Notes on Inequality Measurement : Hardy, Littlewood and Polya, Schur Convexity and Majorization

Michel Le Breton

December 2006

Abstract

Winter School on Inequality and Collective Welfare Theory "Risk, Inequality and Social Welfare" January 10-13 2007, Alba di Canazei (Dolomites)

Outline of the Presentation

1. Generalities. Notations.
2. The Hardy, Littlewood and Polya's Theorem.
3. Schur Convexity and Inequality Measurement
4. Stochastic Dominance.
5. Continuous Distributions.
6. Multivariate majorizations : The Koshevoy's Zonotope.
7. Bivariate Income Distributions : Horizontal Equity and Taxation.

Generalities

I propose an excursion through few mathematical notions arising in inequality measurement and related matters like for instance welfare or poverty measurement, horizontal equity and progressivity in taxation. Inequality measurement is an important branch of applied welfare economics. This area of public/welfare economics is devoted to the development of analytical tools to evaluate the level of inequality attached to the distribution of one or several resources (like for instance, income, wealth, health,...) and the application of these notions to real data.

In this lecture notes, I will mostly devote my attention to the simplest setting. The ingredients of this setting will consist of a finite population N of n units (individuals, households, groups,...) and the distribution of a single divisible and transferable resource say income. An income distribution will be then a vector $x = (x_1, x_2, ..., x_n) \in \Re^n_+$. In this simple framework, each unit is identified by a number (say its social security number). In a more complicated setting, where the population would display some heterogeneity, relevant for the problem under consideration, we would have to describe explicitely the characteristics of the units.

Any comparison of two income distributions rests on interpersonal comparisons of utility and therefore on a specific measurement of social welfare. Any such theory should guide us in answering questions like : Is it "good" from a social perspective to tranfer this amount of resources from this group of units to this other one? We will spend the second section on the key mathematical result establishing the bridge between the statistical practice in inequality measurement and the modern approach of welfare economics. Then, in section 3, I will elaborate on some aspects of the theory of inequality measurement built upon that theorem. In section 4, I will examine the relationships between this area and the theory of stochastic orders developed in the area of decision analysis under uncertainty. Then in section 5, I will show how the theory extends to the case of a population described by a continuum. Sections 6 and 7 overview some recent developments of the theory which ambition to cover more complicated distributional environments.

Many of the developments are extracted from the first 4 chapters of my 20 years old thesis (Le Breton (1986)). I have added few major recent contributions like for instance those of Koshevoy on multivariate extensions. The books of Marshall and Olkin (1979) and Sen (1973) contain most of economic foundations and mathematical results used in standard inequality measurement. Besides, the survey, I plan to spend a significant portion of my talk on the application of a particular stochastic order (third degree stochastic dominance) to inequality measurement. This is based on a recent work coauthored with E. Peluso.

The Hardy, Littlewood and Polya's Theorem.

The Hardy, Littlewood and Polya's theorem is the key mathematical result in the area of inequality measurement. Kolm (1969) was the very first one, followed by Dasgupta, Sen and Starrett (1973), to point out the relevance of this result in establishing the foundations of inequality measurement. This theorem states that that three different partial orders are equivalent. To proceed with the statement, we need to introduce few notions.

A square matrix of order n, $B = (b_{ij})_{1\leq i,j\leq n}$ is bistochastic (or doubly stochastic) if $b_{ij} \geq 0$ $\forall j, i$; $\sum_{i=1}^{n} b_{ij} = 1$ $\forall j$ and $\sum_{j=1}^{n} b_{ij} = 1$ $\forall i$.A square matrix of order n is a permutation matrix if it is bistochastic and has exactly one positive entry in each row and each column. In what follows, we shall denote $\mathcal{B}_n$ (resp. $\mathcal{P}_n$) the set of bistochastic (resp. permutations) matrices of order n. Consider the following three partial preorders. Let x and y be two vectors in $\Re^n$ be such that : $x_i \leq x_{i+1}$ and $y_i \leq y_{i+1}$ for all $i = 1, ..., n-1$.

(1) There is a bistochastic matrix $B \in \mathcal{B}_n$ such that : $y = Bx$.

(2) $\sum_{i=1}^{n} \phi(x_i) \geq \sum_{i=1}^{n} \phi(y_i)$ for all convex functions $\phi : \Re \rightarrow \Re$.

(3) $\sum_{i=1}^{k} x_i \leq \sum_{i=1}^{k} y_i$ for all $k = 1, ..., n-1$ and $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$

It is immediate to see that the binary relations defined by (2) and (3) are preorders. The fact that the first one is also a preorder follows from the properties of the set of bistochatic matrices $\mathcal{B}_n$. This set is stable under multiplication and convex addition.

· The Hardy, Littlewood and Polya 's theorem asserts that the three partial orders are equivalent. I will offer a proof during the talk which is based on the notion of angles introduced by Hardy, Littlewood and Polya (1929) and rediscovered many times since then. Through the proof, we will see that we can make more precise some statements. For instance, an extensive use of the following type of bistochastic matrix will be made. Let $i, j \in \{1, ..., n\}$ and $\lambda \in [0, 1]$ and define as follows the $n \times n$ matrix $T_{i,j}^{\lambda}$:

$$T_{i,j}^{\lambda} = \lambda I + (1-\lambda) P_{i,j}$$

where $P_{i,j}$ is the permutation matrix attached to the permutation of the indices i and j. Under the action of such linear operator, a vector x is transformed into a vector z where $z_k = x_k$ for all $k \neq i, j$, $z_i = x_i + (1-\lambda)(x_j - x_i)$ and $z_i = x_j - (1-\lambda)(x_j - x_i)$. If the vectors are income distributions and $j > i$, then the change from x to z simply describes a single transfer $(1-\lambda)(x_j - x_i)$ from individual j to individual i who is poorer than him. The transfer preserves the rank of i and j iff $\lambda < \frac{1}{2}$. We can show that the matrix in (1) can be taken to be a product of matrices $T_{i,j}^{\lambda}$ where the λ can be selected in such a way that the ranking $1 \leq 2 \leq .... \leq n$ is preserved (not all bistochastic matrices can be expressed like that). With that qualification, condition (1) appears to be a principe of transfers : inequality decreases when such transfers are implemented (This sensitivity condition is known as the Pigou-Dalton 's principle of transfers).

· Condition (2) can be interpreted as a social welfare ranking where $-\phi$ would stand for the individual utilitry function of every individual in the population. The social welfare function which is considered is the utilitarian one. Note however that if we impose $-\phi$ to to be non decreasing, then condition (2) is no longer equivalent to (1) and (3). In (3), the equality $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$ must be replaced by the inequality $\sum_{i=1}^{n} x_i \leq \sum_{i=1}^{n} y_i$.

· The last condition is the classical Lorenz dominance condition. It consists in a simple algorithmic test described by a finite list of linear inequalities. The inequalities have also an immediate interpretation. We first compare the income of the poorest individual in the two distributions. Then we move to the aggregate income of the poorest and second poorest in the two distributions and so on. It is important to point out that the theorem above extends to any two vectors x and y as soon as in condition (1) and (3), x and y are replaced by x^* and y^* where x^* and y^* are the vectors x and y where the coordinates have been rearranged from the lowest to the highest.

· The Lorenz criterion is a well established notion in the statistics. To any vector x, we can attach a curve $L_x : [0,1] \to [0,1]$ where

$$L_x(t) = \frac{\left(t - \frac{k}{n}\right)\left(\sum_{i=1}^{k} x_i^*\right) + \left(\frac{k+1}{n} - t\right)\left(\sum_{i=1}^{k+1} x_i^*\right)}{\sum_{i=1}^{n} x_i} \text{ for all } t \in \left[\frac{k}{n}, \frac{k+1}{n}\right] \text{ and all } k = 0, ..., n-1$$

The function L_x is the Lorenz curve of x. Condition (3) can then be expressed as :

$$L_x(t) \leq L_y(t) \text{ for all } t \in [0,1]$$

i.e. the Lorenz curve of y is pointwise above the Lorenz curve of x.

· In statistics, the Lorenz curve of any probability distribution on $\Re$ is well defined. Let F be the cumulative distribution function of any such distribution and for any $t \in [0,1]$, let :

$$F_R^{-1}(t) = \underset{F(u) \leq t}{Sup}\ u$$

This is the right inverse of F. We could instead, as for instance Gastwirth (1971), consider the left inverse F_L^{-1} defined as follows :

$$F_R^{-1}(t) = \underset{F(u) \geq t}{Inf}\ u$$

The two inverses differ only on a set with Lebesgue measure equal to 0. We denote by F^{-1} this "common" inverse. The Lorenz curve L_F of the probability distribution F is defined as follows :

$$L_F(t) = \frac{\int_0^t F^{-1}(u)du}{\int_{\Re} uF(du)}$$

A probability distribution is identified by its Lorenz curve. We will see in section 4 that the equivalence betwwen (2) and (3) extends to the all class of probability distributions via the Lorenz curves.

· The importance of the Hardy, Littlewood and Polya 's theorem comes from the fact that it is establishes this full equivalence between three different perspectives on inequality measurement : an approach rooted in the social choice and welfare economics tradition, a second one bases on sensitivity to some special types of transfers between units and a third one constructed upon a useful and insightful statistical measure. To some extent, the following literature will often tries to achieve the same goal.

3. Schur Convexity and Inequality Measurement

· A real-valued function f defined on a set $A \subset \mathbb{R}^n$ is said to be Schur-convex on A if :

$$\forall x \in A,\ \forall B \in \mathcal{B}_n \text{ such that } Bx \in A \text{ we have } f(Bx) \leq f(x).$$

· It is strictly Schur-convex on A if

$$\forall x \in A,\ \forall B \in \mathcal{B}_n \text{ such that } Bx \in A \text{ we have } f(Bx) < f(x)$$
$$\text{if } Bx \text{ is not a permutation of } x.$$

· f is Schur-convave (resp. strictly Schur-concave) on A if -f is Schur-convex (resp. strictly Schur-convex) on A.

· A real-valued function f defined on a set $A \in \mathbb{R}^n$ is symmetric on A if :

$$\forall x \in A,\ \forall P \in \mathcal{P}_n \text{ such that } Px \in A \text{ we have } f(Px) = f(x).$$

In inequality measurement theory, different sets A can be considered, depending upon the range of income distributions that we want to cover. When we compare distributions with different aggregate incomes, we must introduce considerations which mix inequality matters with some other principles. In what follows, unless otherwise specified, we will not pay attention to these issues. and focus on the case where $A = S_n = \{x = (x_1,..., x_n) \in \mathbb{R}^n : x_1 \geq 0\ \forall i = 1,..., n \text{ and } \sum_{i=1}^n x_i = 1\}$, the unitary simplex of $\mathbb{R}^n$. An element will be interpreted as a distribution of a single divisible good (whose available quantity is normalized to one) between n individuals.

· A real-valued function f defined on S_n is called an inquality index on S_n if f is continuous and strictly schur-convex.

· It can be verified that any inequality index is symmetric (see Le Breton-Trannoy-Uriarte (1985)).

· Schur-convexity is the key notion in inequality theory. From the Hardy-Littlewood and Polya's theorem, we know that it is equivalent to require monotonicity with respect to the Lorenz order or the Pigou-Dalton principle of transfers. It should be emphasized that Lorenz dominance is equivalent to a finite sequence of Pigou-Dalton transfers. The fact that a function behaves quite well with respect to a single Pigou-Dalton transfer may be a poor indicator of its reaction with respect to a composition of such transfers. This point is well illustrated by Foster and Ok (1999) in the case of the variance of logarithms. This function is not Schur convex and therefore is not an inequality index. However, when we consider a single Pigou-dalton transfer, it behaves badly, exclusively in the case where the highest income is e times greater than the geometric mean of the distribution. With composite transfers this function may conclude that the distribution x is more unequal than the distribution y while x is arbitrarily close to the diagonal and y is arbitrarily close to complete inequality.

The rest of this paragraph, based on Le Breton (2006a) elaborates on this notion and its relations with the usual notion of convexity.

· If is a symmetric and quasi-convex function on S_n, then (Dasgupta-Sen-Starrett (1973) f is Schur-convex.

· A set $A \subset \mathbb{R}^n$ will be Schur-convex if $\forall x \in A$, $\forall B \in \mathcal{B}_\ell, Bx \in A$.

· If A is a symmetric and convex set of $\mathbb{R}^n$ then A is Schur-convex. Under symmetry, Schur-convexity is a (strictly) weaker notion than convexity.

· f is Schur-convex on A iff level sets $\{x \in A \quad f(x) \leq c\}$ are Schur-convex $\forall c \in \mathbb{R}$. In particular if A is a Schur-convex set, the indicator function 1_A of the set A is Schur-convex.

· If A and B are Schur-convex sets of $\mathbb{R}^n$, then $A \cup B$ is a Schur-convex set.

· If A is a Schur-convex set of S_n then A is a symmetric and star-shaped set centered on the point $E = (\frac{1}{n}, \frac{1}{n}, ...\frac{1}{n})$.

Proof: Let $x \in A$ and $\lambda \in [0, 1]$.

We have to show $\lambda x + (1 - \lambda)\lambda \in A$.

But $\lambda x + (1 - \lambda) E$ may be written $(\lambda I_n + (1 - \lambda) M) x$

where I_n is the identity matrix of order and M is the matrix $\frac{1}{n}\begin{pmatrix} 1 & 1...1 \\ & .... \\ & .... \\ 1 & 1...1 \\ 1 & 1...1 \end{pmatrix}$.

As M and $I_n \in \mathcal{B}_n \quad \lambda I_n + (1 - \lambda)M \in \mathcal{B}_n, \quad \lambda x + (1 - \lambda)E \in A$ by Schur-convexity of A.

It is easy to see that there exist symmetric and star-shaped sets centered on E which are not schur-convex. So we can say under symmetric schur-convexity is intermediate between convexity and star-shapedness.

· The Hardy-Littlewood-Polya's theorem leads also to a nice geometric description of the implications of Schur-convexity and emphasizes the fact that the property is truly a monotonicity property (with respect to a partial order) instead of a convexity property.

· There is a vast literature on inequality indices among which an axiomatic literature which aims to provide fondations in order to select some specific index (or family of indices) within the all family. Some indices, like for instance those due to Atkinson, Gini or Theil or properties like for instance decomposability have attracted a lot of attention.

We now move to the study of differentiable inequality indices on $A = S_n$. f is defined to be differentiable if it is differentiable on $r_i \ S_n$ (the relative interior of S_n) in the following sense : S_n is a manifold with boundary, which is homeomorphic to $\tilde{S}_{n-1} = \left\{ (x_1,..., x_{\ell-1}) \in \mathbb{R}_+^{\ell-1} : \sum_{i=1}^{\ell=1} x_i \leq 1 \right\}$. The homeomorphism is simply the projection map $(x_1, ..., x_\ell) \longrightarrow (x_1, x_{\ell-1})$ denoted by ϕ. f is differentiable on $r \ i \ S_\ell$ if $f \ o \ \phi^{-1}$ is differentiable on $\overset{\circ}{\tilde{S}}_{-1}$. Ostrowski's theorem (1952) provides a differential test for Schur-convexity. The regularity condition introduced below is a sufficient condition for strict-Schur convexity.

· A differentiable inequality index f is regular if: $\forall x \in S_n$:

$$x_i \neq x_j \Rightarrow (x_i - x_j)\left(\frac{\partial f}{\partial x_i} - \frac{\partial f}{\partial x_j}\right) > 0$$

· An inequality index f is smooth if $f \in C^{\infty}(S_n, \mathbb{R})$ and f is regular.

The rest of this section is devoted to the proof of an approximation theorem : the set of smooth inequality indices is dense in the set of inequality indices. The proof of the theorem will be deduced from the following sequence of lemmata.

· *Lemma 31:* There exists a sequence of functions $(\varepsilon_k)_{k\geq 1}$ $\mathbb{R}^n \to \mathbb{R}_+$ such that

(i) $\varepsilon_k \in C^{\infty}(\mathbb{R}^n, \mathbb{R})$

(ii) ε_k is schur-concave

(iii) Supp $\varepsilon_k \subset B\left(0, \frac{1}{k}\right) \cap R^n_-$

(iv $\int_{\mathbb{R}^n} \varepsilon_k(x)\ dx = 1$

Proof: We shall make an extensive use of the function $\Psi : \mathbb{R} \to \mathbb{R}$ defined as follows (see figure 31)

$h(x) = g(x) - \sum\limits_{i=1}^{n} (x_i - g(x))^2$ where $g(x) = \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2$

Figure 31

It is easy to verify h is Schur-concave and $h \in C^{\infty}(\mathbb{R}^n, \mathbb{R})$. Then, $\Psi \circ h$ is also Schur-convave and $\Psi \circ h \in C^{\infty}(\mathbb{R}^n, \mathbb{R})$.

Finally, define $\tilde{\varepsilon}_k : \mathbb{R} \to \mathbb{R}$ by :

$$\varepsilon_k(x) = \Psi\left(\frac{1}{k} + \sum_{i=1}^{n} x_i\right)\Psi\left(-\sum_{i=1}^{n} x_i\right)(\Psi \circ h(x))$$

and

$$\begin{array}{ll} \varepsilon_k : \mathbb{R}^n \to \mathbb{R} & \text{by} \\ \varepsilon_k(x) = \frac{1}{C_k}\ \tilde{\varepsilon}_k(x) & \text{where} \quad C_k = \int_{\mathbb{R}^n} \varepsilon_k(x)\ dx \end{array}$$

It is easy to check that ε_k satisfies properties (i), (ii), (iii) and (iv) (see figure 32 : she shaded area represents supp ε_k when $n = 2$).

· *Lemma 32:* Let $f \in C_c^\infty(\mathbb{R}^n, \mathbb{R})$ and $g \in L^1_{\ell oc}(\mathbb{R}^n, \mathbb{R})$ [1]. Then the convolution product of f and g denotes $f * g$ and defined by

$$(f * g)(x) = \int_{\mathbb{R}^\ell} f(x-y)\; g(y)\; dy$$

is well defined and moreover $f * g \in C^\infty\left(\mathbb{R}^\ell, \mathbb{R}\right)$.

Figure 32

· *Lemma 33:* Let f and g be Schur-concave functions defined on $\mathbf{R}^n$. Then $f * g$ (whenever it is defined) is Schur-concave.

Proof: See Eaton and Perlman (1977) or Marshall and Olkin (1974).

Theorem 31: Let f be an inequality index. Then there exists a sequence $(f_k)_{k>1}$ where f_k is a smooth inequality index $\forall k \geq 1$ and such that $f_k \to f$ when $k \to \infty$ uniformly on S_n.

Proof: Let $g = -f$; g is Schur-concave on S_n. We extend g on $\mathbb{R}^n_+$ as follows:

$$\begin{array}{lll} \forall x \in \mathbb{R}^n_+ & \tilde{g}(x) = g\left(\frac{x}{\sum_{i=1}^{\ell} x_i}\right) \quad \sum_{i=1}^{\ell} x_i & \text{if } x \neq 0 \\ \text{and} & \tilde{g}(0) = 0 & \end{array}$$

It is easy to check that $\tilde{g}$ is continuous and Schur-concave on $\mathbb{R}^n_+$. Finally, we extend $\tilde{g}$ on $\mathbb{R}^n$ in the following way:

$$\begin{array}{lll} & \tilde{\tilde{g}}(x) = \min\limits_{y \in S(x)} \tilde{g}(y) & \text{if } \sum_{i=1}^{n} x_i \geq 0 \\ \text{where} & S(x) = \{y \in \mathbb{R}^n_+ & \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} x_i\} \\ \text{and} & \tilde{\tilde{g}}(x) = 0 & \text{if } \sum_{i=1}^{n} x_i < 0 \end{array}$$

$\tilde{\tilde{g}}$ is Schur-concave and belongs to $L^1_{\ell oc}(\mathbb{R}^n, \mathbb{R})$. Now we show that $g * \varepsilon_k \to g$ uniformly when $k \to \infty$ on any compact of $\mathbb{R}^n_+$. From (iv), we deduce :

$$\begin{array}{lll} g * \varepsilon_k(x) - g(x) & = \int_{\mathbb{R}^n} (\tilde{\tilde{g}}(x-y) - \tilde{g}(x))\varepsilon_k(y)dy & \forall x \in \mathbb{R}^n_+ \\ & = \int_{B(o,\frac{1}{k}) \cap \mathbb{R}^n} (\tilde{\tilde{g}}(x-y) - \tilde{g}(x))\varepsilon_k(y)dy & \text{(by (iii))} \\ & = \int_{B(o,\frac{1}{k}) \cap \mathbb{R}^n} (\tilde{g}(x-y) - \tilde{g}(x))\varepsilon_k(y)dy & \end{array}$$

As $\tilde{g}$ is uniformly continuous on $K + B(0,1)$, $\forall \varepsilon > 0$, $\exists\, \delta(\varepsilon) > 0$ such that :

$$\text{For all } x, y \in K + B(0,1) : \|x - y\| \leq \delta(\varepsilon) \implies |\tilde{g}(x) - \tilde{g}(y)| \leqq \varepsilon$$

[1] $C_c^\infty(\mathbb{R}^n, \mathbb{R})$ denotes the set of functions in $C^\infty(\mathbb{R}^n, \mathbb{R})$ with compact support and $L^1_{\ell oc}(\mathbb{R}^n, \mathbb{R})$ denotes the set of functions which are locally integrable.

Thus if $n \geq \frac{1}{\delta(\varepsilon)}$, we deduce :

$$\sup_{x \in K} |\tilde{\tilde{g}} * \varepsilon_k(x) - \tilde{g}(x)| \leq \varepsilon \int_{B(0,\frac{1}{k}) \cap \mathbb{R}^n} \varepsilon_k(y) dy = \varepsilon$$

From lemma 32, $\tilde{\tilde{g}} * \varepsilon_k \in C^\infty(\mathbb{R}^n, \mathbb{R})$ and is Schur-concave by lemma 33. Let $f_n : S_\ell \to \mathbb{R}$ be defined by

$$f_k(x) = -(\tilde{\tilde{g}} * \varepsilon_k) + \frac{1}{k} \left(\sum_{i=1}^{n} (x_i - \frac{1}{n} \sum_{i=1}^{n} x_i)^2 \right)$$

From what precedes, it is easy to verify that f_k is a smooth inequality index and $f_k \to f$ when $k \to \infty$ uniformly on S_n.

4. Stochastic Dominance

Stochastic dominance orders are partial orders defined on subsets of probability distributions over the real numbers. Consider first the case of discrete probability distributions i.e probablity distribution P of the following type[2] :

$$P = \sum_{j=1}^{n} p_j \delta_{xj} \text{ where } x_1 \leq x_2 \leq .... \leq x_n,\ p_j \geq 0\ \forall j = 1, ...n \text{ and } \sum_{j=1}^{n} p_j = 1$$

In risk analysis, P can be interpretated as an uncertain prospect or lottery where the worst outcome is x_1 and occurs with probability p_1, the next worst outcome is x_2 and occurs with probability p_2 and so on. From the point of view of inequality measurement, P can be interpreted as an income distribution in a society. The society is divided into n groups from the poorest group denoted by 1 to the richest group denoted by n. In that interpretation, x_i and p_i denotes respectively the mean outcome and the percentage of the population in group i. The setting considered until now was assuming $p_1 = p_2 = ... = p_n = \frac{1}{n}$. We denote by $\mathcal{P}$ the set of discrete probability distributions.

To define the first three stochastic orders over $\mathcal{P}$, we need the following family of utility functions. $\mathcal{U}_1$ denotes the set of non decreasing real valued functions over $\Re_+$; $\mathcal{U}_2$ denotes the set of non decreasing and concave real valued functions over $\Re_+$ and $\mathcal{U}_3$ denotes the set of differentiable real valued functions over $\Re_+$ whose first derivative is non negative, non increasing and convex. Then for all $P = \sum_{j=1}^{n} p_j \delta_{xj}$ and $Q = \sum_{j=1}^{m} q_j \delta_{yj}$ and all $i = 1, 2, 3$:

[2]For all $t \in \Re$, δ_t denotes the Dirac mass in t.

$$P \succsim_i Q \text{ iff } \sum_{j=1}^{n} p_j u(x_j) \geq \sum_{j=1}^{m} q_j u(y_j) \text{ for all } u \in \mathcal{U}_i$$

The classical results on stochastic dominance are summarized in the result below. Let E_P and F_P denote respectively the first moment of P and the distribution function of probability P i.e. for all $t \in \Re$, $F_P(t) = P(]-\infty, t]$

$\cdot$ Let $P, Q \in \mathcal{P}$. Then $P \succsim_1 Q$ iff $F_P(t) \leq F_Q(t)$ for all $t \in \Re$, $P \succsim_2 Q$ iff $\int_{-\infty}^{t} F_P(u)du \leq \int_{-\infty}^{t} F_Q(u)du$ for all $t \in \Re$, and $P \succsim_3 Q$ iff $\int_{-\infty}^{t}\int_{-\infty}^{r} F_P(u)dudr \leq \int_{-\infty}^{t}\int_{-\infty}^{r} F_Q(u)dudr$ for all $t \in \Re$ and $E_P \geq E_Q$.

The conditions in the above result turn out to be extremely simple when P and Q have the same support and therefore differ exclusively from the point of view of probability weights. For instance, when $P = \sum_{i=1}^{n} p_i \delta_{xi}$ and $Q = \sum_{i=1}^{n} q_i \delta_{xi}$, $P \succsim_2 Q$ iff :

$$p_1 \leq q_1, \; p_1 (x_3 - x_1) + p_2 (x_3 - x_2) \leq q_1 (x_3 - x_1) + q_2 (x_3 - x_2), \; ....$$

However, when the two supports differ the inequalities of the proposition become more intricate. What is the relationship with the Hardy, Littlewood and Polya 's theorem and Lorenz dominance ? If instead of comparing distributions with a common support, we compare distributions with common probability weights, say $P = \sum_{i=1}^{n} p_i \delta_{xi}$ and $Q = \sum_{i=1}^{n} p_i \delta_{yi}$, then $P \succsim_2 Q$ iff :

$$x_1 \leq y_1, \; p_1 x_1 + p_2 x_2 \leq p_1 y_1 + p_2 y_2, ...$$

When the probabilities p_i are all equal, we recognize the Lorenz order. This subset $\mathcal{P}_n$ of probabilities whose support consist of at most n points is in a one to one relationship with the cone K_n :

$$K_n = \left\{ x \in \Re_+^n : x_1 \leq x_2 \leq .... \leq x_n \right\}$$

The stochastic orders on $\mathcal{P}_n$ can be formally defined as follows on K_n. For all $x, y \in K_n$ and all $i = 1, 2, 3$ let :

$$x \succsim_i^t y \text{ if and only if } \frac{1}{n}\sum_{j=1}^{n} \delta_{x_j} \succsim_i \frac{1}{n}\sum_{j=1}^{n} \delta_{y_j}$$

i.e.

$$x \succsim_i^t y \text{ iff } \sum_{j=1}^{n} u(x_j) \geq \sum_{j=1}^{n} u(y_j) \text{ for all } u \in \mathcal{U}_i$$

For all $x \in K_n$ and all $j = 1, ..., n$, let $X_j = \sum_{k=1}^{j} x_k$. In what follows, we will refer to X as being the Lorenz vector attached to x.The following result can be deduced from the previous one or demonstrated directly. The second part is one of the equivalence in Hardy,Littlewood and Polya.

· Let $x, y \in K_n$. Then : $x \succsim_1^t y$ iff $x_j \geq y_j$ for all $j = 1, ...n$ and $x \succsim_2^t y$ iff $X_j \geq Y_j$ for all $j = 1, ...n$

· In the case of continuous distributions, the relationship between stochastic and Lorenz dominance remains valid via the definition of the Lorenz curve introduced in section 2. We can show (Atkinson (1970, Le Breton (1986)) that :

$$P \succsim_2 Q \text{ iff } L_F(t) \geq L_G(t) \text{ for all } t \in [0, 1]$$

Unfortunately, there is no result characterizing third degree stochastic dominance in terms of Lorenz curves. We will devote most of the talk to this question. The rest of this section is based on Le Breton and Peluso (2006). Its main purpose is to examine the properties of the orderings $\succsim_i^t$. It follows from the first result above that both $\succsim_1$, $\succsim_2$and $\succsim_3$satisfy the von Neumann-Morgenstern independence property and therefore $\succsim_1 = \succsim_1^* = \succsim_1^{**}$, $\succsim_2 = \succsim_2^* = \succsim_2^{**}$and $\succsim_3 = \succsim_3^* = \succsim_3^{**}$. It follows also from the second result that both $\succsim_1^t$ and $\succsim_2^t$ are cone preorders. Precisely, $\succsim_1^t = \succsim_{A_1}$and $\succsim_2^t = \succsim_{A_2}$where $A_1 = \{x \in \Re^n : x_i \geq 0 \ \forall i = 1, ..., n\}$ and $A_2 = \{x \in \Re^n : X_i \geq 0 \ \forall i = 1, ..., n\}$. Therefore, they also satisfy the von Neumann-Morgenstern independence property and then $\succsim_1^t = \succsim_1^{t*} = \succsim_1^{t**}$and $\succsim_2^t = \succsim_2^{t*} = \succsim_2^{t**}$.

5. A Continuous Version of Hardy, Littlewood and Polya

This section builds on Le Breton (2006b). Its main purpose is to extend the finite framework to cover the case of continous distributions.To formalize the continuum assumption, we shall assume in the all section that the set of agents is represented by the probability space $([0, 1], \mathcal{B}, \lambda)$ where $\mathcal{B}$ is the σ algebra of Borelian subsets of $[0, 1]$ and λ is the Lebesgue measure on $[0, 1]$.

An income distribution is any measurable function X from $[0, 1]$ to $I\!R_+$ which is integrable with respect to λ. An income distribution X is bounded if there exists a constant C such that $X(t) \leq C$ for λ almost every t in $[0, 1]$.

Thus formally the set of income distributions (resp. bounded income distributions) is the positive cone of $L^1[0,1]$ (resp. $L^\infty[0,1]$). We shall denote by $L^1_+[0,1]$ and $L^\infty_+[0,1]$ these two sets.

As emphasized in the finite case, the two major properties of inequality measurement are symmetry and strict Schur convexity. Let us first introduce, the continuous counterparts of these two properties.

Let X be an arbitrary measurable real-valued function defined on $[0,1]$. It is straightforward to show that the function m_X defined on $I\!R$ by $m_X(x) = \lambda\{t \in [0,1] : X(t) > x\}$ is nonincreasing, right continuous and with values in $[0,1]$. As such, the function m_X admits a right inverse which will be denoted by X^*. To fix ideas and remove certain ambiguities it is convenient to define $X^*(t) = \underset{m_X(x)>t}{sup.X}$ for $t \in]0,1[$. It is nonincreasing and right continuous. The function X^* is called the decreasing rearrangement of X. Indeed it is straightforward to show that two measurable functions X and Y on $[0,1]$ satisfy $X^* = Y^*$ λ a.e on $[0,1]$ if and only if their respective probability distributions on $I\!R$ denote by λ_X and λ_Y are identical. Thus in particular the probability distributions of X and X^* on $I\!R$ are identical. It follows from this observation that if X belongs to $L^1_+[0,1]$ (resp. $L^\infty_+[0,1]$) then X^* belongs also to $L^1_+[0,1]$ (resp. $L^\infty_+[0,1]$).

A real-valued function I defined on $L^1_+[0,1]$ is symmetric if $\forall X \in L^1_+[0,1] : I(X) = I(X^*)$. It is natural to wonder whether this concept of symmetry is totally analogous to the concept of symmetry used in the finite case. With a finite set of agents say $\{1, 2 \ldots, n\}$, two income distributions $X = (x_1, \ldots, x_n)$ and $Y = (y_1, \ldots, y_n)$ are symmetric if there exists a permutation σ on $\{1, 2 \ldots, n\}$ such that $y_i = x_{\sigma(i)}, i = 1, \ldots, n$. The continuous analogue of a permutation is a measure preserving transformation on $[0,1]$ i.e. a measurable function $\sigma : [0,1] \to [0,1]$ such that $\lambda(A) = \lambda(\sigma^{-1}(A)), \forall A \in \mathcal{B}$. It is easy to show that if two real-valued measurable functions X and Y on $[0,1]$ are such that $X = Y$ on σ for a measure preserving transformation σ on $[0,1]$, then $X^* = Y^*$. Unfortunately the conserve is false in general: a counterexample is given by the functions X and Y defined by $X(t) = 1 - t$ and $Y(t) = 2t \pmod 1$, $t \in [0,1]$. Nevertheless it must be noted that Ryff (1970) has proved that for any real-valued measurable function X on $[0,1]$ there exists a measure preserving transformation σ on $[0,1]$ such that $X = X^*$ on σ.

In order to define a continuous version of the property of Schur convexity, we first provide a continuous version of the familiar Lorenz preorder.

Let X and Y be two functions in $L^1[0,1]$. We shall say that X Lorenz dominates Y if :

$$\int_0^s X^*(t)dt \leq \int_0^s Y^*(t)dt \quad \forall s \in [0,1[$$

and

$$\int_0^1 X^*(t)dy = \int_0^1 Y^*(t)dt$$

If the integral inequalities defining Lorenz domination are satisfied by X and Y in $L^1_+[0,1]$ we shall write $X \succsim_L Y$. It is easy to show that for any X and Y in $L^1[0,1] X \sim_L Y$ if and only if $X^* = Y^*$.

We now move to our examination of the right extension of Hardy, Littlewood and Polya's theorem in this context

A linear transformation B from $L^1[0,1]$ is a bistochastic operator if $BX \succsim_L X \quad \forall X \in L^1[0,1]$. The use of the term operator is intended to imply these linear transformations are bounded. Indeed it is easy to verify [see. e.g. Ryff (1963)] that if a linear transformation B from $L^1[0,1]$ to $L^1[0,1]$ is such that $BX \succsim_L X \quad \forall X \in L^1[0,1]$ then it is a contraction for the L^1 norm[3]. Moreover if we consider B in restriction to $L^\infty[0,1]$, it is easy to show that B has its values in $L^\infty[0,1]$ and is a contraction for the L^∞ norm. A representation of bistochastic operators in terms of kernels has been given by Ryff (1963).

The following theorem provides a first characterization of the partial preorder $\succsim$.

Theorem 5.1. [Ryff (1965)]

Let X and Y be two functions in $L^1[0,1]$. Then $X \succsim_L \; Y$ if and only if there exists a bistochastic operator B on $L^1[0,1]$ such that $X = BY$.

If X is an income distribution and $s \in [0,1]$, $\int_0^s X^*(t)\, dt$ represents the amount of income received by the richest s share of the population. Thus if we intend to use a real-valued function I defined on $L^1[0,1]$ in order to perform inequality measurement it appears reasonable to this function to be decreasing with respect to $\succsim_L$ i.e. if X and Y in $L^1[0,1]$ are such that $X \succsim_L \; Y$ then $I(X) \leqq I(Y)$. We may even impose that I be strictly decreasing with respect to $\succsim_L$ i.e. $X \succ_L \; Y$ implies $I(X) < I(Y)$. From theorem 1, it comes that these two

[3]More precisely it is a positive contraction operator on $L^1[0,1]$.

monotonicity requirements are captured by the following definition.

A real-valued function I defined on $L^1[0,1]$ is:

1. Schur-convex if $\forall X \in L^1[0,1] I(BX) \leqq I(X)$ for every bistochastic operator B on $L^1[0,1]$.
2. strictly Schur-convex if $\forall X \in L^1[0,1] I(BX) < I(X)$ for every bistochastic operator B on $L^1[0,1]$ such $(BX)^* \neq X^*$.

This definition of Schur-convexity which is aligned on the definition which is traditionnaly provided in the finite case represents a departure from the definition given for instance by Chong and Rice (1971) and Luxembourg (1967).

From now on, we shall restrict our attention to bounded income distribution. To introduce a continuity requirement we must endow $L^\infty[0,1]$ with a topology. In contrast with the finite case there is no natural topology on $L^\infty[0,1]$. We are going compare three usual topologies on $L^\infty[0,1]$ such that this space is a locally convex linear topological space and motivate the choice of the Mackey topology[4].

The first topology is the topology associated to the norm $\|\cdot\|_\infty$[5]. It can be shown that this metric leads to an "excessive" sensibility of inequality measurement to an additional income for an arbitrary small group of agents). In looking for weaker topologies, we will focus on those which are locally convex and such that the topological dual be $L^1[0,1]$. More precisely, we are going to examine the weaker one which is the weak $\sigma(L^\infty, L^1)$ topology, and the finer one which is the Mackey $\tau(L^\infty, L^1)$ topology.

The weak $\sigma(L^\infty, L^1)$ is too restrictive for our context. Indeed there does not exist real-valued functions on $L^\infty[0,1]$ which are simultaneously symmetric, strictly Schur-convex and $\sigma(L^\infty, L^1)$ continuous. This may seen by considering the following sequence of functions. Let $(X_k)_{k\in I\!N^*}$ be defined by $X_k(t) = \frac{1}{2}$ if $t \in \left[\frac{2j-1}{2k}, \frac{2j}{2k}\right[$ for $j = 1, \ldots, k$ and $X_k(t) = \frac{3}{2}$ if $t \in \left[\frac{2j}{2k}, \frac{2j+1}{2k}\right[$ for $j = 0, \ldots.., k-1$ It is straightforward to show that $(X_k)_{k\in I\!N^*}$ converges (in the $\sigma(L^\infty, L^1)$ topology) to the function $Y \equiv 1\!\!I_{[0,1]}$. Furthermore $X_k^* = \frac{3}{2} 1\!\!I_{[0,1[} + \frac{1}{2} 1\!\!I_{[\frac{1}{2},1]}$ $\forall k \geq 1$. Thus if I is $\sigma(U^\infty, U^1)$ continuous and symmetric on $L^\infty[0,1]$,

[4]For all the relevant material concerning linear topological space, weak and Mackey topologies, we refer to dunford-Schwartz (1966) and Kelley-Namioka (1963).

[5]$\|X\|_\infty \equiv \inf\{c > 0 : |X(t) - \leqq c$ for λ a.e. $t \in [0,1]\}$

we deduce that $I(Y) = I(X^*)$ which contradicts Schur-convexity since $Y \succ_L X_1^*$. This situation is far from being exceptional ; a complete characterization of the functions which are symmetric and $\sigma(L^\infty, L^1)$ continuous is given in Le Breton (2006).

All these considerations suggest to endow $L^\infty[0,1]$ with the Mackey topology $\tau(L^\infty, L^1)$ leading to the following continuous counterpart of the definition provided in the finite case.

An inequality index for bounded income distributions is a real-valued function I defined on $L_+^\infty[0,1]$ such that I is mackey continuous ans strictly Schur-convex.

We shall prove later that any inequality index is symmetric. The remainder of this section is devoted to the proof of some properties of the Mackey topology which will be useful in proving our continuous extension of Hardy, Littlewood and Polya.

Lemma 5.2. The Mackey topology $\tau(L^\infty, L^1)$ is finer that the topology of convergence in probability[6].

Proof

Assume at the contrary that there exists a generalized sequence $(X_\gamma)_{\gamma\in\Delta}$ in $L^\infty[0,1]$ converging to X in the Mackey $\sigma(L^\infty, L^1)$ topology and such that $(X_\gamma)_{\gamma\in\Delta}$ does not converge to X probability.

Then there exists $\varepsilon > 0$ and a generalized subsequence $(X_\gamma)_{\gamma\in\tilde{\Delta}}$ such that $\lambda\{t \in [0,1] : |X_\gamma(t) - X(t)| > \varepsilon\} > \varepsilon, \quad \forall \gamma \in \tilde{\Delta}$.

Consider:

$$f_\gamma(t) = 1\!\!I_{\{X_\gamma - X > \varepsilon\}} - 1\!\!I_{\{X_\gamma - X > \varepsilon\}}, \gamma \in \tilde{\Delta}, t \in [0,1]$$

We denote by $\mathcal{F}$ the circled convex hull of the set $\{f_\gamma\}_{\gamma\in\tilde{\Delta}}$. By Dunford-Pettis's theorem [see e.g. Neveu (1970) proposition IV 2.3], it comes $\mathcal{F}$ is $\sigma(L^1, L^\infty)$ relatively compact since it is equi-integrable. Thus $\mathcal{F}$ is a circled, convex [Dunford-Schwartz (1966) th. 1, p. 413], and $\sigma(L^1, L^\infty)$ compact subset of L^1.

[6]For a definition and some properties of this topology see Kelley-Namioka (1983) p. 55. With this $L^\infty[0,1]$ is a metrizable linear topological space.

From the definition of f_γ, it comes

$$\begin{aligned}\int_0^1 f_\gamma(t)(X_\gamma(t) - X(t))dt &= \int_{\{X_\gamma - X > \varepsilon\}} -X(t)dt + \int_{\{X_\gamma - X < -\varepsilon\}} -X_\gamma(t)dt \\ &\geq \varepsilon^2, \quad \forall \gamma \in \tilde{\Delta}\end{aligned}$$

Thus $\sup_{f\in\mathcal{F}} | \int_0^1 f(t)(X_\gamma(t) - X(t)dt| \geq \varepsilon^2 \quad \forall\gamma \in \tilde{\Delta}$

From the characterization of convergence for the Mackey topology [Kelley-Namioka (1963) th. 18.8] it comes that $(X_\gamma)_{\gamma\in\tilde{\Delta}}$ does not converge to X in the Mackey topology contradicting our assumption.

The following result states a weak converse of lemma 5.2.

Lemma 5.3. Let K be a strongly bounded subset of $L^\infty[0,1]$. In restriction to K the topology of convergence in probability is finer that the Mackey topology $\tau(L^\infty, L^1)$.

Proof

Since the topology of convergence in probability is metrizable and thus first countable it suffices to prove that if $(X_n)_{n>0}$ is a sequence in K converging to X in this topology, then it converges also to X in the Mackey topology $\tau(L^\infty, L^1)$.

Let $C > 0$ be such that $\|Y\|_\infty \leq C \quad \forall Y \in K$ and $\mathcal{F}$ an arbitrary circled, convex and $\sigma(L^1, L^\infty)$ compact subset of $L^1[0,1]$.

For any $f \in \mathcal{F}$ and $\eta > 0$ we have

$$\int_0^1 f(t)(X_n(t) - X(t)dt = \int_{\{|X_n - X| > \eta\}} f(t)(X_n(t) - X(t))dt + \int_{\{|X_n - X| \leq \eta\}} f(t)(X_n(t)) - X(t)dt$$

It comes

$$|\int_0^1 f(t)(X_n(t) - X(t)dt)| \leq 2C \int_{\{|X_n - X| > \eta\}} |f(t)dt + \eta \int_0^1 |f(t)|dt$$

Since $\mathcal{F}$ is $\sigma(L^1, L^\infty)$ compact it is (applying again Dunford-Pettis's theorem) equi-integrable i.e.

$\forall \varepsilon > 0, \exists\ \delta(\varepsilon) > 0$ such that $\lambda(E) \leq \delta(\varepsilon)$ implies $\displaystyle\int_E |f(t)|dt \leq \varepsilon \quad \forall \exists\ \in \mathcal{B}$ and $\forall f \in \mathcal{F}$ and $\exists c' > 0$ such that $\|f\|_1 \leq C' \quad \forall f \in \mathcal{F}$.

Let $\sigma > 0$. Since $(X_n)_{n \geqq 0}$ converges to X in probability, for any $\eta > 0$ there exists $N(\delta, \eta)$ such that $n \geqq N(\delta, \eta)$ implies $\lambda\{t \in [0,1] : |X_n(t) - X(t)| > \eta\} \leqq \delta$.

Thus if $n \geqq N\left(\delta\left(\frac{\varepsilon}{4C}\right), \left(\frac{\varepsilon}{2C}\right)\right)$ it comes:

$$\left|\int_0^1 f(t)(X_n(t) - X(t)dt\right| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \quad \forall f \in \mathcal{F}$$

In combining lemmas 5.2 and 5.3., it follows that the topology of convergence in probability and the Mackey topology coincide on the strongly bounded subsets of $L^\infty[0,1]$[7].

We shall denote $\mathcal{B}^\infty$ the set of bistochastic operators on $L^\infty[0,1]$. It is easy to show that $\mathcal{B}^\infty$ is convex. Furthermore if $B, B' \in \mathcal{B}^\infty$ then $B \circ B' \in \mathcal{B}^\infty$ and if $B \in \mathcal{B}^\infty$ then the adjoint of B is a bistochastic operator on $L^1[0,1]$. Thus $\mathcal{B}^\infty$ is a selfadjoint semi-group of operators on $L^\infty[0,1]$.

For every X in $L^\infty[0,1]$ we shall denote by $\Omega(X)$ the orbit of X under the section of $\mathcal{B}^\infty$ i.e. $\Omega(X) = \{BX, B \in \mathcal{B}^\infty\}$. From theorem 5.1, we know that $\Omega(X)$ is the set of income distributions that Lorenz dominates the income distribution X.

Lemma 5.4. For every X in $L^\infty[0,1]$, $\Omega(X)$ is convex and Mackey closed in $L^\infty[0,1]$.

Proof

1. $\Omega(X)$ is convex: obvious since $\mathcal{B}^\infty$ is convex ;
2. $\Omega(X)$ is Mackey closed.

Since the operators in $\mathcal{B}^\infty$ are contractions we deduce that $\Omega(x)$ is strongly bounded. Since the Mackey topology $\tau(L^\infty, L^1)$ is metrizable in restriction to strongly bounded subsets, we have to show that if $(Y_n)_{n\geq 1}$ is a sequence in $\Omega(X)$ converging for the Mackey

[7] The coincidence does not hold on whole space. Consider for instance the sequence $(X_\eta)_{\eta \geq 1}$ with $X_\eta \equiv \eta\ 1[0, \frac{1}{\eta}]$

topology to Y, then $Y \in \Omega(X)$.

Claim 1: $(Y_n^*)_{n\geq 1}$ converges to Y^* for the Mackey topology.

From lemma 5.2 it comes that $(Y_n)_{n\geqq 1}$ converges in probability to y, and thus in distribution. Then, it is straightforward to show that $(Y_n^*)_{n\geq 1}$converges λ almost surely to Y^*. We deduce from lemma 5.3. that $(Y_n^*)_{n\geqq 1}$ converges for the Mackey topology to Y^*.

Claim 2: $Y^* \in \Omega(X)$

Assume at the contrary $Y^* \notin \Omega(X)$. Then there exists $s \in [0,1]$ such that $\int_0^s Y^*(t)dt > \int_0^s X^*(t)dt$. Consider $f \equiv 1\!\!1_{[0,s]}$. Since from claim 1 $(Y_n^*)_{n\geqq 1}$ converges to Y^* for the Mackey topology, it converges for the $\sigma(L^\infty, L^1)$ topology and thus $\int_0^1 f(t)Y_0^* f(t)Y_n^*(t)dt$ tends to $\int_0^1 f(t)Y^*(t)dt$. when n goes to infinity. This implies that for n sufficiently large we have $\int_0^s Y_n^*(t)dt > \int_0^s X^*(t)dt$. this contradicts the assumption that $Y_n \in \Omega(X)$. Thus $Y^* \in \Omega(X)$. Since $Y \sim_L Y^*$ we deduce from claim 2 that $Y \in \Omega(X)$.

The following result gives a deep information on the geometrical structure of $\Omega(X)$.

Theorem 5.6. [Ryff (1967)]

For every X in $L^\infty[0,1]$, the set of extremal points of $\Omega(X)$ is the set $\{Y \in L^\infty[0,1] : Y \sim_L X\}$[8].

Lemma 5.7. For every X in $L^\infty[0,1]\Omega(X)$ is the Mackey closed convex hull of the set $\{Y \in L^\infty[0,1] : Y \sim_L X\}$.

Proof

From lemma 5.4. we know that $\Omega(X)$ is convex and Mackey closed. Since the topological duals of $L^\infty[0,1]$ for the Mackey topology and the $\sigma(L^\infty, L^1)$ topology are the same, we

[8]Strictly speaking Ryff's theorem is stronger: it is stated for $L^1[0,1]$.

deduce [Dunford-Schwartz (1966) Cor. 14 p. 418] that the closed convex sets are the same for these two topologies. Thus $\Omega(X)$ is $\sigma(L^\infty, L^1)$ closed. Since we have already noticed that it is strongly bounded we deduce from Alaoglu's theorem [Dunford-Schwartz (1966) p. 424] that it is $\sigma(L^\infty L^1)$ compact.
From theorem 5.6 and Krein-Milman's theorem [Dunford-Schwartz (1966) p. 440] we deduce that $\Omega(X)$ is the $\sigma(L^\infty, L^1)$ closed convex hull of the set $\{Y \in L^\infty[0,1] : Y \sim_L X\}$. By using again the argument above, it comes that $\Omega(X)$ is the Mackey closed convex hull of the set $\{Y \in L^\infty[0,1] : Y \sim_L X\}$.

The following result has already been announced.

Lemma 5.8 Every inequality index I on $L_+^\infty[0,1]$ is symmetric.

Proof

Let X, Y belonging to $L_+^\infty[0,1]$ be such that $X \sim_L Y$. Since $\Omega(X) = \Omega(Y)$ is convex it comes $\lambda X + (1-\lambda)Y \in \Omega(X) \quad \forall \lambda \in [0,1]$. For every λ in $]0,1[\lambda X + (1-\lambda)Y$ is not an extreme point of $\Omega(X)$ and thus from theorem 5.6., it comes $(\lambda X + (1-\lambda)Y^* \neq X^*$. Since I is strictly Schur-convex we deduce that $I(\lambda X + (1-\lambda)Y)$ is strictly smaller than $I(X)$ and $I(Y)$.

On the other hand $\|\lambda X + (1-\lambda)Y - X\|_\infty = (1-\lambda)\|X - Y\|_\infty$ and $\|\lambda X + (1-\lambda)Y - Y\|_\infty = \lambda\|X - Y\|_\infty$. Thus when λ tends to 0 (resp. to 1) $\lambda X + (1-\lambda)Y$ converges for the $\| \ _\infty$ norm and consequently for the Mackey topology to Y (resp. to X). Since I is Mackey continuous we deduce that $I(X) \leqq I(Y)$ and $I(Y) \leqq I(X)$ i.e. $I(X) = I(Y)$.

The following result describes an important family of inequality indices.

Lemma 5.9. For every real valued function φ continuous and convex on $I\!R_+$ the function I defined o $L_+^\infty[0,1]$ by $U(X) = \int_0^1 \varphi(X(t))dt$ is an inequality index.

Proof

1. I is Mackey continuous on $L_+^\infty[0,1]$.
   Consider a generalized sequence $(X_\gamma)_{\gamma\in\Delta}$ in $L_+^\infty[0,1]$ converging for the Mackey topol-

ogy to X. Since this implies that it converges for the $\sigma(L^\infty, L^1)$ topology ; thus we deduce from the Banack-Steinhaus's theorem [Kelley-Namioka -1963) th. 12.2] that it is strongly bounded i.e. $\|X_\gamma\|_\infty \leqq C$ and $\|X\|_\infty \leqq C$ for a constant $C > 0$. We made a troncation of φ in C by setting $\varphi_C(X) \equiv \varphi(X)$ if $x \leqq C$ and $\varphi_C(x) = \varphi(C)$ if $x > C$. Since $(X_\gamma)_{\gamma\in\Delta}$ converges to X for the Mackey topology it comes from proposition 2 that it converges to X in distribution. It is clear that $I(X_\gamma) = \int_0^1 \varphi_C(X(t)dt)$; thus since φ_C is continuous and bounded we deduce that $(I(X_\gamma))_{\gamma\in\Delta}$ converges to $I(X)$.

2. I is strictly Schur-convex on $L_+^\infty[0,1]$.
   Since I is convex, symmetric and Mackey continuous on $L_+^\infty[0,1]$, it is Schur-convex on $L_+^\infty[0,1]$. It remains to prove that it is strictly schur-convex. Let $X \in L_+^\infty[0,1]$ and $Y \in \Omega(X)$ with $Y^* \neq X^*$. From theorem 5.6, Y is not an extremal point of $\Omega(X)$. i.e. $\exists Z_1, Z_2 \in \Omega(X), Z_1 \neq Z_2$ such that $Y = \frac{1}{2}(Z_1 + Z_2)$. Since φ is strictly convex we deduce immediately that $I(Y) < \frac{1}{2}(I(Z_1) + I(Z_2))$. Since I is Schur-convex $I(Z_1) \leqq I(X)$ and $I(Z_2) \leqq I(X)$. Thus $I(Y) < I(X)$.

We are now in position to state a suitable continuous version of the theorem of Hardy, Littlewood et Polya.

Theorem 5.10

Let X and Y belonging to $L^\infty[0,1]$. The following properties are equivalent.

1. $Y \succsim_L X$
2. There exists $B \in \mathcal{B}$ such that $Y = BX$
3. Y belongs to the Mackey closed convex hull of the set $\{Z \in L^\infty[0,1] : Z^* = X^*\}$
4. For every convex, symmetric and Mackey continuous real-valued function I on $L^\infty[0,1]$ we have : $I(Y) \leqq I(X)$.

<u>Proof</u>

1. $\Leftrightarrow$ 2. : theorem 5.1.

2. $\Leftrightarrow$ 3. : lemma 5.7.

3. $\Rightarrow$ 4.

Let I be a convex, symmetric and Mackey continuous real-valued function on $L^\infty[0,1]$.

Since $Y \in \overline{CO}\{Z \in L^\infty[0,1] : Z^* = X^*\}$, there exists[9] a generalized sequence $(Z_\gamma)_{\gamma\in\Delta}$ converging to Y and such that $\forall \gamma \in \Delta, Z_\gamma) \sum_{i=1}^{k(\gamma)} \lambda_{i,\gamma}\tilde{X}_{i,\alpha}$ with $\sum_{i=1}^{k(\gamma)} \lambda_{i,\alpha} = 1 \quad 0 \leqq \lambda_i, \leqq 1$ and $\tilde{X}_{i,\gamma^*} = X^*, \quad \forall i = 1, \dots, k(\gamma)$. Since I is convex it comes $I(z_\gamma) \leqq \sum_{i=1}^{k(\gamma)} \lambda_{i,\gamma} I(\tilde{X}_{i,\alpha}), \quad \forall \gamma \in \Delta$, and since it is also symmetric we have $I(Z_\gamma) \leqq I(X) \quad \forall \gamma \in \Delta$. By using the Mackey continuity of I we deduce $I(Y) \leqq I(X)$.

4. $\Rightarrow$ 1.For every $t \in [0,1]$, we consider the function $I_t : L^\infty[0,1] \to I\!R$ defined by $I_t(Z) = \int_0^1 \varphi_t(Z(s)ds)$ with $\varphi_t : I\!R \to I\!R$ defined by $\varphi_t(X) = \max(0, x - X^*(t))$. It is easy to show that I is symmetric and Mackey continuous. Furthermore since φ_t is convex it is also convex.

By applying 4. to I_t, we deduce :

$$I_t(Y^*) = I_t(Y) \leqq I_t(X) = I_t(X^*)$$

Since φ_t is positive we have :

$$I_t(X^*) \geqq \int_0^t \varphi_t(Y^*(s)ds) \geqq \int_0^t (Y^*(s) - X^*(t))ds$$

But by construction :

$$I_t(X^*) = \int_0^t (X^*(s) - X^*(t))ds$$

Thus :

$$\int_0^t Y^*(s)ds \leqq \int_0^t X^*(s)ds\cdot$$

The equality for $t = 1$ follows by considering the linear function I defined by $I(Z) = -\int_0^1 Z(s)ds$.

[9]Since for every subset A of a topological linear space $\overline{CO}A = \overline{CO\ A}$ (see e.g. Dunford-Schwartz (1966) lemma 4 p; 415).

By using arguments totally different from ours, Grothendieck (1955) has proved that condition (4) above is equivalent to the condition :
For every convex,symmetric and $\sigma(L^\infty, L^\infty)$ lower semi-continuous real or $\{+\infty\}$ valued function I on $L^\infty[0,1]$: $I(Y) \leqq I(X)$.

A careful reading of the proofs indicates how this result can be easily deduced from ours. In the case of $L^1[0,1]$, Chong and Rice (1971) and Luxembourg (1967) have established results of the same nature for the weak $\sigma(L^1, L^\infty)$ topology and lower semi-continuity instead of continuity.

Multivariate majorizations : The Koshevoy's Zonotope

The theory of inequality measurement has been developed in the case where individuals or groups differ along a single dimension, say income. Further, it has always been implicitely assumed that individuals were not different among themselves and therefore no specific attention should be paid to the identity and caracteristics of the donor or recipient of a transfer besides their levels of income. The extension of the theory to population of individuals which differ according to many variables is difficult and is far from being achieved despite some recent promising developments.

One line of investigation consists in considering stochastic orders, like those considered in the one dimensional case. These orders are orders on the set (or subsets) of probability distributions over $\Re^m$ where m denotes the number of attributes (characteristics, commodities,...) which are considered. Let F be the distribution function of any such joint distribution on $\Re^m$ and U be a function from $\Re^m$ into $\Re$. Integrating by parts to rearrange the expression

$$\underbrace{\int_{\Re}\int_{\Re}.....\int_{\Re}}_{\text{m times}} U(x_1, x_2, ..x_m)F(dx_1, dx_2, ..dx_m)$$

leads to several stochastic orders (Atkinson and Bourguignon (1982) and Levy and Paroush (1974) are two representive contributions following that line of investigation). For instance, if $m = 2$, F is absolutely continuous with respect to the Lebesgue measure on $\Re^2$ with density f and support in the unit square and U has high order derivatives as much as needed, then :

$$\int_0^1\int_0^1 U(x_1,x_2)f(x_1,x_2)dx_1dx_2 = U(1,1)\int_0^1\int_0^1 f(x_1,x_2)dx_1dx_2 - \int_0^1 \frac{\partial U}{\partial x_1}(x_1,1)F_1(x_1)dx_1 - \int_0^1 \frac{\partial U}{\partial x_2}(1,x_2)F_2(x_2)dx_2 + \int_0^1\int_0^1 \frac{\partial^2 U}{\partial x_1 \partial x_2}(x_1,x_2)F(x_1,x_2)dx_1dx_2$$

where F_i is the marginal distribution on the i^{th} component. The first term will not play any role. The second and third terms bring us back in the one dimensional case and we can apply what we know separately on the two marginals. The last term is really attached to the two dimensional setting. Indeed, take another distribution G with density g such that $F_1 = G_1$ and $F_2 = F_2$. Then :

$$\int_0^1\int_0^1 U(x_1,x_2)f(x_1,x_2)dx_1dx_2 - \int_0^1\int_0^1 U(x_1,x_2)g(x_1,x_2)dx_1dx_2 = \int_0^1\int_0^1 \frac{\partial^2 U}{\partial x_1 \partial x_2}(x_1,x_2)\left(F(x_1,x_2) - G(x_1,x_2)\right)dx_1dx_2$$

The sign of this expression will depend on the respective intensities of correlation of F and G. Under the constraint that the marginals are the same, the condition :

$$F(x_1,x_2) - G(x_1,x_2) \leq 0$$

can be shown to represent indeed the property that F exhibits less correlation than G. Under this condition and the the condition that the sign of the second cross derivative $\frac{\partial^2 U}{\partial x_1 \partial x_2}(x_1,x_2)$ is negative, we deduce that the above integral is positive. This condition on U known as supermodularity leads to several stochastic orders depending upon which assumptions we consider on the class of functions $\mathcal{U}$. For instance, if we assume $\frac{\partial U}{\partial x_1}(x_1,x_2) \geq 0$, $\frac{\partial U}{\partial x_2}(x_1,x_2) \geq 0$, $\frac{\partial^2 U}{\partial x_1 \partial x_2}(x_1,x_2) \leq 0$, we obtain :

$$\int_{\Re^2} U(x_1,x_2)F(dx_1,dx_2) \geq \int_{\Re^2} U(x_1,x_2)G(dx_1,dx_2) \text{ for all } u \in \mathcal{U}$$

iff

$$F(x_1,x_2) - G(x_1,x_2) \leq 0 \text{ for all } (x_1,x_2) \in [0,1]^2$$

If instead, $\mathcal{U}$ consists of all utility functions satisfying, in addition to the above conditions, the extra conditions $\frac{\partial^2 U}{\partial x_1^2}(x_1,x_2) \leq 0$ and $\frac{\partial^2 U}{\partial x_1^2}(x_1,x_2) \leq 0$ (these functions are called functions with nondecreasing increments), then :

$$\int_{\Re^m} U(x_1,x_2)F(dx_1,dx_2) \geq \int_{\Re^m} U(x_1,x_2)G(dx_1,dx_2) \text{ for all } u \in \mathcal{U} \text{ iff :}$$

$$\begin{aligned} F(x_1,x_2) - G(x_1,x_2) &\leq 0 \text{ for all } (x_1,x_2) \in [0,1]^2 \\ \int_0^{x_1} (F_1(u_1) - G_1(u_1))\, du_1 &\geq 0 \text{ for all } x_1 \in [0,1] \\ \text{and } \int_0^{x_2} (F_2(u_2) - G_2(u_2))\, du_2 &\geq 0 \text{ for all } x_2 \in [0,1] \end{aligned}$$

· As soon as we recognize that the integral $\int_0^1 \int_0^1 \frac{\partial^2 U}{\partial x_1 \partial x_2}(x_1,x_2)F(x_1,x_2)dx_1dx_2$ has a structure analogous to $\int_0^1 \int_0^1 U(x_1,x_2)f(x_1,x_2)dx_1dx_2$, we can perform one more round of integration by parts to obtain some more selective stochastic orders. This routine leads however to families $\mathcal{U}$ of functions entailing sign conditions on their third and fourth partial cross derivatives which are not immediate to interpret. Further, the stochastic orders resulting from these families are not themselves immediate to analyse like Lorenz dominance. Note that the above conditions become much more intricate when we move to more than two attributes.

· Le Breton (1986) point out the relevance of Brunk (1964) and Fan and Lorentz (1954) to show that as soon as the two distributions exhibit perfect positive correlation, then, the stochastic order attached to the class of functions having negative cross and direct second order partial derivatives is simply the intersection of the two Lorenz orders.

· The above orders can be examined in restriction to the class of discrete distributions. We can even restrict our attention, as we did in our examination of the unidimensional Lorenz order, to the class of distributions $\frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}$ where $x_i \in \Re^m$ for all $i = 1,...,n$. A distribution can be identified to a $n \times m$ matrix $X = (x_{ik})\, 1 \leq i \leq n, 1 \leq k \leq m$ where x_{ik} denotes the amount of attribute k received by individual i. There are several ways to approach the problem. The Hardy, Littlewood and Polya 's theorem suggests to look at the problem either from the perspective of linear stochastic operators (describing composition of transfers), or from the perspective of dominance, or finally from the perspective of the class of individual utility functions which is considered. Among the many contributions to this line of research, those of Koshevoy (1995,1998) (see also Mosler and Koshevoy (1997)) are quite central.

· Consider an arbitratry multivariate probability distribution F over $\Re^m_+$with finite and strictly positive first moments.Let $\mu_k \equiv \int_{\Re^m_+} x_k F(dx)$ for all $k = 1,...,m$ and $\psi_F(x) = \left(\frac{x_1}{\mu_1},....,\frac{x_m}{\mu_m}\right)$. The Lorenz zonoid of F is the set :

$$LZ(F) \equiv \left\{ z \in \Re_+^{m+1} : z = (z_0, z_1, ..., z_m) = \xi(h) \text{ with } h : \Re_+^{m+1} \to [0,1] \text{ measurable} \right\}$$

where :

$$\xi(h) \equiv \left(\int_{\Re_+^m} h(x)F(dx), \int_{\Re_+^m} h(x)\psi_F(x)\, F(dx) \right)$$

The Lorenz zonoid has the following interpretation. Every unit of the population is assigned a vector x in $\Re_+^m$ and holds therefore a portion $\psi_F(x)$ of the mean endowment. A given measurable function $h : \Re_+^{m+1} \to [0,1]$ may be considered to be a selection of some part of the population : of all those units that have endowment vector x (or portion vector $\psi_F(x)$), the percentage h(x) is selected. Thus, $\int_{\Re_+^m} h(x)F(dx)$ is the size of the population selected by h, and $\int_{\Re_+^m} h(x)\psi_F(x)\, F(dx)$ amounts to be the total portion vector held by this population.

· The nature of the Lorenz zonoid is quite easy to vizualize in the one dimensional case. The dual Lorenz function $\overline{L_F}$ defined by :

$$\overline{L_F}(t) = 1 - L_F(1-t) \text{ for all } t \in [0,1]$$

describes the respective portions of the endowment held by the individuals ordered from the richest to the poorest; for instance t=0.1 corresponds now to the highest decile. The Lorenz zonoid is the convex set whose frontiers are the standard and dual Lorenz functions. The vertical section through t corresponds to the set of feasible shares held by subpopulations representing a fraction t of the total population.

· When the distribution is of the discrete type discussed above i.e. described through a matrix X, then :

$$LZ(F_X) = \left\{ z \in \Re_+^{m+1} : z = \sum_{i=1}^{n} h(i)\widetilde{x}_i,\ 0 \leq h(i) \leq 1 \text{ for all } i = 1, ..., n \right\}$$

where :

$$\widetilde{x}_i = \left(\frac{1}{n}, \frac{x_{i1}}{\sum_{j=1}^n x_{j1}}, ..., \frac{x_{im}}{\sum_{j=1}^n x_{jm}} \right) \text{ for all } i = 1, ..., n$$

or equivalently, is the sum of the line segments $\sum_{i=1}^n h(i)\,[0, \widetilde{x}_i]$. $LZ(F_X)$ is a zonotope contained in the unit cube of $\Re^{m+1}$.

· As already explained, for a given $(z_1, ..., z_m) \in Z(F)$ where :

$$Z(F) = \left\{ y \in \Re_+^m : y = (y_1, ..., y_m) = \int_{\Re_+^m} h(x)\psi_F(x)\, F(dx) \text{ with } h : \Re_+^{m+1} \rightarrow [0,1] \text{ measurable} \right\}$$

$z = (z_0, z_1, ..., z_m) \in LZ(F)$ if and only if z_0 is in the closed interval between the smallest and the largest percentage of the population by which the portion vector $(z_1, ..., z_m)$ is held. This leads to the definition of an inverse Lorenz function L_F :

$$L_F : Z(F) \rightarrow [0,1]\,,\ L_F(y) = Max \{t \in [0,1] : (t, y) \in LZ(F)\}$$

Its graph is the Lorenz surface of F. The following definition is due to Koshevoy.

· The distribution G is not less than the distribution F in the Lorenz zonoid (or multivariate Lorenz) order if $LZ(F) \subset LZ(G)$ holds. It is equivalent to ask that $Z(F) \subset Z(G)$ and $L_F(x) \leq L_G(x)$ for all $x \in Z(F)$. For a given p in $\Re^m$, let F_p be the random variable $x \cdot p$. Koshevoy has demonstrated the following important equivalence :

· $LZ(F) \subset LZ(G)$ iff for all p in $\Re^m$, the Lorenz curve of F_p is above the Lorenz curve of G_p.

· This theorem can be interpreted in terms of prices and expenditures. Given a price vector p, and two distributions F and G of m commodities, F_p and G_p are the corresponding distributions of expenditures. Koshevoy's theorem just says that F has less multivariate inequality than G iff F_p has less univariate inequality than G_p in the sense of Lorenz dominance.

· A shorter proof of Koshevoy's theorem is provided by Dall'Aglio and Scarsini (2001). Note that the price vectors are not restricted to belong to the positive orthant. When p is restricted to belong to $\Re_+^m$, we obtain an order introduced by Kolm (1977) which has not yet been characterized adequately but which is more selective than the Lorenz multivariate order. Koshevoy has also developed efficient algorithms to compare two zonotopes and compared his order to some previous multivariate orders like for instance the one proposed by Taguchi (1972a,b).

· To the best of my knowledge, the geometric approach has not been widely explored. We could consider for instance say that a nxm matrix X exhibits less inequality than a $n \times m$ matrix Y iff : For all $k = 1, ..., m$, there exists a bistochastic matrix B_k such that: $x_{.k} = B_k y_{.k}$ where $x_{.k} = (x_{1k}, ..., x_{nk})$.

This order is quite controversial as it allows to disconnect the transfers across components : it is easy to produce examples where we increase the aggregate inequality. We could then impose the same bistochastic matrix to all attributes. This order has been investigated by Rinott (1973)

Bivariate Income Distributions : Horizontal Equity and Taxation

This question is related in some of its dimensions to the question examined in the previous section. In particular, the question of horizontal equity is formally related to the question of correlation between two distributions. Ordering of taxation schemes according to progressivity is the subject of a seminal contribution by Jakobsson (1976).

References

1. Generalities and Surveys

M. Le Breton *Essais sur les Fondements de l'Analyse Economique de l'Inegalité*, Thèse pour le Doctorat d'Etat, Rennes, 1986.

A.W. Marshall and I. Olkin *Inequalities : Theory of Majorization and its Applications*, Academic Press, New York, 1979.

A.K. Sen *On Economic Inequality*, Clarendon Press, Oxford, 1973.

2.The Hardy, Littlewood and Polya's Theorem

P. Dasgupta, A.K. Sen and D. Starrett, Notes on the measurement of inequality, Journal of Economic Theory, 6 (1973), 180-187.

G.H. Hardy, J.E. Littlewood and G. Polya "Some simple inequalities satisfied by convex functions", *Messenger of Mathematics*, 58 (1929), 145-152.

G.H. Hardy, J.E. Littlewood and G. Polya *Inequalities*, Cambridge University Press, Cambridge, 1934.

S.C. Kolm, The optimal production of social justice, in *Public Economics*, H. Guitton and J. Margolis (Eds), McMillan, London, 1969.

3. Schur Convexity and Inequality Measurement

M.L. Eaton and M.D. Perlman, Reflexion groups, generalized Schur-functions and the geometry of majorization, *Annals of Probability,* 5 (1977), 829-860.

J.E. Foster and E.A. Ok, "Lorenz dominance and the variance of logarithms", *Econometrica*, 67 (1999), 901-907.

M. Le Breton, Approximation theorems in inequality measurement, Mimeo, 2006a.

M. Le Breton, A. Trannoy and J.R. Uriarte, Topological aggregation of inequality preorders, *Social Choice and Welfare*, 2 (1985), 119-129.

A.W. Marshall and I. Olkin, Majorization in multivariate distributions, *Annals of Statistics*, 2 (1974), 1189-1200.

I. Ostrowski, Sur quelques applications des fonctions convexes et concaves au sens de I Schur, *Journal de Mathématiques Pures et Appliquées*, 31 (1952), 253-292.

4. Stochastic Dominance

A.B. Atkinson, On the measurement of inequality, *Journal of Economic Theory*, 2 (1970), 244-263.

P.C. Fishburn, Convex stochastic dominance with continuous distribution functions, *Journal of Economic Theory*, 7 (1974), 143-158.

J.E. Foster and A. Shorrocks, Transfer sensitive inequality measures, *Review of Economic Studies*, 54 (1987), 485-497.

J.L. Gastwirth, A general definition of the Lorenz curve, *Econometrica*, 39 (1971), 1037-1039.

J. Karamata, Sur une inégalité relative aux fonctions convexes, *Publications Mathématiques de l'Université de Belgrade*, 1 (1932), 145-148.

M. Le Breton and E. Peluso, Third-degree stochastic dominance and the von-Neumann-Morgenstern independence property, Mimeo, 2006.

5. Continuous Distributions

K.M. Chong and N.M. Rice, Equimeasurable rearrangements of functions, *Queen's papers in Pure and Applied Mathematics*, 28 (1971), Queen's University, Kingston.

N. Dunford and J.T. Schwartz, *Linear operators, Part 1: General Theory*, Intersciences Publishers Inc, New-York, 1966.

A. Grothendiek, Réarrangements de fonctions et inégalités de convexité dans les algèbres de Von Neumann munies d'une trace, *Séminaire Bourbaki*, 113 (1955), 1-13.

J.L Kelley and I. Namioka, *Linear topological spaces*, D. Van Nostrand Company Inc, New-York, 1963.

M. Le Breton, A Mackey version of a theorem of Hardy, Littlewood and Polya on $L^{\infty}(0,1)$, Mimeo, 2006b.

A.A.J. Luxembourg, Rearrangement invariant Banach functions spaces, in: Proceedings of the Symposium in Analysis, *Queen's Papers in Pure and Applied Mathematics*, 10 (1967), 83–114.

K.R. Pathasarathy 1967, *Probability measures on metric spaces*, Academic Press, New-York, 1967.

J.V. Ryff, Orbits of L^1-functions under doubly stochastic transformations, *Transactions of the American Mathematical Society*, 117 (1965), 92-100.

J.V. Ryff, Extreme points of some convex subsets of $L^1(0,1)$, *Proceedings of the American Mathematical Society*, 18 (1967), 1026-1034.

J.W. Ryff, On the representation of doubly stochastic operators, *Pacific Journal of Mathematics*, 13 (1963), 1379–1386.

J.W. Ryff, 1970, Measure preserving transformations and rearrangements, *Journal of Mathematical Analysis and Applications*, 31 (1970), 449–458.

D. Schmeildler, A. Bibliographical note on a theorem of Hardy, Littlewood and Polya, *Journal of Economic Theory*, **20 (1979)**, 125–128.

6. Multivariate majorizations : The Koshevoy's Zonotope

A.B. Atkinson and F. Bourguignon, The comparison of multidimensioned distributions of economic status, *Review of Economic Studies*, 49 (1982), 183-201.

H.D. Brunk, Integral inequalities for functions with nondecreasing increments, *Pacific Journal of Mathematics*, 14 (1964), 783-793.

M. Dall'Aglio and M. Scarsini, When Lorenz met Lyapunov, *Statistics and Probability Letters*, 54 (2001), 101-105.

K. Fan and G.G. Lorentz, An integral inequality, *American Mathematical Monthly*, 61 (1954), 626-631.

S.K. Kolm, Multidimensional egalitarianisms, *Quarterly Journal of Economics*, 91 (1977), 1-13.

G. Koshevoy, Multivariate Lorenz majorization, *Social Choice and Welfare*, 12 (1995), 93-102.

G. Koshevoy, The Lorenz zonotope and multivariate majorizations, *Social Choice and Welfare*, 15 (1998), 1-14.

G. Koshevoy and K. Mosler, The Lorenz zonoid of a multivariate distribution, *Journal of the American Statistical Association*, 91 (1996), 873-882.

H. Levy and J. Paroush, Towards multivariate efficiency criteria, *Journal of Economic Theory*, 7 (1974), 129-142.

Y. Rinott, Multivariate majorization and rearrangement inequalities with some applications to probability and statistics, *Israel Journal of Mathematics*, 15 (1973), 60-77.

T. Taguchi, On the two-dimensional concentration surface and extensions of concentration coefficient and pareto distribution to the two-dimensional case : I, *Annals of the Institute of Statistical Mathematics*, 24 (1972a), 355-382.

T. Taguchi, On the two-dimensional concentration surface and extensions of concentration coefficient and pareto distribution to the two-dimensional case : II, *Annals of the Institute of Statistical Mathematics*, 24 (1972b), 599-619.

7. Bivariate Income Distributions : Horizontal Equity and Taxation

L.G. Epstein and S.M. Tanny, Increasing generalized correlation : a definition and some economic consequences, *Canadian Journal of Economics*, 13 (1980), 16-34.

U. Jakobsson, On the measurement of the degree of progression, *Journal of Public Economics*, 5 (1976), 161-168.

M. King, "An index of inequality with applications to horizontal equity and social mobility", *Econometrica*, 51 (1983), 99-115.