POLSKA AKADEMIA NAUK, INSTYTUT MATEMATYCZNY

DISSERTATIONES MATHEMATICAE

(ROZPRAWY MATEMATYCZNE)

CCCXXXV

TERESA KOWALCZYK

A unified Lorenz-type approach to divergence and dependence

WARSZAWA 1994

Teresa Kowalczyk
Institute of Computer Science
Polish Academy of Sciences
P.O. Box 22
J. Ordona 21
01-237 Warszawa
Poland
E-mail: tkow@wars.ipipan.waw.pl

Published by the Institute of Mathematics, Polish Academy of Sciences
Typeset in TEX at the Institute
Printed and bound by
drukarnia
herman & herman
02-240 Warszawa, ul. Jakobinów 23, tel: 846-79-66, tel/fax: 49-89-95

P R I N T E D I N P O L A N D

ISSN 0012-3862

CONTENTS

1991 *Mathematics Subject Classification*: 62H30, 62H20, 90A19.

Received 20.7.1993; revised version 28.1.1994.

Abstract

The paper deals with function-valued and numerical measures of absolute and directed divergence of one probability measure from another. In case of absolute divergence, some new results are added to the known ones to form a unified structure. In case of directed divergence, new concepts are introduced and investigated. It is shown that the notions of absolute and directed divergences complement each other and provide a good insight into the extent and the type of discrepancy between two distributions. Consequently, these measures applied together to suitably chosen pairs of distributions prove useful to express such statistical concepts as inequality, dependence, and departures from proportionality.

Introduction

The central notion of the paper is the concentration curve which has been introduced by Cifarelli and Regazzini (1987). Some concepts particularly important for the definition of the divergence curve were proposed by Ali and Silvey (1965, 1966).

The concentration curve is a function-valued measure of the divergence of one probability measure from another. It is defined for arbitrary pairs of probability measures and reflects any kind of discrepancy between them. Thus, it measures *absolute* divergence.

The concentration curve of the probability measure Q with respect to the probability measure P refers to the most powerful test of the null hypothesis $H_0 : P$ against the alternative hypothesis $H : Q$. Roughly speaking, the curve is isometric to a plot of the distribution functions of the most powerful test generated by P and Q, respectively. This is the plot $1 - \alpha$ versus β, where α and β are respectively the probabilities of the errors of the first and second kind. The plot of α versus $1 - \beta$ appears in many textbooks on hypothesis testing (see e.g. Lehmann (1959), Grove (1980)). These plots induced some orderings useful in the testing theory. On the other hand, the plot of α versus β as a measure of divergence of P from Q was proposed by Bromek and Kowalczyk in a paper which appeared in 1990 in the proceedings of a conference held in 1988 in Pittsburgh. In that paper, written parallelly to Cifarelli and Regazzini (1987), stress was laid on properties of the ordering based on the $\alpha - \beta$ plot.

The notion of concentration curve can be used to define other statistical concepts when the curve is applied to suitably chosen pairs of probability measures. One objective of the present paper is to describe such applications in the case of inequality, dependence, and departures from proportionality. In the case of inequality the concentration curve becomes the Lorenz curve, which is a well-known function-valued measure of inequality. It has been frequently used in socio-economic investigations of income and other distributions. The interest in this parameter and its applications is still vivid. Recent contributions to the subject were given e.g. in the works by Arnold (1987) and Foster (1985). The counterparts of the Lorenz curve are used in various fields of applied stochastic science. This is exemplified by the curve related to the so called total-time-on-test-transformation which has an important place in reliability theory (cf. Chandra and Singpurwalla (1981), Klefsjö (1984)).

Links between divergence and inequality are of two kinds. First, the divergence of any two mutually absolutely continuous probability measures is equal to the inequality of the distribution of their likelihood ratio generated by the first measure. This was mentioned in Cifarelli and Regazzini (1987), while Gafrikova and Kowalczyk (1994) used it to study duality of orderings of inequality and divergence. Second, the inequality of a nonnegative variable X with finite expectation may be represented as the divergence between the distribution P_X of X and the distribution λ_{P_X} which assigns to any $B \in \mathcal{B}(\mathbb{R}^+)$ the probability $\int_B x\,dP_X(x)/E(X)$.

It is worth noting that Fogelson (1933) introduced a curve measuring inequality for any nonnegative random variable X with finite expectation as a plot of the distribution function of P_X and λ_{P_X}. Thus, he invented the concentration curve for this particular case.

A Lorenz-type approach to dependence relates to the fact that dependence can be considered as divergence between the joint distribution and the product of the marginal distributions. The related dependence curve was introduced during the conference on dependence in Pittsburgh independently by Bromek and Kowalczyk and by Scarsini (cf. the proceedings edited by Block, Sampson and Savits (1990)). But links between dependence and divergence had been studied before by many authors. Ali and Silvey (1965, 1966) studied measures of dependence based on the likelihood ratio of the joint and product distributions. This subject was also considered by Joe (1985, 1987).

Chapters 1–3 present these topics. Chapter 1 deals with measures of divergence which are used in Chapters 2 and 3 to measure inequality and dependence, respectively. These three chapters contain only few new results but collect material from many papers, some of them by this author, into a systematic and unified structure. An effort has been made to unify the terminology. New results of Chapters 1–3 are given in Sections 2.5 and 3.3.

Chapter 4 deals with evaluating absolute departures from proportional representation. In this case there are two vectors with positive integer components. The first vector represents a partition of a finite population, the second vector represents a related partition of a representation of prescribed size.

Representation can be formed in a number of ways. It can be a sample drawn from the population according to a chosen rule, deterministic or probabilistic. In the paper we are concerned with representations as near as possible to proportional. Thus, we are interested in the minimal elements for the ordering based on divergence curves.

The idea to use the divergence curve for the population and representation to measure departures from proportionality appeared first in Bondarczuk *et al.* (1994), and will be reminded in Sec. 4.1 of the present paper. The remaining two sections of Chapter 4 provide new results concerning the minimal and maximal elements for the ordering based on divergence curves, applied to departures from proportional representation.

Chapter 5 introduces the *directed* divergence. The difference between absolute and directed divergence is best explained in the case of univariate distributions. Then, the directed departure of P from Q tells how much to the "left" of Q is P. A function-valued measure of this tendency of P is the plot of the distribution function of Q with respect to P. This plot is a special case of the directed concentration curve. If the likelihood ratio of Q with respect to P is increasing then the plot becomes the divergence curve of Q from P.

The general definition and properties of the directed concentration curve are given in Sec. 5.1. This curve coincides with the concentration curve of Q w.r.t. P, introduced in Chapter 1 when P and Q are univariate and the direction is indicated by increasing real values. In Sec. 5.2 any two probability measures on the real line are mapped onto a pair (P', Q') on $[0, 1]$ such that P' is uniform and the distribution function of Q' lies on the directed concentration curve of Q w.r.t. P. It is shown that Q is then mapped onto Q' by the same transition probability function which maps P onto the uniform distribution. It follows that (P', Q') represents the class of pairs (P, Q) with the same directed concentration curve.

The concentration curve measuring absolute departures of Q from P can be used jointly with a suitably chosen directed concentration curve to describe not only the extent but also the type of departures. The two curves coincide when directed departures are the only ones present. In the case of bivariate dependence between random variables X and Y, the two curves can be used to measure both the absolute and the monotone (positive or negative) dependence. Positive dependence is the tendency of larger (smaller) values of X to coappear with larger (smaller) values of Y; negative dependence is described analogously.

The dependence of Y on X is often described by means of the regression $r(x) = E(Y \mid X = x)$. Taguchi (1987) in his study on the so-called concentration surface considered the plot of $E(r(X); X \leq x)/E(Y)$ versus $P(X \leq x)$, which he called the *correlation curve*. If Y is nonnegative and $E(Y)$ is finite then this curve is the directed concentration curve for the two distributions. Taguchi also introduced the plot of $E(r(X); X \leq x)/E(Y)$ versus $E(X; X \leq x)/E(X)$ as the *ratio curve*. This plot is a directed concentration curve if both X and Y are nonnegative with finite expectations.

Some new properties of the two curves, obtained under restrictions which turn each of them into a directed concentration curve, are presented in Theorems 5.3.1 and 5.3.2. Some properties of the correlation curve follow from its links with the monotone dependence function for (X, Y) (cf. Kowalczyk (1977)). Each curve can be used to study monotone dependence of Y on X as compared with suitably measured absolute dependence.

Monotone (directed) departures from proportionality are considered in Section 5.4. A comparison of monotone departure with absolute departure leads to conclusions concerning the extent and direction of overrepresentation. The results obtained in this paper throw some new light on the classical divisor methods considered in Baliński and Young (1982).

The last chapter deals with numerical measures which are consistent with the respective function-valued measures of divergence. Section 6.1 deals with the numerical inequality measures which are simultaneously absolute and monotone. Sections 6.2 and 6.3 deal with absolute and directed divergence, respectively. The numerical measures introduced in Section 6.3 are generalized versions of the indices considered in Section 6.2. In particular, formula (6.3.1) defines the directed Pietra index which has not appeared previously in the statistical literature. Numerical measures of dependence and proportional representation are also introduced and investigated.

Summing up, we propose here tools to measure jointly absolute and directed (monotone) divergence, and we use them to generate measures of absolute and directed departures from a prescribed pattern in several areas of statistical modelling.

Further applications are now under investigation. One of them concerns effects of aggregation. An appropriate continuity index (Ciok *et al.* (1994)) could be used to describe and analyze mixed data, resampling techniques etc.

Another direction of further study concerns stratified populations. The idea is to compare each of the strata distributions with the common distribution in the population, using the proposed measures of absolute and directed divergence. This would give an insight into the extent and type of stratification. A preliminary study along these lines was done by Kowalczyk (1990).

Our considerations here are limited to theoretical distributions which correspond to infinite populations. Finite populations are mentioned only with respect to fair representation, while inference based on samples is not tackled at all. We believe, however, that the results obtained so far for general distributions provide a good starting point for developments in these directions. It seems that the measures used here could be easily generalized to finite populations. Moreover, a unified approach to different fields of applications mentioned above should generate estimators equally applicable to all of them.

1. Divergence of probability measures

1.1. Divergence of probability measures connected with two-class classification problems. Let P and Q denote probability measures on the same measurable space $(\Omega, \mathcal{A})$. There is a general feeling (cf. Ali and Silvey (1966)) that some probability distributions are "closer together" than others and consequently that it may be "easier to distinguish" between the distributions of one pair than between those of another. The respective intuitions have been formalized in many ways. Among them, a suggestive formalization refers to the two-class classification problem. In such a problem we deal with a population of objects divided into two classes. Each object in any class has its own description $\omega \in \Omega$ (e.g., it is described by a vector of real-valued features). The descriptions are chosen so that there exists a suitable σ-field $\mathcal{A}$ of subsets of Ω such that the two classes can be presented as some probability measures, say P and Q, on $(\Omega, \mathcal{A})$. The investigator can observe the description(s) ω of an object(s) but its (their) class-membership is not observable. His goal is to recognize for each object where it comes from.

Let a classification rule be a Borel measurable function $\delta : \Omega \to [0,1]$, where $\delta(\omega)$ is the probability of taking the decision that the observed ω is from the first class. Let $a_{12}(\delta)$, $a_{21}(\delta)$ be the probabilities of misclassification:

$$a_{12}(\delta) = \int_{\Omega} (1 - \delta(\omega))\, dP(\omega), \qquad a_{21}(\delta) = \int_{\Omega} \delta(\omega)\, dQ(\omega).$$

The probabilities $a_{12}(\delta)$, $a_{21}(\delta)$ describe the quality of the classification rule δ. Basing on them, we introduce the following natural ordering in the set Δ of all classification rules on Ω:

Definition 1.1.1. We say that a rule δ' is *not worse* than a rule δ ($\delta \preceq \delta'$) if

$$a_{12}(\delta') \le a_{12}(\delta), \qquad a_{21}(\delta') \le a_{21}(\delta). \tag{1.1.1}$$

We will restrict ourselves to the set of rules admissible with respect to the ordering (1.1.1). In order to characterize this set, we introduce the generalized Radon–Nikodym derivative of Q with respect to P: let $Q = Q_{\text{abs}} + Q_{\text{sing}}$ be the Lebesgue decomposition of Q relative to P, where Q_{abs} is absolutely continuous with respect to P ($Q_{\text{abs}} \ll P$) and Q_{sing} is singular with respect to P ($Q_{\text{sing}} \perp P$), and let $N, N^c \subset \mathcal{A}$ be a partition of Ω such that $P(N) = 0$, $Q_{\text{sing}}(N) = Q_{\text{sing}}(\Omega)$. The generalized Radon–Nikodym derivative of Q with respect to P, denoted by $\frac{dQ}{dP}$, is

$$\frac{dQ}{dP}(\omega) = \begin{cases} \dfrac{dQ_{\text{abs}}}{dP}(\omega) & \text{for } \omega \in N^c, \\ \infty & \text{for } \omega \in N. \end{cases}$$

It follows from the Neyman–Pearson Lemma that the set of rules admissible with respect to the ordering (1.1.1) consists of all rules $\delta_{\kappa,s}$ of the form

$$\delta_{\kappa,s}(\omega) = \begin{cases} 1 & \text{if } \dfrac{dQ}{dP}(\omega) < \kappa, \\ s & \text{if } \dfrac{dQ}{dP}(\omega) = \kappa, \\ 0 & \text{if } \dfrac{dQ}{dP}(\omega) > \kappa, \end{cases}$$

for $\kappa \in (0, \infty)$ and $s \in [0,1]$. These rules are called *threshold rules* with respect to $\frac{dQ}{dP}$.

It is convenient to extend the set of admissible rules adding the threshold rules for $\kappa = 0$ and $\kappa = \infty$. We denote the extended set by $\Delta^0_{(P,Q)}$:

$$\Delta^0_{(P,Q)} = \{\delta_{\kappa,s} : \kappa \in [0,\infty],\ s \in [0,1]\}. \tag{1.1.2}$$

The set $\Delta^0_{(P,Q)}$ determines the lower boundary of the so-called risk set, i.e. the closed convex set consisting of points $(a_{12}(\delta), a_{21}(\delta))$ for all classification rules $\delta \in \Delta$. This boundary, which is a continuous, convex and nonincreasing curve joining the points $(0,1)$ and $(1,0)$, will be denoted by $K_{(P,Q)}$ and called the *divergence curve* of Q from P (see Bromek and Kowalczyk (1990)) or the *Neyman–Pearson curve* of Q with respect to P (see Kowalczyk and Mielniczuk (1990)):

$$K_{(P,Q)} = \{(a_{12}(\delta_{\kappa,s}),\ a_{21}(\delta_{\kappa,s})) : \kappa \in [0,\infty],\ s \in [0,1]\}. \tag{1.1.3}$$

Obviously,

$$K_{(P,Q)} = \Big\{\Big(P\big\{\omega \in \Omega : \tfrac{dQ}{dP}(\omega) > \kappa\big\} + (1-s)P\big\{\omega \in \Omega : \tfrac{dQ}{dP}(\omega) = \kappa\big\},$$
$$Q\big\{\omega \in \Omega : \tfrac{dQ}{dP}(\omega) < \kappa\big\} + sQ\big\{\omega \in \Omega : \tfrac{dQ}{dP}(\omega) = \kappa\big\}\Big) : \kappa \in [0,\infty], s \in [0,1]\Big\}.$$

In particular, for any pair of k-valued distributions

$$P = (p_1, \ldots, p_k), \qquad Q = (q_1, \ldots, q_k),$$

the curve $K_{(P,Q)}$ is piecewise linear with vertices

$$\Big(\sum_{r=1}^{l} p_{i_r}, 1 - \sum_{r=1}^{l} q_{i_r}\Big) \qquad \text{for } l = 0, \ldots, k,$$

where $\sum_1^0 = 0$ and $(i_1, \ldots, i_k)$ is a permutation of $(1, \ldots, k)$ such that

$$\frac{q_{i_1}}{p_{i_1}} \ge \ldots \ge \frac{q_{i_k}}{p_{i_k}}.$$

If $Q \ll P$ then $K_{(P,Q)}$ is the graph of a nonincreasing function $K_{(P,Q)}(\cdot)$ defined on $[0,1]$. Otherwise, this function is not defined at 0 and the curve contains an interval of the y-axis (from $(0, Q_{\mathrm{abs}}(\Omega))$ to $(0,1)$).

Apart from convexity and monotonicity, $K_{(P,Q)}$ has the following properties (for proofs see Gafrikova and Kowalczyk (1994)):

1. $K_{(P,Q)}$ and $K_{(Q,P)}$ are related as follows:

$$K_{(Q,P)} = \{(u,v) : (v,u) \in K_{(P,Q)}\}.$$

2. $P = Q$ iff $K_{(P,Q)} = \{(u,v) : v = 1 - u,\ u \in [0,1]\}$ (i.e. $P = Q$ iff $K_{(P,Q)}$ is the segment joining $(0,1)$ and $(1,0)$).

3. $P \perp Q$ iff $K_{(P,Q)} = \{(u,v) : (u = 0, 0 \le v \le 1) \vee (0 \le u \le 1, v = 0)\}$ (i.e. P and Q are singular (in particular, have disjoint supports) iff $K_{(P,Q)}$ consists of the two edges of the unit square emanating from $(0,0)$).

Properties 1, 2, 3 indicate why $K_{(P,Q)}$ is called here the divergence curve of Q from P.

1.2. Concentration curve and its link with the Neyman–Pearson curve. Cifarelli and Regazzini (1987) approach problems of divergence of probability measures on $(\Omega, \mathcal{A})$ as problems of their relative concentration on sets belonging to $\mathcal{A}$. To this end, they choose the generalized Radon–Nikodym derivative $\frac{dQ}{dP}(\omega)$ for a pointwise index of concentration of Q with respect to P. Loosely speaking, the value of $\frac{dQ}{dP}(\omega)$ increases when so does the concentration in ω of Q with respect to P, and $\frac{dQ}{dP}(\omega) \equiv 1$ when $P = Q$. Cifarelli and Regazzini compare the masses of P and Q on subsets of Ω consisting of ω's with sufficiently small concentration (not exceeding a given level). They introduce the set

$$\Big\{\Big(P\big\{\omega : \tfrac{dQ}{dP}(\omega) \le z\big\}, Q\big\{\omega : \tfrac{dQ}{dP}(\omega) \le z\big\}\Big) : z \in [0,\infty]\Big\}. \tag{1.2.1}$$

This set, completed if necessary by linear interpolation, is called the *concentration curve* of Q with respect to P. It will be denoted here by $L_{(P,Q)}$, or $L[P,Q]$ whenever the notation for P or Q is so complicated that the subscript (P,Q) is not convenient (this happens e.g. in Sec. 5.3).

The curve $L_{(P,Q)}$ contains the graph of $(L_{(P,Q)}(t), t \in [0,1])$, where

$$L_{(P,Q)}(t) = \begin{cases} 0 & \text{for } t = 0, \\ Q\big\{\omega : \frac{dQ}{dP}(\omega) < c_t\big\} + c_t\{t - H(c_t-)\} & \text{for } t \in (0,1), \\ Q_{\mathrm{abs}}(\Omega) & \text{for } t = 1, \end{cases}$$
$$H(z) = P\big\{\omega \in \Omega : \tfrac{dQ}{dP}(\omega) \le z\big\},$$
$$c_t = \inf\{z \in \mathbb{R} : H(z) \ge t\},$$
$$H(z-) = H(z - 0).$$

The curve $L_{(P,Q)}$ is convex and nondecreasing in $[0,1]^2$. If P and Q are non-atomic measures, then any set $\big\{\omega \in \Omega : \frac{dQ}{dP}(\omega) \le c_t\big\}$ has P-measure t and Q-measure $L_{(P,Q)}(t)$.

If P and Q are atomic then the curve $L_{(P,Q)}$ consists of segments and the remark above is valid for t corresponding to the vertices of the curve.

Obviously, the concentration curve is linked with the Neyman–Pearson curve by

$$K_{(P,Q)}(t) = L_{(P,Q)}(1-t) \quad \text{for } t \in (0,1],$$
$$K_{(P,Q)}(0^+) = L_{(P,Q)}(1).$$

1.3. Divergence ordering $\preceq_{\mathrm{NP}}$. Let $\mathcal{P}$ be the set of all probability measures defined on the same measurable space $(\Omega, \mathcal{A})$. We introduce an ordering $\preceq_{\mathrm{NP}}$ in $\mathcal{P} \times \mathcal{P}$ with respect to divergence of measures from one another (cf. Bromek and Kowalczyk (1990)).

DEFINITION 1.3.1. We say that

$$(P,Q) \preceq_{\mathrm{NP}} (P',Q'),$$

i.e. the divergence of Q' from P' is *not smaller* than that of Q from P, if for every classification rule δ for (P,Q) there exists a classification rule δ' for (P',Q') such that

$$a'_{12}(\delta') \le a_{12}(\delta), \quad a'_{21}(\delta') \le a_{21}(\delta).$$

The rules δ and δ' in this definition belong to the whole set Δ but, obviously, this set can be restricted to the set $\Delta^0_{(P,Q)}$ of admissible rules in the case of δ and to the set $\Delta^0_{(P',Q')}$ of admissible rules in the case of δ', where $\Delta^0_{(P,Q)}$ and $\Delta^0_{(P',Q')}$ are given by (1.1.2). Therefore, in view of definition (1.1.3), $\preceq_{\mathrm{NP}}$ coincides with the ordering based on divergence curves:

$$(P,Q) \preceq_{\mathrm{NP}} (P',Q') \quad \text{iff} \quad K_{(P,Q)}(t) \geq K_{(P',Q')}(t) \quad \text{for } t \in (0,1].$$

This ordering has the following properties (see Bromek and Kowalczyk (1990) and Gafrikova and Kowalczyk (1994)):

Property 1. $(P,Q) \preceq_{\mathrm{NP}} (P',Q')$ iff $(Q,P) \preceq_{\mathrm{NP}} (Q',P')$.

Property 2. (P,Q) is a smallest element for $\preceq_{\mathrm{NP}}$ iff $P = Q$.

Property 3. (P,Q) is a largest element for $\preceq_{\mathrm{NP}}$ iff $P \perp Q$.

Property 4. Suppose that $y = f(\omega)$ is a measurable transformation from $(\Omega, \mathcal{A})$ onto a measurable space $(Y, \mathcal{G})$. Let Pf^{-1}, Qf^{-1} denote the measures induced by f on Y from P, Q respectively. Then

$$(Pf^{-1}, Qf^{-1}) \preceq_{\mathrm{NP}} (P,Q).$$

(P,Q) and (Pf^{-1}, Qf^{-1}) are equivalent with respect to $\preceq_{\mathrm{NP}}$ iff

$$\frac{dQ}{dP}(\omega) = \frac{d(Qf^{-1})}{d(Pf^{-1})}(f(\omega))$$

for all ω.

As a special case of Property 4 we have:

Property 4′. Let $P = (p_1, \ldots, p_k)$, $Q = (q_1, \ldots, q_k)$ be k-valued distributions and let (P',Q') be $(k-1)$-valued distributions obtained from (P,Q) by pooling any two values of (P,Q). Then

$$(P',Q') \preceq_{\mathrm{NP}} (P,Q).$$

Property 5. Suppose that $\alpha, \beta \in [0,1]$, $\alpha \leq \beta$. Then

$$(\beta P + (1-\beta)Q, \alpha P + (1-\alpha)Q) \preceq_{\mathrm{NP}} (P, \alpha P + (1-\alpha)Q) \preceq_{\mathrm{NP}} (P,Q).$$

Property 6. Suppose that $\alpha \in [0,1]$. Then

$$(P,Q) \preceq_{\mathrm{NP}} (P',Q') \quad \text{iff} \quad (P, \alpha P + (1-\alpha)Q) \preceq_{\mathrm{NP}} (P', \alpha P' + (1-\alpha)Q').$$

Property 7. Let P, Q, Q^ε be k-valued distributions,

$$P = (p_1, \ldots, p_k), \qquad Q = (q_1, \ldots, q_k), \qquad Q^\varepsilon = (q_1^\varepsilon, \ldots, q_k^\varepsilon),$$

such that $q_1/p_1 \leq \ldots \leq q_k/p_k$, and $q_i^\varepsilon = q_i + \varepsilon$, $q_j^\varepsilon = q_j - \varepsilon, q_s^\varepsilon = q_s$ for some

$i < j$, $s \neq i, j$, $s = 1, \ldots, k$, where ε is a nonnegative number such that

$$\varepsilon \le \begin{cases} \dfrac{q_{i+1}p_i - q_i p_{i+1}}{p_i + p_{i+1}} & \text{for } j = i+1, \\ \min\left(p_i\left(\dfrac{q_{i+1}}{p_{i+1}} + \dfrac{q_i}{p_i}\right),\ p_j\left(\dfrac{q_j}{p_j} + \dfrac{q_{j-1}}{p_{j-1}}\right)\right) & \text{for } j > i+1. \end{cases}$$

Then

$$(P, Q^{\varepsilon}) \preceq_{\mathrm{NP}} (P, Q).$$

Three other important properties of $\preceq_{\mathrm{NP}}$, based on the notion of generalized expectation, will be presented in Sec. 2.4.

2. Link between divergence and inequality

2.1. Initial inequality axioms. The notion of inequality of a random variable appears in statistical literature in many contexts; most contributions (including the oldest ones) refer to various economical situations such as welfare or income inequality in a human population. Generally, we deal in practice with two populations of objects of the same kind, described by a variable X, which is additive, nonnegative and has finite mean. It will be convenient to assume for a while that both populations are finite and each of them has n elements. Thus, we deal with two vectors of values of the feature under consideration in each population, say $x = (x_1, \ldots, x_n)$ and $x' = (x'_1, \ldots, x'_n)$. We ask which vector is less "equal" than the other, i.e. for which of them the components are more distant from one another.

An axiomatic approach to comparing inequality of vectors with n nonnegative components is due to Fields and Fey (1978). They formulate three axioms for an ordering $\preceq$ according to inequality in the set of such vectors, where $x \preceq x'$ means that x is less equal than x'. (Fields and Fey used $\succeq$ instead of $\preceq$; we changed this notation to ensure consistency with the rest of this paper). The axioms are as follows:

Axiom 1 (*Scale Irrelevance*). If $x = ax'$, i.e. $x_i = ax'_i$ for $i = 1, \ldots, n$, $a > 0$, then $x \cong x'$ (which means that $x \preceq x'$ and $x' \preceq x$, i.e. x and x' are equally unequal).

This axiom allows us to normalize all vectors, so that $\sum_{i=1}^{n} x_i = 1$. The set of all normalized vectors will be denoted by D_0:

$$D_0 = \Big\{x = (x_1, \ldots, x_n) : x_i \ge 0,\ x_1 \le \ldots \le x_n,\ \sum_{i=1}^{n} x_i = 1\Big\}.$$

Axiom 2 (*Symmetry*). If $(i_1, \ldots, i_n)$ is any permutation of $(1, \ldots, n)$ then

$$(x_{i_1}, \ldots, x_{i_n}) \cong (x_1, \ldots, x_n).$$

Axiom 3 (*Rank-Preserving Equalization*). If $x, x' \in D_0$ and if for some $i < j$ and $\varepsilon > 0$,

$$x_k = x'_k \quad \text{for } k \neq i, j,\ k = 1, \ldots, n,$$
$$x_i = x'_i + \varepsilon, \quad x_j = x'_j - \varepsilon,$$

where

$$\varepsilon \leq \begin{cases} \frac{1}{2}(x'_j - x'_i) & \text{for } j = i + 1, \\ \leq \min(x'_{i+1} - x'_i, x'_j - x'_{j-1}) & \text{for } j > i + 1, \end{cases}$$

then $x \preceq x'$.

Fields and Fey proved that if $x, x' \in D_0$ and if x is obtained from x' by a finite sequence of transformations described in the third axiom, then

$$\begin{aligned} x_1 + \ldots + x_i &\geq x'_1 + \ldots + x'_i \quad \text{for } i = 1, \ldots, n-1, \\ x_1 + \ldots + x_j &> x'_1 + \ldots + x'_j \quad \text{for some } j < n, \end{aligned} \tag{2.1.1}$$

and vice versa: the inequalities (2.1.1) imply that $x \in D_0$ is obtainable from x' by a finite sequence of such transformations.

J. Foster (1985) extended the axioms by the following one aimed at comparing vectors which describe populations of different sizes. Let

$$D = \bigcup_{n=1}^{\infty} D_n, \quad D_n = \Big\{x \in \mathbb{R}^n : \sum_{i=1}^{n} x_i > 0,\ x_i \geq 0,\ i = 1, \ldots, n\Big\}.$$

Axiom 4 (*Population Principle*). If x' is a replication of x (i.e. $x \in D$ and for some $m \geq 2$ we have $x' = (x'_{(1)}, \ldots, x'_{(m)})$, where each $x'_{(i)} = x$) then $x \cong x'$.

2.2. The Lorenz curve for nonnegative random variables. Inequalities (2.1.1) can be interpreted graphically by means of the so-called Lorenz curves for x and x'.

The Lorenz curve was introduced in 1905 for the population $\{x_1, \ldots, x_n\}$ of n individual incomes by setting

$$L_X\left(\frac{i}{n}\right) = \frac{\sum_{j=1}^{i} x_{r_j}}{\sum_{j=1}^{n} x_{r_j}} \quad \text{for } i = 0, \ldots, n,$$

where $x_{r_1} \leq \ldots \leq x_{r_n}$ are the ordered individual incomes in the population. The points $(i/n, L_X(i/n))$ for $i = 0, \ldots, n$ are then linearly interpolated to get the corresponding Lorenz curve. Thus, we have defined the Lorenz curve for a random variable X taking values $x_1, \ldots, x_n$ with probabilities $P(X = x_i) = 1/n$ for $i = 1, \ldots, n$ (if x_i's are not all distinct then the probabilities are changed in an obvious way).

Generally, let $\mathcal{L}$ be the set of all nonnegative random variables with finite nonzero expectations. For any $X \in \mathcal{L}$ with distribution function F_X, the Lorenz curve L_X is

$$L_X(u) = \frac{\int_0^u F_X^{-1}(y)\,dy}{\int_0^1 F_X^{-1}(y)\,dy} \quad \text{for } u \in [0,1], \tag{2.2.1}$$

where $F_X^{-1}(y) = \inf\{t : F_X(t) \ge y\}$ for $0 < y < 1$.

Sometimes it is convenient to use the parametric representation of the Lorenz curve (Arnold (1987)) as the set of points

$$\{(F_X(t),\, F_X^{(1)}(t)) : t \in [0,\infty]\}, \tag{2.2.2}$$

in the unit square, completed if necessary by linear interpolation, where

$$F_X^{(1)}(t) = \frac{1}{E(X)} \int_0^t u\,dF_X(u), \qquad t \in [0,\infty]. \tag{2.2.3}$$

Formula (2.2.2) follows directly from (2.2.1).

Another form of the Lorenz curve, obviously equivalent to (2.2.2), is

$$L_X(u) = \begin{cases} 0 & \text{for } u = 0, \\ \dfrac{E(X; X < x_u) + x_u(u - P(X < x_u))}{E(X)} & \text{for } u \in (0,1), \\ 1 & \text{for } u = 1, \end{cases}$$

where x_u is any quantile of X of order u for $u \in (0,1)$, i.e.

$$P(X < x_u) \le u \le P(X \le x_u).$$

The function $F_X^{(1)}$ is called the *first moment distribution function*. Suppose that X is the length of life in some population. Then $F_X^{(1)}(t)$ denotes the mean life length of an element which dies till t, divided by the mean life time. Now, another partition of the mean life time is also in use. It refers to the *total time on test* (TTT) *transform*. The related distribution function $F_X^{(2)}$ is defined by

$$F_X^{(2)}(t) = \frac{1}{E(X)} \int_0^t (1 - F_X(s))\,ds$$

where $F_X^{(2)}(t)$ denotes the mean length of life truncated at the moment t, divided by the mean life time. The curve $\mathrm{TTT}(p) = F_X^{(2)}(F_X^{-1}(p))$ for $p \in [0,1]$ is a counterpart of the Lorenz curve $L(p) = F_X^{(1)}(F_X^{(-1)}(p))$. The two curves are interrelated in the following way (see e.g. Klefsjö (1984)):

$$L(p) = \mathrm{TTT}(p) - \frac{1}{E(X)}(1-p)F_X^{-1}(p), \qquad p \in [0,1].$$

2.3. Inequality ordering $\preceq_{\mathrm{L}}$. The ordering $\preceq_{\mathrm{L}}$ according to inequality in the set $\mathcal{L}$ (Arnold (1987)) is based on comparing the Lorenz curves.

DEFINITION 2.3.1. For any $X, X' \in \mathcal{L}$, we say that X *does not exhibit more inequality in the Lorenz sense than* X' *does*, and write $X \preceq_{\mathrm{L}} X'$, if

$$L_X(u) \ge L_{X'}(u) \quad \text{for } u \in [0,1].$$

It is easy to check that for the empirical distributions considered in Sec. 2.2 the ordering $\preceq_{\mathrm{L}}$ satisfies Axioms 1–4. Moreover, $\preceq_{\mathrm{L}}$ has the following properties:

PROPERTY 1°. A random variable X is a minimal element for $\preceq_{\mathrm{L}}$ iff X is concentrated at one point $x > 0$ (i.e. X is degenerate). The Lorenz curve for a degenerate random variable coincides with the 45° line in the square $[0,1]^2$.

PROPERTY 2°. Let $X \in \mathcal{L}$ be a discrete k-valued random variable with $P(X = x_i) = \pi_i$ for $i = 1, \ldots, k$, $\sum_{i=1}^k \pi_i = 1$. Let X' be the random variable obtained from X by aggregating any two values, say x_i, x_j, $i, j \in \{1, \ldots, k\}$, to the value

$$\frac{\pi_i}{\pi_i + \pi_j} x_i + \frac{\pi_j}{\pi_i + \pi_j} x_j.$$

Then $X' \preceq_{\mathrm{L}} X$.

PROPERTY 3°. Let $X \in \mathcal{L}$ and $\alpha, \beta \in [0,1], \alpha \le \beta$. Then

$$\frac{\alpha + (1-\alpha)X}{\beta + (1-\beta)X} \preceq_{\mathrm{L}} (1-\alpha)X + \alpha \preceq_{\mathrm{L}} X.$$

PROPERTY 4°. Suppose that $X, X' \in \mathcal{L}$ and $\alpha \in [0,1]$. Then

$$X \preceq_{\mathrm{L}} X' \quad \text{iff} \quad (1-\alpha)X + \alpha \preceq_{\mathrm{L}} (1-\alpha)X' + \alpha.$$

PROPERTY 5°. Let $X \in \mathcal{L}$ be a discrete k-valued random variable with $P(X = x_i) = \pi_i$, $\sum_{i=1}^k \pi_i = 1$, $x_1 < \ldots < x_k$. Let X' be a random variable with k values such that $P(X' = x_i') = \pi_i$, $i = 1, \ldots, k$, where $x_s' = x_s$ for $s \ne i, j$, for some $i < j$, $s = 1, \ldots, k$, and $x_i' = x_i + \varepsilon/\pi_i$, $x_j' = x_j - \varepsilon/\pi_j$ with

$$\varepsilon \le \begin{cases} \dfrac{(x_{i+1} - x_i)\pi_i \pi_{i+1}}{\pi_i + \pi_{i+1}} & \text{for } j = i+1, \\ \min(\pi_i(x_{i+1} - x_i), \pi_j(x_j - x_{j-1})) & \text{for } j > i+1. \end{cases}$$

Then $X' \preceq_{\mathrm{L}} X$.

One of the most important properties of $\preceq_{\mathrm{L}}$ is its characterization by means of convex functions:

PROPERTY 6°. Let $X, X' \in \mathcal{L}, EX = EX'$. Then

$$X \preceq_{\mathrm{L}} X' \quad \text{iff} \quad E(\Phi(X)) \le E(\Phi(X')) \tag{2.3.1}$$

for every convex continuous function Φ.

PROPERTY 7°. I. Let $g : \mathbb{R}^+ \to \mathbb{R}^+$. The following conditions are equivalent:

(i) $g(X) \preceq_{\mathrm{L}} X$ for every $X \in \mathcal{L}$,

(ii) $g(x) > 0$ for every $x > 0$, $g(x)$ is nondecreasing on $[0, \infty)$ and $g(x)/x$ is nonincreasing on $(0, \infty)$.

II. Let $g : \mathbb{R}^+ \to \mathbb{R}^+$. The following conditions are equivalent:

(i) $X \preceq_{\mathrm{L}} g(X)$ for every $X \in \mathcal{L}$,

(ii) $g(x) > 0$ for every $x > 0$, $g(x)$ is nondecreasing on $[0, \infty)$ and $g(x)/x$ is nondecreasing on $(0, \infty)$.

PROPERTY 8°. Suppose that $X, X' \in \mathcal{L}$, $EX = EX'$ and X and X' are absolutely continuous with densities $f_X(x)$ and $f_{X'}(x)$. A sufficient condition for $X \preceq_{\mathrm{L}} X'$ is that $f_X(x) - f_{X'}(x)$ changes sign twice on $(0, \infty)$ and the sequence of signs of $f_X - f_{X'}$ is $- + -$.

Properties 1°, 6°, 8° are proved e.g. in Arnold (1987), and properties 2°–5°, 7° in Gafrikova and Kowalczyk (1994).

Orderings stronger than $\preceq_{\mathrm{L}}$ have been investigated in statistical literature. In particular, reliability theory introduces the star-ordering such that F is star-ordered w.r.t. G (written $F \preceq_* G$) if $G^{-1}(F(x))/x$ is increasing on $0 < x < F^{-1}(1)$. Chandra and Singpurwalla (1981) proved that $F \preceq_* G$ implies $L_F(p) \geq L_G(p)$ for $0 \leq p \leq 1$ if F and G have the same mean.

2.4. Inequality versus divergence. Let $h(\omega)$ be the generalized Radon–Nikodym derivative of Q with respect to P: $h(\omega) = \frac{dQ}{dP}(\omega)$. Let F_i^h $(i = 1, 2)$ be the distribution functions of the transformed measures $P^h = Ph^{-1}$, $Q^h = Qh^{-1}$, respectively, i.e.

$$F_1^h(t) = P^h([0, t]) = P(\omega : h(\omega) \leq t),$$
$$F_2^h(t) = Q^h([0, t]) = Q(\omega : h(\omega) \leq t).$$

Note that $P^h([0, \infty)) = 1$ but $Q^h([0, \infty)) = 1 - Q(h = \infty) = 1 - Q(N)$. From (1.2.1) we see that the concentration curve $L_{(P,Q)}$ is the subset

$$\{(F_1^h(t), F_2^h(t)) : t \in [0, \infty]\} = \Big\{\Big(F_1^h(t), \int_0^t s\, dF_1^h(s)\Big) : t \in [0, \infty]\Big\} \tag{2.4.1}$$

of the unit square, completed if necessary by linear interpolation. Let Z be a random variable defined on $(\Omega, \mathcal{A})$ and let $Z \sim P$ (i.e. Z is distributed according to P). Comparing the sets (2.4.1) and (2.2.2) we obtain an important statement which will be formulated first for $Q \ll P$. Under this assumption $h(Z) \in \mathcal{L}$ since $E_P(h(Z)) = 1$. Then the concentration curve $L_{(P,Q)}$ is the Lorenz curve for the random variable $h(Z)$:

$$L_{(P,Q)}(u) = L_{h(Z)}(u) \quad \text{for } u \in [0, 1]. \tag{2.4.2}$$

When the assumption $Q \ll P$ is omitted, let $\widetilde{h}(\omega) = (dQ_{\mathrm{abs}}/dP)(\omega)$ for $\omega \in N^c$ and $\widetilde{Z}$ be the random variable defined on $\Omega \setminus N$ such that $\widetilde{Z} \sim P$ (note that for $Q \ll P$ we have $Q(N) = 0$ and $\widetilde{h}(\widetilde{Z}) = h(Z)$). Then $\widetilde{h}(\widetilde{Z}) \in \mathcal{L}$ since $E(\widetilde{h}(\widetilde{Z})) = 1 - Q(N)$. The equality (2.4.2) is now generalized to

$$L_{(P,Q)}(u) = (1 - Q(N)) L_{\widetilde{h}(\widetilde{Z})}(u) \quad \text{for } u \in [0, 1). \tag{2.4.3}$$

We see from (2.4.2) that for $Q \ll P$, measuring divergence by means of $L_{(P,Q)}$ is equivalent to measuring inequality for $h(Z)$ by means of $L_{h(Z)}$. Also, there exists an obvious correspondence between the orderings $\preceq_{\mathrm{NP}}$ and $\preceq_{\mathrm{L}}$: if $Q \ll P$, $Q' \ll P'$ then

$$(P,Q) \preceq_{\mathrm{NP}} (P',Q') \quad \text{iff} \quad h(Z) \preceq_{\mathrm{L}} h'(Z'), \tag{2.4.4}$$

where $h = \frac{dQ}{dP}$, $h' = \frac{dQ'}{dP'}$, Z, Z' are random variables defined on Ω, and $Z \sim P$, $Z' \sim P'$.

The equivalence (2.4.4) is an important link between divergence and inequality. It is obvious that under the condition $Q \ll P$ any property of the divergence of Q from P can be reworded as some property of the inequality in the class $\mathcal{L}_1 \subset \mathcal{L}$ of all nonnegative random variables with expectation 1 (since $E(h(Z)) = 1$). For example, the counterparts of properties 2, 4′, 5, 6, 7 of $\preceq_{\mathrm{NP}}$ are properties 1°, 2°, 3°, 4°, 5°, respectively, of the ordering $\preceq_{\mathrm{L}}$ in the class $\mathcal{L}_1$.

Further, the equivalence (2.4.4) implies the characterization (2.3.1) for $X = h(Z)$, $X' = h'(Z')$ where h, h', Z, Z' have the same meaning as in (2.4.4). In the general case (without the assumption $Q \ll P$) we have to use the so-called *generalized expectation* E^* of $\Phi(h(Z))$. This notion was introduced by Ali and Silvey (1966):

$$E^*(\Phi(h(Z))) = \int\limits_{h(z)<\infty} \Phi(h(z))\, dP(z) + Q(N) \lim_{t\to\infty} \frac{\Phi(t)}{t}$$

provided that the right-hand side is meaningful (i.e. $\lim_{t\to\infty} \Phi(t)/t$ exists and the stated expression does not take the indeterminate form $\infty - \infty$). Ali and Silvey show that for any continuous convex function Φ, $E^*(\Phi(h(Z)))$ is either a finite number or ∞. Let us note that $E^*(h(Z)) = 1$. Now, the following characterization of the ordering $\preceq_{\mathrm{NP}}$ may be added to the seven properties stated in Sec. 1.3:

PROPERTY 8. For every convex continuous function Φ for which $E^*(\Phi(h(Z)))$ and $E^*(\Phi(h'(Z')))$ are finite,

$$(P,Q) \preceq_{\mathrm{NP}} (P',Q') \quad \text{iff} \quad E^*(\Phi(h(Z))) \le E^*(\Phi(h'(Z'))). \tag{2.4.5}$$

Moreover, making use of (2.4.3) and Property 7° in Sec. 2.3, we obtain the following:

PROPERTY 9. Let $(P,Q),(P',Q') \in \mathcal{P} \times \mathcal{P}$ and let h, h' be the respective generalized Radon–Nikodym derivatives. Let $h' = g(h)$ where $g : \mathbb{R}^+ \cup \infty \to \mathbb{R}^+ \cup \infty$ is nondecreasing on $\mathbb{R}^+$. Then

(i) $(P,Q) \preceq_{\mathrm{NP}} (P',Q')$ if $g(x)/x$ is nondecreasing on $(0,\infty)$ and $Q(h=\infty) \le Q'(h'=\infty)$,

(ii) $(P',Q') \preceq_{\mathrm{NP}} (P,Q)$ if $g(x)/x$ is nonincreasing on $(0,\infty)$ and $Q(h=\infty) \ge Q'(h'=\infty)$.

Finally, Property 8° of Sec. 2.3 can be used to prove the following property of $\preceq_{\mathrm{NP}}$:

Property 10. Let ν be a real parameter and let $\{P_\nu : \nu \in (a,b)\}$ be a family of mutually absolutely continuous distributions on the real line such that the family of densities $p_\nu(x)$ with respect to a fixed measure ν has monotone likelihood ratio in x (see Lehmann (1959)). Let $a < \nu_1 < \nu_2 < \nu_3 < b$. Then

$$(P_{\nu_1}, P_{\nu_2}) \preceq_{\mathrm{NP}} (P_{\nu_1}, P_{\nu_3}).$$

To end this section, we use the above considerations to indicate the most important link between divergence and inequality. To this end, for any random variable $X \in \mathcal{L}$ defined on $(\Omega, \mathcal{A}, P)$, we compare P with some other distribution on $(\Omega, \mathcal{A})$. Its distribution function $\lambda_P^X(\cdot)$ is given by

$$\lambda_P^X(A) = \frac{\int_A X(\omega)\,P(d\omega)}{\int_\Omega X(\omega)\,P(d\omega)} \quad \text{for } A \in \mathcal{A}. \tag{2.4.6}$$

The notation λ_P^X will be simplified to λ_P whenever $X(\omega) = \omega$. The distribution λ_P^X plays an important role in the present paper.

Note that $\lambda_P^X \ll P$. The density function of λ_P^X w.r.t. P is $X(\omega)/E(X)$, which is equal to the ratio of the densities (w.r.t. P) of λ_P^X and P. It follows that the Lorenz curve of $X/E(X)$, or equivalently of X, coincides with the concentration curve of λ_P^X w.r.t. P.

By Property 4 of $\preceq_{\mathrm{NP}}$ (Sec. 1.3) the concentration curve of λ_P^X w.r.t. P is the same as the concentration curve of these distributions transformed by X. We have $PX^{-1} = P_X$ and $\lambda_P^X X^{-1} = \lambda_{P_X}$ (we write λ_{P_X} instead of $\lambda_{P_X}^{\mathrm{id}}$). Indeed,

$$\lambda_P^X(X^{-1}(B)) = \lambda_{P_X}(B) = \frac{E(X;\ X \in B)}{E(X)}.$$

Thus, the concentration curve of λ_{P_X} w.r.t. P_X coincides with the Lorenz curve L_X. It is worth noting that the definition (2.4.1) of the concentration curve, applied to (P_X, λ_{P_X}), leads to formula (2.2.2) for the Lorenz curve L_X.

2.5. Ratio variables. At the beginning of this chapter it was indicated that in practice the notion of inequality is introduced for variables which are additive, nonnegative and have finite mean in the considered population of objects. Additivity, nonnegativity and finite mean are necessary to form the distribution λ_P^X, which is constructed from means corresponding to particular fractions of the population. On the other hand, two variables X and Y are $\preceq_{\mathrm{L}}$ equivalent (i.e. have identical inequalities) if $X \sim aY$ for some $a > 0$. We shall consider all this in more detail, referring to a measurement scale called ratio.

In measurement theory, a relational structure $\mathcal{R}_0$ on a population Ω_0 is considered together with a relational structure $\mathcal{R}$ on a certain subset $\widetilde{\Omega} \subset \mathbb{R}^k$. A measurement scale is a homomorphism of $\mathcal{R}_0$ into $\mathcal{R}$. An *admissible function* is a mapping $\psi : \widetilde{\Omega} \to \widetilde{\Omega}$ which transforms one scale into another. The set Ψ of all admissible mappings defines the *type of measurement scale*. In particular, when $\widetilde{\Omega} = \mathbb{R}$, the most common types of scales are nominal, ordinal and interval scales, for which Ψ is the set of all injections, increasing and linear increasing mappings, respectively. When $\widetilde{\Omega} = \mathbb{R}^+$, we deal with the ratio scale for which Ψ is the set of the mappings $y = ax$, $a > 0$.

A parameter γ defined on a set $\mathcal{J}$ of random variables will be called an *indicator of the measurement scale type* Ψ *in* $\mathcal{J}$ if:

1°. $\mathcal{J}$ is closed under Ψ, i.e. for each $X \in \mathcal{J}$ and each $\psi \in \Psi$, $\psi(X) \in \mathcal{J}$.
2°. For each $X \in \mathcal{J}$ and each $\psi \in \Psi$, $\gamma(X) = \gamma(\psi(X))$.
3°. Let $g : \mathbb{R} \to \mathbb{R}$ and $\gamma(g(X)) = \gamma(X)$ for every $X \in \mathcal{J}$. Then $g \in \Psi$.

Notice that if γ is an indicator of the scale type Ψ in $\mathcal{J}$, then a transformation of γ, say $f \circ \gamma$, is also an indicator of Ψ in $\mathcal{J}$ only if f is a bijection. Moreover, if a scale Ψ' is weaker than Ψ, i.e. $\Psi \subset \Psi'$, $\Psi \neq \Psi'$, and γ is an indicator of Ψ in $\mathcal{J}$, then γ is *not* an indicator of Ψ' in $\mathcal{J}$.

These remarks justify the following one concerning the relations between statistical theory and practice: if γ is an indicator of a scale type Ψ in a set $\mathcal{J}$ of random variables, then it should not be used in a practical statistical study unless the variables appearing in the study are all measured on a scale not weaker than Ψ. In the practical context it is also worth noting that if γ is an indicator of the scale type Ψ in $\mathcal{J}$, and if $\gamma(X) = \gamma(Y)$ for $X, Y \in \mathcal{J}$, then either $Y \sim \psi(X)$ for some $\psi \in \Psi$, or X and Y are not both measured on the scale Ψ.

The type of measurement scale may be linked with an ordering relating to the considered parameter: if γ is an indicator of a scale Ψ in a set $\mathcal{J}$ and if γ is strictly monotone with respect to some ordering $\preceq$ in $\mathcal{J}$, then

$$X \prec Y \quad \text{iff} \quad \psi(X) < \psi(Y) \quad \text{for } \psi \in \Psi,$$

where $X \prec Y$ means that $X \preceq Y$ and not $X \cong Y$.

Now, let us use the above considerations putting $\mathcal{J} = \mathcal{L}$ (where $\mathcal{L}$ is the set of nonnegative random variables with finite nonzero expectations). It is easy to check that the Lorenz curve is an indicator of the ratio scale in $\mathcal{L}$. Moreover, the Lorenz order satisfies conditions 1°–3° when Ψ is the set of the mappings $y = ax$, $a > 0$.

Random variables from $\mathcal{L}$ will be called "*ratio variables*" in the sequel. It is well known that in practice inequality is evaluated for variables measured on the ratio scale (income, welfare, length of life, various "size" and some "shape" variables, and so on).

3. Link between divergence and dependence

3.1. Preliminary remarks. In this chapter we deal with bivariate distributions only. Therefore we assume that $\Omega = \mathbb{R}^2$, $\mathcal{B}^2$ is the σ-field of Borel sets on the plane, and consider pairs (X, Y) of random variables on $(\mathbb{R}^2, \mathcal{B}^2)$. Let P denote the joint distribution of (X, Y).

We start with some remarks concerning dependence when at least one random variable in the pair (X, Y) is a ratio variable.

According to the definition (2.4.6) we introduce λ_P^X if X is a ratio variable, and λ_P^Y if Y is a ratio variable, where for any $A, A' \in \mathcal{B}(\mathbb{R}^+)$,

$$\lambda_P^X(A \times A') = \frac{\int_A \int_{A'} x\,dP(x,y)}{\int_{\mathbb{R}^+} \int_{\mathbb{R}^+} x\,dP(x,y)} = \frac{E(X; X \in A,\ Y \in A')}{E(X)},$$
$$\lambda_P^Y(A \times A') = \frac{\int_A \int_{A'} y\,dP(x,y)}{\int_{\mathbb{R}^+} \int_{\mathbb{R}^+} y\,dP(x,y)} = \frac{E(Y; X \in A,\ Y \in A')}{E(Y)}.$$

Note that

$$\lambda_P^X(A \times \mathbb{R}^+) = \lambda_{P_X}(A) = \frac{E(X; X \in A)}{E(X)} = \frac{\int_A x f_X(x)\,\nu(dx)}{E(X)},$$
$$\lambda_P^Y(A \times \mathbb{R}^+) = \lambda_{P_X}^{r(X)}(A) = \frac{E(E(Y \mid X); X \in A)}{E(Y)} = \frac{\int_A r(x) f_X(x)\,\nu(dx)}{E(Y)},$$

where $r(x) = E(Y \mid X = x)$ and f_X is the density of X with respect to the given measure ν on $\mathcal{B}(\mathbb{R}^+)$.

Following the schemes appearing in the previous chapter, one could investigate dependence as divergence between two distributions on $(\Omega, \mathcal{B})$, in particular

(P, λ_P^X) when X is a ratio random variable,
(P, λ_P^Y) when Y is a ratio variable,
$(\lambda_P^X, \lambda_P^Y)$ when both variables are ratio variables.

In Taguchi (1987) the triple $(P, \lambda_P^X, \lambda_P^Y)$ was considered in order to introduce a concentration surface.

Among other pairs of distributions which are worth attention when Y is a ratio variable, let us mention here the pair $(P_X, \lambda_{P_X}^{r(X)})$. The curve $L[P_X, \lambda_{P_X}^{r(X)}]$ coincides with $L_{r(X)}$ since the likelihood ratio of $\lambda_{P_X}^{r(X)}$ w.r.t. P_X is equal to $r(x)/E(Y)$ for x such that $f_X(x) > 0$ and $X \sim P_X$. Moreover,

$$L_Y(u) \le L_{r(X)}(u) \quad \text{for } u \in [0,1] \tag{3.1.1}$$

(see Arnold (1987), p. 39).

We will use λ_P^X and λ_P^Y in Sec. 5.3 in the context of the so-called directed concentration curve to be introduced in Sec. 5.1. There we will consider some aspects of monotone stochastic dependence. In this chapter we deal with absolute stochastic dependence between X and Y in the case when no restrictions are made on the measurement scales of the variables. In Sec. 3.2 we introduce an ordering of stochastic dependence and investigate its properties. In Sec. 3.3 we compare this ordering with other orderings concerning dependence, including the so-called quadrant dependence ordering which is used only when the variables X, Y are measured at least on the ordinal scale.

3.2. Dependence ordering $\preceq_{\mathrm{D}}$. Denote by P_0 the product measure on $(\Omega, \mathcal{B})$ corresponding to the marginal distributions P_X, P_Y of P_{XY}: $P_0 = P_X \times P_Y$.

Absolute dependence between X and Y can be treated as divergence of P_{XY} from P_0. This approach based on the Neyman–Pearson curve was proposed by Bromek and Kowalczyk (1990). The authors dealt with a vector $(X_1, \ldots, X_k)$ for $k \geq 2$ and proposed the ordering $\preceq_{\mathrm{D}}$ defined in the bivariate case by

$$(X, Y) \preceq_{\mathrm{D}} (X', Y') \quad \text{if} \quad L_{(P_0, P_{XY})} \geq L_{(P_0', P_{X'Y'})}.$$

The same idea of measuring absolute dependence was simultaneously proposed by Scarsini (1990) (both papers were presented at the same conference). The main properties of the ordering are:

THEOREM 3.2.1. (i) *For any random vectors (X, Y) and (X', Y') defined respectively on Ω, Ω',*

1° *if $f : \Omega \to \mathbb{R}^2$ and $g : \Omega' \to \mathbb{R}^2$ are Borel measurable functions such that $f(x, y) = (f_1(x), f_2(y))$, $g(x, y) = (g_1(x), g_2(y))$ and f_i, g_i are injections then*

$$(X, Y) \preceq_{\mathrm{D}} (X', Y') \quad \textit{iff} \quad f(X, Y) \preceq_{\mathrm{D}} g(X', Y');$$

2° *we have*

$$(X, Y) \preceq_{\mathrm{D}} (X', Y') \quad \textit{iff} \quad (Y, X) \preceq_{\mathrm{D}} (Y', X').$$

(ii) *(X, Y) is a minimal element for $\preceq_{\mathrm{D}}$ iff $P_{XY} = P_0$.*

(iii) *For (X, Y) with continuous marginal distributions, (X, Y) is a maximal element for $\preceq_{\mathrm{D}}$ iff P_{XY} is singular w.r.t. P_0.*

(iv) *Let*

$$(X, Y) \sim N_2(\nu_X, \nu_Y, \sigma_X, \sigma_Y, \varrho), \qquad (X', Y') \sim N_2(\nu_{X'}, \nu_{Y'}, \sigma_{X'}, \sigma_{Y'}, \varrho')$$

where N_2 is the bivariate normal distribution with respective parameters. Then

$$(X, Y) \preceq_{\mathrm{D}} (X', Y') \quad \textit{iff} \quad |\varrho| \leq |\varrho'|.$$

(v) *Let $(X, Y), (X', Y')$ have densities f_{XY}, $f_{X'Y'}$ (with respect to some measure ν) with marginal densities f_X, f_Y, $f_{X'}$, $f_{Y'}$ respectively. Then*

$$(X, Y) \preceq_{\mathrm{D}} (X', Y') \quad \textit{iff} \quad \int \Psi\left(\frac{f_{XY}}{f_X f_Y}\right) f_X f_Y \, d\nu \leq \int \Psi\left(\frac{f_{X'Y'}}{f_{X'} f_{Y'}}\right) f_{X'} f_{Y'} \, d\nu$$

for all continuous convex functions Ψ.

It is evident that the above properties pertain to absolute dependence between X and Y. Properties (i)–(iv) were proved in Bromek and Kowalczyk (1990). Property (v) follows from the characterization (2.4.5).

3.3. Orderings related to $\preceq_{\mathrm{D}}$. Joe (1987) defined a preorder for measurable functions on a measure space which is a generalization of vector majorization. An equivalent form of this definition is the following: let $(\Omega, \mathcal{A}, \nu)$ be a measure space, and let f and g be nonnegative integrable functions on $(\Omega, \mathcal{A}, \nu)$ such that $\int f d\nu = \int g\, d\nu$. We say that f is *majorized* by g (written $f \prec g$) if

$$\int \Phi(f)\, d\nu \leq \int \Phi(g)\, d\nu$$

for all convex, continuous real-valued functions Φ with domain including the ranges of f and g such that $\Phi(0) = 0$ and the integrals exist.

Let $\Omega = \Omega_X \times \Omega_Y$, $\nu = \nu_X \times \nu_Y$, let f_X, f_Y be densities on Ω_X, Ω_Y with respect to ν_X, ν_Y, and $\Pi(f_X, f_Y)$ be the class of densities φ on $(\Omega, \mathcal{A}, \nu)$ such that $\int \varphi\, d\nu_X = f_X$, $\int \varphi\, d\nu_Y = f_Y$. Then the ordering $\prec$, restricted to $\Pi = \Pi(f_X, f_Y)$, can be interpreted as an ordering according to dependence, with g representing stronger dependence than f if $f \prec g$ for $f, g \in \Pi(f_X, f_Y)$.

A special case is the matrix majorization due to Joe (1985), where $\Omega_X = \{1, \ldots, r\}$, $\Omega_Y = \{1, \ldots, c\}$, r and c are positive integers, and ν_X, ν_Y are counting measures.

The orderings $\prec$ and $\preceq_{\mathrm{D}}$ are equivalent if the marginal densities are uniform (see property (v) in Sec. 3.2).

If we fix a probability measure P_0 on $(\Omega, \mathcal{A})$ and consider divergence of probability measures Q from P_0 for $Q \ll P_0$ then

$$(P_0, Q) \preceq_{\mathrm{NP}} (P_0, Q') \quad \text{iff} \quad \frac{dQ}{dP_0} \prec \frac{dQ'}{dP_0}.$$

This is another form of the equivalence (2.4.4) restricted to the case $P = P' = P_0$.

Now we will investigate relations between $\preceq_{\mathrm{D}}$ and the quadrant ordering $\preceq_{\mathrm{QD}}$ which is one of the weakest orderings connected with monotone dependence (cf. Lehmann (1966)). We remind that for $X \sim X'$, $Y \sim Y'$ and for all $x, y \in (-\infty, \infty)$,

$$(X, Y) \preceq_{\mathrm{QD}} (X', Y') \quad \text{iff} \quad \Pr(X \leq x, Y \leq y) \leq \Pr(X' \leq x, Y' \leq y).$$

In general, neither of the two orderings $\preceq_{\mathrm{D}}$ and $\preceq_{\mathrm{QD}}$ implies the other, as shown by the following examples.

Let (X, Y), (X', Y'), (X'', Y'') be pairs of random vectors with values in $\{1, 2, 3\} \times \{1, 2, 3\}$ and distributed as

$$P = \begin{pmatrix} \frac{6}{33} & \frac{1}{33} & 0 \\ \frac{9}{33} & \frac{6}{33} & 0 \\ \frac{2}{33} & \frac{2}{33} & \frac{7}{33} \end{pmatrix}, \quad P' = \begin{pmatrix} \frac{6}{33} & \frac{1}{33} & 0 \\ \frac{10}{33} & \frac{5}{33} & 0 \\ \frac{1}{33} & \frac{3}{33} & \frac{7}{33} \end{pmatrix}, \quad P'' = \begin{pmatrix} \frac{7}{33} & 0 & 0 \\ \frac{7}{33} & \frac{8}{33} & 0 \\ \frac{3}{33} & \frac{1}{33} & \frac{7}{33} \end{pmatrix}.$$

Evidently, P, P' and P'' have the same pairs of marginal distributions. Let P_0 be the product independent distribution corresponding to any of P, P', P''. It is easy to check that in each pair (X, Y), (X', Y'), (X'', Y''), the components of the pair are quadrant dependent. Moreover,

$$P \preceq_{\mathrm{QD}} P' \quad \Big(\text{i.e. for any } i_0, j_0 \in \{1, 2, 3\},\ \sum_{i \leq i_0} \sum_{j \leq j_0} p_{ij} \leq \sum_{i \leq i_0} \sum_{j \leq j_0} p'_{ij}\Big).$$

However, neither $P \preceq_{\mathrm{QD}} P''$ nor $P'' \preceq_{\mathrm{QD}} P$ since

$$p_{11} < p'_{11}, \quad p_{11} + p_{21} > p''_{11} + p''_{21}.$$

On the other hand, the curves $L_{(P_0,P)}$ and $L_{(P_0,P')}$ intersect each other and

$$L_{(P_0,P'')} \le L_{(P_0,P)},$$

so that

$$(X,Y) \not\preceq_{\mathrm{D}} (X',Y'), \quad (X',Y') \not\preceq_{\mathrm{D}} (X,Y), \quad (X,Y) \preceq_{\mathrm{D}} (X'',Y'').$$

These examples supplement the evidence given by many contributors that stochastic dependence is a complicated notion which can be approached on many ways. We still have to look for a consistent set of orderings and families of distributions connected with absolute and monotone dependence. An ordering of absolute dependence should satisfy the condition that, restricted to an appropriately chosen family of monotone dependent pairs (X,Y), it should be equivalent to (or at least weaker or stronger than) an ordering particularly suited to this family of pairs.

Now we will show that in a narrow but important family of quadrant dependent distributions, naturally ordered, this natural ordering is equivalent both to $\preceq_{\mathrm{D}}$ and to $\preceq_{\mathrm{QD}}$.

Let $\mathcal{P} = \bigcup_{\alpha\in[0,1]} \mathcal{P}^+_\alpha \cup \mathcal{P}^-_\alpha$, where

$$\mathcal{P}^+_\alpha = \{P^+_\alpha : P^+_\alpha = \alpha P^+ + (1-\alpha)P_0\}, \quad \mathcal{P}^-_\alpha = \{P^-_\alpha : P^-_\alpha = \alpha P^- + (1-\alpha)P_0\},$$

and P^+, P^- are the upper and lower Fréchet distributions for given continuous marginal distributions, and P_0 is the product of the marginal distributions. We have

$$L_{(P_0,P^+_\alpha)}(t) = L_{(P_0,P^-_\alpha)}(t) = (1-\alpha)t \quad \text{for } t \in [0,1].$$

The family $\mathcal{P}$ is naturally ordered according to α.

Our next example involves the set $\mathcal{P}_{2\times2}$ consisting of pairs of binary random variables. It is known that any two binary random variables X, Y are quadrant dependent. A natural ordering $\preceq_m$ in $\mathcal{P}_{2\times2}$ which is connected with dependence of X and Y is

$$P \preceq_m P' \quad \text{if} \quad p_{ii} \le p'_{ii}, p_{ij} \ge p'_{ij}, \quad i,j = 1,2, i \ne j, \quad \text{for } p_{11}p_{22} \ge p_{12}p_{21}$$
$$\text{or} \quad p_{ii} \ge p'_{ii}, p_{ij} \le p'_{ij}, \quad i,j = 1,2, i \ne j, \quad \text{for } p_{11}p_{22} \le p_{12}p_{21}.$$

It was shown in Bromek and Kowalczyk (1990) that this ordering implies $\preceq_{\mathrm{D}}$. This fact is a nice property of $\preceq_{\mathrm{D}}$.

4. Link between divergence and proportional representation

4.1. Formulation of the problem and definition of the ordering $\preceq_x$. Let Ω be any set, finite or infinite, and let ν be any measure defined on a σ-algebra $\mathcal{A}$ of subsets of Ω, such that $\nu(\Omega)$ is positive and finite. Let $X : \Omega \to [0,\infty)$,

$Y : \Omega \to [0, \infty)$ be $(\Omega, \mathcal{A})$-measurable functions such that

$$0 < \int_\Omega X(\omega)\,\nu(d\omega) < \infty, \qquad 0 < \int_\Omega Y(\omega)\,\nu(d\omega) < \infty$$

and let $P(\cdot) = \nu(\cdot)/\nu(\Omega)$. According to (2.4.6), we introduce

$$\lambda_P^X(A) = \frac{\int_A X(\omega)\,\nu(d\omega)}{\int_\Omega X(\omega)\,\nu(d\omega)}, \qquad \lambda_P^Y(A) = \frac{\int_A Y(\omega)\,\nu(d\omega)}{\int_\Omega Y(\omega)\,\nu(d\omega)}.$$

We shall compare λ_P^X with λ_P^Y by means of the concentration curve $L[\lambda_P^X, \lambda_P^Y]$.

Divergence of λ_P^Y from λ_P^X measures the degree of departure from proportionality of Y to X. Typically, this problem concerns variables X and Y with nonnegative integer values. An important example concerns proportionality of a representation (obtained as a result of an election) to the size of electorate. The population consists of s units $\omega_1, \ldots, \omega_s$ with electorates $x_i = X(\omega_i)$ for $i = 1, \ldots, s$, $x_1 + \ldots + x_s = n$. Suppose that the size of the representation, say m, is selected a priori and let $y_i = Y(\omega_i)$ be the size of the representation of the ith unit. We want to measure the departure from proportionality of the vector $(y_1, \ldots, y_s)$ to $(x_1, \ldots, x_s)$. Ideal proportionality $\big(y_i = \frac{m}{n} x_i$ for $i = 1, \ldots, s\big)$ is rarely possible.

Let ν be the counting measure on $\Omega = \{\omega_1, \ldots, \omega_s\}$. Then

$$\int_\Omega X(\omega)\,\nu(d\omega) = \sum_{i=1}^s x_i = n, \qquad \int_\Omega Y(\omega)\,\nu(d\omega) = \sum_{i=1}^s y_i = m,$$

and λ_P^X, λ_P^Y are defined by

$$\lambda_X = \left(\frac{x_1}{n}, \ldots, \frac{x_s}{n}\right), \qquad \lambda_Y = \left(\frac{y_1}{m}, \ldots, \frac{y_s}{m}\right).$$

Ideal proportionality occurs when $\lambda_X = \lambda_Y$. A departure from proportionality of y's to x's corresponds to divergence of λ_Y from λ_X. In this problem divergence is never maximal since it is not possible to have $\lambda_X \perp \lambda_Y$.

Let

$$\mathbf{Y}(x, m) = \Big\{ y = (y_1, \ldots, y_s) : y_i \in \mathbb{N} \cup \{0\},\ y_i \le x_i,\ \sum_{i=1}^s y_i = m \Big\}.$$

For any fixed vector $x = (x_1, \ldots, x_s)$ with positive integer components and for a positive integer $m \le n$ we have an ordering $\preceq_x$ concerning proportionality of $y \in \mathbf{Y}(x, m)$ to x.

Definition 4.1.1. We say that y is *more proportional* to x than y', written $y \preceq_x y'$, if $(\lambda_X, \lambda_Y) \preceq_{\mathrm{NP}} (\lambda_X, \lambda_{Y'})$.

We recall that

$$(\lambda_X, \lambda_Y) \preceq_{\mathrm{NP}} (\lambda_X, \lambda_{Y'}) \quad \text{iff} \quad L[\lambda_X, \lambda_Y] \ge L[\lambda_X, \lambda_{Y'}].$$

The ordering $\preceq_x$ in $\mathbf{Y}(x,m)$ is the restriction to this set of the relative majorization ordering, considered by Joe (1990) in the set of all vectors with real components and fixed sum. According to Joe, for any vector x with positive components and any $y=(y_1,\ldots,y_s)$, $y'=(y'_1,\ldots,y'_s)$ with real components such that $\sum y_i=\sum y'_i$,

$$y \preceq^{\mathrm{r}}_x y' \quad \text{if} \quad \sum_{i=1}^{s} x_i\psi\left(\frac{y_i}{x_i}\right) \le \sum_{i=1}^{s} x_i\psi\left(\frac{y'_i}{x_i}\right)$$

for all continuous convex functions ψ with domain including y_i/x_i and y'_i/x_i for $i=1,\ldots,s$.

The equivalence of this ordering with $\preceq_x$ follows from (2.3.1) and (2.4.4).

4.2. Minimal elements for $\preceq_x$. As mentioned before, the vector

$$\left(\frac{m}{n}x_1,\ldots,\frac{m}{n}x_s\right) \tag{4.2.1}$$

is the smallest element in $\mathbf{Y}(x,m)$ if all its components are positive integers. However, this element exists only for suitably chosen pairs (x,m). Therefore it is important to look for minimal elements for $\preceq_x$ in $\mathbf{Y}(x,m)$. Intuitively, it is natural to consider as a candidate a vector obtained from (4.2.1) by a suitable rounding up or down of its components.

LEMMA 4.2.1. *Let $u_i=\frac{m}{n}x_i-\left[\frac{m}{n}x_i\right]$ for $i=1,\ldots,s$, $l=m-\sum_{i=1}^s\left[\frac{m}{n}x_i\right]=\sum_{i=1}^s u_i$ and let $\mathcal{I}$ be the set of all permutations $(i_1,\ldots,i_s)$ of $(1,\ldots,s)$ such that*

$$u_{i_1}\ge\ldots\ge u_{i_s}.$$

Then any vector $y^0=(y^0_1,\ldots,y^0_s)$ such that for some $(i_1,\ldots,i_s)\in\mathcal{I}$,

$$y^0_j=\begin{cases}\frac{m}{n}x_j+1-u_j=\left[\frac{m}{n}x_j\right]+1 & \text{for } j=i_1,\ldots,i_l,\\ \frac{m}{n}x_j-u_j=\left[\frac{m}{n}x_j\right] & \text{for } j=i_{l+1},\ldots,i_s,\end{cases}$$

is a minimal element for $\preceq_x$ in $\mathbf{Y}(x,m)$.

Proof. If $l=0$ then (4.2.1) belongs to $\mathbf{Y}(x,m)$ and is the smallest element for $\preceq_x$ in $\mathbf{Y}(x,m)$. Suppose that $l>0$. Let $1<k\le s$ be the number of components of (4.2.1) with nonzero u_i's:

$$u_{i_1}\ge u_{i_2}\ge\ldots\ge u_{i_l}\ge u_{i_{l+1}}\ge\ldots\ge u_{i_k}>u_{i_{k+1}}=\ldots=u_{i_s}=0.$$

Let S be an arbitrary subset of $\{i_1,\ldots,i_k\}$ consisting of l numbers, and let $S^{\mathrm{c}}=\{i_1,\ldots,i_k\}\setminus S$. Denote by $y^S=(y^S_1,\ldots,y^S_s)$ the vector obtained from (4.2.1) by rounding up the components indexed by elements of S, and rounding down the components indexed by elements of S^{c}. The sum of components of y^S is m for any S. This follows from the following equivalent equalities:

$$l=\sum_{j\in S}u_j+\sum_{j\in S^{\mathrm{c}}}u_j,\qquad \sum_{j\in S}(1-u_j)=\sum_{j\in S^{\mathrm{c}}}u_j.$$

We show that y^S minimizes the function assigning to any (x, y) the expression

$$\frac{1}{2}\sum_{i=1}^{s}\left|\frac{x_i}{n}-\frac{y_i}{m}\right|, \tag{4.2.3}$$

which is the maximal departure of the curve $L[\lambda_X, \lambda_Y]$ from the line $y = x$ in $(0, 1)^2$. Indeed,

$$\begin{aligned}\sum_{i=1}^{s}\left|\frac{y_i^S}{m}-\frac{x_i}{n}\right| &= \sum_{j\in S}\left(\frac{y_j^S}{m}-\frac{x_j}{n}\right)+\sum_{j\in S^{\mathrm{c}}}\left(\frac{x_j}{n}-\frac{y_j^S}{m}\right)\\ &= \sum_{j\in S}\frac{1-u_j}{m}+\sum_{j\in S^{\mathrm{c}}}\frac{u_j}{m} \geq \frac{1}{m}\Big(\sum_{r=1}^{l}(1-u_{i_r})+\sum_{r=l+1}^{k}u_{i_r}\Big).\end{aligned}$$

The last inequality is sharp iff $S \neq \{i_1, \ldots, i_l\}$ for every permutation $(i_1, \ldots, i_s) \in \mathcal{I}$. The vector y^S corresponding to such a set is not earlier than the vector $y^{S'}$ corresponding to $S' = \{i_1, \ldots, i_l\}$ for any $(i_1, \ldots, i_s) \in \mathcal{I}$.

To show that y^S for $S = \{i_1, \ldots, i_l\}$ is a minimal element in $\mathbf{Y}(x, m)$, it suffices to prove that for two different permutations belonging to $\mathcal{I}$ either the vectors are the same or they induce curves $L[\lambda_X, \lambda_{Y^S}]$ which are identical or intersect each other.

If $u_{i_l} > u_{i_{l+1}}$ for any $(i_1, \ldots, i_s) \in \mathcal{I}$, then the vectors y^S for $S = \{i_1, \ldots, i_l\}$ and any permutation $(i_1, \ldots, i_s)$ are all equal. Assume now that $u_{i_l} = u_{i_{l+1}}$ for any $(i_1, \ldots, i_s) \in \mathcal{I}$, and let $S' = \{i'_1, \ldots, i'_l\}$ differ from S in one element only. Since the general reasoning is the same, we will only consider this case. There exist j and j' such that $j \neq j'$ and $j \in S$, $j' \in S^{\mathrm{c}}$, $j \in S'^{\mathrm{c}}, j' \in S'$. It follows that y^S and $y^{S'}$ differ at most in components j and j':

$$\begin{aligned} y_j^S &= \frac{m}{n}x_j + 1 - u, \qquad & y_{j'}^S &= \frac{m}{n}x_{j'} - u,\\ y_j^{S'} &= \frac{m}{n}x_j - u, & y_{j'}^{S'} &= \frac{m}{n}x_{j'} + 1 - u,\end{aligned}$$

here $u = u_{i_l} = u_{i_{l+1}}$. Let $x_j \leq x_{j'}$. The inequalities

$$\frac{u}{x_j} \geq \frac{u}{x_{j'}}, \qquad \frac{1-u}{x_j} \geq \frac{1-u}{x_{j'}}$$

imply that

$$\frac{\frac{m}{n}x_j - u}{x_j} \leq \frac{\frac{m}{n}x_{j'} - u}{x_{j'}}, \qquad \frac{\frac{m}{n}x_{j'} + 1 - u}{x_{j'}} \leq \frac{\frac{m}{n}x_j + 1 - u}{x_j},$$

which is equivalent to

$$\frac{y_j^{S'}}{x_j} \leq \frac{y_{j'}^S}{x_{j'}} < \frac{m}{n} < \frac{y_{j'}^{S'}}{x_{j'}} \leq \frac{y_j^S}{x_j}. \tag{4.2.4}$$

Since the slopes of the piecewise linear curve $L[\lambda_X, \lambda_Y]$ are equal to the respective quotients $(y_j/m)/(x_j/n)$, the inequalities (4.2.4) imply that

(i) the curves $L[\lambda_X, \lambda_{Y^S}]$ and $L[\lambda_X, \lambda_{Y^{S'}}]$ coincide if $x_j = x_{j'}$,

(ii) if $x_j < x_{j'}$ then all inequalities in (4.2.4) are sharp, so that the two curves L intersect.

The vectors y^0 defined by (4.2.2) were considered in Baliński and Young (1982). They are called there Hamilton's rules as they were used by Hamilton in apportioning seats among the states in the United States election. Baliński and Young mentioned the fact that these rules minimize the function (4.2.3). Note that (4.2.3) is of the form $\sum_i x_i \psi(h_i)$ for $h_i = y_i/x_i$ where ψ is a convex continuous function. In view of (2.4.5), this suffices to prove that the Hamilton rule is a minimal element for $\preceq_x$ in the case when this rule is unique. Lemma 4.2.1 extends this assertion to the general case.

The vectors obtained by the Hamilton rule may also be interpreted as those vectors from $\mathbf{Y}(x, m)$ which give a distribution λ_Y such that the transfer of probability mass from λ_Y to obtain λ_X is minimal.

Obviously, the Hamilton vectors are not the only minimal elements for $\preceq_x$. This property is also shared by vectors obtained by some other rules of proportional apportioning mentioned in Baliński and Young (1982). These rules have been invented as intuitively "most closest" to ideal proportional representation since they minimize some measure of departure from proportionality. In particular, we have the rules proposed by Adams, Jefferson, Hill and Webster. We shall not describe each rule in detail, restricting ourselves to the following:

$$\begin{aligned}
y_{\text{Adams}} &= \arg\Big(\max_y \min_{1\le i\le s} \frac{y_i}{x_i}\Big),\\
y_{\text{Jeff}} &= \arg\Big(\min_y \max_{1\le i\le s} \frac{y_i}{x_i}\Big),\\
y_{\text{Hill}} &= \arg\Big(\min_y \sum_{i=1}^{s} y_i \Big(\frac{x_i}{y_i} - \frac{n}{m}\Big)^2\Big),\\
y_{\text{Web}} &= \arg\Big(\min_y \sum_{i=1}^{s} x_i \Big(\frac{y_i}{x_i} - \frac{m}{n}\Big)^2\Big).
\end{aligned}$$

All these vectors are minimal elements for $\preceq_x$ in the case when they are unique. For the first two methods the proof follows directly from their interpretation involving the curve $L[\lambda_X, \lambda_Y]$: the Adams rule maximizes the slope of the first segment of L while the Jefferson rule minimizes the slope of the last segment of the curve. For the next methods, the proof follows from the fact that the minimized functions are of the form $\sum x_i \psi(h_i)$ for $h_i = y_i/x_i$ and some convex function ψ.

At the moment, we have neither a proof nor even an intuitive view whether the vectors obtained by the rules proposed by Adams, Jefferson, Hill and Webster are minimal elements for $\preceq_x$ when they are not unique. It was not possible to find a non-unique solution of any of these rules which would not also be a non-unique Hamilton vector so that they were minimal elements due to Lemma 4.2.1.

The rules considered in this section will be discussed once more in Sec. 5.4.

4.3. Maximal elements for $\preceq_x$. Intuitively it is clear that the departure from proportionality will be maximal when some electorates get the maximal possible number of representatives, and some other electorates get the minimal possible numbers. We now provide a proof of this statement.

Let Z be a random variable taking values $1, \ldots, s$ with probabilities $\lambda_X(i) = x_i/n$ for $i = 1, \ldots, s$. For any vector $y = (y_1, \ldots, y_s) \in \mathbf{Y}(x, m)$ let h_y be the function on $\{1, \ldots, s\}$ defined by

$$h_y(i) = \frac{ny_i}{mx_i} \quad \text{for } i = 1, \ldots, s.$$

We have $0 \le h_y(i) \le n/m$ for $i = 1, \ldots, s$, and $y \in \mathbf{Y}(x, m)$. Let

$$\mathbf{Y}_1(x, m) = \{y \in \mathbf{Y}(x, m) : y_i = 0 \text{ or } y_i = x_i \text{ and, for at most one index } i_0 \in \{1, \ldots, s\},\ 0 < y_{i_0} < x_{i_0}\}.$$

If $y \in \mathbf{Y}_1(x, m)$, then $h_y(Z)$ takes on at most three values: 0, n/m, and $h \in (0, n/m)$. Thus we have $\mathbf{Y}_1(x, m) \subset \mathbf{Y}_3(x, m)$ where

$$\mathbf{Y}_3(x, m) = \{y \in \mathbf{Y}(x, m) : h_y \text{ takes on at most three values: } 0, n/m \text{ and } h \in (0, n/m)\}.$$

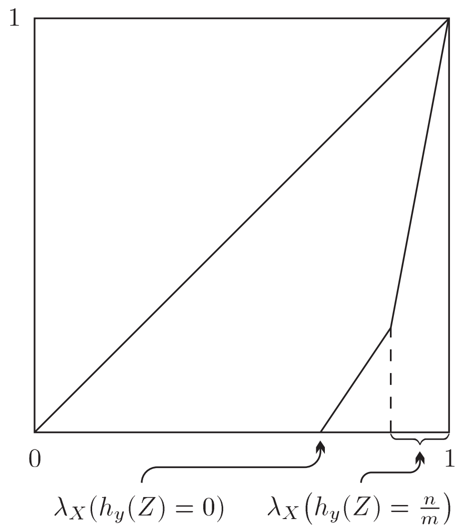

Fig. 1. The curve $L[\lambda_X, \lambda_Y]$ for $y \in \mathbf{Y}_3(x, m)$

LEMMA 4.3.1. *If $y, y' \in \mathbf{Y}_3(x, m)$ then the inequalities*

$$\lambda_X(h_y(Z) = 0) \le \lambda_X(h_{y'}(Z) = 0), \qquad \lambda_X\left(h_y(Z) = \frac{n}{m}\right) \le \lambda_X\left(h_{y'}(Z) = \frac{n}{m}\right)$$

imply that $y \preceq_x y'$.

The proof follows immediately from Fig. 1.

THEOREM 4.3.1. *The set of maximal elements for $\preceq_x$ in $\mathbf{Y}(x, m)$ is a subset of $\mathbf{Y}_1(x, m)$.*

P r o o f. In view of Lemma 4.3.1 it is enough to show that the set of maximal elements is a subset of $\mathbf{Y}_3(x,m)$. We shall show that a vector which does not belong to $\mathbf{Y}_3(x,m)$ is not a maximal element for $\preceq_x$.

Let $\{i_1,\ldots,i_k\}$ be the largest subset of $\{1,\ldots,s\}$ such that

$$0<\frac{y_{i_1}}{x_{i_1}}\le\ldots\le\frac{y_{i_k}}{x_{i_k}}<1.$$

Since $y\notin\mathbf{Y}_3(x,m)$ we have

$$\frac{y_{i_1}}{x_{i_1}}<\frac{y_{i_k}}{x_{i_k}}.$$

Let $y'=(y'_1,\ldots,y'_s)$ be defined by

$$y'_j=y_j \quad \text{for } j\ne i_1,i_k \quad \text{and}$$
$$y'_{i_1}=0,\quad y'_{i_k}=y_{i_1}+y_{i_k}\quad \text{if}\quad y_{i_1}+y_{i_k}\le x_{i_k},$$
$$y'_{i_1}=y_{i_1}+y_{i_k}-x_{i_k},\quad y'_{i_k}=x_{i_k}\quad \text{if}\quad y_{i_1}+y_{i_k}>x_{i_k}.$$

Since $L[\lambda_X,\lambda_Y]\ge L[\lambda_X,\lambda_{Y'}]$ and $L[\lambda_X,\lambda_Y]\ne L[\lambda_X,\lambda_{Y'}]$, obviously y is not a maximal element.

If $y\in\mathbf{Y}_1(x,m)$ is such that for every $i=1,\ldots,s$ either $y_i=0$ or $y_i=x_i$ then y is the largest element for $\preceq_x$ in $\mathbf{Y}(x,m)$. For such a vector we have

$$\lambda_X(h_y(Z)=0)=1-\frac{m}{n},\qquad \lambda_X\left(h_y(Z)=\frac{n}{m}\right)=\frac{m}{n}.$$

There may exist more than one largest element but all of them lead to the same curve $L[\lambda_X,\lambda_Y]$. If no largest element exists, there may exist more than one maximal element and the corresponding curves may intersect.

EXAMPLES. 1. If $i_0\in\{1,\ldots,s\}$ is such that $m\le\min\{x_1,\ldots,x_s\}=x_{i_0}$, then the maximal (and largest) vector $y=(y_1,\ldots,y_s)$ for $\preceq_x$ has components

$$y_{i_0}=m,\qquad y_i=0\quad \text{for } i\ne i_0.$$

2. If $i_0,i_1\in\{1,\ldots,s\}$ are such that

$$x_{i_0}=\min\{x_1,\ldots,x_s\}<m<\min\{\{x_1,\ldots,x_s\}\setminus\{x_{i_0}\}\}=x_{i_1}$$

then there exist two maximal vectors y,y' with components

$$y_{i_0}=x_{i_0},\qquad y_{i_1}=m-x_{i_0},\qquad y_j=0\quad \text{for } j\ne i_0,i_1,$$
$$y'_{i_1}=m,\qquad y'_j=0\quad \text{for } j\ne i_1.$$

5. Directed concentration of probability measures

5.1. Directed concentration curve. For any measurable space $(\Omega,\mathcal{A})$, let P, Q be probability measures defined on it, and let $\preceq_\varphi$ be an ordering in Ω introduced by means of a given function $\varphi:\Omega\to[-\infty,\infty]$:

$$\omega_1\preceq_\varphi\omega_2\quad \text{if}\quad \varphi(\omega_1)\le\varphi(\omega_2).$$

We will also consider the equivalence

$$\omega_1 \cong_\varphi \omega_2 \quad \text{if} \quad \omega_1 \preceq_\varphi \omega_2 \text{ and } \omega_2 \preceq_\varphi \omega_1,$$

and strict ordering:

$$\omega_1 \prec_\varphi \omega_2 \quad \text{if} \quad \omega_1 \preceq_\varphi \omega_2 \text{ and not } \omega_1 \cong \omega_2.$$

Obviously, $\omega_1 \cong_\varphi \omega_2$ iff $\varphi(\omega_1) = \varphi(\omega_2)$, and $\omega_1 \prec_\varphi \omega_2$ iff $\varphi(\omega_1) < \varphi(\omega_2)$.

If $\Omega \subset \mathbb{R}$, the ordering $\preceq_\varphi$ is often identified with inequality $\leq$ in $\mathbb{R}$.

We will consider concentration of P and Q on the sets

$$A_z^\varphi = \{\omega \in \Omega : \varphi(\omega) \leq z\} \quad \text{for } z \in [-\infty, \infty].$$

To this end, we introduce a curve $C^\varphi_{(P,Q)}$, called the *φ-directed concentration curve* of Q with respect to P, which is defined to be the set

$$\{(P(A_z^\varphi), Q(A_z^\varphi)) : z \in [-\infty, \infty]\}$$

contained in $[0,1]^2$, completed if necessary by the points $(0,0)$, $(1,1)$, and by linear interpolation. The curve $C^\varphi_{(P,Q)}$ is nondecreasing (i.e. it is the graph of a nondecreasing relation), but not necessarily convex. It lies above or below the line $y = x$ in $[0,1]^2$.

It is convenient to assign to this curve a function $C^\varphi_{(P,Q)}(\cdot)$ on $[0,1]$ such that $(t, C^\varphi_{(P,Q)}(t))$ lies on the curve for any $t \in [0,1]$ and

$$\begin{aligned} C^\varphi_{(P,Q)}(t) &= C^\varphi_{(P,Q)}(t+) \quad \text{for } t \in [0,1), \\ C^\varphi_{(P,Q)}(1) &= C^\varphi_{(P,Q)}(1-). \end{aligned}$$

The superscript φ will be omitted for $\varphi(x) = x$, which can happen only when the distributions P and Q are concentrated on $\mathbb{R}$. In this case we use the term *directed* (instead of *φ-directed*) *concentration curve*. Moreover, we will also use the notation $C[P,Q]$ instead of $C_{(P,Q)}$ whenever P and Q are written in a complicated way (e.g., $P = \lambda^{r(X)}_{P_X}$ or $P = \lambda_{P_X}$).

There exists a counterpart of $C^\varphi_{(P,Q)}$ which has an important interpretation in discriminant analysis. It is the set

$$\{(x,y) : (1-x, y) \in C^\varphi_{(P,Q)}\}$$

(i.e. this curve is related to $C^\varphi_{(P,Q)}$ in the same way as the Neyman–Pearson curve is related to the concentration curve). Each point on the counterpart of $C^\varphi_{(P,Q)}$ is formed by the probabilities of wrong decisions,

$$(a_{12}(\delta^\varphi_{\kappa,s}), a_{21}(\delta^\varphi_{\kappa,s})),$$

corresponding to the decision rule based on φ:

$$\delta^\varphi_{\kappa,s} = \begin{cases} 1 & \text{if } \varphi(\omega) < \kappa, \\ s & \text{if } \varphi(\omega) = \kappa, \\ 0 & \text{if } \varphi(\omega) > \kappa, \end{cases}$$

for $\kappa \in [-\infty, \infty]$ and $s \in [0,1]$.

It is evident that for suitably chosen φ, the φ-directed concentration curve coincides with the corresponding concentration curve (cf. property (iii) below).

The curve $C^{\varphi}_{(P,Q)}$ may be used to describe the stochastic ordering $\le_{\rm st}$ of $P\varphi^{-1}$, $Q\varphi^{-1}$:

$$P\varphi^{-1} \le_{\rm st} Q\varphi^{-1} \Leftrightarrow C^{\varphi}_{(P,Q)}(t) \le t \text{ for } t \in [0,1],$$

which is equivalent to

$$P\varphi^{-1}[-\infty, z] = P\{\omega : \varphi(\omega) \le z\} \ge Q\{\omega : \varphi(\omega) \le z\} = Q\varphi^{-1}[-\infty, z]$$

for all $z \in [-\infty, \infty]$.

The measures $P\varphi^{-1}$, $Q\varphi^{-1}$ are equivalent iff $C^{\varphi}_{(P,Q)}(t) = t$ for $t \in [0,1]$. If there exists $z \in (-\infty, \infty)$ such that $P(A^{\varphi}_z) = 1, Q(A^{\varphi}_z) = 0$, we say that Q is *completely right of* P with respect to $\preceq_\varphi$; the curve $C^{\varphi}_{(P,Q)}$ consists then (and only then) of the two edges of the unit square emanating from $(1,0)$. Similarly, if there exists $z \in (-\infty, \infty)$ such that $P(A^{\varphi}_z) = 0$, $Q(A^{\varphi}_z) = 1$, we say that Q is *completely left of* P with respect to $\preceq_\varphi$; $C^{\varphi}_{(P,Q)}$ consists then (and only then) of the two edges emanating from $(0,1)$.

If P, Q are measures on $\mathbb{R}$ and $\varphi(x) = x$ then

$$P \le_{\rm st} Q \quad \text{iff} \quad C^{\varphi}_{(P,Q)}(t) \le t \quad \text{for } t \in [0,1].$$

The following properties of $C^{\varphi}_{(P,Q)}$ are immediately implied by its definition:

THEOREM 5.1.1. (i) *If $\Omega \subset \mathbb{R}^+$ and φ is strictly increasing, then*

$$C^{\varphi}_{(P,Q)} = C^{\rm id}_{(P,Q)} \qquad \textit{where } {\rm id}(x) = x.$$

(ii) *If $\Omega \subset \mathbb{R}^+$ and φ is strictly decreasing, then*

$$C^{\varphi}_{(P,Q)} = \{(x,y) : (1-x, 1-y) \in C^{\rm id}_{(P,Q)}\}.$$

(iii) *The likelihood ratio dQ/dP is nondecreasing with respect to the ordering $\preceq_\varphi$ iff*

$$C^{\varphi}_{(P,Q)} = L_{(P,Q)}.$$

(iv) *The likelihood ratio dQ/dP is nonincreasing with respect to $\preceq_\varphi$ iff*

$$C^{\varphi}_{(P,Q)}(t) = L^{-}_{(P,Q)}(t) \qquad \textit{for } t \in [0,1],$$

where

$$L^{-}_{(P,Q)}(t) = 1 - L_{(P,Q)}(1-t).$$

(v) *Let P, Q be measures on $\mathbb{R}$ with distribution functions F, H, respectively, such that $Q \ll P$ and let $\varphi(x) = x$. Then*

$$C_{(P,Q)}(u) = \int_0^u \frac{dQ}{dP}(F^{-1}(v))\,dv,$$

where $F^{-1}(v) = \inf\{t : F(t) \ge v\}$.

THEOREM 5.1.2. *Let P, Q be measures on $\mathbb{R}$ and let $\varphi(x) = x$. Then*

(i) $C_{(P,Q)} = L_{(P,Q)}$ *iff* $C_{(P,Q)}$ *is convex*,
(ii) $C_{(P,Q)} = L^{-}_{(P,Q)}$ *iff* $C_{(P,Q)}$ *is concave.*

COROLLARY 5.1.1. *Under the assumptions of Theorem* 5.1.1(v) *we have*

$$L_{(P,Q)}(u) = \int_0^u (F^h)^{-1}(t)\,dt,$$

where $h = \frac{dQ}{dP}$, $F^h = Fh^{-1}$.

Proof. By Theorem 5.1.1(iii) and (v),

$$L_{(P,Q)}(u) = C^h_{(P,Q)} = C[P^h, Q^h] = \int_0^u \frac{dH^h}{dF^h}((F^h)^{-1}(t))\,dt,$$

where $P^h = Ph^{-1}$, $Q^h = Qh^{-1}$, $H^h = Hh^{-1}$. The proof is complete since

$$\frac{dH^h}{dF^h}(t) = t,$$

which follows from the equality

$$H^h(t) = \int_0^s t\,dF^h(t).$$

In Cifarelli and Regazzini (1987), the curve $L^{-}_{(P,Q)}$ appearing in (iv) was called the upper concentration function of Q with respect to P. Moreover, they showed that if P and Q are nonatomic probability measures on $(\Omega, \mathcal{A})$ then the range of the vector probability measure (P,Q) is a closed, convex subset $\mathcal{S}$ of $\mathbb{R}^2$; if $A \in \mathcal{A}$ and $P(A) = t$, then

$$L_{(P,Q)}(t) = \min\{Q(A) : A \in \mathcal{A}, P(A) = t\},$$
$$L^{-}_{(P,Q)}(t) = \sup\{Q(A) : A \in \mathcal{A}, P(A) = t\}.$$

It follows that the range of (P,Q) coincides with the closed subset of $\mathbb{R}^2$ bounded by the graphs of the concentration function and of the upper concentration function of Q with respect to P. Thus, for any $\varphi : \Omega \to [-\infty, \infty]$ we have

$$L_{(P,Q)}(t) \le C^{\varphi}_{(P,Q)}(t) \le L^{-}_{(P,Q)}(t) \quad \text{for } t \in [0,1]. \tag{5.1.1}$$

To illustrate, consider $C^{\varphi}_{(P,Q)}$ when $(P,Q) = (P_X, \lambda^X_P)$ for a ratio random variable X defined on $\mathbb{R}^+$. It is then natural to put $\varphi(x) = x$. It follows that for this pair (P,Q) and for $\varphi = \mathrm{id}$, $C^{\varphi}_{(P,Q)} = L_{(P,Q)} = L_X$, which means that the directed concentration curve appearing naturally in inequality analysis coincides with the concentration curve for (P_X, λ^X_P).

The curves $C^{\varphi}_{(P,Q)}$ are used to introduce an ordering $\preceq_{\mathrm{C}}$ analogous to $\preceq_{\mathrm{NP}}$. Let $\mathcal{P}$ and $\mathcal{Q}$ be two families of distributions defined on $(\Omega, \mathcal{A})$ and let $\preceq_{\varphi}$ be a fixed ordering in Ω for a given function φ. In $\mathcal{P} \times \mathcal{Q}$ we introduce an ordering $\preceq_{\mathrm{C}}$ according to the degree of the directed concentration of Q with respect to P.

The ordering is based on the comparison of the directed concentration curves corresponding to the two pairs being compared: for $P, P' \in \mathcal{P}$, $Q, Q' \in \mathcal{Q}$,

$$(P,Q) \preceq_{\mathrm{C}} (P',Q') \quad \text{iff} \quad C^{\varphi}_{(P,Q)} \geq C^{\varphi}_{(P',Q')}.$$

If $\mathcal{P}$ is equal to $\mathcal{Q}$ and consists of all probability measures on $(\Omega, \mathcal{A})$ then every pair (P,Q) with Q completely left to P is a smallest element for $\preceq_{\mathrm{C}}$, and every pair (P,Q) with Q completely right to P is a largest element for $\preceq_{\mathrm{C}}$. If $\varphi = \mathrm{id}$ and $\Omega = \mathbb{R}$ then $\preceq_{\mathrm{C}}$ is equal to $\leq_{\mathrm{st}}$ for pairs of measures:

$$(P,Q) \leq_{\mathrm{st}} (P',Q') \quad \text{if} \quad Q \text{ is more left to } P \text{ than } Q' \text{ to } P'.$$

5.2. Grade transformation of a random variable. Let X and Y be continuous random variables on $(\mathbb{R}, \mathcal{B})$ with distribution functions F and H, respectively, such that $H \ll F$. Then the graph of the distribution function of the random variable $F(Y)$ coincides with the directed concentration curve for the pair (X,Y) (we use the notation $C_{(X,Y)}$ instead of $C_{(P,Q)}$ when $X \sim P$, $Y \sim Q$):

$$C_{(X,Y)}(u) = H(F^{-1}(u)) \quad \text{for } u \in (0,1),$$

where $F^{-1}(u) = \inf\{t : F(t) \geq u\}$. The random variable $F(Y)$ will be called the *grade transformation* of Y with respect to X, and $C_{(X,Y)}$ will be called the *grade distribution function* of Y with respect to X. These two notions will now be generalized to arbitrary pairs of real-valued variables X and Y.

Definition 5.2.1 (cf. Szczesny (1991)). For any random variables X, Y on $(\mathbb{R}, \mathcal{B})$ we say that Y_X is a *grade transformation* of Y with respect to X if it is a random variable on $[0,1]$ with distribution function

$$H^*(u) = \int_{\mathbb{R}} F^*(x,u)\, dH(x),$$

where

$$F^*(x,u) = \begin{cases} 1 & \text{if } F(x+) \leq u, \\ (u - F(x-))/(F(x+) - F(x-)) & \text{if } F(x-) \leq u < F(x+), \\ 0 & \text{if } F(x-) > u. \end{cases}$$

Obviously, a grade transformation of X with respect to X is uniform on $[0,1]$. The function F^* is the restriction to the set $\mathbb{R} \times \{(0,u) : u \in (0,1)\}$ of a suitable transition probability function defined on the cartesian product of $\mathbb{R}$ and the Borel field of subsets of $[0,1]$. This random transformation of Y onto Y_X by means of F^* is in the continuous case realized by F; in this case

$$F^*(x,u) = \begin{cases} 1 & \text{if } u \geq F(x), \\ 0 & \text{otherwise}, \end{cases}$$

$$H^*(u) = H(F^{-1}(u)), \qquad Y_X = F(Y).$$

If X and Y are discrete random variables with distributions $(p_1,\ldots,p_k)$, $(q_1,\ldots,q_k)$ then

$$F^*(i,u)=\begin{cases}(1/p_i)\nu([0,u]\cap[\sum_{j=1}^{i-1}p_j,\sum_{j=1}^{i}p_j])\\ \qquad\text{for } i\in\{1,\ldots,k\} \text{ such that } p_i>0 \text{ and } u\in(0,1),\\ 0 \quad \text{for } i\in\{1,\ldots,k\} \text{ such that } p_i=0 \text{ and } u<\sum_{j=1}^{i-1}p_j,\\ 1 \quad \text{for } i\in\{1,\ldots,k\} \text{ such that } p_i=0 \text{ and } u\ge\sum_{j=1}^{i-1}p_j,\end{cases}$$

where ν is the Lebesgue measure. It follows that if $p_i>0$ for $i=1,\ldots,k$ then

$$\begin{aligned}H^*(u)&=\sum_{i=1}^{k}F^*(i,u)q_i\\ &=\begin{cases}(q_1/p_1)u \quad \text{for } u\le p_1,\\ \sum_{j=1}^{i-1}q_j+(q_i/p_i)(u-\sum_{j=1}^{i-1}p_j)\\ \qquad\qquad \text{for } \sum_{j=1}^{i-1}p_j<u\le\sum_{j=1}^{i}p_j \text{ and } i=2,\ldots,k.\end{cases}\end{aligned}$$

It follows from Szczesny (1991) that for any random variables X, Y with distributions F, H on $(\mathbb{R},\mathcal{B})$ the graph of the distribution function H^* of the grade transformation of Y with respect to X lies on the directed concentration curve $C_{(F,H)}$.

Moreover, for any random variable X with distribution F the Lorenz curve L_X is the distribution function of the grade transformation of $X^{(1)}$ with respect to X, where $X^{(1)}$ is the random variable with distribution $F^{(1)}$ given by (2.2.3).

5.3. Correlation and ratio curves. We mentioned in Sec. 3.1 that dependence can be investigated as divergence between two distributions on $(\Omega,\mathcal{B})$; in particular: (P,λ_P^X) when X is a ratio random variable; (P,λ_P^Y) when Y is a ratio random variable; $(\lambda_P^X,\lambda_P^Y)$ when both variables are ratio variables. Now, we are interested in *directed* divergence of Y from X: we want to know whether there exists a tendency that smaller values of Y coappear with smaller values of X (or that smaller values of Y coappear with greater values of X). Therefore, we should consider the measurable space $((\mathbb{R}^+)^2,\mathcal{B}^2)$ and the function $\varphi(x,y)=x$. It follows that

$$(x_1,y_1)\preceq_\varphi(x_2,y_2) \quad \text{if} \quad x_1\le x_2$$

and $A_z^\varphi=\{(x,y):x\le z\}$ for $z\in[0,\infty]$.

We now consider the following curves concerning directed divergence:

(1) $C^\varphi_{(P,Q)}$ for $P=P_{XY}$, $Q=\lambda^Y_{P_{XY}}$, given by

$$\{(P_{XY}(A_z^\varphi),\lambda^Y_{P_{XY}}(A_z^\varphi)):z\in[0,\infty]\},$$

(2) $C^\varphi_{(P,Q)}$ for $P=\lambda^X_{P_{XY}}$, $Q=\lambda^Y_{P_{XY}}$, given by

$$\{(\lambda^X_{P_{XY}}(A_z^\varphi),\lambda^Y_{P_{XY}}(A_z^\varphi)):z\in[0,\infty]\},$$

each curve completed if necessary by linear interpolation.

These two curves were mentioned by Taguchi (1987) in his study on the concentration surface as some projection curves corresponding to this surface. The first curve was called the *correlation curve* and the second the *ratio curve*.

It is easy to see that the correlation and ratio curves are the sets

$$\Big\{\Big(P_X(X\le z), \frac{1}{E(Y)}E(Y;X\le z)\Big): z\in[0,\infty]\Big\}$$
$$= \Big\{\Big(P_X(X\le z), \frac{1}{E(Y)}E(E(Y\mid X);X\le z)\Big): z\in[0,\infty]\Big\}$$

and

$$\Big\{\Big(\frac{1}{E(X)}E(X;X\le z), \frac{1}{E(Y)}E(Y;X\le z)\Big): z\in[0,\infty]\Big\}$$
$$= \Big\{\Big(\frac{1}{E(X)}E(X;X\le z), \frac{1}{E(Y)}E(E(Y\mid X);X\le z)\Big): z\in[0,\infty]\Big\},$$

respectively. Consequently, the correlation curve coincides with the directed divergence curve $C^{\varphi}_{(P,Q)}$ for $\varphi(x)=x$, $P=P_X$ and $Q=\lambda^{r(X)}_{P_X}$, where $r(x)=E(Y|X=x)$. Similarly, the ratio curve is the directed divergence curve $C^{\varphi}_{(P,Q)}$ for $\varphi(x)=x$, $P=\lambda_{P_X}$, $Q=\lambda^{r(X)}_{P_X}$. We simplify the notation for the correlation and ratio curves, putting

$$C_{\mathrm{cor}}[(X,Y)] = C[P_X,\lambda^{r(X)}_{P_X}],\qquad C_{\mathrm{ratio}}[(X,Y)] = C[\lambda^{X}_{P_X},\lambda^{r(X)}_{P_X}].$$

We have

$$C_{\mathrm{cor}}[(X,Y)]$$
$$= \Big\{\Big(\int_0^z f_X(x)\,\nu_X(dx), \frac{1}{E(Y)}\int_0^z r(x)f_X(x)\,\nu_X(dx)\Big): z\in[0,\infty]\Big\},$$
$$C_{\mathrm{ratio}}[(X,Y)]$$
$$= \Big\{\Big(\frac{1}{E(X)}\int_0^z xf_X(x)\,\nu_X(dx), \frac{1}{E(Y)}\int_0^z r(x)f_X(x)\,\nu_X(dx)\Big): z\in[0,\infty]\Big\},$$

completed if necessary by linear interpolation.

The density ratios for the univariate distributions compared by means of the correlation and ratio curves are, respectively,

$$\frac{r(x)}{E(Y)}\quad\text{and}\quad\frac{r(x)}{xE(Y)}$$

for any x such that $f_X(x)>0$.

The properties of the two curves are presented in Theorems 5.3.1 and 5.3.2. Some of the properties of the correlation curve follow from its connections with the monotone dependence function introduced for pairs (X,Y) of nondegenerate random variables, in the continuous case in Kowalczyk and Pleszczyńska (1977) and in the general case in Kowalczyk (1977), as

$$\mu^+_{Y,X}(u) = \frac{uE(Y) - E(Y; X < x_u) - r(x_u)(u - P(X < x_u))}{uE(Y) - E(Y; Y < y_u) - y_u(u - P(Y < y_u))},$$
$$\mu^-_{Y,X}(u) = \frac{uE(Y) - E(Y; X < x_u) - r(x_u)(u - P(X < x_u))}{E(Y; Y > y_{1-u}) - uE(Y) + y_{1-u}(u - P(Y > y_{1-u}))},$$

and

$$\mu_{Y,X}(u) = \begin{cases} \mu^+_{Y,X}(u) & \text{if } \mu^+_{Y,X}(u) \ge 0, \\ \mu^-_{Y,X}(u) & \text{if } \mu^-_{Y,X}(u) < 0, \end{cases}$$

for $u \in (0,1)$.

By straightforward transformations,

$$(5.3.1)\quad C_{\mathrm{cor}}[(X,Y)](u) = \frac{E(Y; X < x_u) + E(Y \mid X = x_u)(u - P(X < x_u))}{E(Y)}$$
$$= \begin{cases} u(1 - \mu^+_{Y,X}(u)) + \mu^+_{Y,X}(u)L_Y(u) & \text{if } \mu^+_{Y,X}(u) \ge 0, \\ u(1 + \mu^-_{Y,X}(u)) - \mu^-_{Y,X}(u)L^-_Y(u) & \text{if } \mu^-_{Y,X}(u) < 0 \end{cases}$$

for $u \in (0,1)$.

THEOREM 5.3.1. (i) $C_{\mathrm{cor}}[(X,Y)](u) \le u$ *for all* $u \in [0,1]$ (*i.e. the correlation curve lies under the* 45° *line*) *iff*

$$E(Y \mid X \le x) \le E(Y) \quad \textit{for all } x,$$

and $C_{\mathrm{cor}}[(X,Y)](u) \ge u$ *for all* $u \in [0,1]$ *iff*

$$E(Y \mid X \le x) \ge E(Y) \quad \textit{for all } x.$$

(ii) $L_Y \le L_{r(X)} \le C_{\mathrm{cor}}[(X,Y)] \le L^-_{r(X)} \le L^-_Y$, *where for any nonnegative random variable* Z,

$$L^-_Z(u) = 1 - L_Z(1-u) \quad \textit{for } u \in [0,1].$$

(iii) *We have*

$$C_{\mathrm{cor}}[(X,Y)] = \begin{cases} L_{r(X)} & \textit{iff } r \textit{ is nondecreasing}, \\ L^-_{r(X)} & \textit{iff } r \textit{ is nonincreasing}, \\ L_Y & \textit{iff there exists a nondecreasing function } g \\ & \textit{such that } Y = g(X), \\ L^-_Y & \textit{iff there exists a nonincreasing function } g \\ & \textit{such that } Y = g(X). \end{cases}$$

(iv) *Assume that there exists an increasing function* g^+ *such that* $g^+(X) \sim Y$; *then for each* $\varrho \in [0,1]$ *the necessary and sufficient condition for*

$$C_{\mathrm{cor}}[(X,Y)](u) = u(1-\varrho) + \varrho L_Y(u)$$

is that the regression function m^+ *of* Y *on* $g^+(X)$ *is of the form*

$$m^+(x) = \varrho x + (1-\varrho)E(Y).$$

Analogously, if there exists an increasing function g^- *such that* $g^-(X) \sim (-Y)$, *then for each* $\varrho \in [-1,0]$ *the necessary and sufficient condition for*

$$C_{\rm cor}[(X,Y)](u) = u(1+\varrho) - \varrho L_Y^-(u)$$

is that regression function m^- *of* Y *on* $g^-(X)$ *is of the form*

$$m^-(x) = \varrho x + (1+\varrho)E(Y).$$

(v) *Let* (X,Y), (X',Y') *satisfy* $X' \sim X$, $Y' \sim Y$. *Then the inequality*

$$C_{\rm cor}[(X,Y)] \ge C_{\rm cor}[(X',Y')] \tag{5.3.2}$$

holds if and only if

$$E(Y \mid X \le x) \ge E(Y' \mid X' \le x) \quad \text{for all } x \in \mathbb{R}^+.$$

Moreover, equality holds in (5.3.2) *if and only if*

$$E(Y \mid X = x) = E(Y' \mid X' = x) \quad \text{for all } x \in \mathbb{R}^+.$$

(vi) *For any* $u \in (0,1)$, *the equality*

$$C_{\rm cor}[(X,Y)](u) = L_Y(u)$$

holds if and only if either

$$P(X \ge x_u, Y < y_u) = P(X \le x_u, Y > y_u) = 0$$

and $P(X < x_u) < u < P(X \le x_u)$, *or*

$$P(X > x_u, Y < y_u) = P(X \le x_u, Y > y_u) = 0$$

and $P(X \le x_u) = u$, *or*

$$P(X \ge x_u, Y < y_u) = P(X < x_u, Y > y_u) = 0$$

and $P(X < x_u) = u < P(X \le x_u)$.

A similar condition can be formulated for

$$C_{\rm cor}[(X,Y)](u) = L_Y^-(u).$$

(vii) *If the sequence* $(P_n,\ n = 1,2,\ldots)$ *of distributions of points* (X_n,Y_n) *is weakly convergent to the distribution* P *of a pair* (X,Y) *such that for some* $a \in \mathbb{R}^+$, $P(Y \le a) = 1$, *then*

$$\lim_{n\to\infty} C_{\rm cor}[(X_n,Y_n)](u) = C_{\rm cor}[(X,Y)](u) \quad \text{for every } u \in (0,1).$$

Proof. (i) $C_{\rm cor}[(X,Y)](u) \le u$ for all $u \in (0,1) \Leftrightarrow \mu^+_{Y,X}(u) \ge 0$ for all $u \in (0,1) \Leftrightarrow E(Y|X \le x) \ge E(Y)$ for all x.

Similarly, $C_{\rm cor}[(X,Y)](u) \ge u$ for all $u \in (0,1) \Leftrightarrow \mu^-_{Y,X}(u) \le 0$ for all $u \in (0,1) \Leftrightarrow E(Y|X \le x) \ge E(Y)$ for all x.

(ii) See 3.1.1 and (5.1.1).

(iii) The first two equalities follow from Theorem 5.1.1(iii)–(iv). Further, in view of (5.3.1), $C_{\text{cor}}[(X,Y)] \equiv L_Y \Leftrightarrow \mu^+_{Y,X} \equiv 1 \Leftrightarrow$ there exists a nondecreasing function g such that $Y = g(X)$.

Similarly, $C_{\text{cor}}[(X,Y)] \equiv L^-_Y \Leftrightarrow \mu^-_{Y,X} \equiv -1 \Leftrightarrow$ there exists a nonincreasing function g such that $y = g(X)$.

(iv)–(vi) These follow immediately from (5.3.1) and from the respective properties of the monotone dependence function (Kowalczyk (1977), p. 354).

COROLLARY 5.3.1. (i) *If* $P(Y \le y \mid X = x_1) \ge$ (*resp.* $\le$) $P(Y \le y \mid X = x_2)$ *for all* $x_1 < x_2$, *then*

$$C_{\text{cor}}[(X,Y)] = L[P_X, \lambda^{r(X)}_{P_X}] \quad (\textit{resp.}\ = L^-[P_X, \lambda^{r(X)}_{P_X}]).$$

(ii) *Let* (X^+, Y^+) (*resp.* (X^-, Y^-)) *be the pair distributed according to the upper* (*lower*) *Fréchet bound of* (X,Y), *and let* r^+ (*resp.* r^-) *be the regression function of* Y^+ (*resp.* Y^-) *on* X^+ (*resp.* X^-). *Then*

$$\begin{aligned} L_{Y^+} \le L_{r^+(X^+)} = C_{\text{cor}}[(X^+,Y^+)] \le C_{\text{cor}}[(X,Y)] \\ \le C_{\text{cor}}[(X^-,Y^-)] = L^-_{r^-(X^-)} \le L^-_{Y^-}. \end{aligned}$$

(iii) $L[P_X, \lambda^{r(X)}_{P_X}](u) \equiv C_{\text{cor}}[(X,Y)](u) \equiv L^-[P_X, \lambda^{r(X)}_{P_X}](u) \equiv u$ *iff*

$$E(Y|X = x) = E(Y) \quad \textit{for all } x.$$

Proof. (i) The first (resp. second) inequality implies that r is nondecreasing (resp. nonincreasing); then the equality of C and L (resp. L^-) follows from Theorem 5.3.1.

(ii) follows from the inequalities

$$E(Y^+ \mid X^+ \le x) \le E(Y \mid X \le x) \le E(Y^- \mid X^- \le x) \tag{5.3.3}$$

and from the fact that r^+ is nondecreasing while r^- is nonincreasing.

THEOREM 5.3.2. *For the ratio curve* $C_{\text{ratio}}[(X,Y)]$ *we have*:

(i) $C_{\text{ratio}}(u) = C_{\text{cor}}(L_X^{-1}(u))$.

(ii) *If* $r(x)/x$ *is nondecreasing*, *then*

$$C_{\text{ratio}}[(X,Y)] = L[\lambda^X_{P_X}, \lambda^{r(X)}_{P_X}].$$

(iii) *If* $r(x)/x$ *is nonincreasing*, *then*

$$C_{\text{ratio}}[(X,Y)] = L^-[\lambda^X_{P_X}, \lambda^{r(X)}_{P_X}].$$

(iv) *In the set of pairs of positive random variables* (X,Y) *with fixed marginals*, *the ordering based on ratio curves is equivalent to the ordering of monotone dependence*:

$$C_{\text{ratio}}[(X,Y)] \le C_{\text{ratio}}[(X',Y')] \quad \textit{iff} \quad E(Y \mid X \le x) \le E(Y' \mid X' \le x).$$

So, if (X^+,Y^+), (X^-,Y^-) *are random variables distributed according to upper and lower Fréchet distributions, respectively, then*

$$C_{\text{ratio}}[(X^+,Y^+)] \le C_{\text{ratio}}[(X,Y)] \le C_{\text{ratio}}[(X^-,Y^-)].$$

(v) $C_{\text{ratio}}(t) = t$ *iff* $E(Y|X=x) = ax$ *for some* $a>0$. *In particular, for* $Y = aX$ *the ratio curve lies on the* 45° *line.*

(vi) *If* $E(Y|X=x) = E(Y)$ *then the respective ratio curve is isometric to the Lorenz curve* L_X:

$$C_{\text{ratio}}(u) = L_X^{-1}(u) \quad \textit{for } u \in (0,1).$$

Proof. (i) follows immediately from the definitions of these curves.

(ii)–(iii) follow from Theorem 5.1.1(iii)–(iv) and from the equality

$$\frac{d\lambda_{P_X}^{r(X)}}{d\lambda_{P_X}^{X}}(x) = \frac{r(x)}{xE(Y)}$$

for x such that $f_X(x) > 0$.

(iv) The first part of (iv) follows immediately from the comparison of $C_{\text{ratio}}[(X,Y)](u)$ with $C_{\text{ratio}}[(X',Y')](u)$. The second part follows from the inequalities (5.3.3).

(v) $C_{\text{ratio}}(u) = u$ iff $\int_0^t (E(X)r(x) - E(Y)x) f_X(x)\,\nu_X(dx) = 0$ for every $t \in \mathbb{R}^+$ iff $r(x) = xE(Y)/E(X)$.

(vi) follows from Corollary 5.3.1(iii) and the property (i) of the theorem.

5.4. Directed departure from proportionality. In Chapter 4 we investigated the problem of proportionality of elected representation. The measures introduced there referred to absolute departure from proportionality, with no regard to the trend of over- or under-representation according to the size of the electorate. More precisely, our procedures were invariant with respect to any ordering of electorates. Now, we will measure directed departure from proportionality, involving electorate sizes. This approach was investigated in Ciok *et al.* (1992) and in Bondarczuk *et al.* (1994). We retain the notation of Sec. 4.1.

Let $\Omega = \{\omega_1,\ldots,\omega_s\}$ where $\omega_1 \prec \ldots \prec \omega_s$. For a given vector indicating the sizes of electorates,

$$x = (x_1,\ldots,x_s) = (X(\omega_1),\ldots,X(\omega_s)), \qquad \sum_i x_i = n,$$

and given the representative size $m \le n$, we introduce in $\mathbf{Y}(x,m)$ an ordering $\preceq_{\mathrm{C}}^{x}$ which corresponds to "left departure" from proportionality with respect to x.

DEFINITION 5.4.1. For $y, y' \in \mathbf{Y}(x,m)$ we say that y *exhibits more left departure from proportionality* to x than y' (written $y \preceq_{\mathrm{C}}^{x} y'$) if

$$C^{\varphi}_{(\lambda_X,\lambda_Y)} \ge C^{\varphi}_{(\lambda_X,\lambda_{Y'})}$$

for $\varphi(\omega_i) = i$.

The ordering $\preceq_{\mathrm{C}}^{x}$ is equivalent to the stochastic ordering of the measures λ_Y, $\lambda_{Y'}$, where

$$\lambda_Y \le_{\mathrm{st}} \lambda_{Y'} \quad \text{iff} \quad \sum_{i=1}^{r} y_i \ge \sum_{i=1}^{r} y'_i \quad \text{for } r = 1, \ldots, s. \tag{5.4.1}$$

This formalizes the intuitive notion of y being "left" to y'.

If $x_1 \le \ldots \le x_s$ then "left departure" from proportionality means that y tends to overrepresent small electorates. In $\mathbf{Y}(x, m)$ there exist smallest and largest vectors for the ordering $\preceq_{\mathrm{C}}^{x}$, i.e. vectors whose components are maximally transformed to the left and to the right, respectively. The smallest vector $y^{\min}$ has components

$$y_i^{\min} = \min\Big(m - \sum_{j=1}^{i-1} y_j^{\min}, x_i\Big), \qquad i = 1, \ldots, s, \qquad \sum_{1}^{0} = 0, \tag{5.4.2}$$

and the largest vector $y^{\max}$ has components

$$y_{s-i}^{\max} = \min\Big(m - \sum_{j=s-i+1}^{s} y_j^{\max}, x_{s-i}\Big), \qquad i = 0, \ldots, s-1, \qquad \sum_{s+1}^{s} = 0. \tag{5.4.3}$$

The curve $C^{\varphi}_{(\lambda_X, \lambda_{Y^{\min}})}$ is concave, and

$$C^{\varphi}_{(\lambda_X, \lambda_{Y^{\min}})}(t) = L^{-}_{(\lambda_X, \lambda_{Y^{\min}})}(t) \quad \text{for } t \in [0, 1].$$

On the other hand, $C^{\varphi}_{(\lambda_X, \lambda_{Y^{\max}})}$ is convex, and

$$C^{\varphi}_{(\lambda_X, \lambda_{Y^{\max}})}(t) = L_{(\lambda_X, \lambda_{Y^{\max}})}(t) \quad \text{for } t \in [0, 1].$$

Thus, for any $y \in \mathbf{Y}(x, m)$,

$$C^{\varphi}_{(\lambda_X, \lambda_{Y^{\max}})}(t) \le C^{\varphi}_{(\lambda_X, \lambda_Y)}(t) \le C^{\varphi}_{(\lambda_X, \lambda_{Y^{\min}})}(t) \quad \text{for } t \in [0, 1].$$

Ideal proportionality is represented by the vector

$$y = \Big(\frac{m}{n} x_1, \ldots, \frac{m}{n} x_s\Big)$$

whose components are positive integers; in this case we have

$$C^{\varphi}_{(\lambda_X, \lambda_Y)}(t) = t \quad \text{for } t \in [0, 1].$$

Now we will investigate once more the methods of Adams, Dean, Hill, Webster and Jefferson, mentioned previously in Sec. 4.2. They were presented in Baliński and Young (1982) as the so-called divisor methods. Each divisor method corresponds to a real-valued strictly increasing function d defined on the set $\mathbb{N}$ of positive integers such that for any $a \in \mathbb{N}$, $a \le d(a) \le a + 1$. The function d describes a particular way of rounding any nonnegative number z up or down to the neighbouring integer. This result of rounding, denoted by $[z]_d$, is an integer a satisfying $d(a-1) \le z \le d(a)$; it is unique unless $z = d(a)$, in which case it takes on either of the values a or $a+1$. We say that a vector y *is obtained from* x *and* m *by a divisor method corresponding to* d if

$$y_i = \left[\frac{m}{n} x_i\right]_d \quad \text{for } i = 1, \ldots, s, \quad \text{and} \quad \sum_{i=1}^{s} y_i = m.$$

The divisor methods considered in Baliński and Young (1982) are:

$$\begin{aligned} &\text{Adam's rule:} && d^{\mathrm{A}}(a) = a, \\ &\text{Dean's rule:} && d^{\mathrm{D}}(a) = \frac{a(a+1)}{a+1/2}, \\ &\text{Hill's rule:} && d^{\mathrm{H}}(a) = \sqrt{a(a+1)}, \\ &\text{Webster's rule:} && d^{\mathrm{W}}(a) = a + \frac{1}{2}, \\ &\text{Jefferson's rule:} && d^{\mathrm{J}}(a) = a + 1. \end{aligned}$$

Baliński and Young introduced an ordering, say $\preceq_{\mathrm{BY}}$, according to which a method M' is more left-biased than a method M ($M' \preceq_{\mathrm{BY}} M$) if for any fixed s, x, m, and any pair ($y = (y_1, \ldots, y_s)$, $y' = (y'_1, \ldots, y'_s)$) such that y is obtained under M and y' under M', the implication

$$i < j, x_1 \le \ldots \le x_s \Rightarrow y'_i > y_i \text{ or } y'_j \le y_j \tag{5.4.4}$$

holds for any $i, j \in \{1, \ldots, s\}$. They proved that the five methods of apportionment (Adams's through Jefferson's) are ordered according to $\preceq_{\mathrm{BY}}$:

$$M^{\mathrm{A}} \preceq_{\mathrm{BY}} M^{\mathrm{D}} \preceq_{\mathrm{BY}} M^{\mathrm{H}} \preceq_{\mathrm{BY}} M^{\mathrm{W}} \preceq_{\mathrm{BY}} M^{\mathrm{J}}.$$

This is due to the fact that (see Baliński and Young (1982))

$$\frac{d^{\mathrm{A}}(a)}{d^{\mathrm{A}}(b)} > \frac{d^{\mathrm{D}}(a)}{d^{\mathrm{D}}(b)} > \frac{d^{\mathrm{H}}(a)}{d^{\mathrm{H}}(b)} > \frac{d^{\mathrm{W}}(a)}{d^{\mathrm{W}}(b)} > \frac{d^{\mathrm{J}}(a)}{d^{\mathrm{J}}(b)}$$

for all integers $a > b \ge 0$.

It is worth noting that (5.4.4) is a very strong formalization of the intuitive notion of the vector y' being "left" to the vector y. In particular, it is stronger than (5.4.1), since

$$y' \preceq_{\mathrm{BY}} y \Rightarrow y' \preceq_{\mathrm{C}}^{x} y$$

and the two orderings are not equivalent (cf. Ciok *et al.* (1992)).

6. Numerical measures relating to divergence

6.1. Numerical inequality measures. We start with a short reference to the finite population case considered in Sec. 2.1 and, in particular, to the axioms concerning measures of inequality on $D = \bigcup_{n=1}^{\infty} D_n$, where

$$D_n = \Big\{x \in \mathbb{R}^n : x_i \ge 0,\ \sum_{i=1}^{n} x_i > 0\Big\}.$$

An index I defined on D with values in any set of comparison points ordered by some binary relation $\preceq$ (which is reflexive, transitive and antisymmetric) is said to be *consistent* with these axioms if

(i) $I(x) = I(ax)$ for every $a > 0$,
(ii) $I(x_1, \dots, x_n) = I(x_{i_1}, \dots, x_{i_n})$,
(iii) $x \preceq y$, $x \not\equiv y \Rightarrow I(x) < I(y)$,
(iv) $I(z_{(1)}, \dots, z_{(m)}) = I(x_1, \dots, x_n)$ where $z_{(i)} = (x_1, \dots, x_n)$ for $i = 1, \dots, m$.

Axioms (i)–(iv) are satisfied by the Lorenz curve, which is a function-valued measure of inequality. Foster (1985) proved that a parameter I satisfies (i)–(iv) if and only if I is *Lorenz consistent*, i.e. for all $x, y \in D$,

$$L_X \ge L_Y, L_X \ne L_Y \Rightarrow I(x) < I(y), \quad \text{and} \quad L_X = L_Y \Rightarrow I(x) = I(y), \tag{6.1.1}$$

where X and Y are discrete random variables corresponding to equally probable outcomes for each object in the population.

Moreover, I satisfies (i)–(iii) if and only if I is Lorenz consistent on each D_n (this refers to the case when the population size is known).

Now we turn back to the general notation, suitable both for finite and infinite populations. Evidently, the Lorenz consistency property (6.1.1) can be easily generalized. This property holds for the well-known numerical measure of inequality, called the Gini inequality index. It is defined as twice the area between the Lorenz curve and the diagonal:

$$G(X) = 2 \int_0^1 (t - L_X(t))\, dt.$$

Among many possible formulas for $G(X)$, we mention the following:

$$G(X) = 2 \int_0^1 \int_0^1 |t - t'|\, dP_X(t)\, dP_X(t').$$

Assume that X is continuous with distribution function F, and put $X' = F(X^{(1)})$, where $X^{(1)}$ is the random variable with distribution $F^{(1)}$ defined by (2.2.3) in Sec. 2.2. Then

$$G(X) = \frac{2}{E(X)} \operatorname{cov}(X, F(X)) = \frac{2}{E(X')} \operatorname{cov}(X', L_X(X')). \tag{6.1.2}$$

The first equality is proved e.g. in Lerman and Yitzaki (1984) and the second is a particular case of the first for the random variable X' with distribution function L_X. The random variable $X' = F(X^{(1)})$ was shown in Sec. 5.2 to be the grade transformation of $X^{(1)}$ with respect to X for a continuous random variable X, and it was denoted there by $X_X^{(1)}$. Now, let us drop the assumption of continuity and let $X' = X_X^{(1)}$. Then it is of course still true that

$$G(X) = \frac{2}{E(X')} \operatorname{cov}(X', L_X(X')).$$

It follows from the first part of (6.1.2) that for $Y \ll X$,

$$G(Y_X) = \frac{2}{E(Y_X)} \operatorname{cov}(Y_X, C_{(P,Q)}(Y_X)).$$

Another numerical measure which is widely used although it is not Lorenz consistent is the Pietra inequality index:

$$D(X) = \max_{t\in[0,1]} (t - L_X(t)) = \frac{1}{2E(X)} \int_0^1 |t - EX| \, dP_X(t).$$

This index satisfies a condition weaker than (6.1.1), namely

$$L_X \geq L_Y \Rightarrow D(X) \leq D(Y).$$

Generally, in view of (2.3.1), inequality indices of the form $I(X) = E(\Phi(X))$ for some function Φ which is convex and continuous on $[0, \infty)$ satisfy this weaker condition:

$$X \preceq_{\mathrm{L}} Y \Rightarrow I(X) \leq I(Y).$$

For the Gini and Pietra indices, the functions Φ are

$$\Phi_{\mathrm{Gini}}(t) = \frac{1}{4} \int |t - t'| \, dP_X(t'), \qquad \Phi_{\mathrm{Pietra}}(t) = |t - 1|.$$

6.2. Numerical measures of divergence. This section concerns divergence considered in Chapter 1. Indices concerning directed divergence introduced in Chapter 5 will be dealt with in the next section. To distinguish between the two types of indices, those appearing in this section will be referred to as measures of absolute divergence.

In the case of absolute divergence, the counterpart of the Lorenz consistency property (6.1.1) is strict monotonicity with respect to the ordering $\preceq_{\mathrm{NP}}$. An index $\tau : \mathcal{P} \times \mathcal{Q} \to \mathbb{R}^+$ is said to be *strictly consistent with respect to* $\preceq_{\mathrm{NP}}$ if

$$L_{(P,Q)} \geq L_{(P',Q')}, \ L_{(P,Q)} \neq L_{(P',Q')} \Rightarrow \tau(P, Q) < \tau(P', Q').$$

Similarly, the weak Lorenz consistency of τ is equivalent to monotonicity with respect to $\preceq_{\mathrm{NP}}$:

$$L_{(P,Q)} \geq L_{(P',Q')} \Rightarrow \tau(P, Q) \leq \tau(P', Q').$$

We can exploit the links between divergence and inequality, presented in Sec. 2.4. Let $Q \ll P$, $Z \sim P$, $h = dQ/dP$, and let $E(\Phi(U))$ be an inequality index for $U = h(Z)$. Then the index τ given by

$$\tau(P, Q) = E\Phi(h(Z))$$

is an index of divergence of Q from P.

Generally (omitting the assumption $Q \ll P$), the divergence indices of the form

$$\tau(P,Q) = E^*\Phi(h(Z)) \tag{6.2.1}$$

are monotone with respect to $\preceq_{\mathrm{NP}}$. The *Gini divergence index*

$$G(P,Q) = \frac{1}{2} \int \int |h(t) - h(t')|\, dP(t)\, dP(t') + Q(h = \infty)$$

is strictly monotone with respect to $\preceq_{\mathrm{NP}}$. The *Pietra divergence index*

$$D(P,Q) = \frac{1}{2} \int |h(t) - 1|\, dP(t) + Q(h = \infty)$$

is monotone with respect to $\preceq_{\mathrm{NP}}$ but not strictly monotone.

Divergence measures can be represented in many ways. For discrete distributions on $\{1,\ldots,s\}$ defined by the vectors $p = (p_1,\ldots,p_s)$ and $q = (q_1,\ldots,q_s)$ the Gini divergence index (which will be denoted by $G(p,q)$) is given by

$$G(p,q) = \frac{1}{2}\sum_{i,j} |p_i q_j - q_i p_j| = 1 - \sum_{j=1}^{s} q_{i(j)}\Big(p_{i(j)} + 2\sum_{r=j+1}^{s} p_{i(r)}\Big), \tag{6.2.2}$$

where $(i(1),\ldots,i(s))$ is a permutation of $(1,\ldots,s)$ such that

$$\frac{q_{i(1)}}{p_{i(1)}} \le \ldots \le \frac{q_{i(s)}}{p_{i(s)}}.$$

The right-hand side of (6.2.2) relates to the geometrical interpretation of the Gini index as twice the area between the Lorenz curve and the 45° line. Many other equivalent forms are also in use.

The Pietra divergence index is given by

$$D(p,q) = \frac{1}{2}\sum_{i=1}^{s} |p_i - q_i|.$$

Examples of other divergence indices which are monotone with respect to $\preceq_{\mathrm{NP}}$ but are not of the form (6.2.1) are given in Ali and Silvey (1966), for example

$$\tau(P,Q) = 1 - \alpha^*,$$

where $\alpha^* = a_{12}(\delta_{\kappa,s})$ (see Sec. 1.1) for κ, s such that $a_{12}(\delta_{\kappa,s}) = a_{21}(\delta_{\kappa,s})$.

6.3. Numerical measures of directed divergence. In this section we consider numerical measures of divergence of Q from P directed according to φ, which are connected with the φ-directed concentration curve $C^{\varphi}_{(P,Q)}$ (and with the ordering $\preceq_{\mathrm{C}}$), just as the respective numerical measures of absolute divergence are connected with the curve $L_{(P,Q)}$ (and with the ordering $\preceq_{\mathrm{NP}}$). Let $\tau(P,Q)$ be a numerical measure of absolute divergence (e.g., $G(P,Q)$ or $D(P,Q)$) and let $\tau^{\varphi}(P,Q)$ be the corresponding numerical measure of directed divergence. It should take values in $[-1,1]$ and have the following properties:

(i) $\tau^\varphi(P,Q)$ is monotone with respect to $\preceq_{\mathrm{C}}$, i.e.

$$(P,Q) \preceq_{\mathrm{C}} (P',Q') \Rightarrow \tau^\varphi(P,Q) \le \tau^\varphi(P',Q'),$$

(ii) $\tau^\varphi(P,Q) = \tau(P,Q)$ iff the ratio $\frac{dQ}{dP}$ is nondecreasing with respect to $\preceq_\varphi$,
(iii) $\tau^\varphi(P,Q) = -\tau(P,Q)$ iff the ratio $\frac{dQ}{dP}$ is nonincreasing with respect to $\preceq_\varphi$,
(iv) $\tau^\varphi(P,Q) = +1$ iff Q is completely right to P,
(v) $\tau^\varphi(P,Q) = -1$ iff Q is completely left to P,
(vi) $\tau^\varphi(P,Q) = 0$ if $P = Q$.

Properties (i)–(vi) indicate that $\tau^\varphi(P,Q)$ would be a valuable supplement to $\tau(P,Q)$. The pair of indices $(\tau^\varphi(P,Q), \tau(P,Q))$ serves to evaluate how strongly Q differs from P and to what extent this departure is explained by the fact that Q is "right" ("left") to P.

The *directed Gini divergence index* $G^\varphi(P,Q)$ is defined as twice the difference between the two areas: the first one between the 45° line and the part of the curve C^φ which lies above it, and the second between the 45° line and the remaining part of C^φ:

$$G^\varphi(P,Q) = 2\int_0^1 (t - C^\varphi_{(P,Q)}(t))\,dt.$$

It satisfies postulates (i)–(vi).

It seems that the directed version of the Pietra divergence index has not been introduced yet. We propose the following definition, formulated first under the assumption that P and Q are defined on $\Omega = \mathbb{R}^+$, have densities p, q with respect to the Lebesgue measure, and $\varphi(x) = x$ for $x \in \mathbb{R}^+$.

Let F, H be the distributions induced by the measures P, Q, respectively. The Pietra index $D(P,Q)$ can be represented in this case as

$$\begin{aligned} D(P,Q) = D(p,q) &= \frac{1}{2}\int_{\mathbb{R}^+} |p(t) - q(t)|\,dt \\ &= \frac{1}{2}\Big(\int_{\{t:H(t)\ge F(t)\}} |p(t) - q(t)|\,dt + \int_{\{t:H(t)\le F(t)\}} |p(t) - q(t)|\,dt\Big). \end{aligned}$$

The first term of the right-hand side represents the mass of probability in the distribution Q which should be transferred from left to right in order to transform Q onto P. The second term has an analogous interpretation for the transfer from right to left. Thus, the *directed Pietra index* is defined to be

$$D_{\mathrm{d}}(P,Q) = \frac{1}{2}\int_{\{t:H(t)\ge F(t)\}} |p(t) - q(t)|\,dt - \frac{1}{2}\int_{\{t:H(t)\le F(t)\}} |p(t) - q(t)|\,dt. \tag{6.3.1}$$

Since the curve $C_{(P,Q)}$ is the set

$$\{(F(t), H(t)) : t \in (0,\infty)\},$$

the sets $\{t : H(t) \geq F(t)\}$ and $\{t : H(t) \leq F(t)\}$ are of course determined by the segments of $C_{(P,Q)}$ lying above the diagonal $y = x$ and below it, respectively. If $C_{(P,Q)}$ lies below (resp. above) the diagonal then $D_{\mathrm{d}}(P,Q) = D(P,Q)$ (resp. $= -D(P,Q)$).

We now turn to the case when P and Q are atomic, defined on $\{1,\ldots,s\}$ by $p = (p_1,\ldots,p_s)$, $q = (q_1,\ldots,q_s)$. Let

$$F(i) = \sum_{j=1}^{i} p_j, \quad H(i) = \sum_{j=1}^{i} q_j, \quad \text{for } i = 0,\ldots,s, \quad \sum_{1}^{0} = 0,$$

and let $D_{\mathrm{L}}(P,Q)$ and $D_{\mathrm{R}}(P,Q)$ be the (discrete) counterparts of the "left" and "right" components of the sum (6.3.1) with the same interpretation:

$$\begin{aligned} D_{\mathrm{L}}(P,Q) = \frac{1}{2}\sum_{i=1}^{s} &|p_i - q_i|\psi_{[H(i-1)\geq F(i-1),H(i)\geq F(i)]} \\ &+ \sum_{i=1}^{s}(H(i-1) - F(i-1))\psi_{[H(i-1)>F(i-1),H(i)<F(i)]} \\ &+ \sum_{i=1}^{s}(H(i) - F(i))\psi_{[H(i-1)<F(i-1),H(i)>F(i)]}, \\ D_{\mathrm{R}}(P,Q) = \frac{1}{2}\sum_{i=1}^{s} &|p_i - q_i|\psi_{[H(i-1)\leq F(i-1),H(i)\leq F(i)]} \\ &+ \sum_{i=1}^{s}(F(i) - H(i))\psi_{[H(i-1)>F(i-1),H(i)<F(i)]} \\ &+ \sum_{i=1}^{s}(F(i-1) - H(i-1))\psi_{[H(i-1)<F(i-1),H(i)>F(i)]}. \end{aligned}$$

where $\psi_{[\cdot]}$ is the indicator function.

The indices D_{L} and D_{R} are formed as follows: if $H(i-1) \geq F(i-1)$, $H(i) \geq F(i)$ (i.e. the points $(F(i-1),H(i-1))$, $(F(i),H(i))$ lie above the diagonal $y = x$), then the difference $|p_i - q_i|$ enters $D_{\mathrm{L}}(P,Q)$; if these points lie below the diagonal, then $|p_i - q_i|$ enters $D_{\mathrm{R}}(P,Q)$; if $F(i-1) < H(i-1)$, $F(i) > H(i)$ (i.e. $(F(i-1),H(i-1))$ lies above the diagonal, and $(F(i),H(i))$ lies below it) then $|p_i - q_i| = (H(i-1) - F(i-1)) + (F(i) - H(i))$; $(H(i-1) - F(i-1))$ enters $D_{\mathrm{L}}(P,Q)$ while $(F(i) - H(i))$ enters $D_{\mathrm{R}}(P,Q)$. This justifies the definition

$$D^{\varphi}(P,Q) = D_{\mathrm{R}}(P,Q) - D_{\mathrm{L}}(P,Q).$$

This index has values in $[-1,1]$ and satisfies (i)–(v).

6.4. Numerical measures of dependence. We concentrate here on the Gini index of monotone dependence based on the correlation curve (cf. Sec. 5.3). According to Sec. 5.3, the correlation curve of Y with respect to X is the directed divergence curve for the pair of distributions $(P_X, \lambda_{P_X}^{r(X)})$, where r is the regression function of Y on X. Then the directed Gini index for this pair of distributions is a measure of monotone dependence of Y on X. But in this case the situation is slightly different from that described in Sec. 6.3. As said in Sec. 5.3, all correlation curves of Y on X appear in the area bounded by two curves: L_Y and L_Y^-. This means that measures of dependence must be connected with measures of inequality of Y.

It follows from Theorem 5.3.1(iii) that for $\varphi(x,y) = x$ we have

$$-G(Y) \le -G(r(X)) \le G^{\varphi}(P_X, \lambda_{P_X}^{r(X)}) \le G(r(X)) \le G(Y).$$

Therefore we define the *correlation curve Gini index* as

$$\gamma(X,Y) = \frac{G^{\varphi}(P_X, \lambda_{P_X}^{r(X)})}{G(Y)}.$$

PROPOSITION 6.4.1. *The correlation curve Gini index γ has the following properties*:

(i) $-1 \le \gamma(X,Y) \le 1$.

(ii) $\gamma(X,Y) = +1$ *iff there exists a nondecreasing function g such that* $Y = g(X)$.

(iii) $\gamma(X,Y) = -1$ *iff there exists a nonincreasing function g such that* $Y = g(X)$.

(iv) $\gamma(X,Y) = 0$ *if* $E(Y \mid X = x) = E(Y)$.

(v) *Under the assumptions of Theorem* 5.3.1(iv),

$$\gamma(X,Y) = \varrho \quad \text{if} \quad m^+(x) = \varrho x + (1-\varrho)E(Y) \quad \text{and} \quad \varrho \in [0,1] \quad \text{or}$$
$$\text{if} \quad m^-(x) = \varrho x + (1+\varrho)E(Y) \quad \text{and} \quad \varrho \in [-1,0].$$

(vi) *Let* (X,Y), (X',Y') *satisfy* $X' \sim X$, $Y' \sim Y$. *Then*

$$E(Y \mid X \le x) \ge E(Y' \mid X' \le x) \Rightarrow \gamma(X,Y) \le \gamma(X',Y').$$

(vii) *We have*

$$\gamma(X,Y) = \begin{cases} \dfrac{G(r(X))}{G(Y)} & \text{iff } r \text{ is nondecreasing}, \\ -\dfrac{G(r(X))}{G(Y)} & \text{iff } r \text{ is nonincreasing}. \end{cases}$$

(viii) *If X, Y are continuous random variables then*

$$\gamma(X,Y) = \frac{\operatorname{cov}(Y, F_X(X))}{\operatorname{cov}(Y, F_Y(Y))}.$$

Proof of (viii). We compute

$$G^{\varphi}(P_X,\lambda_{P_X}^{r(X)}) = 1-2\int_0^\infty \lambda_{P_X}^{r(X)}(u)\,dP_X(u) = -1+2\int_0^\infty P_X(u)\,d\lambda_{P_X}^{(r(X)}(u)$$

$$= -1+\frac{2}{E(Y)}\int_0^\infty F_X(u)r(u)f_X(u)\,du$$

$$= \frac{2}{E(Y)}\operatorname{cov}(F_X(X),r(X)) = \frac{2}{E(Y)}\operatorname{cov}(F_X(X),Y).$$

We note that Schechtman and Yitzhaki (1987) considered the quantity

$$\Gamma(X,Y) = \frac{\operatorname{cov}(Y,F_X(X))}{\operatorname{cov}(Y,F_Y(Y))}$$

as a measure of association between two random variables with a continuous bivariate distribution $F_{X,Y}$.

6.5. Numerical measures of departures from proportional representation. In this section we will consider once more the problem of proportional representation using the notation introduced in Sec. 4.1. Election data are there described by vectors $x=(x_1,\ldots,x_s)$ and $y=(y_1,\ldots,y_s)$ where $\sum_i x_i = n$, $\sum_i y_i = m$, $y_i \le x_i$, x_i, y_i are positive integers. Let

$$p_i = \frac{x_i}{n},\quad q_i = \frac{y_i}{m},\qquad i=1,\ldots,s.$$

According to formula (6.2.2) the Gini divergence index adjusted to proportional representation data is

$$G(x,y) = 1-\frac{1}{nm}\sum_{j=1}^{s} y_{i(j)}\Big(x_{i(j)}+2\sum_{r=j+1}^{s} x_{i(r)}\Big),$$

where $(i(1),\ldots,i(s))$ is a permutation of $(1,\ldots,s)$ such that

$$\frac{y_{i(1)}}{x_{i(1)}} \le \ldots \le \frac{y_{i(s)}}{x_{i(s)}}.$$

The notation $G(x,y)$ is used instead of $G(p,q)$ since in this way we retain the information on n and m.

Analogously, for the Pietra divergence index we have

$$D(x,y) = \frac{1}{2}\sum_{i=1}^{s}\left|\frac{y_i}{m}-\frac{x_i}{n}\right|.$$

Both indices take values in the interval $[0,1)$. They are 0 if and only if the proportionality is ideal: $y_i = \frac{m}{n}x_i$. But in general $\frac{m}{n}x_i$ are not integers, and therefore G and D take values in a subset of $(0,1)$.

It would be interesting to find, for any given (x,m) and for a chosen numerical divergence index, the minimal (maximal) elements for the ordering $\preceq_x$ (defined in Sec. 4.2 (4.3)) for which this index attains its smallest (largest) possible value. We are only able to do this in the case of the Pietra divergence index $D(x,\cdot)$.

It follows from Sec. 4.2 that the smallest value of $D(x,\cdot)$ is only taken on at *each* vector $y \in \mathbf{Y}(x,m)$ which is obtained by the Hamilton rule (formula 4.2.2). On the other hand, we will show below that the largest value of $D(x,\cdot)$ is taken on at some (not necessarily all) elements of $\mathbf{Y}_1(x,m)$. According to Sec. 4.3, this set contains all maximal elements for $\preceq_x$. We do not exclude the possibility that the largest value of $D(x,\cdot)$ is also attained at some $y \notin \mathbf{Y}_1(x,m)$.

Let $y^1 \in \mathbf{Y}_1(x,m)$ be the maximal element for $\preceq_x$. If for every $j=1,\ldots,s$, $y^1_j = 0$ or $y^1_j = x_j$ then of course y^1 is the largest element, and

$$D(x,y^1) = 1 - \frac{m}{n}.$$

Assume that there exists $i_0 \in \{1,\ldots,s\}$ such that the components of y^1 satisfy

$$y^1_{i_0} = m_0 < x_{i_0}, \quad y^1_j = 0 \quad \text{or} \quad y^1_j = x_j \quad \text{for } j \neq i_0.$$

It is easy to show that in this case

$$D(x,y^1) = 1 - \frac{m}{n} - \min\left(\frac{m_0}{m} - \frac{m_0}{n}, \frac{x_{i_0}-m_0}{n}\right).$$

Let

$$y^1_{\max} = \arg\min_{y\in\mathbf{Y}_1(x,m)} \left(\min\left(\frac{m_0}{m} - \frac{m_0}{n}, \frac{x_{i_0}-m_0}{n}\right)\right),$$

$$D_{\mathrm{c}}(x,y) = \frac{D(x,y) - D(x,y^0)}{D(x,y^1_{\max}) - D(x,y^0)}.$$

The index D_{c} has the following properties:

(i) $0 \le D_{\mathrm{c}}(x,y) \le 1$,
(ii) $y \preceq_{\mathrm{C}} y' \Rightarrow D_{\mathrm{c}}(x,y) \le D_{\mathrm{c}}(x,y')$,
(iii) $D_{\mathrm{c}}(x,y) = 0$ iff $y = y^0$,
(iv) $y = y^1_{\max} \Rightarrow D_{\mathrm{c}}(x,y^1_{\max}) = 1$.

We have not been able to obtain similar results for $G(x,\cdot)$. However, it is known (Duncan and Duncan (1955)) that

$$1 - \sqrt{1 - G(x,y)} \le D(x,y) \le G(x,y).$$

But the smallest value of $D(x,\cdot)$ is achieved for any y^0 obtained under the Hamilton rule. Thus for any y,

$$G(x,y) \ge D(x,y^0) = \frac{1}{2m}\Big(\sum_{r=1}^{l}(1-u_{i_r}) + \sum_{r=l+1}^{s} u_{i_r}\Big),$$

where

$$u_i = \frac{m}{n}x_i - \left[\frac{m}{n}x_i\right], \quad i=1,\ldots,s, \quad l = \sum_{i=1}^{s} u_i$$

and $(i_1, \ldots, i_s)$ is a permutation of $(1, \ldots, s)$ such that

$$u_{i_1} \geq \ldots \geq u_{i_s}.$$

Example (see Ciok *et al.* (1992)). Let $x = (10, 25, 35, 50, 65, 115)$, $m = 30$. The following vectors are obtained by the Hamilton rule:

$$\begin{aligned} y^{(1)} &= (1, 3, 4, 5, 6, 11), \\ y^{(2)} &= (1, 3, 3, 5, 7, 11), \\ y^{(3)} &= (1, 3, 3, 5, 6, 12), \\ y^{(4)} &= (1, 2, 4, 5, 6, 12), \\ y^{(5)} &= (1, 2, 4, 5, 7, 11), \\ y^{(6)} &= (1, 2, 3, 5, 7, 12). \end{aligned}$$

We have

$$G(x, y^{(i)}) = \begin{cases} .0433 & \text{for } i = 1, 6, \\ .0467 & \text{for } i = 2, 3, 4, 5, \end{cases}$$

$$D(x, y^{(i)}) = .0333 \quad \text{for } i = 1, \ldots, 6.$$

It is worth noting that in the above example the vectors $y^{(i)}$ for $i = 1, \ldots, 6$ are solutions of the Webster method, $y^{(1)}$ is the solution of the Adams, Dean and Hill methods, and $y^{(6)}$ is the solution of the Jefferson method.

It remains to introduce the directed counterparts of the numerical measures of departures from proportional representation, considered in Sec. 6.3. In this case the function φ could be chosen according to the size of the x's, i.e. $\varphi(\omega_i) = x_i$; this would allow us to detect overrepresentation (underrepresentation) with respect to the size of electorates. This problem was tackled in Bondarczuk *et al.* (1994) where the directed version of $G(x, y)$ was introduced:

$$G^{\varphi}(x, y) = 1 - \frac{1}{nm} \Big(\sum_{i=1}^{s} x_i y_i + 2 \sum_{j=2}^{s} \sum_{i=1}^{j-1} y_i x_j \Big),$$

where $x_1 \leq \ldots \leq x_s$. Analogously we obtain $D^{\varphi}(x, y)$.

It is interesting to find the smallest and largest values of the directed numerical measure τ^{φ}. It takes the smallest value for the vector $y_{\min}$ with components given by (5.4.2), and the largest value for $y_{\max}$, with components given by (5.4.3).

Acknowledgements. I am grateful to Professor Elżbieta Pleszczyńska for her encouragement to write this paper.

References

S. M. Ali and S. D. Silvey (1965), *Association between random variables and the dispersion of a Radon–Nikodym derivative*, J. Roy. Statist. Soc. Ser. B, 27, 100–107.

S. M. Ali and S. D. Silvey (1965), *A further result on the relevance of the dispersion of a Radon–Nikodym derivative to the problem of measuring association*, ibid., 108–110.

—, — (1966), *A general class of coefficients of divergence of one distribution from another*, ibid., 28, 131–142.

B. C. Arnold (1987), *Majorization and the Lorenz Order: a Brief Introduction*, Lecture Notes in Statist. 43, Dekker.

M. Baliński and H. P. Young (1982), *Fair Representation*, Yale Univ. Press, New Haven.

R. C. Blitz and J. A. Brittain (1964), *An extension of the Lorenz diagram to the correlation of two variables*, Metron 23 (1964), 137–143.

H. Block, A. Sampson and T. Savits (eds.) (1990), *Topics in Statistical Dependence*, IMS Lecture Notes Monograph Ser., Inst. Math. Statist., Hayward.

Z. Bondarczuk, T. Kowalczyk, E. Pleszczyńska and W. Szczesny (1994), *Evaluating departures from fair representation*, Appl. Stochastic Models Data Anal., to appear.

T. Bromek, T. Kowalczyk and E. Pleszczyńska (1988), *Measurement scales in evaluation of stochastic dependence*, in: S. Das Gupta and J. K. Ghosh (eds.), Proc. Internat. Conf. on Advances in Multivariate Statistical Analysis, Indian Statistical Institute, Calcutta, 83–96.

T. Bromek and T. Kowalczyk (1990), *A decision approach to ordering stochastic dependence*, in: A. Sampson (ed.), Topics in Statistical Dependence, IMS Lecture Notes Monograph Ser., Inst. Math. Statist., Hayward, 103–109.

M. Chandra and N. D. Singpurwalla (1981), *Relationship between some notions which are common to reliability and economics*, Math. Oper. Res. 6, 113–121.

D. M. Cifarelli and E. Regazzini (1987), *On a general definition of concentration function*, Sankhyā Ser. B 49, 307–319.

A. Ciok, T. Kowalczyk, E. Pleszczyńska and W. Szczesny (1994), *Inequality measures in data analysis*, Archiwum Informatyki Teoretycznej i Stosowanej, to appear.

A. Ciok, T. Kowalczyk and W. Szczesny (1992), *Comparing methods of fair representation*, IPI PAN, preprint, 718.

O. D. Duncan and B. Duncan (1955), *A methodological analysis of segregation indexes*, Amer. Sociological Rev., 210–217.

J. Fellman (1976), *The effect of transformations on Lorenz curves*, Econometrica 44 (4), 823–824.

G. S. Fields and J. C. H. Fey (1978), *On inequality comparisons*, Econometrica 46, 303–316.

S. Fogelson (1933), *Miary koncentracji i ich zastosowania* [*Measures of concentration and their applications*], Kwart. Statyst. 10(1), 149–197.

J. E. Foster (1985), *Inequality measurement*, in: Proc. Sympos. Appl. Math. 33, 31–68.

V. Gafrikova and T. Kowalczyk (1994), *Links between measuring divergence and inequality*, Metron, to appear.

D. M. Grove (1980), *A test of independence against a class of ordered alternatives in a $2 \times C$ contingency table*, J. Amer. Statist. Assoc. 75, 454–459.

H. Joe (1985), *An ordering of dependence for contingency tables*, Linear Algebra Appl. 70, 89–103.

— (1987), *Majorization, randomness and dependence for multivariate distribution*, Ann. Probab. 15, 1217–1225.

— (1990), *Majorization and divergence*, J. Math. Anal. Appl. 148, 287–305.

B. Klefsjö (1984), *Reliability interpretations of some concepts from economics*, Naval Res. Logist. 31, 301–308.

T. Kowalczyk (1977), *General definition and sample counterparts of monotonic dependence functions of bivariate distributions*, Math. Oper. Statist. Ser. Statist. 8, 351–365.

— (1990), *On measuring heterogeneity in $m \times k$ contingency tables*, in: Proc. DIANA III, Conference of Discriminant Analysis and Other Methods of Data Classification, Bechyne, 111–121.

T. Kowalczyk and J. Mielniczuk (1990), *Neyman–Pearson curves, properties and estimation*, preprint 683, IPI PAN.

T. Kowalczyk and E. Pleszczyńska (1977), *Monotonic dependence functions of bivariate distributions*, Ann. Statist. 5, 1221–1227.

T. Kowalczyk, E. Pleszczyńska and W. Szczesny (1991), *Evaluation of stochastic dependence*, in: Statistical Inference: Theory and Practice, Theory Decis. Lib. Ser. B: Math. Statist. Methods 17, Reidel, 106–136.

E. L. Lehmann (1959), *Testing Statistical Hypotheses*, Wiley, New York.

— (1966), *Some concepts of dependence*, Ann. Math. Statist. 37, 1137–1153.

R. Lerman and S. Yitzaki (1984), *A note on the calculation and interpretation of the Gini index*, Econom. Lett. 15, 363–368.

C. E. Rao (1982), *Diversity and dissimilarity coefficients*: *a unified approach*, Theoret. Population Biol. 21, 24–43.

A. Raveh (1989), *Gini correlation as a measure of monotonicity and two of its usages*, Comm. Statist. Theory Methods 18 (4), 1415–1423.

E. Regazzini (1990), *Concentration comparisons between probability measures*, Instituto per le Applicazioni della Matematica e dell'Informatica, preprint 90.15, Milano.

M. Scarsini (1990), *An ordering of dependence*, in: A. Sampson (ed.), Topics in Statistical Dependence, IMS Lecture Notes Monograph Ser., Inst. Math. Statist., Hayward, 403–414.

E. Schechtman and S. Yitzaki (1987), *A measure of association based on Gini's mean difference*, Comm. Statist. Theory Methods 16 (1), 207–231.

W. Szczesny (1991), *On the performance of a discriminant function*, J. Classification 8, 201–215.

T. Taguchi (1987), *On the structure of multivariate concentration — some relationships among the concentration surface and two variate mean difference and regressions*, Comput. Statist. Data Anal. 6, 307–334.

N. White (1986), *Segregation and diversity measures in population distribution*, Population Index 52, 198–221.

Index of symbols

F_X	distribution function of X
F_X^{-1}	$\inf\{t : F_X(t) \geq y\}$, $0 < y < 1$
$r(x) = E(Y\|X = x)$	the value at the point x of the regression function of the random variable Y on the random variable X
$F_X^{(1)}$	$F_X^{(1)}(t) = (1/E(X)) \int_0^t u\, dF_X(u)$ (Sec. 2.2); called the first moment distribution function
$F_X^{(2)}$	$F_X^{(2)}(t) = (1/E(X)) \int_0^t (1 - F_X(s))\, ds$ (Sec. 2.2)
$X \sim P$	X is distributed according to P
$X \sim Y$	X and Y have the same distribution
$h(\omega) = (dQ/dP)(\omega)$	generalized Radon–Nikodym derivative of Q w.r.t. P
L_X	Lorenz curve for X (Sec. 2.2)
L_X^-	upper Lorenz curve for X (Sec. 5.3)
$K_{(P,Q)}$	divergence curve of Q from P (Sec. 1.1)
$L_{(P,Q)}$	concentration curve of Q w.r.t. P (Sec. 1.2)
$C^{\varphi}_{(P,Q)}$	φ-directed concentration curve of Q w.r.t. P (Sec. 5.1)
$C_{(P,Q)}$	(or $C_{(X,Y)}$ if $X \sim P, Y \sim Q$) the special case of $C^{\varphi}_{(P,Q)}$ for $\varphi(\omega) = \omega$
C_{cor}	correlation curve (Sec. 5.3)
C_{ratio}	ratio curve (Sec. 5.3)
$\preceq_{\mathrm{st}}$	stochastic ordering
$\preceq_{\mathrm{QD}}$	quadrant dependence ordering (Sec. 3.3)
$\preceq_{\mathrm{L}}$	Lorenz ordering (Sec. 2.3)
$\preceq_{*}$	star ordering (Sec. 2.3)
$\preceq_{\mathrm{D}}$	dependence ordering (Sec. 3.2)
$\preceq_{\mathrm{NP}}$	NP-ordering (Sec. 1.3)
$\preceq_{\mathrm{C}}$	ordering according to directed concentration curve (Sec. 5.1)
$G(\cdot)$	Gini inequality index (Sec. 6.1)
$D(\cdot)$	Pietra inequality index (Sec. 6.1)
$G(\cdot,\cdot)$	Gini divergence index (Sec. 6.2)
$D(\cdot,\cdot)$	Pietra divergence index (Sec. 6.2)