Concentration of Multivariate Statistical Tables

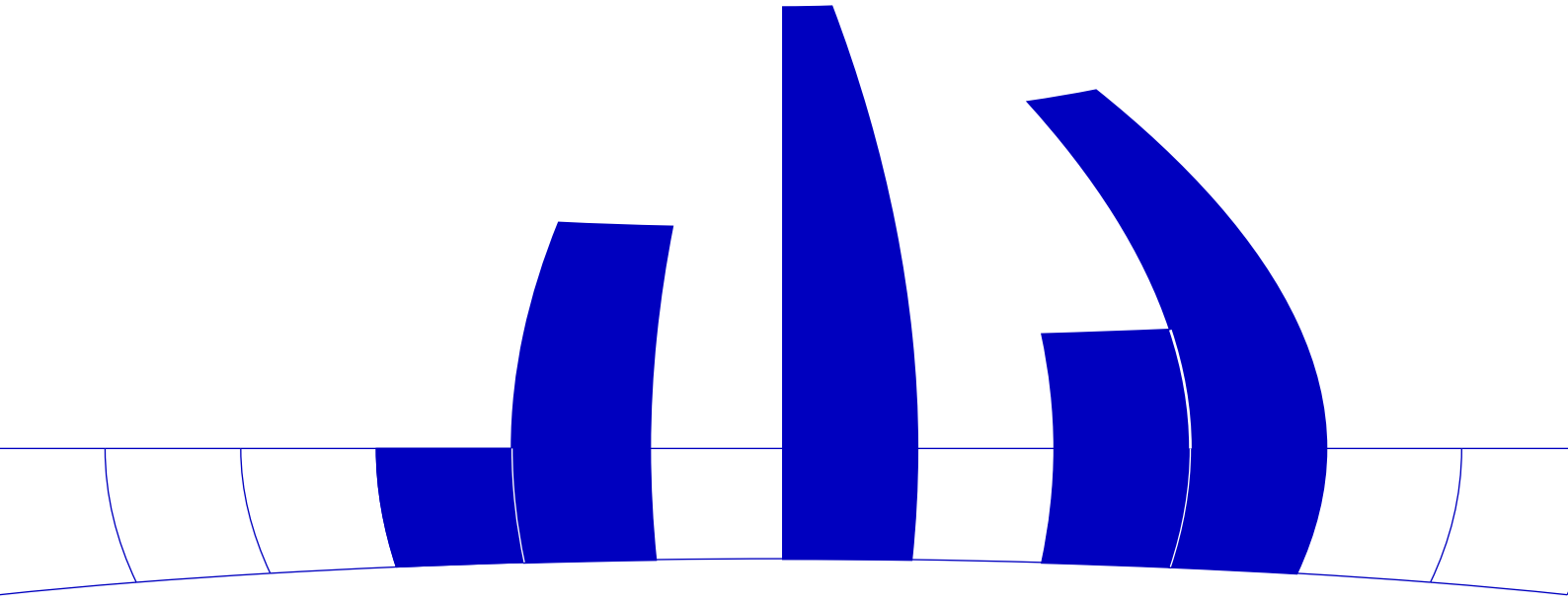
Helmut Strasser

Institut für Statistik
Wirtschaftsuniversität Wien

Forschungsberichte

Bericht 6
April 1990

http://statistik.wu-wien.ac.at/



Concentration of Multivariate Statistical Tables

Helmut Strasser *

November 1991

Abstract

In this paper we lay the foundation of the concentration measurement for statistical tables with more than two columns. A concentration function and a coefficient of concentration are defined which can be used in a similar way as the Lorenz diagram and the Gini coefficient in case of tables with two columns. For computational purposes we derive an explicit formula and give an algorithm. The mathematics behind our approach is formally equivalent to the statistical theory of the comparison of experiments.

1 Introduction

Given the empirical distribution of an extensive quantitative variable it is a basic statistical problem going back to Lorenz, [8], to consider its concentration. The basic mathematical reference for the the subject is Marshall and Olkin, [9].

In the most simple case the empirical distribution consists of a data list $x = (x_1, x_2, \ldots, x_N)$ of nonnegative numbers $x_i \geq 0$ satisfying $\sum_{i=1}^N x_i = 1$. The set of all such sequences constitutes the unit simplex in $\mathbb{R}^N$ and is denoted by S_N. The simple concept of majorization is most easily explained by the principle of transfers (cf. Dalton, [3]). One says that a sequence $y \in S_N$ arises by a transfer from $x \in S_N$ if there exists a pair of indices i, j and a number $\alpha \in [0, 1]$ such that

$$\begin{aligned} y_i &= (1-\alpha)x_i + \alpha x_j, \\ y_j &= \alpha x_i + (1-\alpha)x_j. \end{aligned}$$

Such a transfer is levelling since

$$\min\{x_i, x_j\} \leq \min\{y_i, y_j\} \leq \max\{y_i, y_j\} \leq \max\{x_i, x_j\}.$$

This is the background of

(1.1) DEFINITION *A sequence $x \in S_n$ majorizes a sequence $y \in S_N$ ($x \succ y$) if y arises from x by a sequence of transfers.*

Let us denote the sorted version of a sequence $x \in S_N$ by $(x_{(1)}, x_{(2)}, \ldots, x_{(N)})$. The following fundamental result is due to Hardy, Littlewood and Polya, [5].

(1.2) THEOREM *The relation $x \succ y$ is valid iff*

$$\sum_{i=1}^k x_{(i)} \leq \sum_{i=1}^k y_{(i)} \quad \text{for every } k = 1, 2, \ldots, N.$$

*Prof. Dr. Helmut Strasser, Department of Statistics, University of Economics and Business Administration, Augasse 2–6, A–1090 Vienna

If the points $(\frac{k}{N}, \sum_{i=1}^{k} x_{(i)})$, $k = 0, 1, 2, \ldots, N$, are connected by straight lines then we obtain the so–called Lorenz diagram (cf. Ferschl, [4], Definition 3.26), which is the graph of the function

$$L_x(t) = \sum_{i=1}^{[Nt]} x_{(i)} + (Nt - [Nt])x_{[Nt]+1}, \quad t \in [0, 1].$$

It is clear that $x \succ y$ iff $L_x \leq L_y$. The order relation[1] of majorization can thus be studied considering graphs of functions.

Let $x = (x_1, x_2, \ldots, x_N)$ be a sequence in S_N, where the values $(q_1, q_2, \ldots, q_n)$ occur with the relative frequencies $(f_1, f_2, \ldots, f_n)$. The table

$$A = \begin{pmatrix} f_1 & f_1 q_1 \\ f_2 & f_2 q_2 \\ \vdots & \vdots \\ f_n & f_n q_n \end{pmatrix} \tag{1.3}$$

contains full information concerning the sorted sequence x and can thus be used as basis for the computation of the Lorenz diagram.

In the field of statistical concentration measurement one is mainly concerned with tables

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{m1} & a_{m2} \end{pmatrix}$$

with two columns in S_m (cf. Ferschl, [4], pp. 130 ff.). Let us denote the set of all such tables by $\mathcal{T}_2$. The rows of such a table arise from subsets of the underlying statistical population, called cases, and the columns belong to extensive quantitative variables. In most applications the first column contains the relative frequencies.

The concentration of a table A is usually described by its Lorenz diagram which is understood as if the table arises from a single sequence in the sense of (1.3). Practically, the ratios $q_i = a_{i2}/a_{i1}$ are computed and sorted:

$$q_{i_1} \leq q_{i_2} \leq \ldots \leq q_{i_m}.$$

If the points $(\sum_{i=1}^{i_k} a_{i1}, \sum_{i=1}^{i_k} a_{i2})$, $k = 0, 1, 2 \ldots, m$, are connected by straight lines then we obtain the Lorenz diagram L_A of the table A (Ferschl, [4], p.130).

The Lorenz diagram defines an order relation among all tables in $\mathcal{T}_2$.

(1.4) DEFINITION *Let A and B be tables in $\mathcal{T}_2$. The table A is at least as concentrated as the table B ($A \succ B$) if $L_A \leq L_B$.*

The relation between majorization and concentration is obvious: A sequence $x \in S_n$ majorizes a sequence $y \in S_n$ iff

$$\begin{pmatrix} \frac{1}{n} & x_1 \\ \frac{1}{n} & x_2 \\ \vdots & \vdots \\ \frac{1}{n} & x_n \end{pmatrix} \succ \begin{pmatrix} \frac{1}{n} & y_1 \\ \frac{1}{n} & y_2 \\ \vdots & \vdots \\ \frac{1}{n} & y_n \end{pmatrix}.$$

Thus, majorization of sequences corresponds to the concentration order between tables whose first column contains the uniform distribution.

[1] By *order relation* we always mean a partial ordering, not a linear ordering.

It should be noted that the concept of concentration is much more general than the concept of majorization. It is possible to compare tables with different numbers of rows and it is not at all necessary that one column contains the uniform distribution. The concentration order is able to compare tables where each column arises from any extensive quantitative variable.

In the present paper we consider the problem of concentration measurement for tables with more than two columns. We will try to lay a theoretical foundation leading to concepts which are computationally accessible.

From an abstract point of view the concentration problem of tables is mathematically equivalent to the sufficiency problem of statistical decision theory. We give some details in section 5. Our solution of the concentration problem uses in an essential way mathematical ideas of the statistical theory of experiments. This theory has been initiated by Blackwell, [1]. Main results for the finite case are due to Torgersen, [12]. Important sources are Heyer, [7], and Torgersen, [13].

In section 2 we discuss several equivalent concepts of concentration for tables with two columns. One of these concepts is the starting point for the definition of multivariate concentration in section 3. It is shown that our concept of multivariate concentration can be studied by the analysis of a certain function called the concentration function. For tables with two columns our concentration function essentially equals the conjugate convex function of the Lorenz curve. As an application we show in section 4 how the amount of concentration of a multivariate table can be measured by a concentration coefficient which is analogous to the usual Gini coefficient. An explicit formula as well as an algorithm for this coefficient of concentration is eleborated and its numerical performance is illustrated. Section 5 contains those proofs which are included for reasons of completeness.

2 Concentration of Tables with Two Columns

The aim of the present paper is to define a concentration order for tables with more than two columns. The definition (1.4) cannot be extended to this case. Therefore in this section we will discuss some equivalent mathematical interpretations of the concentration order of tables with two columns. One of these interpretations will be the starting point of our extension to tables with more than two columns.

Let

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{m1} & a_{m2} \end{pmatrix}$$

be an arbitrary table in $\mathcal{T}_2$. In order to motivate our next definition let us discuss some intuitive ideas.

(2.1) DISCUSSION Assume that the variable of the first column is the relative frequency. The fact that the variable of the second column is highly concentrated could be described as follows: There is a partition of the set $\{1, 2, \ldots, m\}$ into two subsets $M_1,\ M_2$, which splits the the total mass of each variable

$$\begin{aligned} 1 = \sum_{i=1}^{m} a_{i1} = \sum_{i \in M_1} a_{i1} + \sum_{i \in M_2} a_{i1}, \\ 1 = \sum_{i=1}^{m} a_{i2} = \sum_{i \in M_1} a_{i2} + \sum_{i \in M_2} a_{i2}, \end{aligned} \tag{2.2}$$

in such a way that

$$\begin{aligned} &\sum_{i \in M_1} a_{i1} \text{ is large}, \quad \sum_{i \in M_2} a_{i1} \text{ is small}, \\ &\sum_{i \in M_1} a_{i2} \text{ is small}, \quad \sum_{i \in M_2} a_{i2} \text{ is large}. \end{aligned}$$

Unfortunately, this concept does not yet lead to a satisfactory mathematical theory. However, the difficulty is overcome if the partitions into subsets used in (2.2) are replaced by weighted partitions:

$$1 = \sum_{i=1}^{m} a_{i1} = \sum_{i=1}^{m} \xi_i a_{i1} + \sum_{i=1}^{m} (1-\xi_i) a_{i1},$$
$$1 = \sum_{i=1}^{m} a_{i2} = \sum_{i=1}^{m} \xi_i a_{i2} + \sum_{i=1}^{m} (1-\xi_i) a_{i2},$$

where $(\xi_1, \xi_2, \ldots, \xi_m) \in [0,1]^m$ is a sequence of weights. Then high concentration means that the weights can be chosen such that

$$\sum_{i=1}^{m} \xi_i a_{i1} \text{ is large}, \quad \sum_{i=1}^{m} (1-\xi_i) a_{i1} \text{ is small},$$
$$\sum_{i=1}^{m} \xi_i a_{i2} \text{ is small}, \quad \sum_{i=1}^{m} (1-\xi_i) a_{i2} \text{ is large}.$$

It will be shown that in this way we arrive at an explanation of the concentration order in $\mathcal{T}_2$.

For reasons of convenience we introduce a particular name for the condition we have in mind. Let

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{m1} & a_{m2} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ \vdots & \vdots \\ b_{n1} & b_{n2} \end{pmatrix}$$

be tables in $\mathcal{T}_2$.

(2.3) DEFINITION *The relation $A \supseteq_K B$ holds iff the following conditions is fulfilled:*
For every sequence of weights $(\eta_1, \eta_2, \ldots, \eta_n) \in [0,1]^n$ there exists a sequence of weights $(\xi_1, \xi_2, \ldots, \xi_m) \in [0,1]^m$ such that

$$\sum_{i=1}^{m} \xi_i a_{i1} \geq \sum_{j=1}^{n} \eta_j b_{j1}$$
$$\sum_{i=1}^{m} (1-\xi_i) a_{i2} \geq \sum_{j=1}^{n} (1-\eta_j) b_{j2}.$$

The following theorem states that several mathematical definitions of a concentration order for tables with two columns are equivalent.

(2.4) THEOREM *The following assertions are equivalent:*

1. *$A \succ B$, (i.e. $L_A \leq L_B$).*
2. *$A \supseteq_K B$.*
3. *For every convex function $f : [0, \infty) \to \mathbb{R}$ we have*

$$\sum_{i=1}^{m} a_{i1} f\left(\frac{a_{i2}}{a_{i1}}\right) \geq \sum_{i=1}^{n} b_{i1} f\left(\frac{b_{i2}}{b_{i1}}\right).$$

4. *There is a transition, i.e. a column stochastic $n \times m$–matrix U satisfying $B = UA$.*

This theorem is well–known. In a measure theoretic framework it is a basic result of statistical decision theory. For completeness we will give a proof in section 5.

Condition (4) deserves additional explanation. This condition is the general version of the "principle of transfers" which was the starting point for the concept of majorization. In fact, transfers can be written as transitions (see Marshall and Olkin, [9], 2B). But a transition is a more general object than a transfer. For example, any grouping of data can be written as a transition: A transition U is a grouping of data if every column of U contains exactly one entry equal to 1 (the other entries have to be 0). A grouping combines rows of a table by adding the entries. It is clear that grouping diminishes the concentration of a table. This is formally expressed by the equivalence $(2) \Leftrightarrow (4)$.

Let us discuss this theorem with regard to the theory of concentration. We are interested in the question how the concepts and assertions of theorem (2.4) can be extended to tables with more than two columns.

The relation $A \succ B$ used in condition (1) cannot be extended to tables with more than two columns since for such tables a Lorenz curve cannot be defined. The concepts used in conditions (2), (3) and (4), however, can be extended without any complication. For example Pflug, [10], Definition 4, applies the convex function criterion of condition (3) for defining a concentration order for multivariate variables.

It is known from Blackwell, [1], [2], that the equivalence $(3) \Leftrightarrow (4)$ is valid also for tables with more than two columns. However, the extension of condition (2) yields a coarser order relation than that of (3) or (4). Thus, it seems most natural to define a concentration order for tables with more than two columns by an extension of the conditions (3) or (4).

But we will make another proposal. The reason for our proposal is the fact known from decision theory that the order structures derived from (3) or (4) are difficult to treat numerically.

The basic message of the present paper is that a concentration order for tables with more than two columns which is based on a suitable extension of $\supseteq_K$ (i.e. condition (2)) has a major advantage: It is then possible the reduce the comparison of tables to the comparison of graphs of functions. This gives us an analytical instrument which can be handled in a similar way as the Lorenz curve is used for tables with two columns. We will carry out this program in the subsequent sections.

3 Multivariate Concentration

Comparing tables with more than two columns we may choose between several non–equivalent approaches. An assertion as that of theorem (2.4) seems not to be valid, in general. We will compare two alternative concepts of an order relation for tables which are oriented at conditions (2) and (4) of theorem (2.4).

Let us begin with the generalization of condition (2) of theorem (2.4).

(3.1) DISCUSSION Let

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1r} \\ a_{21} & a_{22} & \dots & a_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mr} \end{pmatrix}$$

be a table with r columns in S_m. The rows correspond to subsets of the underlying statistical population, later on called cases, and each column contains the values of an extensive quantitative variable defined on those cases.

The fact that the variables are highly concentrated on disjoint subsets of cases could be described as follows: There is a partition of the set of cases $\{1, 2, \dots, m\}$ into r subsets $M_1, M_2, \dots, M_r$, such that different column distributions are concentrated mainly on different sets. In other words: The partition $(M_1, M_2, \dots, M_r)$ should be such that the numbers

$$\sum_{i \in M_1} a_{i1}, \sum_{i \in M_2} a_{i2}, \dots, \sum_{i \in M_r} a_{ir}$$

are large.

A mathematical satisfactory definition, however, requires to replace the partitions into subsets by weighted partitions.

(3.2) DEFINITION *A weighted partition of the set* $\{1,2,\dots,n\}$ *is a set of functions* $\phi_s : \{1,2,\dots,n\} \longrightarrow [0,1],\ s = 1,2,\dots,r$ *satisfying* $\sum_{s=1}^{r} \phi_s = 1$.

Recall, that in section 2 for $r = 2$ we denoted $\phi_1(i) = \xi_i$ and $\phi_2(i) = 1 - \xi_i$, $i = 1,2,\dots,m$. (In mathematical terminology a weighted partition is called a partition of the unity.)

Now we are in a position to give a mathematically precise definition of the comparison of tables in the sense of condition (2) of theorem (2.4). For notational convenience let $\mathcal{T}_r$ be the set of tables with r columns.

(3.3) DEFINITION *Let* $A \in \mathcal{T}_r$ *and* $B \in \mathcal{T}_r$ *be tables. Assume that* A *has* m *columns and* B *has* n *columns. The table* A *is more concentrated than the table* B *(denoted by* $A \supseteq_K B$*) if the following condition is fulfilled:*

For any weighted partition $(\psi_1, \psi_2, \dots, \psi_r)$ *of* $\{1,2,\dots,n\}$ *there exists a weighted partition* $(\phi_1, \phi_2, \dots, \phi_r)$ *of* $\{1,2,\dots,m\}$*, such that*

$$\sum_{i=1}^{m} a_{is}\phi_s(i) \geq \sum_{j=1}^{n} b_{js}\psi_s(j),\ s = 1,2,\dots,r. \tag{3.4}$$

The relation $\supseteq_K$ is an order relation. We will show that this order relation can be characterized by the pointwise comparison of certain function graphs. For this we relate each table $A \in \mathcal{T}_r$ to a so–called concentration function.

The concentration function will be defined on the unit simplex

$$S_r = \left\{(\alpha_s) \in [0,1]^r : \sum_{s=1}^{r} \alpha_s = 1\right\}.$$

For any $a = (a_s) \in \mathbb{R}^r$ let

$$f_a(\alpha_1, \alpha_2, \dots, \alpha_r) = \sum_{s=1}^{r} \alpha_s a_s - \max_{1 \leq s \leq r} \alpha_s a_s \quad \text{if} \quad (\alpha_1, \alpha_2, \dots, \alpha_r) \in S_r. \tag{3.5}$$

(3.6) DEFINITION *Suppose that* $A \in \mathcal{T}_{n,r}$ *and let* $a_1, a_2, \dots, a_n$ *be the rows of* A*. Then the function*

$$K_A(\alpha) = \sum_{i=1}^{n} f_{a_i}(\alpha)\ , \quad \alpha \in S_r,$$

is called the concentration function of A*.*

The following assertion is the first main result of this paper. The concentration function characterizes the order relation $\supseteq_K$:

(3.7) THEOREM *Let* $A, B \in \mathcal{T}_r$*. Then* $A \supseteq_K B$ *iff* $K_A \leq K_B$*.*

Proof: Let $A \in \mathcal{T}_{m,r}$ and $B \in \mathcal{T}_{n,r}$. Let us denote by Φ and Ψ the sets of weighted partitions of $\{1,2,\dots,m\}$ resp. of $\{1,2,\dots,n\}$. For $\phi \in \Phi$ we define a column stochastic matrix $U_\phi = (\phi_s(i))$ and similarly U_ψ if $\psi \in \Psi$.

The assertion $A \supseteq_K B$ can be stated as follows:

(3.8) For every $\psi \in \Psi$ there is some $\phi \in \Phi$ such that

$$(U_\phi A)_{s,s} \geq (U_\psi B)_{s,s} \quad \text{if } s = 1,2,\dots,r.$$

It follows from the Minimax Theorem that (3.8) is equivalent to:

$$\sup_{\phi\in\Phi}\sum_{s=1}^{r}\alpha_s(U_\phi A)_{s,s} \geq \sup_{\psi\in\Psi}\sum_{s=1}^{r}\alpha_s(U_\psi B)_{s,s} \text{ for all } \alpha\in S_r. \tag{3.9}$$

Let us consider the parts of the preceding inequality more thoroughly. For all $\alpha\in S_r$ we have

$$\begin{aligned}\sup_{\phi\in\Phi}\sum_{s=1}^{r}\alpha_s(U_\phi A)_{s,s} &= \sup_{\phi\in\Phi}\sum_{s=1}^{r}\sum_{i=1}^{m}\phi_s(i)a_{is}\\ = \sup_{\phi\in\Phi}\sum_{i=1}^{m}\sum_{s=1}^{r}\alpha_s\phi_s(i)a_{is} &= \sum_{i=1}^{m}\max_{1\leq s\leq r}\alpha_s a_{is}.\end{aligned}$$

The relation to the concentration function arises by

$$\begin{aligned}K_A(\alpha) &= \sum_{i=1}^{m}\Big(\sum_{s=1}^{r}\alpha_s a_{is} - \max_{1\leq s\leq r}\alpha_s a_{is}\Big)\\ &= 1-\sum_{i=1}^{m}\max_{1\leq s\leq r}\alpha_s a_{is}\\ &= 1-\sup_{\phi\in\Phi}\sum_{s=1}^{r}\alpha_s(U_\phi A)_{s,s}.\end{aligned}$$

A similar formula is valid for table B. It follows that (3.9) is equivalent to $K_A\leq K_B$. □

(3.10) REMARKS

1. For tables $A,B\in\mathcal{T}_2$ we have

   $$L_A\leq L_B \Leftrightarrow A\succ B \Leftrightarrow A\supseteq_K B \Leftrightarrow K_A\leq K_B.$$

   Thus, for tables with two columns the concentration function plays the same role as the Lorenz diagram. Of course, the concentration function is another function than the Lorenz function, but it characterizes the concentration of tables in the same way as the Lorenz diagram does. The advantage of the concentration function is due to the fact that it can be applied also to tables with more than two columns.

2. To satisfy those readers who believe in the Lorenz curve as an inviolable sanctuary let us indicate a strong relationship between the Lorenz curve and the concentration function.

   Recall that the Lorenz curve is a convex function. For every convex function Φ there is a so–called conjugate convex function Φ^* being defined by

   $$\Phi^*(x)=\sup_y\{xy-\Phi(y)\}.$$

   It is a well–known fact of convex analysis that problems may become easier if an convex function is replaced by its conjugate convex function. (In classical mechanics the Lagrange function and Hamiltonian are conjugate convex functions.)

   Let L_A^* be the conjugate convex function of the Lorenz curve L_A. Then it can be shown that

   $$K_A(1-\alpha,\alpha)=(1-\alpha)-\alpha L_A^*\Big(\frac{1-\alpha}{\alpha}\Big).$$

   (We give a proof of this relation at the end of section 5). This relation shows that the concentration function arises from the conjugate Lorenz curve by a simple transformation which maps the unit interval $[0,1]$ onto the interval $[0,\infty]$. In fact, defining $T(\alpha)=\frac{1-\alpha}{\alpha}$ we get

   $$T\Big(K_A(1-\alpha,\alpha)\Big)=1+L_A^*\Big(T(\alpha)\Big).$$

3. Let us call a table a uniform table if its columns are equal. It is easy to see that all uniform tables are equivalent and have minimal concentration. We want to compute the concentration function of a uniform table.

   Let A be a uniform table. Since all uniform tables are equivalent it is sufficient to consider only the table $A = (1, 1, \ldots, 1)$. The concentration function is

   $$K_A(\alpha) = 1 - \max_{1 \le s \le r} \alpha_s \ , \quad \alpha \in S_r.$$

4. Let us call a table a singular table if the sets where its columns have positive entries are pairwise disjoint. It is easy to see that all singular tables are equivalent and have maximal concentration. We want to compute the concentration function of a singular table.

   Let A be a singular table. Since all singular tables are equivalent it is sufficient to consider only the table

   $$A = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}.$$

   In this case we have $K_A = 0$.

5. If $A \in \mathcal{T}_{n,r}$ is an arbitrary table then its concentration function is the sum of

   $$f_a(\alpha_1, \alpha_2, \ldots, \alpha_r) = \sum_{s=1}^{r} \alpha_s a_s - \max_{1 \le s \le r} \alpha_s a_s,$$

   where a varies among all rows of A. To get an idea of a single function f_a the following remarks could be helpful.

   Each function f_a is an concave function on the unit simplex S_r. The graph of f_a piecewise linear, i.e. it consists of pieces of hyperplanes. If we split up the unit simplex into the subsets

   $$M_t = \{\alpha \in S_r : \alpha_t a_t = \max_{1 \le s \le r} \alpha_s a_s\} \ , \quad t = 1, 2, \ldots, r$$

   then

   $$f_a(\alpha_1, \alpha_2, \ldots, \alpha_r) = \sum_{s \ne t} \alpha_s a_s \text{ whenever } \alpha \in M_t.$$

Now we turn to the generalization of condition (4) of theorem (2.4). To explain the terminology we note that the equation $B = UA$ means that the table B can be constructed with the information contained in A. Hence, A contains sufficient information for B, or simply A is sufficient for B. This concept of sufficiency is compatible with the corresponding concept known in decision theory.

(3.11) DEFINITION *Let $A \in \mathcal{T}_r$ and $B \in \mathcal{T}_r$ be tables. Assume that A has m columns and B has n columns. The table A is sufficient for the table B (denoted by $A \supseteq B$) if there is a transition, i.e. a column stochastic $n \times m$–matrix U satisfying $B = UA$.*

The relation $\supseteq$ also defines an order relation among all tables with a fixed number of columns. Let us indicate how to get an intuitive impression of the assertion $A \supseteq B$. For this we denote by $a_1, a_2, \ldots, a_m$ the rows of A and by $b_1, b_2, \ldots, b_n$ the rows of B. Then the equation $B = UA$ means that

$$b_i = \sum_{j=1}^{m} u_{ij} a_j \ , \ i = 1, 2, \ldots, n,$$

i.e. each row of B is obtained as a linear combination of rows of A with weights in $[0,1]$. It is clear that such an operation might cause loss of information. Recall that groupings are the most important examples of transitions used in descriptive statistics in order to simplify tables.

The relation between concentration and sufficiency is illuminated by the following two assertions.

(3.12) THEOREM *If A is sufficient for B then A is at least as concentrated as B.*

Proof: Let $A \in \mathcal{T}_{m,r}$ and $B \in \mathcal{T}_{n,r}$. By assumption there is a transition $U = (u_{kl})$ such that

$$b_{ks} = \sum_{i=1}^{m} u_{ki} a_{is} \text{ for } k = 1, 2, \ldots, n \text{ and } s = 1, 2, \ldots, r.$$

Let $(\psi_1, \psi_2, \ldots, \psi_r)$ be a weighted partition of $\{1, 2, \ldots, n\}$. Then we have

$$\sum_{k=1}^{n} b_{ks} \psi_s(k) = \sum_{k=1}^{n} \sum_{i=1}^{m} u_{ki} a_{is} \psi_s(k) = \sum_{i=1}^{m} a_{is} \sum_{k=1}^{n} u_{ki} \psi_s(k).$$

If we define

$$\phi_s(i) := \sum_{k=1}^{n} u_{ki} \psi_s(k),$$

then $(\phi_1, \phi_2, \ldots, \phi_r)$ is a weighted partition of $\{1, 2, \ldots, m\}$ which satisfies (3.4). □

The next assertion makes things even more precise: If a transition causes loss of information then concentration becomes smaller, too. This is the second main result of this paper.

(3.13) THEOREM *Suppose that A is sufficient for B but not conversely. Then $K_A(\alpha) < K_B(\alpha)$ for some point $\alpha \in S_r$.*

Proof: For any row $a \in [0,1]^r$ let f_a be defined by (3.5). Let us show that

$$f_{a_1+a_2} \geq f_{a_1} + f_{a_2},$$
$$f_{a_1+a_2} = f_{a_1} + f_{a_2} \;\Rightarrow\; a_1 \parallel a_2.$$

The inequality follows from the facts that $\alpha \mapsto f_a(\alpha)$ is concave and positively homogeneous. In order to prove the implication of equality we note that

$$f_{a_1+a_2}(\alpha, 1-\alpha, 0, \ldots, 0) = f_{a_1}(\alpha, 1-\alpha, 0, \ldots, 0) + f_{a_2}(\alpha, 1-\alpha, 0, \ldots, 0)\,,\ \alpha \in [0,1],$$

implies

$$\frac{a_{11}}{a_{12}} = \frac{a_{21}}{a_{22}}.$$

For reasons of symmetry we obtain

$$\begin{aligned} f_{a_1+a_2} = f_{a_1} + f_{a_2} \quad &\Rightarrow \quad \frac{a_{1i}}{a_{1j}} = \frac{a_{2i}}{a_{2j}} \text{ for } i, j = 1, 2, \ldots, m \\ &\Rightarrow \quad a_1 \parallel a_2. \end{aligned}$$

Let $b = \sum_{j=1}^{m} u_j a_j$ where $u_j \in [0,1]$. It is easy to see that the following slightly more general assertions are valid:

$$f_b \geq \sum_{j=1}^{n} u_j f_{a_j},$$
$$f_b = \sum_{j=1}^{n} u_j f_{a_j} \;\Rightarrow\; a_i \parallel a_j \text{ if } u_i > 0 \text{ and } u_j > 0.$$

Let $B = UA$, i.e.

$$b_i = \sum_{j=1}^{m} u_{ij} a_j \ , \ i = 1, 2, \ldots, n.$$

We claim: If $B \not\supseteq A$ then there is at least one i_0 such that

$$f_{b_{i_0}} > \sum_{j=1}^{m} u_{i_0 j} f_{a_j}.$$

It this were not the case then every b_i would be a linear combination of parallel rows of A, thus itself parallel to those rows of A. Then we could reconstruct the rows of A from the rows of B which would contradict to the assumption $B \not\supseteq A$.

Hence the assumption $B \not\supseteq A$ implies

$$\begin{aligned} K_B &= \sum_{i=1}^{n} f_{b_i} > \sum_{i=1}^{n} \sum_{j=1}^{m} u_{ij} f_{a_j} = \\ &= \sum_{j=1}^{m} \Big(\sum_{i=1}^{n} u_{ij} \Big) f_{a_j} = K_A. \end{aligned}$$

□

4 The Coefficient of Concentration

Concentration is a phenomenon of high dimension. The structure of concentration differences between tables has many degrees of freedom and cannot be summarized by a single number. Therefore the concentration structure of a table must be mapped by a high dimensional mathematical object, e.g. by the concentration function.

So–called measures of concentration, i.e. numbers summarizing a table by a single number, can only describe a single aspect of the whole phenomenon. Such an interesting aspect could be the amount of concentration.

In the following let us denote by A_0 any uniform table and by A_∞ any singular table. For tables with two columns the amount of concentration is usually measured by the Gini coefficient.

(4.1) REMARK Let us recall the idea and the basic properties of the Gini coefficient.

Let $A \in \mathcal{T}_2$. The Gini coefficient of A is defined by

$$G(A) = 1 - \frac{\int_0^1 L_A(\alpha)\, d\alpha}{\int_0^1 L_{A_0}(\alpha)\, d\alpha} \ .$$

The basic properties are

$$\begin{array}{lll} G(A_0) = 0, \ G(A_\infty) = 1, & & \\ A \supseteq_K B & \Rightarrow & G(A) \geq G(B), \\ A \supseteq_K B, \ A \not\sim_K B & \Rightarrow & G(A) > G(B). \end{array}$$

An explicit formula for the Gini coefficient is

$$G(A) = \sum_{i=1}^{n} \det \begin{pmatrix} a_{\pi(i),1} & A_1(i-1) \\ a_{\pi(i),2} & A_2(i-1) \end{pmatrix}$$

We will define a measure for the amount of concentration for an arbitrary table $A \in \mathcal{T}_r$ in a similar way as the Gini coefficient is defined in case $r = 2$.

(4.2) DEFINITION *Let $A \in \mathcal{T}_r$, $r \geq 2$. Then*

$$k(A) := 1 - \frac{\int_{S_r} K_A(\alpha)\, d\alpha}{\int_{S_r} K_{A_0}(\alpha)\, d\alpha}$$

is called the coefficient of concentration (COC) of A.

It should be noted that the integrals occurring in the definition are surface integrals.

Obviously, for tables with two columns the COC does not coincide with the Gini coefficient. However, the COC has similar properties.

(4.3) THEOREM *The COC has the following properties:*

$$\begin{array}{lll} k(A_0) = 0,\ k(A_\infty) = 1, & & \\ A \supseteq_K B & \Rightarrow & k(A) \geq k(B), \\ A \supseteq_K B,\ A \not\sim_K B & \Rightarrow & k(A) > k(B). \end{array}$$

Proof: This is an obvious consequence of theorem (3.7). □

(4.4) REMARK Concerning the numerical computation the COC has a major advantage compared with the Gini coefficient:

The COC is a sum of terms each of which depends only on a single line of the table. Any change of single lines of the table has only influence on the respective additive term and leaves the remaining terms unchanged.

The Gini coefficient is also a sum but each term of the sum depends on the whole table since a sorting mechanism is to be involved. Hence, a change of a single line requires sorting of the table.

Let us turn to the computational aspects of the COC. We will derive an explicit formula. Note that

$$\int_{S_r} K_A(\alpha)\, d\alpha = \sum_{i=1}^{n} \int_{S_r} f_{a_i}(\alpha)\, d\alpha$$

where $a_1, a_2, \ldots, a_n$ denote the rows of A. Thus, we have to consider integrals of the type

$$\int_{S_r} f_a(\alpha)\, d\alpha \quad \text{where } a \in [0,1]^r.$$

Let $a = (a_1, a_2, \ldots, a_r) \in [0,1]^r$. Some of the components of a may be equal to zero. For reasons of symmetry we may assume that $a_1 = a_2 = \ldots = a_{l-1} = 0$ and $a_i > 0$ for $i = l, l+1, \ldots, r$. Let $\mathcal{S}$ be the set of all permutations of $\{l, l+1, \ldots, r\}$. The unit simplex S_r can be splitted into the subsets

$$M_{i_l \ldots i_r} = \{\xi \in S_r : \xi_{i_l} a_{i_l} \leq \xi_{i_{l+1}} a_{i_{l+1}} \leq \ldots \leq \xi_{i_r} a_{i_r}\},\ (i_l \ldots i_r) \in \mathcal{S}.$$

These sets cover the unit simplex. Their intersections are of measure zero. This implies that

$$\int_{S_r} f_a(\alpha)\, d\alpha = \sum_{(i_l \ldots i_r) \in \mathcal{S}} \int_{M_{i_l \ldots i_r}} f_a(\alpha)\, d\alpha.$$

Denoting

$$\int_{M_{l \ldots r}} f_a(\alpha)\, d\alpha =: g_l(a_l, a_{l+1}, \ldots, a_r)$$

we obtain

$$\int_{S_r} f_a(\alpha)\,d\alpha = \sum_{(i_l\dots i_r)\in\mathcal{S}} g_l(a_{i_l}, a_{i_{l+1}}, \dots, a_{i_r}). \tag{4.5}$$

It is therefore sufficient to get an explicit formula for $g_l(a_l, a_{l+1}, \dots, a_r)$.

(4.6) THEOREM *Suppose that $a_i > 0$ and denote $b_i := \dfrac{1}{a_i}$ for $i = l, l+1, \dots, r$. Then*

$$g_l(a_l, a_{l+1}, \dots, a_r) = \frac{\sqrt{r}}{r!} \cdot \prod_{j=l}^{r-1} \frac{b_j}{b_j + \dots + b_r} \cdot \sum_{j=l}^{r-1} \frac{r-j}{b_j + \dots + b_r} \tag{4.7}$$

Proof: The integral

$$g_l(a_l, a_{l+1}, \dots, a_r) = \int_{M_{l\dots r}} f_a(\alpha)\,d\alpha$$

is a surface integral of dimension $r-1$. Consider the mapping

$$\begin{array}{lcll}
\xi_1 & = & \phi_1(\eta_1 \dots \eta_{r-1}) & = \eta_1 \\
\xi_2 & = & \phi_2(\eta_1 \dots \eta_{r-1}) & = \eta_2 \\
\dots & & & \\
\xi_{r-1} & = & \phi_{r-1}(\eta_1 \dots \eta_{r-1}) & = \eta_{r-1} \\
\xi_r & = & \phi_r(\eta_1 \dots \eta_{r-1}) & = 1 - \sum_{i=1}^{r-1} \eta_i
\end{array}$$

If we denote

$$M = \left\{ \eta \in [0,1]^{r-1} : 0 \le \eta_l a_l \le \eta_{l+1} a_{l+1} \le \dots \le \eta_{r-1} a_{r-1} \le \left(1 - \sum_{i=1}^{r-1} \eta_i\right) a_r \right\}$$

then we have $M_{l\dots r} = \phi(M)$. The Gram determinant of ϕ is (denote $e := \sum_{i=1}^{r-1} e_i$)

$$\det(D_i\phi \cdot D_j\phi)_{1\le i,j\le r-1} = \det \begin{pmatrix} 2 & 1 & \dots & 1 \\ 1 & 2 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 2 \end{pmatrix}$$

$$\begin{array}{lcll}
= & \det(e_1 + e, e_2 + e, \dots, e_{r-1} + e) & & \\
= & \det(e_1 \dots e_{r-1}) + \sum_{i=1}^{r-1} \det(e_1 \dots e_{i-1}\, e\, e_{i+1} \dots e_{r-1}) & = & r
\end{array}$$

Thus, we obtain

$$\int_{M_{l\dots r}} f_a(\alpha)\,d\alpha = \sqrt{r} \int_M \left(\sum_{i=l}^{r-1} a_i \eta_i \right) d\eta.$$

Next, we apply the following transformation:

$$\begin{array}{lcl}
x_1 & = & \eta_1 \\
x_2 & = & \eta_1 + \eta_2 \\
\dots & & \\
x_{l-1} & = & \eta_1 + \eta_2 + \dots + \eta_{l-1} \\
b_l x_l & = & (\eta_1 + \dots + \eta_{l-1}) b_l + \eta_l (b_l + \dots + b_r) \\
\dots & & \\
b_{r-1} x_{r-1} & = & (\eta_1 + \dots + \eta_{r-2}) b_{r-1} + \eta_{r-1} (b_{r-1} + b_r)
\end{array}$$

This transformation satisfies

$$x_i - x_{i-1} = (b_i + \cdots + b_r)\left(\frac{\eta_i}{b_i} - \frac{\eta_{i-1}}{b_{i-1}}\right) \quad \text{for } i = l, l+1, \ldots, r-1.$$

It is easy to see that the transformation maps the set M onto

$$M^* = \{x \in \mathbb{R}^{r-1} : 0 \le x_1 \le x_2 \le \ldots \le x_{r-1} \le 1\}.$$

The functional determinant is

$$\prod_{j=l}^{r-1} \frac{b_j + \cdots + b_r}{b_j}.$$

Since

$$\begin{aligned} \sum_{i=l}^{r-1} a_i \eta_i &= (r-l)\frac{\eta_l}{b_l} + (r-l-1)\left(\frac{\eta_{l+1}}{b_{l+1}} - \frac{\eta_l}{b_l}\right) + \cdots \\ &= \frac{r-l}{b_l + \cdots + b_r}(x_l - x_{l-1}) + \frac{r-l-1}{b_{l+1} + \cdots + b_r}(x_{l+1} - x_l) + \cdots \end{aligned}$$

we arrive at

$$\int_{M_{l\ldots r}} f_a(\alpha)\, d\alpha = \sqrt{r} \cdot \prod_{j=l}^{r-1} \frac{b_j}{b_j + \cdots + b_r} \cdot \sum_{j=l}^{r-1} \frac{r-j}{b_j + \cdots + b_r} \int_{M^*} (x_j - x_{j-1})\, dx.$$

The asserted formula follows because it is easy to see that

$$\int_{M^*} x_j\, dx = \frac{j}{r!} \quad \text{if } j = 1, 2, \ldots, r.$$

□

(4.8) EXAMPLES

1. For a comparison with the Gini coefficient it might be interesting to have an explicit form of the COC for tables with two columns. Let

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \cdots & \\ a_{n1} & a_{n2} \end{pmatrix}.$$

   Then we have

$$k(A) = 1 - 2 \cdot \sum_{i=1}^{n} \frac{a_{i1} a_{i2}}{a_{i1} + a_{i2}}.$$

2. Instead of deriving complicated formulas for particular values of $r = 3, 4, \ldots$ let us give an algorithm for the computation of the COC. The algorithm is written in PASCAL. Below we discuss the performance of the algorithm.

```
type
   rowtype= array [1..10] of real;
   tabletype= array [1..100] of rowtype;

function integral(r : integer; a : rowtype) : real; {evaluates 4.5}
   type
      settype= set of 1..10;
   var
      subset : settype;
```

```
    permutation : array [1..10] of integer;
    l, level, i : integer;
    b : rowtype;
    value : real;

function evaluation(l : integer) : real; {evaluates 4.7}
    var
        denominator, product, sum : real;
        j : integer;
    begin {evaluation}
        denominator := b[permutation[r - l + 1]];
        product := 1;
        sum := 0;
        for j := r - l downto 1 do
        begin
            denominator := denominator + b[permutation[j]];
            product := product * b[permutation[j]]/denominator;
            sum := sum + (r - j - l + 1)/denominator
        end;
        evaluation := product * sum
    end; {evaluation}

procedure recursion; {generates all permutations}
    var
        position, element, i : integer;
    begin {recursion}
        position := 0;
        level := level + 1; {level of recursion}
        repeat
            repeat position := position + 1
            until (position > r - l + 1) or not (position in subset);
            if position <= r - l + 1 then
            begin
                subset := subset + [position];
                permutation[level] := position;
                if level < r - l + 1
                then recursion
                else value := value + evaluation(l);
            end;
            subset := subset - [position];
        until position > r - l + 1;
        level := level - 1
    end; {recursion}

begin {integral}
    l := 0;
    for i := 1 to r do
        if a[i] > 0
        then begin l := l + 1; b[l] := 1/a[i] end;
    l := r - l + 1;
    value := 0;
    subset := [];
    level := 0;
    recursion;
```

```
        for i := 1 to r do value := value/i;
        integral := value * sqrt(r)
    end; {integral}

function coc(n, r : integer; a : tabletype) : real;
    var
        sum : real;
        i : integer;
        e : rowtype;
    begin
        sum := 0;
        for i := 1 to n do sum := sum + integral(r, a[i]);
        for i := 1 to r do e[i] := 1;
        coc := 1 - sum/integral(r, e)
    end;
```

3. The performance of the algorithm is illustrated by table (4.9). For $r = 2, 3, 4, 5$ we generated randomly five tables with 100 rows and computed the COC. The computations were performed on a AT–Personal Computer (80386/87 with 16 MHz). We see that the computing time (seconds) is proportional to $r!$ which is due to the generation of all permutations of r.

$r =$	2	3	4	5
1	0.27	0.82	3.52	18.73
2	0.28	0.82	3.51	18.73
3	0.27	0.83	3.51	18.73
4	0.28	0.83	3.52	18.68
5	0.28	0.82	3.52	18.73

(4.9)

4. Let us illustrate the validity of the COC by a numerical example. We will randomly generate tables which differ slightly concerning their expected amount of concentration. It will turn out that the COC is able to indicate that difference.

   First let us describe method of random table generation.

   Let $T_\infty = (t_{ij})$ be that 20×4–table such that

$$t_{ij} = \begin{cases} 1 \text{ if } i \bmod r = j - 1, \\ 0 \text{ otherwise} \end{cases}$$

   By T_0 we denote the table with $t_{ij} \equiv 1$. If T_∞ is normed to column sums equal to 1 then a singular table arises. From T_0 we would obtain a uniform table.

   Let $\lambda \in [0, 1]$ and let R be a random 20×4–matrix containing independent $U(0, 1)$–distributed entries. Then we compute $\lambda T_\infty + (1 - \lambda)T_0 + \epsilon R$ and by norming to column sums equal to 1 we obtain a table denoted by $A(\lambda, \epsilon)$.

   From the construction it is clear that $A(1, 0)$ is a singular table, $A(0, 0)$ is a uniform table, and $\lambda \mapsto A(\lambda, 0)$ is a continuous curve joining $A(1, 0)$ and $A(0, 0)$. For $\epsilon > 0$ we obtain a noisy version $\lambda \mapsto A(\lambda, \epsilon)$ of that curve.

   Table (4.10) contains the COC for 5 realizations of $A(0.95, 0.5)$ and of $A(0.90, 0.5)$. We see that the

COC indicates the difference of concentration.

$\lambda = 0.95$	$\lambda = 0.90$
0.4717	0.3928
0.4625	0.4257
0.4187	0.4270
0.4865	0.4016
0.4845	0.4227

(4.10)

5 Some Proofs

In this section we present elementary proofs of theorem (2.4) $(1) \Leftrightarrow (2)$ and $(2) \Rightarrow (4) \Rightarrow (3) \Rightarrow (2)$ which are adapted to the finite discrete case being the topic of this paper.

(5.1) REMARK To satisfy those readers who are interested in the connection of theorem (2.4) with statistical decision theory let us give some hints in this direction.
A table in $\mathcal{T}_2$ can be viewed as a pair of probability distributions called a binary experiment. If $A = (P, Q)$ then $L_A(\alpha)$ is the power of the Neyman–Pearson test of level α for the simple testing problem (P, Q). The comparison of tables corresponds to the comparison of binary experiments in statistical decision theory (cf. Strasser, [11], and Torgersen, [13]).
Let us give some explicit references. The property $A \supseteq_K B$ corresponds to Definition 15.1 in Strasser, [11], for the case $\epsilon = 0$. The equivalence $(1) \Leftrightarrow (2)$ is proved in Strasser, [11], Theorem 15.10. The equivalence $(3) \Leftrightarrow (4)$ is due to Blackwell, [2], (see also Marshall and Olkin, [9], 14 A). It has been generalized to the famous Blackwell–Sherman–Stein theorem of measure theory and to the randomization criterion of LeCam. A proof of the convex function criterion $(2) \Leftrightarrow (3)$ if given in Strasser, [11], Theorem 17.1. A proof of the randomization criterion $(2) \Leftrightarrow (4)$ can be found in Strasser, [11], section 55.

For convenience we introduce some notation.

Let A be a table with two columns and m rows. Suppose that the table contains no rows equal to zero. Choose a permutation π such that

$$\frac{a_{\pi(1),2}}{a_{\pi(1),1}} \leq \frac{a_{\pi(2),2}}{a_{\pi(2),1}} \leq \ldots \leq \frac{a_{\pi(m),2}}{a_{\pi(m),1}}.$$

Define the partial sums

$$\begin{aligned} A_1(k) &= \sum_{i=1}^{k} a_{\pi(i),1}\ ,\ k = 1, 2, \ldots, m, \\ A_2(k) &= \sum_{i=1}^{k} a_{\pi(i),2}\ ,\ k = 1, 2, \ldots, m. \end{aligned}$$

Recall that the Lorenz diagram of the table A is the graph of the function $L_A : [0,1] \to [0,1]$ which is defined by linear interpolation of the points $(A_1(k), A_2(k)),\ k = 1, 2, \ldots, m$.

The proof of (2.4), $(1) \Leftrightarrow (2)$ is based on an extremal property of the Lorenz diagram. Let us mention that in the statistical decision theory the following lemma is called the Neyman–Pearson Lemma.

(5.2) LEMMA *The Lorenz diagram of A satisfies for every $\alpha \in [0,1]$*

$$L_A(\alpha) = \inf\left\{ \sum_{i=1}^{m} \xi_i a_{i2} : \sum_{i=1}^{m} \xi_i a_{i1} \geq \alpha, (\xi_i) \in [0,1]^m \right\}.$$

Proof: Let $\alpha \in [0,1]$. There is a vector $(\xi_i) \in [0,1]^m$ such that

$$L_A(\alpha) = \sum_{i=1}^{m} \xi_i a_{i2} \quad \text{and} \quad \sum_{i=1}^{m} \xi_i a_{i1} = \alpha. \tag{5.3}$$

To see this choose $l \in \{1, 2, \ldots, m\}$ such that

$$L_A(\alpha) = A_2(l-1) + \Big(\alpha - A_1(l-1)\Big)\frac{a_{\pi(l),2}}{a_{\pi(l),1}} \quad \text{whenever} \quad A_1(l-1) \leq \alpha \leq A_1(l)$$

and define

$$\begin{aligned} \xi_{\pi(j)} &= 1 && \text{if } j \leq l-1, \\ \xi_{\pi(l)} &= \frac{\alpha - A_1(l-1)}{a_{\pi(l),1}}, \\ \xi_{\pi(j)} &= 0 && \text{if } j \geq l+1. \end{aligned}$$

It is easy to see that (ξ_i) satisfies (5.3).
Let $(\eta_i) \in [0,1]^m$ be any vector satisfying $\sum_{i=1}^m \eta_i a_{i1} \geq \alpha$. Then

$$\sum_{i=1}^{m} (\eta_i - \xi_i)\left(\frac{a_{\pi(l),2}}{a_{\pi(l),1}} - \frac{a_{i2}}{a_{i1}}\right) a_{i1} \leq 0.$$

It follows that

$$\sum_{i=1}^{m} \eta_i a_{i2} - \sum_{i=1}^{m} \xi_i a_{i2} \geq \frac{a_{\pi(l),2}}{a_{\pi(l),1}}\left(\sum_{i=1}^{m} \eta_i a_{i1} - \sum_{i=1}^{m} \xi_i a_{i1}\right) \geq 0.$$

□

Proof: (of Theorem ((2.4), $(1) \Leftrightarrow (2)$). Condition (2) of Theorem (2.4) can be stated as follows:
For every vector $(\eta_j) \in [0,1]^n$ there is a vector $(\xi_i) \in [0,1]^m$ such that

$$\begin{aligned} \sum_{i=1}^{m} \xi_i a_{i1} &\geq \sum_{j=1}^{n} \eta_j b_{j1}, \\ \sum_{i=1}^{m} \xi_i a_{i2} &\leq \sum_{j=1}^{n} \eta_j b_{j2}. \end{aligned}$$

In other words this means that for every $\alpha \in [0,1]$

$$\begin{aligned} L_A(\alpha) &= \inf\left\{\sum_{i=1}^{m} \xi_i a_{i2} : \sum_{i=1}^{m} \xi_i a_{i1} \geq \alpha,\ (\xi_i) \in [0,1]^m\right\} \\ &\leq \inf\left\{\sum_{j=1}^{n} \eta_j b_{j2} : \sum_{j=1}^{n} \eta_j b_{j1} \geq \alpha,\ (\eta_j) \in [0,1]^n\right\} = L_B(\alpha). \end{aligned}$$

□

Proof: (of Theorem (2.4) $(2) \Rightarrow (4)$).
By assumption (2) for every vector $\eta \in [0,1]^n$ there exists a vector $\xi \in [0,1]^m$ such that

$$\begin{aligned} \sum_{j=1}^{m} \xi_j a_{j1} &\geq \sum_{i=1}^{n} \eta_i b_{i1}, \\ \sum_{j=1}^{m} \xi_j a_{j2} &\geq \sum_{i=1}^{n} \eta_i b_{i2}. \end{aligned} \tag{5.4}$$

There is even the following stronger assertion valid: If $\eta \in [p,q]^n$, where $[p,q] \subseteq [0,1]$, then we may also choose $\xi \in [p,q]^m$. This is due to the fact that the inequalities (5.4) remain valid if ξ_j and η_i are transformed by $x \mapsto p + (q-p)x$.

It follows that we may choose $\xi \in [\min \eta_i, \max \eta_i]^m$. Thus, each ξ_i is a convex linear combination of the components of η:

$$\xi_j = \sum_{i=1}^{n} \eta_i \, u_{ij} \, , \quad j = 1, \ldots, m, \tag{5.5}$$

where $U = (u_{ij})$ is a column–staochastic $n \times m$–matrix. Let us denote by $\mathcal{U}$ the set of all column–stochastic $n \times m$–matrices. The representation (5.5) is inserted into (5.4):

$$\sum_{j=1}^{m} \xi_j a_{js} = \sum_{j=1}^{m} \sum_{i=1}^{n} \eta_i u_{ij} a_{js} = \sum_{i=1}^{n} (UA)_{is} \eta_i \quad \text{for } s = 1, 2.$$

For $\alpha \in [0,1]$, $U \in \mathcal{U}$ and $\eta \in [0,1]^n$ let

$$\Phi(\alpha, U, \eta) := (1-\alpha)\Big(\sum_{j=1}^{n} (UA - B)_{i1} \eta_i \Big) + \alpha \Big(\sum_{i=1}^{n} (UA - B)_{i2} \eta_i \Big).$$

Then assumption (2) can be phrased as

$$\sup_{\alpha \in [0,1]} \sup_{\eta \in [0,1]^n} \inf_{U \in \mathcal{U}} \Phi(\alpha, U, \eta) \leq 0.$$

Applying the Minimax theorem the requirements of which are easily checked we obtain

$$\inf_{U \in \mathcal{U}} \sup_{\alpha \in [0,1]} \sup_{\eta \in [0,1]^n} \Phi(\alpha, U, \eta) \leq 0.$$

Since for any vector $x \in \mathbb{R}^n$

$$\sup_{\eta \in [0,1]^n} \sum_{i=1}^{n} x_i \eta_i = \frac{1}{2} \sum_{i=1}^{n} |x_i|$$

we finally get

$$\inf_{U \in \mathcal{U}} \max \left\{ \sum_{i=1}^{n} |(UA - B)_{i1}|, \sum_{i=1}^{n} |(UA - B)_{i2}| \right\} = 0.$$

Hence there is some $U \in \mathcal{U}$ such that $UA = B$. □

Proof: (of Theorem (2.4) $(4) \Rightarrow (3)$). The proof is straightforward from

$$\begin{aligned} \sum_{i=1}^{n} b_{i1} f\Big(\frac{b_{i2}}{b_{i1}}\Big) &= \sum_{i=1}^{n} b_{i1} f\Big(\sum_{j=1}^{m} \frac{u_{ij} a_{j1}}{b_{i1}} \frac{a_{j2}}{a_{j1}} \Big) \\ \leq \sum_{i=1}^{n} b_{i1} \sum_{j=1}^{m} \frac{u_{ij} a_{j1}}{b_{i1}} f\Big(\frac{a_{j2}}{a_{j1}}\Big) &= \sum_{j=1}^{m} a_{j1} f\Big(\frac{a_{j2}}{a_{j1}}\Big). \end{aligned}$$

□

Proof: (of theorem (2.4) $(3) \Rightarrow (2)$). Let a be a row of a table $A \in \mathcal{T}_2$. We have

$$f_a(\alpha_1, \alpha_2) = a_{i1} \Big(\alpha_1 + \alpha_2 \frac{a_{i2}}{a_{i1}} - \max \Big\{ \alpha_1, \alpha_2 \frac{a_{i2}}{a_{i1}} \Big\} \Big),$$

Since

$$x \mapsto \alpha_1 + \alpha_2 x - \max\{\alpha_1, \alpha_2 x\}$$

is a concave function it follows from (3) that $K_A \leq K_B$ which by Theorem (3.7) implies $A \supseteq_K B$. □

Now we turn to the proof of the fact that for tables with two columns the Lorenz curve and the concentration function are related as conjugate functions. First we need some basic facts on conjugate convex functions.

(5.6) DISCUSSION Let $\Phi : [0,1] \to [0,1]$ be a convex function. The conjugate convex function is defined by

$$\Phi^*(y) = \sup_x \{xy - \Phi(x)\}.$$

There is an explicit representation of the conjugate convex function based on Young's inequality.
The convex function Φ has an integral representation (Hewitt and Stromberg, [6], (18.43))

$$\Phi(x) = \int_0^x \phi(s)\, ds$$

where ϕ is a nondecreasing function. Let ϕ^{-1} be the generalized inverse of ϕ in the sense

$$\phi^{-1}(t) := \sup\{x : \phi(x) \leq y\}.$$

(Confer Witting, [14], 1.2.1, where ϕ^{-1} is denoted by $\tilde{\phi}^{-1}$.) Then the conjugate convex function Φ^* can be written as

$$\Phi^*(y) = \int_0^y \phi^{-1}(t)\, dt.$$

This is due to Young's inequality

$$xy \leq \int_0^x \phi(s)\, ds + \int_0^y \phi^{-1}(t)\, dt,$$

where equality holds iff $y = \phi(x)$ or $x = \phi^{-1}(y)$. (A proof of Young's inequality for bijective continuous functions ϕ is given in Hewitt and Stromberg, [6], Theorem 13.2. The easy geometric argument is valid also if ϕ is not continuous.)

(5.7) THEOREM *If L_A denotes the Lorenz curve of a table A with two columns then the concentration function satisfies*

$$K_A(1-\alpha, \alpha) = (1-\alpha) - \alpha\, L_A^*\left(\frac{1-\alpha}{\alpha}\right), \text{ if } \alpha \in [0,1].$$

Proof: We apply the notation introduced at the beginning of this section.
Let

$$\phi(s) := \frac{a_{\pi(k),2}}{a_{\pi(k),1}} \text{ whenever } A_1(k-1) \leq s < A_1(k).$$

Then the integral representation of the Lorenz curve is

$$L_A(x) = \int_0^x \phi(s)\, ds.$$

The generalized inverse ϕ^{-1} is

$$\phi^{-1}(t) = \sum \left\{ a_{i1} : \frac{a_{i2}}{a_{i1}} \leq t \right\}.$$

This is the distribution function of the probability measure μ which gives probability a_{i1} to the point $\frac{a_{i2}}{a_{i1}}$ for $i = 1, 2, \ldots, m$. In view of Discussion (5.6) we have to show that

$$K_A(1-\alpha, \alpha) = (1-\alpha) - \alpha \int_0^{\frac{1-\alpha}{\alpha}} \phi^{-1}(t)\, dt. \tag{5.8}$$

The proof of (5.8) is given in Strasser, [11], Lemma (16.3). For completeness let us repeat the proof. The concentration function can be written as

$$\begin{aligned} K_A(1-\alpha,\alpha) &= \sum_{i=1}^{m}\left((1-\alpha)a_{i1}+\alpha a_{i2}-\max\left\{(1-\alpha)a_{i1},\alpha a_{i2}\right\}\right) \\ &= \sum_{a_{i1}\neq 0} a_{i1}\left(\alpha\frac{a_{i2}}{a_{i1}}-\max\left\{0,-(1-\alpha)+\alpha\frac{a_{i2}}{a_{i1}}\right\}\right) \\ &= \int_0^\infty \psi_\alpha(x)\,\mu(dx), \end{aligned}$$

where

$$\psi_\alpha(x) := \alpha x - \Big(1-\alpha(1+x)\Big)^{-}.$$

Integration by parts yields for every $a>0$

$$\int_0^a \psi_\alpha(x)\,\mu(dx) = \psi_\alpha(a)\phi^{-1}(a) - \psi_\alpha(0)\phi^{-1}(0) - \int_0^a \psi'_\alpha(x)\phi^{-1}(x)\,dx.$$

For $a\to\infty$ we obtain (5.8). □

References

[1] Blackwell, D. (1951) *Comparison of Experiments.* Proc. 2^{nd} Berkeley Symp. Math. Statistics Prob., 93–102.

[2] Blackwell, D. (1953) *Equivalent Comparisons of Experiments.* Ann. Math. Statistics 24, 265–272.

[3] Dalton, H. (1920) *The measurement of the inequality of incomes.* Econom. J. 30, 348–361.

[4] Ferschl, F. (1980) *Deskriptive Statistik.* Physika–Verlag, Wien

[5] Hardy, G.H., J.E. Littlewood and G.Polya (1934). *Inequalities.* Cambridge Univ. Press, London.

[6] Hewitt E. and K. Stromberg (1965). *Real and Abstract Analysis*. Springer, Berlin.

[7] Heyer, H. (1982) *Theory of Statistical Experiments.* Springer, Berlin.

[8] Lorenz, M.O. (1905) *Methods of measuring concentration of wealth.* J. Amer. Statist. Assoz. 9, 209–219.

[9] Marshall, A.W. and I. Olkin (1979) *Inequalities: Theory of Majorization and Its Applications.* Academic Press, New York.

[10] Pflug, G. (1979) *Statistische Konzentrationsmaße – ein mathematischer Überblick.* Allgem. Statist. Archiv 63, 240–260.

[11] Strasser, H. (1985) *Mathematical Theory of Statistics.* Walter de Gruyter, Berlin.

[12] Torgersen, E.N. (1970) *Comparison of Experiments when the Parameter Space is Finite.* Z. Wahrscheinlichkeitstheorie verw. Geb. 16, 219–249.

[13] Torgersen, E.N. (1991) *Comparison of Statistical Experiments.* Cambridge Univ. Press, Cambridge.

[14] Witting H. (1985) *Mathematische Statistik I.* Teubner, Stuttgart.