

Partial and Total Orderings on Contingency Tables in the Fréchet Class

Ordinamenti Parziali e Totali su Tabelle di Contingenza nella Classe di Fréchet

Marcella Poliscichio

Dipartimento di Metodi Quantitativi per l'Economia

Università di Milano Bicocca

marcella.poliscichio@unimib.it

Riassunto: Lo studio della dipendenza fra due caratteri può essere condotto ipotizzando differenti provenienze dei dati. Nel presente lavoro si è fatto riferimento alla classe di Fréchet, ovvero l'insieme delle distribuzioni congiunte di frequenza di n individui classificati secondo due o più caratteri con distribuzioni marginali prefissate. Inoltre si sono considerati caratteri con scala nominale. Con l'intento di ordinare le tabelle di contingenza della suddetta classe secondo il grado di connessione, sono stati illustrati ordinamenti parziali di dipendenza basati sul concetto di maggiorazione e di trasferimenti che aumentano/diminuiscono la connessione. In particolare, fissando l'attenzione su un ordinamento parziale definito in base al comportamento delle contingenze, si è indagato sulla coerenza di alcune misure sintetiche di dipendenza rispetto a tale ordinamento parziale.

Keywords: Fréchet Class, Partial Ordering, Total Ordering, Principle of Transfers, Contingency

1. Introduction and Notation

Let A and B be two nominal variables with categories denoted respectively by $\{a_1, \dots, a_i, \dots, a_r\}$ and $\{b_1, \dots, b_j, \dots, b_c\}$. Let n be the number of units classified according to A and B . Let n_{ij} denote the frequency of the cell (a_i, b_j) , with $i=1, \dots, r$; $j=1, \dots, c$; and $\sum_i \sum_j n_{ij} = n$. Let $n_{i\cdot}$ denote the marginal frequency of category a_i of A , for $i=1, \dots, r$, and $n_{\cdot j}$ the marginal frequency of category b_j of B , for $j=1, \dots, c$. We assume $n_{i\cdot} \geq 1$ and $n_{\cdot j} \geq 1$ for $i=1, \dots, r$; $j=1, \dots, c$. We say that A and B are independent if $n_{ij} = (n_{i\cdot} n_{\cdot j})/n$ for $i=1, \dots, r$; $j=1, \dots, c$. Of course between A and B there is dependence if at least one of the previous equalities does not hold.

The study of dependence may be carried out in different ways, considering the following situations:

- a) the marginal frequencies of both nominal variables are considered fixed;
- b) the marginal frequencies of one nominal variable are fixed;
- c) only the population size n is fixed, while the marginal frequencies are not constrained.

In this paper we refer to the case a) and our purpose is to study the dependence of A and B in the Fréchet class (1951). If we consider a set of assigned marginal frequencies $\{n_{i.}, i=1, \dots, r; n_{.j}, j=1, \dots, c\}$, the Fréchet class $T(n_{i.}, n_{.j})$ is the set of all contingency tables with the fixed marginal frequencies.

Let $T\{n_{ij}\}$ be a generic table of $T(n_{i.}, n_{.j})$, more synthetically, we will denote it by an $r \times c$ matrix $\mathbf{T}$ whose elements n_{ij} are the joint frequencies of A and B. All tables $\mathbf{T}$ of $T(n_{i.}, n_{.j})$ have the same independence frequencies $(n_{i.}n_{.j})/n = \hat{n}_{ij}$ for $i=1, \dots, r; j=1, \dots, c$, so we denote it by $T\{\hat{n}_{ij}\}$ or $\hat{\mathbf{T}}$. The table $T\{\hat{n}_{ij}\}$ does not necessarily belong to $T(n_{i.}, n_{.j})$ because the joint independence frequencies can be no-integer numbers. To locate and single out all tables $\mathbf{T}$ of $T(n_{i.}, n_{.j})$ see, among the others, Leti (1970), Greselin (2001).

The analysis of dependence between A and B can be realized by the comparisons of the frequencies n_{ij} of the table $\mathbf{T}$ with the corresponding frequencies $\hat{n}_{ij}$ of the independence table $\hat{\mathbf{T}}$. To realize these comparisons, we can use the difference or the ratio of frequencies. So we define the “contingency” $c_{ij} = n_{ij} - \hat{n}_{ij}$, $i=1, \dots, r; j=1, \dots, c$; and the “relative contingency” $\rho_{ij} = c_{ij}/\hat{n}_{ij}$, for $i = 1, \dots, r; j = 1, \dots, c$.

Our interest is to graduate the different contingency tables according to their degree of dependence in such a way that tables more distant from the independence table $\hat{\mathbf{T}}$ have a higher degree of dependence.

Naturally we may define the opposite extreme situation, the maximal dependence, and graduate the tables with respect to it. Unfortunately, given a set of marginal frequencies, such situation exists, but it may be not unique. For this reason we must base our analysis on the former approach.

The introduction of partial orderings of dependence in the tables of the Fréchet class is a valid tool to graduate a subset of tables. Moreover, total orderings of all tables of the Fréchet class may be introduced with the requirement to be coherent with some partial orderings.

2. Partial Orderings on the Fréchet Class

In this section we relate some partial ordering relations between tables of the Fréchet class based on the common point of view of majorization ordering. (Hardy, Littlewood, Polya, 1934; Marshall, Olking, 1979).

Definition 1: Let $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ be vectors in R^n with $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$. Then $\mathbf{x}$ majorizes $\mathbf{y}$, denoted by $\mathbf{x} \geq^M \mathbf{y}$, if $\sum_{i=1}^k x_{(i)} \leq \sum_{i=1}^k y_{(i)}$, for $k=1, \dots, n-1$, where $x_{(1)} \leq \dots \leq x_{(n)}$ and $y_{(1)} \leq \dots \leq y_{(n)}$.

It is well known that, assuming $x_{(1)} \geq 0$ and $y_{(1)} \geq 0$, the majorization ordering is equivalent to the Lorenz ordering based on the Lorenz curve and to the Pigou-Dalton's

principle of transfers. So $\mathbf{x} \geq^M \mathbf{y}$ if and only if $\mathbf{x} \geq^L \mathbf{y}$, that is the Lorenz curve based on the n elements of $\mathbf{x}$ is lower than or equal to the Lorenz curve of the elements of $\mathbf{y}$. Moreover $\mathbf{x} \geq^M \mathbf{y}$ if and only if the ordered elements of $\mathbf{y}$ may be obtained from those of $\mathbf{x}$ with a finite number of transfers that reduce the concentration.

A generalization of this ordering applied to contingency tables of the Fréchet class has been introduced by Joe (1985).

Definition 2: Let $\mathbf{T}_1$ and $\mathbf{T}_2$ be elements of $T(n_i, n_j)$. Then $\mathbf{T}_1$ majorizes $\mathbf{T}_2$ and denote it by $\mathbf{T}_1 \geq^J \mathbf{T}_2$, if $\text{vec}(\mathbf{T}_1) \geq^M \text{vec}(\mathbf{T}_2)$, where $\text{vec}(\mathbf{T})$ is the $r \times c$ vector obtained by piling the columns of $\mathbf{T}$.

Joe's ordering has two drawbacks. First of all, in this ordering there is no certainty that we are comparing the joint frequency of each cell of table $\mathbf{T}_1$ with the corresponding one of table $\mathbf{T}_2$. Indeed, when we arrange the elements of a table in non-decreasing order we have lost their reference to the cells. In other words, we are treating the joint frequencies of A and B just like numbers of a matrix. We are ignoring that a frequency distribution is the result of a set function that associates to each pair of categories of A and B the number of units having simultaneously the considered categories. In this way we are thinking that it is meaningful to compare a cell of the distribution $\mathbf{T}_1$ with a different cell of another distribution $\mathbf{T}_2$. But we do not share this opinion.

The other drawback of Joe's ordering is that the definition does not take into account the fixed marginal frequencies, even if the comparison is made in the Fréchet class.

The underlying idea of this approach is to analyse how the n units of the population are divided into the $r \times c$ cells of the tables. In this way the two extreme situation of minimum and maximum concentration of the n units among the $r \times c$ cells of the table are respectively the table with the same frequency n/rc in each cell and the set of tables with only one positive frequency equal to n . Nevertheless being in the Fréchet class, this analysis of the concentration of frequencies must be conditioned to the fixed marginal distributions. So it may be that the two above-said extreme situations are not compatible with the marginal distributions. Indeed the total concentration of the n units in a unique cell is not compatible with marginals such that $n_i \geq 1$ and $n_j \geq 1$ for $i=1, \dots, r$; $j=1, \dots, c$.

Moreover, it is possible that there is not a unique table $\mathbf{T}$ of the Fréchet class which is majorized by all the others of the class. This unique table exists if the marginal frequencies are uniformly divided into the cells.

Another partial ordering based on majorization ordering and fixed marginal frequencies has been introduced by Scarsini (1990). We have here adapted the definition to the Fréchet class based on integer numbers instead of probability mass.

Definition 3: Let $\mathbf{T}\{n_{ij}\}$ be a table of the Fréchet class $T(n_i, n_j)$. Let ρ_{ij}^* , for $i = 1, \dots, r$; $j = 1, \dots, c$, be the ratio between the frequency n_{ij} and the independence frequency $\hat{n}_{ij}$. The Lorenz curve of $\mathbf{T}\{n_{ij}\}$, denoted by L_T , is defined by the concentration curve obtained by the ratios ρ_{ij}^* whose weights are given by the independence frequencies $\hat{n}_{ij}$.

Definition 4: Let $\mathbf{T}_1$ and $\mathbf{T}_2$ be elements of the Fréchet class $T(n_i, n_j)$. Let L_1 and L_2 be respectively the Lorenz curves of $\mathbf{T}_1$ and $\mathbf{T}_2$. Between A and B there is more dependence in $\mathbf{T}_1$ than in $\mathbf{T}_2$, denoted by $\mathbf{T}_1 \geq^D \mathbf{T}_2$, if and only if L_1 is lower than or equal to L_2 .

Obviously the use of the ratios ρ_{ij}^* in the definition of this ordering makes itself a partial ordering which takes into account the fixed marginal frequencies. Moreover, with respect to Joe's ordering, the independence frequencies have a primary role in the ordering of the elements of a contingency table. In fact, the comparison of two cells by the ratio ρ_{ij}^* may order them in an opposite direction with respect to the ordering based on the frequencies n_{ij} . This result may depend on a different ordering of the independence frequencies in relation to the corresponding frequencies of the table under discussion. Naturally, being ratios $\rho_{ij}^* = \rho_{ij} + 1$, the same ordering of the cells of a table may be obtained using the relative contingencies. Moreover the definition of the Lorenz curve of a table $\mathbf{T}$ using the ratios ρ_{ij}^* weighted by the independence frequencies causes the existence of a unique minimal table equal to the independence table $\hat{\mathbf{T}}$. Of course if we are in the Fréchet class it may be that $\hat{\mathbf{T}} \notin T(n_i, n_j)$.

The partial ordering D presents the same defect of the previous ordering in relation to the absence of the reference to the cells.

New partial orderings of dependence based on the concept of “transfers of frequencies” overcome this defect and are defined in the Fréchet class (Greselin, Zenga, 2001).

Definition 5: Let $\mathbf{T}_1 \{n_{ij}\}$ be an element of the Fréchet class $T(n_i, n_j)$. Let (h;k), (h;g), (l;k), (l;g), with $l \neq h$ and $k \neq g$, be four cells such that:

$$n_{hk} \geq \hat{n}_{hk} + 1 \quad n_{lg} \geq \hat{n}_{lg} + 1 \quad 0 \leq n_{hg} \leq \hat{n}_{hg} - 1 \quad 0 \leq n_{lk} \leq \hat{n}_{lk} - 1.$$

Let $\mathbf{T}_2 \{n'_{ij}\}$ be the table of $T(n_i, n_j)$ whose frequencies are:

$$n'_{hk} = \hat{n}_{hk} - 1, \quad n'_{lg} = \hat{n}_{lg} - 1, \quad n'_{hg} = \hat{n}_{hg} + 1, \quad n'_{lk} = \hat{n}_{lk} + 1, \quad n'_{ij} = n_{ij} \text{ for all the other cells.}$$

The transformation $\mathbf{T}_1 \rightarrow \mathbf{T}_2$ is a frequency transfer that decreases the directional dependence.

In other words, the directional dependence decreases by transferring one frequency from n_{hk} to n_{hg} and one frequency from n_{lg} to n_{lk} . With the aim to define an ordering relation of dependence, we must wonder what conditions must be satisfied so that the distribution $\mathbf{T}_1$ reflects more dependence than $\mathbf{T}_2$.

Definition 6: Let $\mathbf{T}_1 \{n_{ij}\}$ and $\mathbf{T}_2 \{n'_{ij}\}$ be two tables of $T(n_i, n_j)$. Between A and B there is more directional dependence in $\mathbf{T}_1$ than in $\mathbf{T}_2$, denoted by $\mathbf{T}_1 \geq^{DD} \mathbf{T}_2$, if and only if the contingencies satisfy the following conditions:

a) c_{ij} and c'_{ij} have the same sign for $i = 1, \dots, r; j = 1, \dots, c$;

b) $|c_{ij}| \geq |c'_{ij}|$ for $i=1, \dots, r; j=1, \dots, c$; with strict inequality for at least one cell.

It is evident that the directional dependence ordering is equivalent to consider an appropriate transformation from a table to another one of the Fréchet class that is a frequency transfer with decreasing directional dependence.

The directional dependence ordering is very restrictive because it requires $r \times c$ comparisons related to the $r \times c$ cells of the tables.

Moreover, for each comparison is necessary to evaluate the signs and the absolute values of the contingencies.

So Greselin and Zenga have weakened the criterion considering only the absolute value of contingencies.

Definition 7: Let $\mathbf{T}_1 \{n_{ij}\}$ and $\mathbf{T}_2 \{n'_{ij}\}$ be two tables in $T(n_{i.}, n_{.j})$. Between A and B there is more intensity of dependence in $\mathbf{T}_1$ than in $\mathbf{T}_2$, denoted by $\mathbf{T}_1 \geq^{ID} \mathbf{T}_2$, if and only if $|c_{ij}| \geq |c'_{ij}|$ for $i=1, \dots, r; j = 1, \dots, c$; with strict inequality for at least one cell. Between A and B there is the same intensity of dependence in $\mathbf{T}_1$ and $\mathbf{T}_2$, denoted by $\mathbf{T}_1 \approx^{ID} \mathbf{T}_2$, if and only if $|c_{ij}| = |c'_{ij}|$ for $i = 1, \dots, r; j = 1, \dots, c$.

Remark 1 The partial ordering DD implies the partial ordering ID.

Another obvious consideration is that the DD or ID orderings can be equivalently defined on the relative contingencies $|\rho_{ij}| = |c_{ij}| / \hat{n}_{ij}$, having all the tables of the Fréchet class the same independence frequencies $\hat{n}_{ij}$.

3. Total Orderings on the Fréchet Class

A partial ordering defined on the Fréchet class does not order all the tables of the class. For such purpose we need a total ordering that synthetizes the degree of dependence of a table with a number, a dependence index.

Dependence measures can be obtained in accordance with the different approaches to the study of dependence.

From a general point of view, total dependence orderings can be obtained by operating appropriate synthesis of

- quantities defined comparing n_{ij} with $\hat{n}_{ij}$;
- comparisons between all the partial distributions of one variable;
- comparisons between the partial distributions of one variable and the respective marginal distribution.

Now partial and total orderings can be connected analysing the behavior of a total ordering with respect to a partial ordering. In particular our interest is to study the behavior of dependence indexes with respect to the intensity dependence ordering.

Definition 8: Let $\mathbf{T}_1$ and $\mathbf{T}_2$ be tables in the Fréchet class $T(n_{i.}, n_{.j})$ with $\mathbf{T}_1 \geq^{ID} \mathbf{T}_2$. We say that the general dependence index I is coherent with the partial ordering ID if and

only if $I(T_1) > I(T_2)$, being $I(T_1)$ and $I(T_2)$ the index I valued respectively for T_1 and T_2 .

Of course equivalent definitions may be obtained using other partial orderings. Considering dependence indexes obtained by a synthesis of quantities defined comparing n_{ij} with $\hat{n}_{ij}$, just like contingencies $|c_{ij}|$ or relative contingencies $|\rho_{ij}|$, we deduce that all dependence indexes that are non-negative and strictly increasing function of (relative) contingencies are coherent with the ID ordering.

So the weighted arithmetic mean of $|\rho_{ij}|$ with weights $\hat{n}_{ij}$:

$$M_1(|\rho|) = \frac{1}{n} \sum_i \sum_j |\rho_{ij}| \hat{n}_{ij} = \frac{1}{n} \sum_i \sum_j |c_{ij}|$$

satisfies this conditions.

Also the weighted quadratic mean of $|\rho_{ij}|$ with weights $\hat{n}_{ij}$:

$$M_2(|\rho|) = \left[\frac{1}{n} \sum_i \sum_j |\rho_{ij}|^2 \hat{n}_{ij} \right]^{1/2} = \left[\frac{1}{n} \sum_i \sum_j c_{ij}^2 / \hat{n}_{ij} \right]^{1/2} = \left[\frac{X^2}{n} \right]^{1/2},$$

where X^2 is the classic Pizzetti-Pearson's statistics, is coherent with the ID ordering.

Comparing the partial distributions of A with respect to the categories of B, we may consider the global dependence index of second order (Castellano, 1962):

$${}^2C_{g,A/B} = \left[\sum_{j=1}^c \sum_{h=1}^c \left(\sum_{i=1}^r \left| \frac{n_{ij}}{n_j} - \frac{n_{ih}}{n_h} \right|^2 \right) \frac{n_j n_h}{n^2} \right]^{1/2}$$

or equivalently (Zanella, 1988, § 2.3.3):

$${}^2C_{g,A/B} = \sqrt{2} \left\{ \left[1 - \sum_{i=1}^r (n_{i.} / n)^2 \right] - \sum_{j=1}^c \frac{n_j}{n} \left[1 - \sum_{i=1}^r (n_{ij} / n_j)^2 \right] \right\}^{1/2}.$$

This index may be rewritten as a function of the relative contingencies with the above-said properties:

$${}^2C_{g,A/B} = \sqrt{2} \left\{ \sum_{i=1}^r \left[\sum_{j=1}^c (\rho_{ij} + 1)^2 \frac{n_j}{n} - 1 \right] \left(\frac{n_{i.}}{n} \right)^2 \right\},$$

so it is coherent with the ID ordering.

When we compare the partial distributions of A with the respective marginal distribution using the total dependence index of second order (Gini, 1955), we have the same result, in fact

$${}^2C_{T,A/B} = \left\{ \sum_{j=1}^c \frac{n_j}{n} \left[\sum_{i=1}^r \left(\frac{n_{ij}}{n_j} - \frac{n_{i.}}{n} \right)^2 \right] \right\}^{1/2} = \frac{1}{\sqrt{2}} {}^2C_{g,A/B}.$$

A simple but also meaningful measure of dependence coherent with the intensity dependence ordering may be expressed by the relative position that the table under question assumes among all the tables of the Fréchet class.

Let $\mathbf{T}$ be a table of the Fréchet class $T(n_{i.}, n_{.j})$. Let $T(n_{i.}, n_{.j})_{ID/T}$ be the subset of the Fréchet class composed by all tables comparable with $\mathbf{T}$ according to the ID ordering. A measure of dependence for the table $\mathbf{T}$ is

$$F_{ID}(\mathbf{T}) = \frac{\text{card}\{\mathbf{C} \in T(n_{i.}, n_{.j})_{ID/T} : \mathbf{C} <^{ID} \mathbf{T}\}}{\text{card}\{\mathbf{C} \in T(n_{i.}, n_{.j})_{ID/T}\}}$$

(Greselin, Zenga; 2001).

The dependence index $F_{ID}(\mathbf{T})$ has the quality of an immediate comprehension and the coherence with the partial ordering considered. Unfortunately it may be valued only for the subset of the tables of the Fréchet class comparable with $\mathbf{T}$.

Nevertheless this is an appropriate way to define a dependence measure, in fact it may be generalized to any partial ordering given on the Fréchet class. Moreover, the same procedure may be used for any total dependence ordering.

So, let the function $f : T(n_{i.}, n_{.j}) \rightarrow R$ be a dependence measure, the relative position of the table $\mathbf{T}$ among the others of the class, that is

$$I_f(\mathbf{T}) = \frac{\text{card}\{\mathbf{C} \in T(n_{i.}, n_{.j}) : f(\mathbf{C}) \leq f(\mathbf{T})\}}{\text{card}\{\mathbf{C} \in T(n_{i.}, n_{.j})\}}$$

may be assumed as a normalized dependence measure.

4. Conclusions

With the purpose of analyzing the dependence of the contingency tables of the Fréchet class and graduate them according to the degree of dependence, we have made some choices. First of all, we retain that the degree of dependence of a contingency table must be valued with respect to the independence distribution. Moreover, we think that the introduction of a partial ordering on the Fréchet class is an helpful tool to graduate the tables. So, we have considered some partial orderings with the common point of view of majorization and therefore the principle of transfers. Nevertheless we do not share the opinion that the application of the majorization ordering to contingency tables can be done treating the frequencies of a table just like the elements of a matrix. In fact the frequencies are strictly linked to a couple of categories of the variables, so we retain that it is reasonable to compare a cell of a contingency table with the same cell of another contingency tables of the Fréchet class. These opinions lead to the introduction of transfers of frequencies among cells, leaving fixed the marginal frequencies. Obviously, these transfers must decrease or increase the degree of dependence measured, for examples, by the difference between the frequencies of a table of the Fréchet class and the corresponding independence frequencies. In this way, we illustrate the DD partial ordering and, weakening the ranking criterion, the ID ordering. With the intention to analyze the behavior of some measures of dependence in relation to the ID ordering, we prove that the Mortara's index, the Pearson's X^2 , the Castellano's index of global dependence and Gini's index of total dependence are coherent with the ID ordering. Another index with the same feature, but also with the quality of an immediate comprehension, is the relative frequency of tables that precede, in the ID ordering, the table under question, among all tables of the Fréchet class comparable with it according to the ID ordering. This approach for defining a dependence measure may be extended

to any total ordering of the tables in the Fréchet class. So, it is useful to introduce a normalized dependence measure defined as the relative frequency of tables with the value of the dependence index less than or equal to that of the table in discussion, among all the tables of the Fréchet class.

References

- Castellano V. (1962) *Istituzioni di Statistica*, Ilardi, Roma.
- Fréchet M. (1951) Sur le tableaux de corrélation dont les marges sont donnés, *Annales de l'Université de Lion*, III Fasc., 4, Sciences.
- Gini C. (1955) *Corso di Statistica*, Veschi, Roma.
- Greselin F. (2001) Counting and enumerating rectangular arrays with given margins, submitted for publication to *Discrete Mathematics*, Elsevier Publ.
- Greselin F. and Zenga M. (2001) Partial and Total Ordering on Tables in the Fréchet class, *Dipartimento di Metodi Quantitativi per l'Economia*, Rapporto di Ricerca N. XXXVIII.
- Hardy J. H., Littlewood J. E. and Polya G. (1934) *Inequalities*, Cambridge University Press, Cambridge.
- Joe H. (1985) An Ordering of Dependence for Contingency Tables, *Linear Algebra and Its Applications*, 70, 89-103.
- Leti G. (1970) La Distribuzione della Tabelle della Classe di Fréchet, *Metron*, vol. XXVIII, n. 1-4.
- Marshall A.W. and Olkin I. (1979) *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York.
- Scarsini M. (1990) An Ordering of Dependence, in: *Topics in Statistical Dependence*, Block H. W., Sampson A. R., Savits T. H. (Eds.), Hayward Institute of Mathematical Statistics.
- Zanella A. (1988) *Lezioni di Statistica. Parte II: Struttura di Dati in due o più Dimensioni. La Connessione*, Vita e Pensiero, Milano.