Analyse & Kritik 29/2007 (© Lucius & Lucius, Stuttgart) p. 206–225

Wolfgang Eichhorn/Manfred Krtscha

Multivariate Lorenz Majorization and Heterogeneity Measures: An Axiomatic Approach

Abstract: This work introduces two new curves that are multivariate generalizations of the "classical" Lorenz curve. All data of d-variate distributions can be visualized by drawing these curves in the plane, whereas Koshevoy's and Mosler's generalization by a lift zonoid in $\mathbb{R}^{d+1}$ can only be drawn for $d = 2$. The generalizations of the Lorenz curve induce partial orderings of d-variate distributions. Furthermore, two inequality or heterogeneity measures that are consistent with the induced rankings are proposed. They can be considered as new generalizations of the univariate Gini coefficient. For deciding which of the two measures is more appropriate for measuring a sort of convergence concerning different countries of an union or of regions of a country, we establish systems of axioms. Although these systems are reflecting natural properties, several of the axioms are new. Moreover, by means of these axioms well-known inequality measures are tested, too.

0. Introduction

Consider n units i, where $i = 1, 2, \dots n$ (individuals or households or different states of a union or ...), each of which is endowed with an attribute x_i (annual income, property value or gross national product, all of which are normalized by the number of inhabitants). If, for instance, x_i is the income of an individual i, we call $x = (x_1, \dots, x_i, \dots, x_n)$ an income vector or an income distribution. From an egalitarian viewpoint, a distribution y is improved if a richer individual i gives a part $(1-\alpha)\cdot s$, $0 < \alpha < 1$, of his surplus $s = y_i - y_j$ to a poorer individual j. A transfer of this kind is called a Pigou-Dalton transfer, and we say that a distribution x is **majorized** by the "worse" distribution y, if x has originated from y by one or more Pigou-Dalton transfers.

Assuming $i = 1$ and $j = 2$ for this case, we can write

$$(x_1, x_2, \dots, x_n) = (y_1, y_2, \dots, y_n) \cdot \begin{pmatrix} \alpha & 1-\alpha & 0 & \dots & 0 \\ 1-\alpha & \alpha & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix},$$

$0 < \alpha < 1$.

This transformation can also be generated by a doubly stochastic matrix P, so that $x = y \cdot P$, where P is neither a permutation matrix nor the identity.

Moreover, if $\alpha = 0$ and $\alpha = 1$ are also permitted, this transformation is called a T-transformation. A consequence of this fact is that no region of the well-known Lorenz curve of y lies above the Lorenz curve of x. Also the Gini coefficient of y is equal or greater than the Gini coefficient of x, as the Gini coefficient is a consistent extension of the described majorization.

A new problem arises if each unit i is endowed with two or more quantities y_{1i} (personal income, say), y_{2i} (property, say), …, y_{di} (leisure time, say), even if each quantity distribution is assumed to be improved by a Pigou-Dalton transfer.

In order to illustrate this problem we will consider the normalized quantity y_1 as personal income and the normalized quantity y_2 as personal property, where y_1 and y_2 have the same welfare estimation. We assume that an individual with income one and property zero obtains the property $\frac{1}{2}$ from a person with property one and income zero. This transfer would then enlarge the economic inequality, if both persons did not also share their incomes, since it is assumed that income and property can be easily aggregated. However, if y_1 was the normalized leisure time which could not so easily be aggregated with property and is sometimes a compliment of income or property, we could say that the diversity of both persons was diminished by this Pigou-Dalton transfer as they are more similar afterwards. In any case, we can avoid this problem of interpretation if we decide that only simultaneous Pigou-Dalton transfers improve the two dimensional distribution. That means the surplus s_1 in income or leasure time of person i is shared in the same way with the person j as the surplus s_2 in property. Observe that it is thereby unimportant if the surplus s_2 comes from person i or j.

Assuming again the case $i = 1, j = 2$, for $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_{11} & \dots & x_{1n} \\ x_{21} & \dots & x_{2n} \end{pmatrix}$ we have

$$\begin{pmatrix} x_{11} & \dots & x_{1n} \\ x_{21} & \dots & x_{2n} \end{pmatrix} = \begin{pmatrix} y_{11} & \dots & y_{1n} \\ y_{21} & \dots & y_{2n} \end{pmatrix} \begin{pmatrix} \alpha & 1-\alpha & 0 & \dots & 0 \\ 1-\alpha & \alpha & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix},$$

$0 \le \alpha \le 1$, or $\binom{x_1}{x_2} = \binom{y_1}{y_2} \cdot P$, where P is a doubly stochastic matrix.

These ideas lead to two different definitions for majorization (see Marshall and Olkin 1979) which coincide only for one attribute or for $n = 2$. A further kind of majorization, called a rowwise majorization, considers each row separately with regard to Pigou-Dalton transfers so that separate permutations within a row do not influence this majorization, whereas the columns presenting the individuals are disfigured. One avoids the disfiguration by the following majorization: Comparing all one dimensional weighted arithmetic mean distributions of the rows of X and Y, which are denoted by $\bar{y} = (\bar{y}_1, \dots, \bar{y}_n), \bar{x} = (\bar{x}_1, \dots, \bar{x}_n)$ with $\bar{y}_i = \sum_{j=1}^{d} \alpha_j y_{ij}, \bar{x}_i = \sum_{j=1}^{d} \alpha_j x_{ij}$, $\sum_{j=1}^{d} \alpha_j = 1, i = 1, 2, \dots, n,\ j = 1, 2, \dots, d,$

where the weights $\alpha = (\alpha_1, \ldots, \alpha_d)$ can be chosen arbitrarily, one defines that X is "directionally majorized" by Y if no region of every Lorenz curve of $\bar{y}$ lies above the corresponding Lorenz curve of $\bar{x}$.

As the weights $\alpha_1, \ldots, \alpha_d$ can be chosen arbitrarily one also avoids the often problematic weightselection, which should be performed in accordance with the economic estimation or welfare judgement of the different attributes.

Using this definition we can moreover assume that every distribution of an attribute x_i is equivalent to all distributions of $\lambda \cdot x_i$, $\lambda > 0$. For that reason we can assume that all matrices X and Y have normalized row sums, where the exogeneously determined weight α_i of the attribute x_i is taken in account by an α_i-fold replication of x_i. From now on we will in future consider normalized $d \times n$ matrices X and Y where the d rows show the distributions of equally estimated attributes of n units. (Observe that replications of the attribute reflect the welfare importances of the attributes.)

For this case we will specialize the already explained majorization by comparing only the Lorenz curves of the rows $x_1, \ldots, x_d$ and the mean row $\bar{x} = \frac{1}{d}(\sum_{j=1}^{d} x_{1j}, \ldots, \sum_{j=1}^{d} x_{nj})$ of X with the corresponding Lorenz curves originating from Y. If no region of these $d + 1$ Lorenz curves emanating from Y lies above the corresponding Lorenz curves emanating from X, we define X as being less inequal than Y.

This spezialized new majorization allows more comparisons because it is not as restrictive as the previously explained majorization. Moreover, it is consistent with the "correlation increasing axiom" proposed by Tsui (1999). However, we will substitute this axiom of Tsui by two natural requirements for eluding Bourguignon's and Chakravarty's (2003) critizism.

To obtain a total order we will in chapter 3 define two new multidimensional generalizations G_1 and G_2 of the univariate Gini coefficient being compatible with different majorizations which will be explained in chapter 1.

It should be mentioned that some economists who are working on welfare research believe that the problem of measuring the inequality of a multivariate distribution would be easily solved if the weights $\alpha_1, \ldots, \alpha_d$ of the attributes were known by determining the trade-off between the different characteristics. Then the valuation vector $\alpha = (\alpha_1, \ldots, \alpha_d)$ can be considered as a price vector by which the wealth position of the individual is determined. In the second step an application of an univariate index of inequality leads to a multi-attribute index's value. However, this two-step method ignores the heterogeneity of each column or individual, and measures only the aggregated inequality of the individuals. We do not proceed this method because the economic inequality of a multidimensional distribution where the attributes are weighted equally could be zero although the columns are very heterogen.

For this reason and because we do not judge economic inequality as injustice or lack of welfare, the neutral word *heterogeneity*, which can be interpreted both in a negative sense as *inequality* and in a positive sense as *diversity* or *dissimilarity*, will be preferred in this paper. Moreover, by this terminology we will distinguish our paper from papers using a social evaluation function for the inequality measurement as it is done by many authors who are mentioned in the working paper of Maria Ana Lugo (2005).

At last it will be shown that the suggested multivariate Gini coefficient G_1 respects the heterogeneities of all attributes more than the heterogeneity of the mean, whereas the measure G_2 respects the heterogeneity of the mean more than the heterogeneities of the attributes. However for measuring the convergence to the equal distribution in practice, the measures G_1 and G_2 are both suitable, as shown by an example in the appendix.

1. A New Majorization

According to Marshall and Olkin (1979) we consider a nonnegative $d \times n$ matrix $Y = (y_{ij})$ where each of the n columns $\boldsymbol{y}'_1, \boldsymbol{y}'_2, \ldots, \boldsymbol{y}'_n$ shows one of the d attributes of the n individuals we compare[1]. We assume cardinally measurable nonnegative attributes, for instance personal income, property, leisure and so on. (The attributes need not be independent because of the possible replications we mentioned in our introduction.) At first we will recapitulate five common definitions for the majorization of a $d \times n$ matrix X by an other $d \times n$ matrix Y.

Definition 1.1 *We say "X is chain majorized by Y", denoted by $X \ll Y$, if $X = Y \cdot P$, where P is produced by a finite number of T-transformations meaning that exactly two columns $\boldsymbol{x}'_i, \boldsymbol{x}'_j$ of X come from two columns $\boldsymbol{y}'_k, \boldsymbol{y}'_\ell$ of Y by the transformation*

$$(\boldsymbol{x}'_i, \boldsymbol{x}'_j) = (\boldsymbol{y}'_k, \boldsymbol{y}'_\ell) \cdot T$$

$$\text{where } T = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{pmatrix}, \;\; 0 \le \alpha \le 1, \quad i, j, k, \ell \in \{1, ..., n\}.$$

Because of $T = \alpha I + (1-\alpha)Q$, where I is the identity and Q is a permutation matrix, our definition is not different from Marshall/Olkin's definition. It can be visualized by the possible position of the vectors $\boldsymbol{x}'_i, \boldsymbol{x}'_j$ if for instance $\boldsymbol{y}'_k, \boldsymbol{y}'_l$ are given and $\alpha = 0.4$.

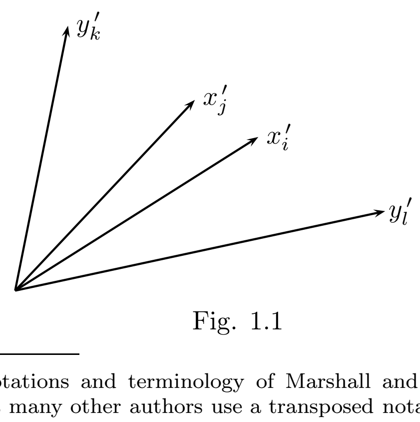

Fig. 1.1

[1] We follow the notations and terminology of Marshall and Olkin, Koshevoy 1995 and Savaglio 2002, whereas many other authors use a transposed notation.

Definition 1.2 *We say "X is majorized by Y", denoted by $X < Y$, if $X = Y \cdot P$, where the $n \times n$ matrix P is doubly stochastic.*
The special case that X and Y are $1 \times n$-matrices (i. e. row vectors x and y), is included and defines the meaning of "$x < y$". (Joe and Verducci 1993 call this majorization "Uniform majorization".)

Definition 1.3 *We say "X is rowwise majorized by Y", denoted by $X <^{row} Y$, if for each row $\boldsymbol{x_i}$ of X and the corresponding row $\boldsymbol{y_i}$ of Y the equation $\boldsymbol{x_i} = \boldsymbol{y_i} \cdot P_i$ holds , where P_i is a doubly stochastic $n \times n$ matrix.*

This definition is equivalent to the following property

$$\boldsymbol{a}_k X < \boldsymbol{a}_k Y \ , \quad k = 1, 2, \ldots, d,$$

with $\boldsymbol{a}_k := (0, \ldots, 0, 1, 0, \ldots, 0)$, all coordinates being zero without the k-th coordinate which is one.
The following first implication is well-known and can be easily proved as the product of two doubly stochastic matrices is always a doubly stochastic matrix.

Theorem 1.1 $X \ll Y \Longrightarrow X < Y \Longrightarrow X <^{row} Y$.

Marshall and Olkin show that the converse of the first implication is not true for $d \geq 2$ and $n \geq 3$. They show that with the aid of the doubly stochastic matrix

$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ which is **not** a product of T-transformation matrices.

The converse of the second implication is wrong for $d \geq 2$, which can be already seen from the example given in the introduction.
Between $X < Y$ and $X <^{row} Y$ there is an other majorization defined by Bhandari (1988) which is called "Directional majorization".

Definition 1.4 *We say "X ist directionally majorized by Y", denoted by $X <^D Y$, if $\boldsymbol{a}\ X < \boldsymbol{a}\ Y$ for any row vector $\boldsymbol{a} \in \mathbb{R}^d$.*
The Directional majorization $<^D$ is also called "price majorization" (see Mosler 1994, where a survey on majorization in economic disparity measures is given).
A consequence of $X <^D Y$ is that every weighted arithmetic mean of the rows of Y is more dispersed than the weighted arithmetic mean of the rows of X using the same weights, which can be interpreted as prices.
The majorization $<^D$ is equivalent to another majorization called "Lorenz majorization", which means that the "Lorenz zonotope" $LZ(X)$ is enclosed by the Lorenz zonotope $LZ(Y)$ (see Koshevoy 1995). The following well-known implications can easily be proved, too.

Theorem 1.2 $X < Y \Longrightarrow X <^D Y \Longrightarrow X <^{row} Y$.

Proof.
$X < Y$ means that there exists a doubly stochastic matrix P with $X = Y \cdot P$. Hence we have $aX = aYP$. Because of $aYP < aY$ for all arbitrary vectors a, we have $X <^D Y$, such that the first implication is proved. The second implication is a trivial consequence of both definitions.

$\square$

The inverse of the first implication is not true for $n \geq 4$, which can be seen by an example given in the appendix where we will prove that for $n = 2$ all majorizations $\ll$, $<$, $<^D$ are equivalent. Moreover there will be proved that for $n = 3$ the relation $X <^D Y$ implies $X < Y$.

The converse of the second implication is wrong for $d \geq 2$. This can be seen by the example given in the introduction if we take the vector $a = (1, 1)$ which implies $(1.5, 0.5) \not< (1, 1)$.

Between $X <^D Y$ and $X <^{row} Y$ we will now define a new majorization which is a special case of "$<^D$" considering only matrices X and Y, normalized by the welfare importance.

Definition 1.5 *We say "X is slightly Lorenz majorized by Y", denoted by* $X <_L Y$*, if* $\boldsymbol{a}_k X < \boldsymbol{a}_k Y$ *for* $\boldsymbol{a}_k = (0, \dots, 0, 1, 0, \dots, 0)$, $k = 1, 2, \dots, d$*, and* $a_{d+1}X < a_{d+1}Y$*, where* $\boldsymbol{a}_{d+1} := \left(\frac{1}{d}, \frac{1}{d}, \dots, \frac{1}{d}\right)$.

This means that every row of X is derived from the corresponding row of Y by means of separate T-transformations, and the same is true for the arithmetic mean $\bar{x}$ of all rows.

It is important that this order is consistent with the "correlation increasing majorization axiom" proposed by Tsui (1999), which demands: Switching the attributes of two individuals i and j so that individual i gets $y_{ki} = \min\{x_{ki}, x_{kj}\}$ for all $k = 1, \dots, d$ and individual j gets $y_{kj} = \max\{x_{ki}, x_{kj}\}$ for all $k = 1, \dots, d$, increases not only the correlation between the individuals i and j in X but also the inequality grade of X. This switching is called "correlation increasing transfer" defining the order $X <_C Y$. (For this definition and notation see Thibault Gajdos and John A. Weymark 2005.) It is obvious that this transfer increases also the spread of $a_{d+1}X$ so that the implication$X <_C Y \Longrightarrow X <_L Y$ holds. We also note

Theorem 1.3 $X <^D Y \Longrightarrow X <_L Y \Longrightarrow X <^{row} Y$.
The converse of both implications are wrong. ($X \not<_L Y \Longrightarrow X \not<^D Y$ is logically equivalent to the first implication.)
The following example shows that $X <_L Y$ does not imply $X <^D Y$.

Example 1.1 Let $X = \begin{pmatrix} 3 & 1 \\ 1 & 3 \\ 4 & 0 \end{pmatrix}$, $Y = \begin{pmatrix} 4 & 0 \\ 0 & 4 \\ 4 & 0 \end{pmatrix}$ be matrices of distribution.

Then we have $X <_L Y$ as $\boldsymbol{a}_k X < \boldsymbol{a}_k Y,\ k = 1,2,3$, and $\boldsymbol{a}_4 X = \left(\frac{8}{3}, \frac{4}{3}\right) = \boldsymbol{a}_4 Y$, whereas $\boldsymbol{b}_4 X = (5,3) > (4,4) = \boldsymbol{b}_4 Y$ for $\boldsymbol{b}_4 = (0,1,1)$. Hence $X \not<^D Y$.
This example shows that the intuitively assumed heterogeneity is supported by $<_L$ and not by $<^D$. The following example will show that $X <^{row} Y$ does not imply $X <_L Y$.

Example 1.2 Let

$$X = \begin{pmatrix} 0.1 & 0.9 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, \ Y = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

be matrices of distribution.
Then we have $X <^{row} Y$ and $X \not<_L Y$ as

$$\boldsymbol{a}_4 X = \left(\frac{0.1}{3}, \frac{2.9}{3}\right) \not< \left(\frac{1}{3}, \frac{2}{3}\right) = \boldsymbol{a}_4 Y \ .$$

It is true that the first attribute in X is not as broadly spread as the first attribute in Y, however the individuals in X appear to be more inequal than the individuals in Y, since one is inclined to aggregate the attributes belonging to each individual. Although the correlation of the attributes in X seems greater than in Y, the correlation increasing axiom does not work because there are no switches.

A further example containing all possible switches shall test the intuitively believed heterogeneity.

Example 1.3 Let

$$X_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \ X_2 = \begin{pmatrix} 3 & 2 & 1 \\ 1 & 2 & 3 \end{pmatrix}, \ X_3 = \begin{pmatrix} 1 & 3 & 2 \\ 1 & 2 & 3 \end{pmatrix},$$

$$X_4 = \begin{pmatrix} 3 & 1 & 2 \\ 1 & 2 & 3 \end{pmatrix}, \ X_5 = \begin{pmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \end{pmatrix}, \ X_6 = \begin{pmatrix} 2 & 1 & 3 \\ 1 & 2 & 3 \end{pmatrix}$$

be matrices of distribution. Observe that the heterogeneities of all six matrices are considered as equivalent with regard to $<^{row}$. This coincides with the view that the attributes are totally separated from the individuals. On the other hand there is no comparison by means of $\ll, <, <^D$ possible. However, the correlation increasing axiom decides $X_2 <_C X_4 <_C X_3 <_C X_1$ and $X_2 <_C X_5 <_C X_6 <_C X_1$ so that X_2 is the best distribution and X_1 the worst distribution. Moreover, we have $X_4 <_C X_6$, but between X_4 and X_5 or X_3 and X_6 there are no decisions. With regard to $<_L$ we have also to compare the rows $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})X_i$: $(1,2,3)$, $(2,2,2), (1,2.5,2.5)$, $(2,1.5,2.5)$, $(1.5,2.5,2)$, $(1.5,1.5,3)$, in order to obtain the heterogeneity position.

From this viewpoint the heterogeneity of X_2 is less than the heterogeneities of all other matrices, and X_1 has the highest heterogeneity as all other rows $a_3 X_i$ come from a T-transformation of $(1,2,3)$.

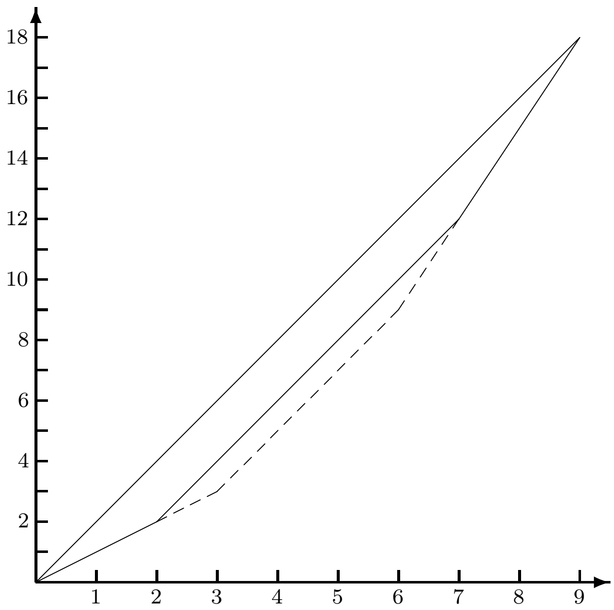

Fig. 1.2

Moreover, we obtain $X_4 <_L X_5$ and $X_5 <_L X_4$ so that both matrices have the same heterogeneity. The comparisons are completed by $X_5 <_L X_6$. However, the decision between X_3 and X_6 is not possible by means of $<_L$.

The order which is effected by $<_L$ can also be visualized by the Lorenz curves of the 3×2 elements of the X_i, extended by the elements of the third rows. This idea will now be explained in detail.
Denoting the extended matrices by $\bar{X}_i$ we get

$$\bar{X}_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad \bar{X}_2 = \begin{pmatrix} 3 & 2 & 1 \\ 1 & 2 & 3 \\ 2 & 2 & 2 \end{pmatrix}, \quad \bar{X}_3 = \begin{pmatrix} 1 & 3 & 2 \\ 1 & 2 & 3 \\ 1 & 2.5 & 2.5 \end{pmatrix},$$

$$\bar{X}_4 = \begin{pmatrix} 3 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 1.5 & 2.5 \end{pmatrix}, \bar{X}_5 = \begin{pmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \\ 1.5 & 2.5 & 2 \end{pmatrix}, \bar{X}_6 = \begin{pmatrix} 2 & 1 & 3 \\ 1 & 2 & 3 \\ 1.5 & 1.5 & 3 \end{pmatrix}.$$

An ascending order of the nine elements of every matrix $\bar{X}_i$, $i = 1, 2, \ldots, 6$, produces the row vectors

$$\begin{aligned} x_1 &= (1,1,1,2,2,2,3,3,3), & x_2 &= (1,1,2,2,2,2,2,3,3), \\ x_3 &= (1,1,1,2,2,2.5,2.5,3,3), & x_4 &= (1,1,1.5,2,2,2,2.5,3,3), \\ x_5 &= (1,1,1.5,2,2,2,2.5,3,3), & x_6 &= (1,1,1.5,1.5,2,2,3,3,3) \ . \end{aligned}$$

The Lorenz curves of these six distributions $x_1, \ldots, x_6$ are only different because of the elements coming from the third rows of $\bar{X}_i$, and these elements decide the order of the x_i and X_i according $<_L$, respectively. So the Lorenz curve of x_1 is at least as low as all other Lorenz curves. Fig. 1.2 shows the Lorenz curves of x_1 and x_2.

Drawing all six Lorenz curves one can see that $X_2 <_L X_4 =_L X_5 <_L X_3 <_L X_1$ holds, but there is no decision between X_3 and X_6. In order to get a decision it is obvious that we will analogously to the univariate Gini coefficient G define a multivariate Gini coefficient G_1 as the ratio of the well known areas of the polygone and the triangle. Doing so we get $G_1(X_3) = \frac{34}{9 \cdot 18} < \frac{35}{9 \cdot 18} = G_1(X_6)$.

2. Multidimensional Heterogeneity Measures

Inspired by, but different from Tsui (1995) and Tsui (1999) where properties of multidimensional inequality measures are formulated as axioms, we demand: A heterogeneity measure $I : X \to \mathbb{R}_+$, where a $d \times n$ matrix X represents the situation of n individuals possessing d nonnegative cardinally measurable attributes, should fulfil the following axioms.

(A_0) **Continuity axiom**
I is continuous.

(A_1) **Normalization axiom**
$0 \le I(X) \le 1$ for all X and
$I(X) = 0$ iff the n columns of X are identical.

(A_2) **Anonymity axiom**
a) $I(X)$ does not change if the rows of X are permutated.
b) $I(X)$ does not change if the columns of X are permutated.

(A_3) **Transfer axiom**, *aggravated by (A_3^*)*
$I(X) \le I(Y)$ for $X \ll Y$, (A_3^*) : $I(X) \le I(Y)$ for $X <_L Y$.

(A_4) **Zero Homogeneity axiom**
$I(X) = I(CX)$, where $C = diag(c_1, \ldots, c_d)$ is a diagonal matrix with $c_k > 0,\ k = 1, 2, \ldots, d$, so that one can make all positive row-sums identical, especially one.

(A_5) **Reduction axiom**, *aggravated by (A_5^*)*

$$I(\begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{11} & x_{12} & \ldots & x_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{11} & x_{12} & \ldots & x_{1n} \end{pmatrix}) \stackrel{(A_5)}{=} cI(\begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \end{pmatrix})$$
$$\stackrel{(A_5^*)}{=} G(x_{11}, x_{12}, \ldots, x_{1n})$$

where c is a constant depending on the dimension d, and G is the normal univariate Gini coefficient.

(A_6) **Antisplitting axiom**
$I(X)$ must not be generally insensitive to a separate permutation within a row while the other rows are unchanged.

Some remarks on these axioms will explain and add comment to these requirements. (Note that for $d = 1$ the axioms $(A_0), (A_2), (A_3)$, (A_4) are also required for concentration measures (see Koshevoy and Mosler 1999.)

With regard to (A_1):
If one takes the volume of the zonoid as an heterogeneity measure, in order to extend the univariate Gini coefficient, the inverse implication of (A_1) is not fulfilled. That is the reason why Koshevoy and Mosler (1996) suggested the "expanded" volume as a "volume Gini mean difference" which according to our transposed notation is defined by

$$M_V(X) = \frac{1}{2^d - 1} \sum_{s=1}^{d} \frac{1}{n^{s+1}} \sum_{1 \le i_1 < \ldots < i_s \le d} \; \sum_{1 \le r_1 < \ldots < r_{s+1} \le n} \left|\det\begin{pmatrix} \mathbf{1} \\ A_{i_1 \ldots i_s}^{r_1 \ldots r_{s+1}} \end{pmatrix}\right|$$

where $\mathbf{1}$ is a row of ones, and $A_{i_1 \ldots i_s}^{r_1 \ldots r_{s+1}}$ is the matrix obtained from the rows $i_1, \ldots, i_s$ and the columns $r_1, \ldots, r_{s+1}$.
Moreover, as already mentioned in the introduction, any two-step inequality measure does not fulfil the inverse implication of (A_1) if X is normalized.
With regard to (A_2):
These natural axioms are required for the rows and the columns because the attributes or individuals can be arranged or ordered arbitrarily.

With regard to (A_3) and (A_3^*):
The requirement (A_3) is indispensable for every multidimensional heterogeneity measure. If the attributes are totally independent so that no aggregation of the attributes makes sense, we have to drop $<_L$ and consequently (A_3^*) which considers the arithmetic mean of the attributes. But if we can assume that the attributes could substitute each other partially we can require (A_3^*) in order to reach a decision in cases comparable with example 1.3.

With regard to (A_4):
As the Gini coefficient is a relative heterogeneity measure it is clear that any generalization of the Gini coefficient must fulfil a special zero-homogeneity.
The objection that the same estimation of every attribute is too specific can be rejected by the agreement that a doubly important attribute is represented by two identical rows.

With regard to (A_5) and (A_5^*):
In the axiomatic research of univariate heterogeneity measures there is no doubt that "cloning" of a population should not change the heterogeneity measure of the population. For this purpose Tsui (1999) demands the Replication Invariance (RI): $I^{nr}(X, X, \ldots, X) = I^n(X)$ for any r. On the other hand if all rows $\boldsymbol{x}_i$ of the matrix X are $\lambda_i \boldsymbol{x}_1$ (that means by assuming (A_4), all rows $\boldsymbol{a}_i$ of the normalized $\bar{X}$ are identical) we have a "cloning" of the first attribute. From this point of view the cloning of one attribute should not change the multivariate heterogeneity measure, too. As a consequence it is necessary to make $c = 1$ for all d as an aggravation of (A_5). However this requirement would cause a problem if we want to generalize the univariate Gini coefficient $G(x)$. The reason for this problem is that the Gini coefficient is normalized by the mean of the only row x and we do not want to normalize the multivariate coefficient $I(X)$ by the mean

of only one row, but rather by the mean of all x_{ij}. However, by the latter mean and the demand $c = 1$ in (A_5) we obtain a contradiction since the mean of all elements x_{ij} taken from the two matrices in (A_5) are different. For that reason we remove the requirement for $c = 1$ and instead include the other aggravation (A_5^*) of (A_5) which guarantees a generalization of the normal Gini coefficient.

For instance, demanding (A_5) with $c = 1$ the first generalization of the Gini coefficient in Koshevoy and Mosler (1997) called the "distance-Gini mean difference" which (according to our proposal) is defined by

$$M_D(X) = \frac{1}{2n^2 d} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\sum_{s=1}^{d} (x_{si} - x_{sj})^2 \right)^{1/2},$$

should be changed. A possible change is

$$\tilde{M}_D(X) = \frac{1}{2n^2 d} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{s=1}^{d} |x_{si} - x_{sj}|$$

which fulfils (A_5) with $c = 1$.
(The measures $2dM_D$ and $2d\tilde{M}_D$ are desirable because of the attractive interpretation as the mean of the Euclidean distance or the mean distance, induced by the ℓ_1-norm, of two individuals who were taken one by one from the population by the same chance with probability $1/n$, see Rao 1982.)
Observe that for $d = 1$ $\tilde{M}_D(X)$ coincides with $M_D(X)$ and $M_V(X)$.
However, a better change of M_D in order to fulfil (A_5^*) is the multiplication of $M_D(X)$ with $\frac{\sqrt{d}}{\mu}$ where $\mu := \frac{1}{n^2} \sum_{i,j=1\ldots n}^{n} x_{ij}$. We prefer the measure $\frac{\sqrt{d}}{\mu} M_D(X)$ as it fulfils (A_6), too, whereas $\frac{1}{\mu}\tilde{M}_D$ cannot fulfil (A_6).

With regard to (A_6):
Let us regard only two individuals with d attributes. If every permutation of the coordinates within a row which is not caused by a permutation of both columns does not change the heterogeneity measure, we cancel out any individuality. As we do not agree with this effect, (A_6) is required to be adhered to.

3. New generalizations of the Gini coefficient

Definition 3.1

1. *In order to define the generalization G_1 of the Gini coefficient G, we start with the original $d \times n$ matrix $X = (x_{ij})$, which can be normalized so that the row-sums are equal to one (without loss of generality). The resultant matrix is denoted by $\bar{X} = (a_{ij})$.*
2. *We extend $\bar{X}$ by a row $\boldsymbol{a}_{d+1}$ which is the arithmetic mean of the other rows. The resultant matrix is denoted by $\bar{\bar{X}}$.*

3. *We calculate the normal Gini coefficient of the* $(d+1) \times n$ *elements of* $\bar{\bar{X}}$ *and call it the Generalized multivariate Gini coefficient* G_1*, which is defined by*

$$G_1(X) = \frac{2 \cdot \sum_{j=1}^{m} j a_j - (m+1) \sum_{j=1}^{m} a_j}{m \sum_{j=1}^{m} a_j} \quad ,$$

where $m = n \times (d+1)$ *and the* a_j *are the* m *ascendingly ordered* a_{ik}*,* $i = 1, 2, \ldots, d+1,$ $k = 1, 2, \ldots, n$ *of the extended matrix* $\bar{\bar{X}}$*.*

Theorem 3.1
The heterogeneity measure G_1 *fulfils the axioms* $(A_0), \ldots, (A_6)$ *as well as* (A_3^*) *and* (A_5^*)*.*

Proof.
(A_5^*) is fulfilled as the normal Gini coefficient fulfils the so-called Replication axiom, i. e. "the heterogeneity value of n data is unchanged if these data are cloned d-times."
(A_1) is fulfilled, since the value of the normal Gini coefficient lies between zero and one and is zero and only zero if the n data are identical. Moreover, the idendity of all rows of X means the identity of all $n \times d$ elements of $\bar{X}$.
(A_4) is fulfilled because of the definition of $G_1(X)$.
The axioms (A_0) and (A_2) are fulfilled because of the definition of $G_1(X)$ and as the normal Gini coefficient has these properties. This is also true for the aggravated transfer axiom (A_3^*) and the last axiom (A_6). □

One disadvantage of the axiom (A_3^*) is caused by the majorization $<_L$ which graphically does not show the heterogeneity of the attributes of each individual as the data of each individual are no longer connected after having arranged all $n \times (d+1)$ data of $\bar{\bar{X}}$ in increasing order. That is the reason why we suggest another heterogeneity measure[2]. For this purpose we start with the steps 1 and 2, already defined in Definition 3.1. Then we add three further steps:

3. We draw the Lorenz curve belonging to the row $\boldsymbol{a}_{d+1}$ of X, that is the mean of the other rows, and call it the "mean Lorenz curve".

4. The attributes of each of the n individuals are ascendingly ordered, and the data divided by d are drawn as n separate Lorenz curves lying beyond the mean Lorenz curve. These n separate Lorenz curves are consequently drawn with the scales $1/d$ of the original coordinates.

5. The ratio of the area between the n convex Lorenz curve pieces and the line of zero disparity and the area of the triangle bordered by the line of zero disparity and the x-axis we denote $G_2(X)$ which is consequently defined by

[2] This separable measure was suggested by Wolfgang Bischoff, whom we thank very much for several private communications.

Definition 3.2
G_2 is the sum of the Gini coefficient of the n mean attributes $a_{d+1,j}$ and the n Gini coefficients of the attributes of each individual multiplied with the weight $\frac{1}{d}\sum_{i=1}^{d} a_{ij}/n\sum_{j=1}^{n} a_{d+1,j}$ implying the formula

$$G_2(X) = \frac{2\sum_{j=1}^{n} j a_{d+1,j} - (n+1)\sum_{j=1}^{n} a_{d+1,j}}{n\sum_{j=1}^{n} a_{d+1,j}} + \frac{\frac{1}{d\cdot d}\sum_{j=1}^{n}\left(2\sum_{i=1}^{d} i a_{ij} - (d+1)\sum_{i=1}^{d} a_{ij}\right)}{n\sum_{j=1}^{n} a_{d+1,j}} ,$$

where the n numbers $a_{d+1,j} \in \bar{\bar{X}},\ j = 1, 2, \ldots, n$ are ascendingly ordered, and the d numbers $a_{ij} \in \bar{X},\ \ i = 1, 2, \ldots, d$ are arranged in increasing order for every fixed $j \in \{1, 2, \ldots, n\}$.

G_2 is, because of this interpretation, very close to the univariate Gini coefficient G, which is so popular that Shorrocks (1984) seems to deplore that it does not belong to the class of decomposable inequality measures he characterizes. (He mentions that the Gini coefficient ist decomposable under all *non-overlapping* partitions of the income distribution and suggests to extend the research in this direction. However, this way cannot lead to a characterization.)
It is remarkable that the matrices X_3 and X_6 in example 1.3 have the same heterogeneity number $G_2(X_3) = 7/36 = G_2(X_6)$.
$G_2(X_3)$ and $G_2(X_6)$ are also illustrated by the polygonial areas in the following figures.

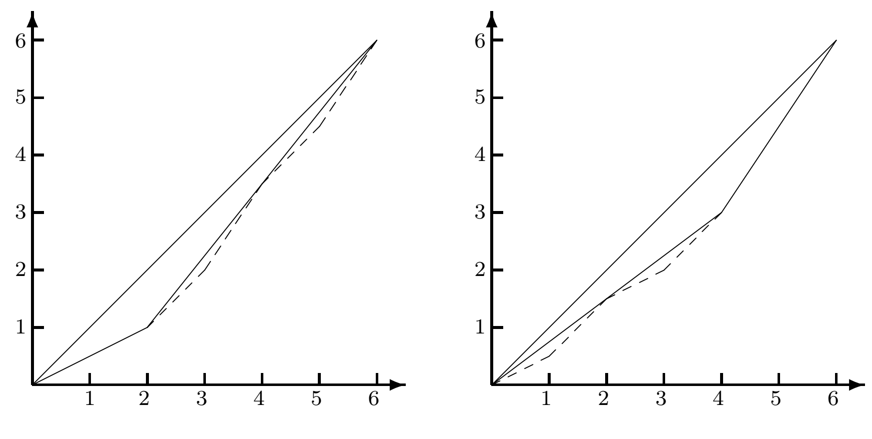

Fig. 3.1

One can verify that $G_2(X)$ fulfils the axioms $(A_0), (A_1), (A_2), (A_3), (A_4), (A_5^*), (A_6)$. We have to accept that (A_3^*) is not generally fulfilled by $G_2(X)$

which can be demonstrated by the following example.

Example 3.1
Let matrices of distributions be

$$X = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 11 \\ 2 & 0 & 10 \\ \frac{7}{2} & \frac{17}{2} & 0 \end{pmatrix} , \quad Y = \begin{pmatrix} 1 & 0 & 11 \\ 2 & 0 & 10 \\ 3 & 9 & 0 \end{pmatrix} ,$$

where both matrices have the mean row vector $\boldsymbol{a}_4 = (2, 3, 7)$, so that we have $X <_L Y$.
At first we calculate
$G_1(Y) = G((0, 0, 0, 1, 2, 2, 3, 3, 7, 9, 10, 11) = \frac{308}{12 \cdot 48} = 0,534$,
$G_1(X) = G(0, 0, \frac{1}{2}, \frac{1}{2}, 2, 2, 3, \frac{7}{2}, 7, \frac{17}{2}, 10, 11) = \frac{305}{12 \cdot 48} = 0.529$,
implying $G_1(X) < G_1(Y)$ in accordance with (A_3^*).
Because of $G_2(Y) = \frac{134}{18 \cdot 18} = 0.4135 < \frac{135}{18 \cdot 18} = 0.4166 = G_2(X)$, the axiom (A_3^*) is not fulfilled by G_2 (which is only possible for $n \geq 3$).
Comparing with the measure $\frac{\sqrt{d}}{\mu} M_D$ we note that it supports the decision of G_1, but it shows a "bad reaction" as to the example 1.3

$$M_D(\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}) = M_D(\begin{pmatrix} 3 & 2 & 1 \\ 1 & 2 & 3 \end{pmatrix}) < M_D(\begin{pmatrix} 1 & 3 & 2 \\ 1 & 2 & 3 \end{pmatrix})$$

meaning that by a permutation of the "most heterogeneous" distribution the heterogeneity number can grow. However, we think that a mixing of a "well ordered" multidimensional distribution - unimportant if the attributes are independent - should never increase the heterogeneity number. We will therefore require this property as a new, but natural weak axiom, being a consequence of axiom (A_3^*) as well as a consequence of the correlation increasing axiom.

$(\tilde{A}_3)$ **Mixing axiom**.

The heterogeneity number $I(\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{11} & x_{12} & \dots & x_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{11} & x_{12} & \dots & x_{1n} \end{pmatrix})$
will not be increased by any permutation within a row which disturbs the former order of the attributes.

This "mixing" axiom can be motivated by a consideration coming from physics: We know that disturbing an order increases the entropy. Reaching the maximal entropy of a closed system makes all participants "equal". We can also say: "Decreasing order makes a decreasing inequality".
One can easily prove that the previously suggested measure M_D fulfils the natural axioms $(A_1), (A_2), (A_3), (A_4), (A_5^*), (A_6)$ if we multiply it with $\frac{\sqrt{d}}{\mu}$, but example 1.3 shows that M_D and M_V do not fulfil the weak axiom $(\tilde{A}_3)$. For that reason we will give preference to the new measure G_2 which fulfils$(\tilde{A}_3)$. The properties of $G_1(X)$ and $G_2(X)$ are now summarized by

Theorem 3.2
The first generalization $G_1(X)$ of the Gini coefficient fulfils the system of axioms

$$\text{(a) } (A_0), (A_1), (A_2), (A_3^*), (A_4), (A_5^*), (A_6).$$

The second generalization $G_2(X)$ of the Gini coefficient fulfils the system of axioms

$$\text{(b) } (A_0), (A_1), (A_2), (A_3), (\tilde{A}_3), (A_4), (A_5^*), (A_6)$$

which demands less than the system (a).

The proof of this theorem works equally as well if the unweighted arithmetic mean a_{d+1} in Definition 3.1 and Definition 3.2 is replaced by any weighted arithmetic mean of the rows $a_1, \ldots, a_d$, and we will denote the so generalized measures $\tilde{G}_1$ and $\tilde{G}_2$.

Proof.
It should firstly be noted that (a) and the majority of (b) are analogously proved as in Theorem 3.1. For simplification we assume that all rows of X and Y are already normalized.
In order to prove that $\tilde{G}_2$ fulfils $(\tilde{A}_3)$ we observe that in the special matrix X with identical rows the weighted mean x_{d+1} is the row $(x_{11}, x_{12}, \ldots, x_{1n})$ and $\tilde{G}_1(X) = \tilde{G}_2(X)$. Then any permutation of a row, which changes X reduces the spread of x_{d+1}. Thereby the number $\tilde{G}_1(X)$ is reduced and, because of $\tilde{G}_2(X) \leq \tilde{G}_1(X)$, an increase of $\tilde{G}_2(X)$ is impossible. Hence $\tilde{G}_2$ fulfils $(\tilde{A}_3)$.
Finally, we have to prove in (b) that $\tilde{G}_2(X)$ fulfils (A_3). As $X \ll Y$ implies $X <^D Y$ the mean row $\tilde{x}_{d+1}$ originating from X is in any case less broadly spread than the mean row $\tilde{y}_{d+1}$ originating from Y. Now we consider the generalized formula in Definition 3.2 and assume that X comes from Y by a T-transformation concerning the first two columns. Then the formula $\tilde{G}_2(X)$ contains the term

$$\sum_{i=1}^{d}(2i-d-1)[\alpha y_{i1} + (1-\alpha)y_{i2}] + \sum_{i=1}^{d}(2i-d-1)[(1-\alpha)y_{i1} + \alpha y_{i2}],$$

which does not depend on $\alpha \in [\alpha, 1]$. As the T-transformation diminishes the Gini coefficient of the mean vector $(\tilde{y}_{d+1,1}, \ldots, \tilde{y}_{d+1,n})$ for $0 < \alpha < 1$ the rest of the numerator in the formula is diminished, whereas the denominator in the formula is always unchanged. Since further T-transformations do not increase the value $\tilde{G}_2(Y)$, the proof of (b) is completed. Thus we have $\tilde{G}_2(X) \leq \tilde{G}_2(Y)$.
$\square$

4. Conclusions

We feel it is necessary to reiterate that the antisplitting axiom (A_6) requires that the attributes of the individuals must not be separated from the individuals. Such separation could be avoided if the attributes are aggregated by any mean. And this mean is restricted by (A_3^*) to the arithmetic mean, whereas the mean is not affected by (A_3). Therefore it can be freely chosen.
(For instance we could also prove Theorem 3.2 if the weighted arithmetic mean $\tilde{y}_{d+1}$ vector was substituted by

$$\left(\sqrt{\textstyle\sum_{i=1}^{d} y_{i1}}, \sqrt{\textstyle\sum_{i=1}^{d} y_{i2}}, \dots, \sqrt{\textstyle\sum_{i=1}^{d} y_{in}}\right) \qquad \text{or}$$

$$\left(\textstyle\sum_{i=1}^{d} y_{i1}, \sum_{i=1}^{d} y_{i2}, \dots, \sum_{i=1}^{d} y_{in}\right).)$$

However requiring only (A_3) this axiom should be fulfilled in addition to the weak requirement ($\tilde{A}_3$) which cannot be fulfilled by modifications of the previously introduced measures M_D and M_V fulfilling (A_6). On the other hand any modification of $\tilde{M}_D$ fulfilling ($\tilde{A}_3$) cannot fulfil (A_6). These facts are an additional motivation for the use of the new measures G_1 and G_2.
Comparing Figures 1.2 and 3.1 we obtain more information from Figure 3.1. Moreover G_2, which is visualized by Fig. 3.1, does not prefer X_3 or X_6. There is no doubt that the geometrical interpretation of G_2 is more easily understood than the interpretation by a lift zonoid.

For that reason we believe that politicians who are interested in the convergence of the regions of their responsibility have a better insight by the geometrical interpretation of G_2. This aspect is important in states where the legal power is divided by a Federal Constitution as it is in the Federal Republic of Germany.

On the other hand the aggregation of the attributes has a greater influence on the inequality or heterogeneity number if we use G_2 instead of G_1. That can be demonstrated by the "doubtful" situation in the example mentioned in the introduction, which does not fulfil the premises of $(A_3), (A_3^*)$ and $(\tilde{A}_3)$:

$$Y = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } X = \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

There we have $G_1(X) = \frac{13}{36} < \frac{16}{36} = G_1(Y)$ and, because of the more widely spread mean row, we have $G_2(X) = \frac{3}{8} > \frac{2}{8} = G_2(Y)$.

It is obvious that the weight of a welfare aspect aggregating the attributes can, in extreme cases, change the heterogeneity position of X and Y. However, for measuring the convergence to the equal distribution in **practice** the measures G_1 and G_2 are equally suitable as shown in the appendix.

Appendix

In this appendix the concluding remarks are supported by means of an example which is given by Koshevoy (1995, 98):

Suppose, the bundle $(10\frac{2}{3}, 14\frac{2}{3})$ is distributed among four participants, where the matrices of distributions are

$$A = \begin{pmatrix} 1 & 2\frac{2}{3} & 3 & 4 \\ 6 & 3\frac{2}{3} & 3 & 2 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1\frac{2}{3} & 2 & 3.5 & 3.5 \\ 5 & 4\frac{2}{3} & 2.5 & 2.5 \end{pmatrix}.$$

Estimating both attributes as equally important we have to compare the correspondent matrices

$$X = \begin{pmatrix} 66 & 176 & 198 & 264 \\ 288 & 176 & 144 & 96 \end{pmatrix} \text{ and } Y = \begin{pmatrix} 110 & 132 & 231 & 231 \\ 240 & 224 & 120 & 120 \end{pmatrix}.$$

Then we calculate $G_1(X) = 0.17282$ and $G_1(Y) = 0.1450$ which is 83.91 % of $G_1(X)$. Further we calculate $G_2(X) = 0.04936$ and $G_2(Y) = 0.04261$ which is 86.33 % of $G_2(X)$.

As the mean distributions $\bar{x} = (177, 176, 171, 180)$ and $\bar{y} = (175, 178, 175.5, 175.5)$ are both only a little spread, the absolute heterogeneity numbers $G_2(X)$ and $G_2(Y)$ are much smaller than $G_1(X)$ and $G_1(Y)$, whereas the ratios $G_2(Y)/G_2(X)$ and $G_1(Y)/G_1(X)$, indicating the convergence to the equal distribution, are almost the same.

Now we will show the already indicated equivalence of the majorizations $\ll, <, <^D$ for $n = 2$, meaning that the distributions of only two units are compared:

Proof:
It is obvious and already mentioned in Marshall/Olkin (1979, 431) that $\ll$ and $<$ are equivalent for $n = 2$, so we have only to prove that for $n = 2$ the majorization $<^D$ implies the majorization $<$.

At first we have to notice that

$$\begin{pmatrix} x_{11} & x_{12} \\ \vdots & \vdots \\ x_{d1} & x_{d2} \end{pmatrix} = X <^D Y = \begin{pmatrix} y_{11} & y_{12} \\ \vdots & \vdots \\ y_{d1} & y_{d2} \end{pmatrix}$$

is equivalent to

$$\begin{pmatrix} \lambda_1 x_{11} & \lambda_1 x_{12} \\ \vdots & \vdots \\ \lambda_d x_{d1} & \lambda_d x_{d2} \end{pmatrix} <^D \begin{pmatrix} \lambda_1 y_{11} & \lambda_1 y_{12} \\ \vdots & \vdots \\ \lambda_d y_{d1} & \lambda_d y_{d2} \end{pmatrix}$$

for all $\lambda_i \neq 0$ as instead of multiplying X and Y with $a = (a_1, \ldots, a_d)$ we can also multiply with $a_\lambda = (\lambda_1 a_1, \ldots, \lambda_d a_d)$ without changing the majorization $aX < aY$ because the vector a can be arbitrarily chosen.

If a row of Y is (y_{i1}, y_{i1}), then the corresponding row (x_{i1}, x_{i2}) of X must also be $\lambda(y_{i1}, y_{i1})$ because of $X <^D Y$. As all rows of X and Y of this kind have no influence on the relations $<^D$ and $<$, such rows will be excluded in the following considerations.

Assuming that the first rows of X and Y are not excluded, we take $a = (1, 0, \ldots, 0)$, and have

$$(x_{11}, x_{12}) < (y_{11}, y_{12})$$

that means the existence of a matrix $P_\alpha = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{pmatrix}$ for $\alpha \in [0,1]$ such that

$$(x_{11}, x_{12}) = (y_{11}, y_{12}) P_\alpha. \tag{1}$$

Now we will show that also $(x_{i1}, x_{i2}) = (y_{i1}, y_{i2}) P_\alpha$ holds for $i = 2, \ldots, d$, where for all $\lambda_i \in \mathbb{R}$ we have $(y_{i1}, y_{i2}) \neq \lambda_i (y_{11}, y_{12})$. Then we can unite these d equations to $X = Y \cdot P_\alpha$, and $X < Y$ is proved:

The assumption that there exists a row $i \in \{2, \ldots, d\}$ with for instance

$$\lambda(x_{21}, x_{22}) = \lambda(y_{21}, y_{22}) P_\beta, \quad \lambda \neq 0, \tag{2}$$

where $P_\beta = \begin{pmatrix} \beta & 1-\beta \\ 1-\beta & \beta \end{pmatrix}$ is different from P_α, leads to the equation

$$(x_{11} - \lambda x_{21}, x_{12} - \lambda x_{22}) = (y_{11}, y_{12}) P_\alpha - \lambda (y_{21}, y_{22}) P_\beta \ . \tag{3}$$

On the other hand by multiplying X and Y with $a = (1, -\lambda, 0, \ldots, 0)$ we must have a matrix $P_\gamma = \begin{pmatrix} \gamma & 1-\gamma \\ 1-\gamma & \gamma \end{pmatrix}$ with

$$\begin{aligned} (x_{11} - \lambda x_{21}, x_{12} - \lambda x_{22}) &= (y_{11} - \lambda y_{21}, y_{12} - \lambda y_{22}) P_\gamma \\ &= (y_{11}, y_{12}) P_\gamma - \lambda (y_{21}, y_{22}) P_\gamma, \end{aligned} \tag{4}$$

where, different from α and β, γ depends on λ.

Hence the thereby implied equation

$$(y_{11}, y_{12}) P_\gamma - \lambda (y_{21}, y_{22}) P_\gamma = (y_{11}, y_{12}) P_\alpha - \lambda (y_{21}, y_{22}) P_\beta$$

leads to the equation

$$(y_{11}, y_{12})(\gamma - \alpha) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \lambda (y_{21}, y_{22})(\gamma - \beta) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

implying

$$(\gamma - \alpha)(y_{11} - y_{12}) = (\gamma - \beta) \lambda (y_{21} - y_{22}). \tag{5}$$

Because of $y_{11} - y_{12} \neq 0$ and $\lambda(y_{21} - y_{22}) \neq 0$ a consequence of (5) is that for $\gamma = \alpha$ or $\gamma = \beta$ the equality $\alpha = \beta = \gamma$ must hold.

On the other hand the equation (5) is equivalent to

$$\gamma(y_{11} - y_{12} - \lambda(y_{21} - y_{22})) = \alpha(y_{11} - y_{12}) - \beta\lambda(y_{21} - y_{22}) \tag{6}$$

Defining $\lambda := \frac{y_{11}-y_{12}}{y_{21}-y_{22}}$ we have to distinguish two cases:

1. The right side of (6) is zero.
   Then we can choose $\gamma = \alpha$ implying $\gamma = \beta$ because of (5).
2. The right side of (6) is not zero.
   Then we get a contradiction in (6), so that this case must be excluded. □

In order to prove that for $n = 3$ the majorization $<^D$ implies the majorization $<$ it is advantageous to use the equivalence of the "Lorenz majorization" and the "directional majorization" meaning that the Lorenz zonotop $LZ(X)$ lies in the Lorenz zonotop $LZ(Y)$ (see Koshevoy 1995). Abstaining from the normalization of the height of both zonotops we can thereby conclude
$X = (x_1', x_2', \dots, x_n') <^D Y = (y_1', y_2', \dots, y_n') \Longleftrightarrow$

$$\underset{0 \le \alpha_i \le 1, i = 1, \dots, n}{\left\{\alpha_1 \binom{1}{x_1'} + \alpha_2 \binom{1}{x_2'} + \dots + \alpha_n \binom{1}{x_n'}\right\}} \subset \underset{0 \le \beta_i \le 1, i = 1, \dots, n}{\left\{\beta_1 \binom{1}{y_1'} + \beta_2 \binom{1}{y_2'} + \dots + \beta_n \binom{1}{y_n'}\right\}}$$

For $n = 3$ this inclusion means $x_j' = \lambda_{1j}y_1' + \lambda_{2j}y_2' + \lambda_{3j}y_3'$, $0 \le \lambda_{ij}$, $\sum_{i=1}^{3} \lambda ij = 1$, $i, j = 1, 2, 3$.

In this case we have $X = Y \cdot P$ with $X = (x_1', x_2', x_3')$, $Y = (y_1', y_2', y_3')$ and the bistochastic matrix

$$P = \begin{pmatrix} \lambda_{11} & \lambda_{12} & 1 - \lambda_{11} - \lambda_{12} \\ \lambda_{21} & \lambda_{12} & 1 - \lambda_{21} - \lambda_{22} \\ 1 - \lambda_{11} - \lambda_{21} & 1 - \lambda_{12} - \lambda_{22} & \lambda_{11} + \lambda_{12} + \lambda_{21} + \lambda_{22} - 1 \end{pmatrix}$$

meaning $X < Y$. □

For $n = 4$ the implication "$X <^D Y \Longrightarrow X < Y$" can be wrong if not every column vector of Y (in $\mathbb{R}^2$) is a boundary point in the convex hull of the column vectors (see Bhandari 1988). This is the case in the following extremely simple example

$$X = \begin{pmatrix} 0.25 & 0.25 & 0.5 & 0.5 \\ 0 & 0.5 & 0.5 & 0.5 \end{pmatrix} <^D Y = \begin{pmatrix} 0 & 0.5 & 0 & 1 \\ 0 & 0.5 & 1 & 0 \end{pmatrix},$$

where the nonexistence of a doubly stochastic matrix that transforms Y to X follows from Lemma 3 in Koshevoy (1995). The example given by Koshevoy

(1995, 98) should also show that $B <^D A$ and $B \nless A$ is possible, but $LZ(B) \subset LZ(A)$ is wrong. That is the reason why an other example is necessary to show that for $n \geq 4$ the case "$X <^D Y$ and $X \nless Y$" is possible.

Bibliography

Bhandari, S. K. (1988), Multivariate Majorization and Directional Majorization; positive Results, in: *Sankhya*: *The Indian Journal of Statistics 50*, 199–204

Bourguignon, F./F. B. Chakravarty (2003), The Measurement of Multidimensional Property, in: *Journal of Economic Inequality 1*, 25–49

Gajdos, Th./J. A. Weymark (2005), Multidimension Generalized Gini Incides, in: *Economic Theory 26*, 471–496

Joe, H./J. Verducci (1993), Multivariate Majorization by Positive Combinations, in: *Stochastic IMS Lecture Notes, Monograph Series 22*, 159–181

Koshevoy, G. (1995), Multivariate Lorenz Majorization, in: *Soc Choice Welf 12*, 93–102

—/K. Mosler (1996), The Lorenz Zonoid of a Multivariate Distribution, in: *J Amer Statist Assoc 91(434)*, 873–882

—/— (1997), Multivariate Gini Indices, in: *Journal of Multivariate Analysis 60*, 252–276

—/— (1999), Measuring Multidimensional Concentration: A Geometric Approach, in: *Allg. Statistisches Archiv 83*, 173–189

Lugo, M. A. (2005), *Working Paper Series ECINEQ 2005–14*

Marshall, A. W/I. Olkin (1979), *Inequalities: Theory of Majorization and its Applications*, New York

Mosler, K. (1994), Majorization in Economic Disparity Measures, in: *Linear Algebra and Its Applications 199*, 91–114

Rao, C. R. (1982), Diversity and Dissimilarity Coefficients: A Unified Approach, in: *Theoret Population Biol 21*, 24–43

Savaglio, E. (2002), Multidimensional Inequality: A Survey, in: *QUADERNI Università degli Studi di Viena*, 1–27

Shorrocks, A. F. (1984), Inequality Decomposition by Population Subgroups, in: *Econometrica 52*, 1369–1385

Tsui, K. Y. (1995), Multidimensional Generalizations of the Relative and Absolute Inequality Indices: The Atkinson-Kolm-Sen Approach, in: *J Econom Theory 67*, 251–265

— (1999), Multidimensional Inequality and Multidimensional Generalized Entropy Measures: An Axiomatic Derivation, in: *Soc Choice Welf 16*, 145–157