo, L. (2006). *Statistical inference based on divergence measures*. Boca Raton, USA: Chapman & Hall/CRC Press.

Rényi, Alfréd (1961). On measures of entropy and information. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, volume 1: contributions to the theory of statistics*, Volume 4, pp. 547–562. University of California Press.

Shannon, Claude E (1951). Prediction and entropy of printed english. *Bell system technical journal 30*(1), 50–64.

Wesolowski, J (1996). A new conditional specification of the biva riate po isson cond it iona is distribution'. *Statistica Neerlandica 50*(3), 390–393.

Zografos, K and S Nadarajah (2005). Expressions for rényi and shannon entropies for multivariate distributions. *Statistics & Probability Letters 71*(1), 71–84.