Comparative analysis of inequality of opportunity: description versus explanation

Nathalie Bulle *

CNRS, Methods of Sociological Research Group (*GEMAS*)

Paris

France

Abstract

Explanations of changes in social, geographic, or other subgroups' relative access to a discrete good G are not well supported by any classical measure of inequality, even by existing 'margin-free' indices. The purpose of this paper is to propose a method allowing analysis of generative mechanisms of inequalities. The inequality coefficient defined here permits to compare inequality in the selection process underlying observed opportunities and is independent of marginal distributions (overall access to good G and fraction of the population in the various subgroups). This coefficient offers, in general circumstances, a new interpretation for one familiar statistics of association comparing the differences in proportion of the columns of the 2×2 contingency table.

Keywords : *Inequality of opportunity, selection process, explanation, segregation, joint density.*

1. Explaining inequality of opportunity

1.1 *The difficulty to take on the problem of explanation*

The problem raised here is central to any field where comparisons of inequality of access to a discrete good G underpin explanations of societal differences or changes over time, such as labor discrimination, schooling inequalities, social mobility, urban geography, etc. Our claim is that explanations cannot be empirically sustained until one points to discrepancies in social processes generating observed inequalities. For instance, such processes may be highly differentiated between subgroups, but their effects on actual relative opportunity of access to G may be more or less acute, depending on G's scarcity.

* *E-mail*: nathalie.bulle@wanadoo.fr

Journal of Interdisciplinary Mathematics
Vol. 12 (2009), No. 4, pp. 565–588

Despite the great variety of inequality of opportunity measures[1], there is none which addresses this very issue. Additionally, in this comparative perspective, differences in the magnitude of inequality measures have to be independent of the whole rate of access to G and distribution of subgroups within the population. In every domain requiring interpretations of inequalities, one focus has been to build "margin-free" indices[2]. As is argued here, the measures developed do not tackle explanation until the present.

Let us take a numerical example. Suppose we have a population divided between two subgroups S1 and S2. Individuals from the first subgroup S1 have less chances to accede to a good G (a definite level of education for instance) than individuals from the second subgroup S2. Nevertheless the situation changes with the opening of access to G and requires interpretation. Tables 1 and 2 describe the respective distributions regarding access to G of individuals from the two subgroups in two different periods P1 and P2.

Table 1

Distribution of access to G of individuals from subgroups S1 **and** S2 **in period P1**

	Access to G	No access to G	
S2	100	450	550
S1	10	540	550
	110	990	1100

Table 2

Distribution of access to G of individuals from subgroups S1 **and** S2 **in period P2**

	Access to G	No access to G	
S2	340	210	550
S1	100	450	550
	440	660	1100

[1]Social science research has made clear, especially with Atkinson's [1], and Kolm's [6] works on distributions such as income, that value judgments concerning the nature of inequality are inherent in any measure of inequality. See Hutchens [5] for extension to analysis of inequality in the distribution of people across groups.

[2]For segregation analysis, see Duncan and Duncan [3], Blackburn, Siltanen and Jarman [2], and Watts [9].

The access rate to G of individuals from S1 raises from 1,8% to 18,2% between P1 and P2, while the access rate of individuals from S2 raises from 18,2% to 61,8%. The ratio of proportions, comparing access rates, diminishes from 10 to 3,4. From this first point of view, inequality of opportunity decreases. Nevertheless, such a ratio cannot express the same intensity of inequality independently of marginal distributions. For instance, if the fraction of the population having access to G still increases, the ratio of proportion can remain stable (with an access rate of individuals from S2 ten times superior to that of individuals from S1) only if this fraction remains inferior to 55% (all the 550 individuals from S2 access to G and 55 from S1). On the other hand, the exclusion rate of individuals from S2 diminishes from 81,8% to 38,2%, while the exclusion rate of individuals from S1 decreases only from 98,2% to 81,8%. Thus, the ratio of exclusion rates increases from 1,2 to 2,1. From this second point of view, inequality of opportunity increases. As it was the case for the ratio of access rates, the ratio of exclusion rates cannot express the same intensity of inequality independently of marginal distributions. For instance, if the fraction of the population having access to G decreases, the ratio of exclusion rates can remain stable (with an exclusion rate of individuals from S1 1,2 time superior to that of individuals from S2) only if this fraction remained superior to 8,3% (all the 550 individuals from S1 are excluded from G and $458 = 550/1,2$ from S2). The odds ratio, which represents the product of the two previous ratios of proportions, diminishes between P1 and P2 from 12 to 7,3. From this third point of view, inequality of opportunity decreases. Odds ratios, unlike ratios of proportions, are insensitive to marginal distributions. This insensitivity is usually characterized by the fact that their value does not change when we multiply throughout any raw or column of a contingency table by a constant. For this reason, actual trends of research rely on log-linear models using odds ratios[3] in order to compare changes in inequalities from one population to another. However insensitivity to marginal variation does not mean that odds ratios are good tools for explaining differences in inequality of opportunity between populations. For instance, in the case of educational inequalities, odds ratios provide a measurement of links between social categories and educational attainments net of marginal distributions. Nevertheless,

[3]Odds ratios are ratios of two odds (ratios of the number of people incurring an event to the number of people who have non-events): they compare here individuals' opportunities from a social group C_k to individuals' opportunities from another social group C_i of access versus no access to G.

they do not measure variations in the selection process which generate observed inequalities. They do not take into account the changes in meaning of educational attainments in terms of selection constraints while the distribution of schooling levels varies within populations.

The same problem arises in research on social mobility. The concepts of absolute versus relative rates of mobility (founded on odds ratios calculations) were substituted to the concepts of structural versus circulation mobility. Nevertheless these concepts did not permit one to solve the previous problem at stake, i.e., to control for "forced" mobility which is constrained by discrepancies in occupational structure to analyze intrinsic variations in the social processes underlying occupational opportunity.[4]

More generally, an index's insensitivity to the distributions of margins allows margin-free comparisons regarding the precise aspect of inequalities the index measures. It does not necessarily allow comparisons of social processes generating observed inequalities. Thus, interpretation raises implicitly theoretical assumptions usually without any clear justification of the links between deductions and empirical measures. This causes recurrent disagreements regarding conclusions about the social processes in question. Explanation needs an index aiming at comparing inequality in the selection process underlying access to G.

We first propose to reformulate the concept of insensitivity to marginal distributions in order to open it to new issues, in particular explanation-oriented measures. We then introduce the concept of inequality in the selection process. The following is dedicated to the development of a measure, independent of marginal distributions, applied to this aspect of inequality.

1.2 *Insensitivity to marginal distributions: the necessity of a broad definition*

Table 3
Distribution of access to G of individuals from complementary C_i and $\overline{C}_i$ subgroups

	Access to G	No access to G	
$\in \overline{C}_i$	α_{ij}	β_{ij}	$\alpha_{ij} + \beta_{ij} = 1 - m_i$
$\in C_i$	γ_{ij}	δ_{ij}	$\gamma_{ij} + \delta_{ij} = m_i$
	$\alpha_{ij} + \gamma_{ij} = x_j$	$\beta_{ij} + \delta_{ij} = 1 - x_j$	1

[4] See Sorensen [8], Sobel [7], Harrison [4].

Insensitivity to marginal distributions: a definition. Let us consider the 2×2 contingency table showing the relationship between access to a specific good G and belonging to a defined C_i social subgroup, at a given stage x_j of G's diffusion (cf. Table 3). When insensitivity to marginal distributions is considered, we mean that the intensity of inequality measured by the coefficient at stake does not depend on the margins' values of the contingency table. One condition is that in each context defined by the contingency table's margins, the same magnitude of inequality may be observed. Thus, one further condition is that the coefficient's extreme potential values must not depend on margins.

In a more specific, but classical sense, insensitivity to marginal distributions means that the index does not change if any row or column of the contingency table is multiplied throughout by a constant. The index must be based on ratios so that changes in margins do not affect the meaning of the index' magnitude in terms of relative inequality, i.e., the index value remains stable if proportionalities are respected. Note that, according to the broad definition given above, it must be possible to preserve proportionalities whatever are the margins, so that the coefficient's extreme potential values must not depend on margins.

Discussion. In the restrictive, but conventional, sense of the concept of insensitivity to marginal distributions, changes in rows or columns are considered independently of one another. The broad sense of the concept may apply to other aspects of inequality where the status quo in terms of inequality involves interrelations between rows and columns.

1.4 *Inequality in the selection process*

Generally, inequality of access to a discrete good G can be ascribed to the following:

(1) Net results of the selection process in a broad sense. This concerns the effects of all of the factors influencing individuals' opportunities of access to G, but taking no account of individuals' actual access.

(2) Diffusion of G in society, i.e., the overall proportion of individuals who accede to G. Inequality with respect to (1) is inequality in the selection process defined as follows.

Inequality of opportunity in the selection process: a definition. We define inequality in the selection for access to a discrete good G as a measure permitting to compare the results of the selection process for access to G in a reference mark independent of variations of overall access to G (such as the deciles, centiles, etc.).

Discussion. In order to compare selection process results in different contexts of good G's diffusion in society, we suppose that access to G is the result of an ordering of the whole population and of allocation of G to the best ranked individuals, until equalizing overall rate of access to G. Such ranking permits us to refer to a fixed reference mark of relative opportunity, such as the percentiles ranks of the population. Inequality of subgroups distributions in such a reference mark represents what we define as "inequality in the selection process".

The present research rests on the assumption that apprehending this aspect of inequality is essential for explaining observed inequalities and that it is possible to propose a margin-free index which solves this problem.

2. Distribution of opportunity in the selection process

2.1 *Definition of joint densities $\widetilde{f}(x, C_i)$ describing opportunity distribution in the selection process*

Definition. It is assumed that access to a discrete good G has been derived from a latent continuous variable g which theoretically permits one to rank the whole population according to individual opportunities of access. By convention, a lower value of g will mean a greater opportunity of access to G. The variable g can be interpreted as a distance to G revealing overall effect of various handicaps in the process of access to G.

The population is of mass 1 and divided in k different social subgroups: C_i is social subgroup i and $\overline{C}_i$ is the complementary aggregated subgroup of C_i within the population.

The k joint densities are $f(g, C_i)$ and the k joint cumulative distributions are $F(g, C_i)$ where C_i represents a nominal variable distinguishing individuals from C_i in the whole population.

Assume that the support of g is $[\underline{g}, \overline{g}]$, its density is $h(g)$ and its cumulative distribution is $H(g)$ (supposed to be strictly increasing so that its inverse function exists).

Define $x = H(g)$ the $100\,x$-th percentile of the distribution of g.

x represents a continuous variable varying from 0 to 1.

Define the k joint cumulative distributions as $\widetilde{F}(x, C_i) = F(H^{-1}(x), C_i)$.

Define $\widetilde{f}(x, C_i) = \frac{d}{dx}\widetilde{F}(x, C_i)$ the k joint densities.

2.2. *Properties of joint densities* $\widetilde{f}(x, C_i)$

For each curve $\widetilde{f}(x, C_i)$ we have the following properties:

2.2.1. $\widetilde{f}(x, C_i) \geq 0$ *since it is a density.*

2.2.2. $\widetilde{f}(x, C_i) + \widetilde{f}(x, \overline{C}_i) = 1$.

Proof. By definition, $\widetilde{F}(x)$ represents the fraction of the whole population in the first $(100x)$ % whose opportunity of access to G is the greatest, that is x:

$$\widetilde{F}(x) = \widetilde{F}(x, C_i) + \widetilde{F}(x, \overline{C}_i) = x \text{ so that } \widetilde{f}(x) = \widetilde{f}(x, C_i) + \widetilde{f}(x, \overline{C}_i) = 1\,.$$

2.2.3. $\widetilde{f}(x, C_i)$ *is traced within a square.*

Proof. x varies between 0 and 1. According to 2.2.1 and 2.2.2, $\widetilde{f}(x, C_i)$ varies between 0 and 1.

2.2.4. $\int_0^1 \widetilde{f}(x, C_i)dx = m_i$

Proof. $\widetilde{F}(1, C_i)$ represents the fraction of the whole population in C_i, that is m_i:

$$\widetilde{F}(1, C_i) = \int_0^1 \widetilde{f}(x, C_i)dx = m_i$$

2.2.5. *When the distribution of C_i is independent of x, we have $\widetilde{f}(x, C_i) = m_i$.*

Proof. From Bayes's identity: $\widetilde{f}(x, C_i) = \widetilde{f}(x/C_i) \times m_i$

Then, in case of independence of x and C_i we have:

$\widetilde{f}(x, C_i) = \widetilde{f}(x) \times m_i$ and from 2.2.2 $\widetilde{f}(x) = 1$ so that $\widetilde{f}(x, C_i) = m_i$

3. Overall coefficient of inequality of opportunity in the selection process

3.1 *Construction of the* $\widetilde{d}(x, C_i)$ *straight line segments*

We now consider virtual constructs in the aim of comparing opportunity in the selection process for access to a discrete good G. Following 2.1, it is assumed that the access of individuals from various C_i subgroups to G, at each stage x_j of G's diffusion, has been derived from underlying continuous joint densities $\widetilde{f}(x, C_i)$. Then we assume that the curves

$\widetilde{f}(x, C_i)$ are:

(1) Either broken line segments in cases where $\widetilde{f}(x, C_i) = 0$ or $\widetilde{f}(x, C_i) = 1$ on a segment of $[0, 1]$. Define $\widetilde{d}(x, C_i)$ the straight line segment such that $\widetilde{f}(x, C_i) = \widetilde{d}(x, C_i)$ where $0 < \widetilde{f}(x, C_i) < 1$.

(2) Or straight line segments so that $\widetilde{f}(x, C_i) = \widetilde{d}(x, C_i)$ on $[0, 1]$.

For each 2×2 contingency table showing the relationship between accessing to G and belonging to a social subgroup C_i, at a given stage x_j of G's diffusion, there exists a virtual joint density $\widetilde{f}(x, C_i)$ as defined above such that access of individuals from C_i to G could have been derived from $\widetilde{f}(x, C_i)$. Such existence can be showed graphically.

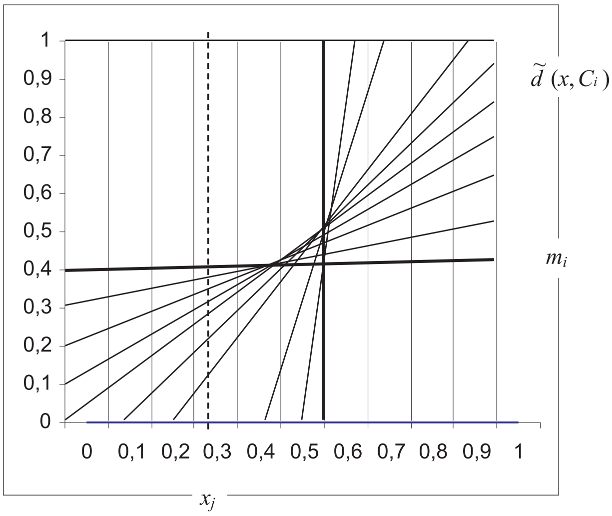

Figure 1
Family of virtual joint densities $\widetilde{f}(x, C_i)$

$\widetilde{d}(x, C_i)$ is the straight line segment defined in the square where $\widetilde{f}(x, C_i)$ is inscribed such that $\widetilde{f}(x, C_i) = \widetilde{d}(x, C_i)$ where $0 < \widetilde{f}(x, C_i) < 1$; m_i is the fraction of the whole population in C_i and x_j the fraction of the whole population accessing to G. According to 2.2.4, we have $\int_0^1 \widetilde{f}(x, C_i)dx = m_i$. The family of joint densities $\widetilde{f}(x, C_i)$ that fit this condition is represented on Figure 1, in the case where the straight line segments $\widetilde{d}(x, C_i)$ has a positive slope. The complementary family of straight line segments $\widetilde{d}(x, C_i)$ with negative slope is symmetric to this family to the axis $x = 1/2$. As showed on Figure 1, at any stage x_j of G's diffusion, for each possible distribution of access to G between C_i and

$\overline{C}_i$, there exists a virtual joint density $\widetilde{f}(x, C_i)$ which can be associated to this distribution.

Definition of $(\widetilde{a}_i)$ **coefficients.** Define the coefficients $(\widetilde{a}_i)$ as the slopes of the $\widetilde{d}(x, C_i)$ straight line segments, where $0 < \widetilde{f}(x, C_i) < 1$.

3.2 *General properties of the* $(\widetilde{\boldsymbol{a}}_i)$ *coefficients*

3.2.1. *If the slope of* $\widetilde{d}(x, C_i)$ *is* $\widetilde{a}_i$, *and* $\overline{C}_i$ *the complementary aggregated subgroup of* C_i *the slope of* $\widetilde{d}(x, \overline{C}_i)$ *is* $(-\widetilde{a}_i)$.

Proof. According to 2.2.2, $\widetilde{f}(x, C_i) + \widetilde{f}(x, \overline{C}_i) = 1$. Thus, the curves are symmetric with respect to the axis $y = 1/2$.

3.2.2. $\widetilde{a}_i < 0$ (> 0) *iif for any* x_j *defining the whole rate of access to* G, *the access rate of individuals from* C_i *is higher* (*lower*) *than overall access to* G, x_j.

Proof. Let r_{ij} be the fraction of the social subgroup C_i accessing to G at the stage x_j of G's diffusion, we have: $r_{ij} = \frac{1}{m_i} \times \int_0^{x_j} \widetilde{f}(x, C_i) dx$.

General case: $\widetilde{d}(x, C_i) = \widetilde{f}(x, C_i)$ on the whole interval $[0, 1]$ so that $\widetilde{d}(x, C_i) = \widetilde{a}_i \times x + m_i - \widetilde{a}_i/2$: $\forall\ x_j \in [0, 1]$, $r_{ij} > x_j \Leftrightarrow \frac{\widetilde{a}_i}{2m_i}(x_j - 1) > 0$.

Proofs are trivial in the specific cases where $\widetilde{f}(x, C_i) = 0$ or $\widetilde{f}(x, C_i) = 1$ on a segment of $[0, 1]$.

3.2.3. *The* $(\widetilde{a}_i)$ *coefficients are insensitive to marginal distributions in the broad sense* (*defined in* 1.2).

Proof. Insensitivity with regard to x_j variations stems from definition of $\widetilde{a}_i$ in a reference mark independent of variations of overall access to G, x_j.

According to its definition in 3.1, $\widetilde{a}_i$ characterizes the same intensity of inequality in the whole set of parallel straight line segments of slope $\widetilde{a}_i$-representing inequality of opportunity distributions (on the segment of $[0, 1]$ where $0 < \widetilde{f}(x, C_i) < 1$) within the square where $\widetilde{f}(x, C_i)$ is inscribed. From 2.2.5, we deduce that in a context in which inequality in the selection process is independent of C_i, $\widetilde{d}(x, C_i)$ would be fitted to the horizontal straight line segment $y = m_i$ on $[0, 1]$ so that $\widetilde{a}_i = 0$. In a context with full inequality of opportunity, the disadvantaged subgroup C_i would tend to make up the bottom intervals inter-percentiles of the population so that $\widetilde{d}(x, C_i)$ would tend to be aligned to the vertical axis $x = 1 - m_i$ (according to 3.2.2 $\widetilde{a}_i > 0$). In the reverse situation, the whole subgroup C_i would tend to make up the top intervals inter-percentiles of the population so that the $\widetilde{d}(x, C_i)$ would tend to be aligned to the vertical axis $x = m_i$ (with $\widetilde{a}_i < 0$). Consequently, whatever are the values of m_i and $x_j, \widetilde{a}_i \in]-\infty, +\infty[$. Coefficient $\widetilde{a}_i$'s values are not constrained by marginal values, i.e., coefficient $\widetilde{a}_i$ is insensitive to marginal distributions.

3.3 *Specific properties of the* $(\widetilde{a}_i)$ *coefficients in the general case*

General case:

$$\forall\, i \;\; \frac{|\widetilde{a}_i|}{2} \leq m_i \leq 1 - \frac{|\widetilde{a}_i|}{2} \tag{1}$$

None of the $\widetilde{d}(x, C_i)$ straight lines segments (defined in 3.1) intersect the top or the bottom of the square where the $\widetilde{f}(x, C_i)$ curves are inscribed (see demonstration in annex). Thus $\forall\, i$, $\widetilde{d}(x, C_i) = \widetilde{f}(x, C_i)$ on the whole interval $[0, 1]$.

3.3.1. *According to* (1),

$$-1 \leq \widetilde{a}_i \leq +1 \tag{2}$$

Proof. According to (1), $\frac{|\widetilde{a}_i|}{2} \leq m_i$ and $\frac{|\widetilde{a}_i|}{2} \leq 1 - m_i$ so that $|\widetilde{a}_i| \leq 1$.

3.3.2. *If subgroups are aggregated,* $C_k = \cup C_i$ *the slope of* $\widetilde{d}(x, C_k)$ *is equal to the sum of the slopes of*

$$\widetilde{d}(x, C_i) : \widetilde{a}_k = \sum \widetilde{a}_i. \tag{3}$$

Proof. As defined in 3.1, the joint density function $\widetilde{f}(x, C_k)$ underlying access to G of individuals from $C_k = \cup C_i$, must satisfy the following condition:
At each stage x_j of G's diffusion,

$$\begin{aligned}\int_0^{x_j} \widetilde{f}(x, C_k)dx &= \sum \int_0^{x_j} \widetilde{d}(x, C_i)dx \\ &= \sum \int_0^{x_j} \left[a_i \times \left(x - \frac{1}{2}\right) + m_i\right] dx\end{aligned}$$

The straight line segment $\widetilde{d}(x, C_k) = (\sum \widetilde{a}_i) \times (x - 1/2) + (\sum m_i)$, satisfies the condition on $[0, 1]$. Conversely $\widetilde{a}_k = \sum \widetilde{a}_i$ and $m_k = \sum m_i$ satisfies the general condition (1) above: $\frac{|\widetilde{a}_k|}{2} \leq m_k \leq 1 - \frac{|\widetilde{a}_k|}{2}$ so that $\widetilde{f}(x, C_k) = \widetilde{d}(x, C_k)$ on $[0, 1]$ and we deduce $\widetilde{a}_k = \sum \widetilde{a}_i$.

Proof. Suppose that $\widetilde{a}_k$ and m_k don't satisfy condition (1) so that $\widetilde{d}(x, C_k)$ intersect the top or the bottom of the square where the $\widetilde{f}(x, C_k)$ curve is inscribed. The case $\widetilde{f}(x, C_k) = 0$ on a non empty segment of $[0, 1]$

is rejected as by hypothesis in this section (general case (1) applies), $\forall\ i$ $1 \le i \le n$, $\forall\ x \in]0,1]$, $\widetilde{d}(x, C_i) > 0$. Suppose $\widetilde{f}(x, C_k) = 1$ on a segment of $[0,1]$, we deduce from 3.2.1 that $\widetilde{f}(x, \overline{C}_k) = 0$ on a segment of $[0,1]$. Such an hypothesis is rejected for the reason previously stated: by hypothesis in this section, $\forall\ i\ \ 1 \le i \le n$, $\forall\ x \in]0,1]$, $\widetilde{d}(x, C_i) > 0$.

3.3.3. $\sum_{i=1}^{k} \widetilde{a}_i = 0$

Proof. The equation of the straight line underlying inequality in the selection process for access of the whole population to G is $\widetilde{d}(x) = 1$, so that, according to 3.3.2, $\sum_{i=1}^{k} \widetilde{a}_i = 0$

3.3.4. $\mathrm{Max}(|\sum \widetilde{a}_i|) = \widetilde{a}_g \le 1$: *the maximum* $\widetilde{a}_g$ *of the absolute value of* $\sum \widetilde{a}_i$ *is necessarily inferior or equal to one.*

Proof. According to general case (1) and additivity of coefficients (3), we have $1/2\,\widetilde{a}_g = 1/2 \sum_{\widetilde{a}_i > 0} \widetilde{a}_i \le \sum_{a_i > 0} m_i$ and $1/2\,\widetilde{a}_g = 1/2 \sum_{\widetilde{a}_i < 0} (-\widetilde{a}_i) \le \sum_{\widetilde{a}_i < 0} m_i = 1 - \sum_{\widetilde{a}_i > 0} m_i$ so that $\widetilde{a}_g \le 1$.

3.4 *Interpretation of general coefficient* $\widetilde{a}_g$ *as an overall measure of inequality in the selection process for access to* G

Define the two groups $\cup C_i\ \ (\widetilde{a}_i \le 0) = C_g^-$ and $\cup C_i\ \ (\widetilde{a}_i \ge 0) = C_g^+$ as the sets of subgroups C_i where, according to 3.2.2, individuals have opportunity of access to G respectively higher and lower than the average. According to assumptions in 3.1, we consider a virtual construct permitting to compare opportunity of access to G. We assume that access of individuals from C_g^+ to good G has been derived from underlying continuous joint density $\widetilde{f}(x, C_g^+)$ supposed to be linear on the segment of $[0,1]$ where $0 < \widetilde{f}(x, C_g^+) < 1$: $\widetilde{f}(x, C_g^+) = \widetilde{d}(x, C_g^+)$. The coefficient $\widetilde{a}_g$ represents the slope of $\widetilde{d}(x, C_g^+)$. According to 3.3.2 and 3.3.3, if general condition (1) is followed, $\sum_{\widetilde{a}_i > 0} \widetilde{a}_i = \sum_{\widetilde{a}_i < 0} (-\widetilde{a}_i) = \widetilde{a}_g$.

Discussion. We propose to check $\widetilde{a}_g$'s qualities as an index of inequality of opportunity through conditions imposed to segregation indices. Segregation is defined as a measure of the unevenness of distribution of individual characteristics between organizational units. The validity of a measure as an index of segregation relies on the following four general criteria[5]:

[5] As listed by Watts [9].

size invariance, organizational equivalence, symmetry, and the principle of transfers in its weak form. One further requisite concerns the index independence to marginal distributions. We explicit below the meaning of these properties in segregation analysis and adapt them for inequality of opportunity analysis. Segregation analysis focuses on dissimilarity between units U_i (types of job, urban areas etc.) in terms of individual characteristics (men/women, whites/nonwhites, etc.), which depends on distribution of individual characteristics between units. Inequality of opportunity analysis focuses on dissimilarity between subgroups C_i in terms of opportunity, which depends on distribution of opportunity between subgroups. The status of units U_i in segregation analysis parallels that of subgroups C_i in inequality of opportunity analysis. Note that, in segregation analysis, advantaged (disadvantaged) U_i units are characterized by a percentage of advantaged (disadvantaged) individuals higher than average. This identification can be compared to the status of advantaged (disadvantaged) C_i subgroups defined above by a percentage of individuals accessing to G (excluded from G) higher than average. Nevertheless, the status of segregation variable (distinguishing individual characteristics) differs to that of opportunity variable (distinguishing access and non access to G) on one point. Individual characteristics must be formally equivalent for measuring segregation (that is the condition of the symmetry principle detailed below). On the other hand, "access" and "non access" to G are not necessarily given a formally equivalent status when defining indices of inequality of opportunity.

3.4.1. Size invariance *refers to the invariance of the index when the populations are changed proportionately, so that* $I(\lambda N) = I(N)$ *where* λ *is a positive scalar.*

Proof. $\widetilde{a}_g$ has been defined on the basis of fractions m_i of the whole population in C_i subgroups.

3.4.2. Organizational equivalence *refers to invariance of the index when two of the units that have an identical pattern of segregation are combined or when a single unit is divided into units with identical segregation patterns. This criterion should be reinterpreted as* classification equivalence *referring to* C_i *subgroups.*

Proof. According to its definition in 4, the coefficient $\widetilde{a}_g$ opposing opportunity of individuals belonging to C_g^- and C_g^+ subgroups is not affected by any composition of the C_i subgroups within these two sets.

Note that if general condition (1) is followed, according to 3.4 we have $\sum_{\widetilde{a}_i>0} \widetilde{a}_i = \sum_{\widetilde{a}_i<0} (-\widetilde{a}_i) = \widetilde{a}_g$.

When two subgroups show identical level of inequality in the selection process are combined, according to 3.3.2, the resulting coefficient of inequality of aggregated group is equal to the sum of the inequality coefficients of each subgroup, thus $\widetilde{a}_g$ remains invariant. It also remains invariant when a single subgroup is divided into further subgroups with identical level of inequality.

3.4.3. *In segregation analysis, symmetry means that the index expresses segregation with the same magnitude if data regarding advantaged and disadvantaged individual characteristics are commuted in the index definition. Reversing the respective status of segregation and opportunity variables here, symmetry means that the index expresses inequality of opportunity with the same magnitude if data regarding access to G and exclusion from G are commuted in the index definition so that we consider equally the opportunity of access to G and the risk of exclusion from G.*

Proof. According to 3.4 the straight line segment $\widetilde{d}(x, C_g^+)$ characterizes inequality of access of individuals from C_g^+ to good G as well as exclusion from G of the same individuals. Commutating data regarding opportunity of access to G and risks of exclusion from G, which corresponds to a reversal of the ordering of individuals within the population, does not affect $\widetilde{a}_g$'s magnitude.

Discussion: In segregation analysis, the object of the symmetry criteria prevents contradictory diagnostics if there are two values for the index. Dissimilarity between subgroups C_i in terms of opportunity of access or in terms of risk of exclusion should also refer, for coherence, to the same aspect of inequality of opportunity. This property is a consequence of the definition of $\widetilde{a}_g$ in a reference mark independent of variations of overall access to G. Symmetry is not followed by several classical indices of inequality of opportunity: the two ratios of proportion, for instance, don't follow this condition.

Note that situations regarding advantaged (disadvantaged) units in segregation analysis are implicitly symmetrical since there is one value for the index characterizing overall segregation opposing the two complementary groups of units. Accordingly, the respective situations of C_g^- and C_g^+ regarding inequality in the selection process are symmetrical, so that we have one overall value measuring this aspect of inequality of opportunity within the population under study: $\widetilde{a}_g$.

3.4.4. *According to the* principle of transfers *in its strong form, segregation declines when for example, within the set of the disadvantaged units, a member with disadvantaged characteristics moves from a less advantaged unit to a more advantaged one and is replaced by the member with advantaged characteristics from the latter unit,* ceteris paribus. *According to the weak form of the principle, when the exchange occurs between an advantaged and a disadvantaged unit, then the index should decline.*

As for organizational equivalence, this principle has to be reinterpreted to refer to exchanges of opportunity of access to G (of exclusion from G) between the subgroups C_i. The coefficient $\widetilde{a}_g$ satisfies the principle of transfers in its weak form.

Proof. According to the construction of inequality coefficients defined in 3.1, $\widetilde{a}_g$ $(-\widetilde{a}_g)$ characterizes C_g^+ (C_g^-) relative opportunity distribution underlying access to G, so that if one more individual from C_g^+ accedes to G and one less from C_g^-, $\widetilde{a}_g$, tends to diminish.

3.4.5. *Insensitivity to marginal distributions*

Proof. Insensitivity to marginal variation is demonstrated in 3.2.3 for $(\widetilde{a}_i)$ coefficients comparing relative inequality of two complementary subgroups C_i and $\overline{C}_i$, and then applies for $\widetilde{a}_g$.

In conclusion, the coefficient $\widetilde{a}_g$ is a valid measure of inequality of opportunity. Besides, $\widetilde{a}_g$ is defined in a reference mark independent of variations of overall access to G and is insensitive to the distribution of the C_i subgroups within the population. Thus the $\widetilde{a}_g$ coefficient represents an overall measure of inequality in the selection process as defined in 1.4 and is insensitive to margins (x_j and m_g).

4. Empirical determination of the $(\widetilde{a}_g)$ coefficient

4.1 *Determination of the $\widetilde{d}(x, C_g^+)$ straight line segment*

We know the fraction m_g of the population in the C_g^+ subgroup as well as the fraction r_g of the C_g^+ subgroup having access to G. We can thus construct the straight line $d(x, C_g^+)$ such that $\int_0^1 d(x, C_g^+)dx = m_g$ and $\int_0^{x_j} d(x, C_g^+)dx = rg \times m_g$. Define a_g the slope of this straight line. It can be easily shown that if the curve $\widetilde{f}(x, C_g^+)$ defined in 3.4 is a broken line on $[0,1]$, that is $\widetilde{f}(x, C_g^+) = 0$ or $\widetilde{f}(x, C_g^+) = 1$ on a non-empty segment of $[0,1]$, i.e. $\widetilde{d}(x, C_g^+)$ intersects the base or the top of the square

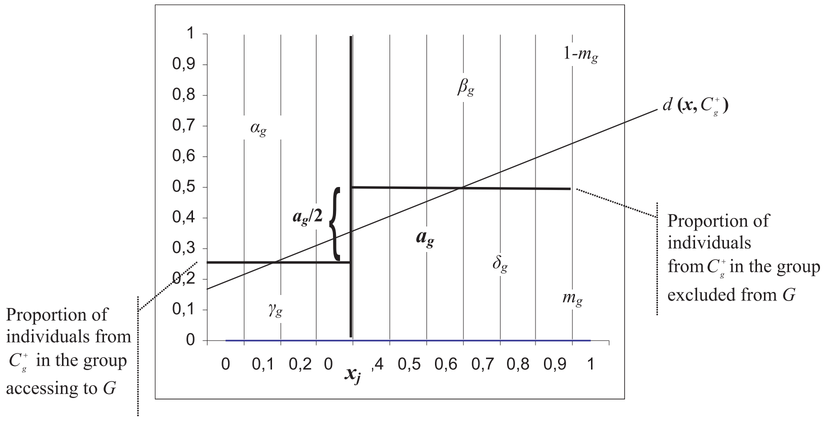

Figure 2

Graphic representation of contingency table data and construction of $d(x, C_s^+)$

where $\widetilde{f}(x, C_g^+)$ is inscribed, then the condition $\frac{a_g}{2} \leq m_g \leq 1 - \frac{a_g}{2}$ is not followed, i.e. $d(x, C_g^+)$ intersects the base or the top of the square where $\widetilde{f}(x, C_g^+)$ is inscribed (see annex).

We are then facing two cases:

(1) *Condition* (1) *is verified*: $\frac{a_g}{2} \leq m_g \leq 1 - \frac{a_g}{2}$. The straight line $d(x, C_g^+)$ does not intersect the base or the top of the square where the virtual joint density $\widetilde{f}(x, C_g^+)$ is traced. We deduce that the $\widetilde{f}(x, C_g^+)$ curve is a straight line segment on $[0,1]$ (since if it were a broken line segment the contrary would be observed): $\widetilde{f}(x, C_g^+) = \widetilde{d}(x, C_g^+)$ on $[0,1]$ so that the straight line $d(x, C_g^+)$ merges with the virtual construct $\widetilde{d}(x, C_g^+)$ underlying access of individuals from C_g^+ to good G. Hence $a_g = \widetilde{a}_g$.

Proof. $\widetilde{f}(x, C_g^+) = \widetilde{d}(x, C_g^+)$ on $[0,1]$ so that (according to 3.1)

$$\int_0^{x_j} \widetilde{d}(x, C_g^+)dx = r_g m_g \quad \text{and} \quad \int_0^1 \widetilde{d}(x, C_g^+)dx = m_g\,.$$

The straight line segment $d(x, C_g^+)$ is the only one that satisfies this condition so that $d(x, C_g^+) = \widetilde{d}(x, C_g^+)$ on $[0,1]$.

(2) *Condition* (1) *is not verified*: The straight line $d(x, C_g^+)$ intersects the base or the top of the square where the joint density is traced. Hence,

the $\widetilde{f}(x, C_g^+)$ curve is a broken line segment on $[0,1]$; $d(x, C_g^+)$ does not merge with $\widetilde{d}(x, C_g^+)$ underlying access of individuals from C_g^+ to good G.

4.2 *Calculation of the* $\widetilde{a}_g$ *coefficient*

As shown in 4.1, if condition (1) is verified for (a_g, m_g), then the $d(x, C_g^+)$ straight line merges with the virtual $\widetilde{d}(x, C_g^+)$ straight line segment. In the reverse case, the calculation of $\widetilde{a}_g$ remains possible if the fraction of the C_g^+ subgroup having access to G is not nil (and if the fraction of the C_g^- subgroup having no access to G is not nil). For this more complex and less common case, the calculation of $\widetilde{a}_g$ is presented in annex. In the following we suppose that condition (1) is verified.

4.2.1 *Calculation*

m_g represents the fraction of the population in social subgroup C_g^+, and r_g represents the fraction of the subgroup C_g^+ with access to G.

According to 4.1, define the straight line $d(x, C_g^+) = a_g \times x + m_g - a_g/2$ so that

$$\int_0^1 d(x, C_g^+)dx = m_g \quad \text{and}$$

$$\int_0^{x_j} d(x, C_g^+)dx = r_g \times m_g \text{ with } x_j \neq 0 \text{ and } x_j \neq 1.$$

We have:

$$\widetilde{a}_g = a_g = \frac{2 \times m_g \times (x_j - r_g)}{(1 - x_j) \times x_j}. \tag{4}$$

4.2.2 *According to* 4.2.1

$$\widetilde{a}_g/2 = \frac{m_g - m_g r_g}{1 - x_j} - \frac{m_g r_g}{x_j}$$

Referring to parameters in Figure 2:

$$\widetilde{a}_g = 2 \times \left[\frac{\delta_g}{1 - x_j} - \frac{\gamma_g}{x_j}\right].$$

The $\widetilde{a}_g/2$ coefficient is calculated as a familiar statistic of association, the difference of proportions comparing columns of the contingency table[6]

[6]See Blackbur, Siltanen and Jarman [2] for a comparison with basic segregation indexes.

Returning to the example given on tables 1 and 2 in 1.1, we can calculate $\widetilde{a}_g$:

In period P1, $\widetilde{a}_g = 2\left[\dfrac{550}{990} - \dfrac{10}{110}\right] = 0,91$.

In period P2, $\widetilde{a}_g = 2\left[\dfrac{450}{660} - \dfrac{100}{440}\right] = 0,91$.

We check that $\dfrac{a_g}{2} \leq m_g \leq 1 - \dfrac{a_g}{2}$: $0,45 \leq 0,5 \leq 0,55$

The conclusion is that inequality in the selection process remained stable between periods P1 and P2. The changes revealed by the variations of different measures of inequality as the two ratios of proportions and their product (an odds ratio) represent the effect of the opening of access to G on these measures while the results of generative mechanisms underlying inequality are the same in the two periods.

5. Intrinsic inequality in the selection process

A definition. Suppose there is a hierarchical ranking of n discrete goods G_k (constituting a vertical ordering of spatial units, types of jobs, social status, education levels, etc.) such that all individuals accessing to any higher ranked good G_{k+1} would have access to the lower ranked good G_k . Inequality coefficients in the selection process are $\widetilde{a}_g^k$ with $1 \leq k \leq n$. Intrinsic inequality in the selection process for access to the various goods $\underset{0\leq k\leq n}{G_k}$ is represented by the function $\widetilde{A}_g(x_j)$. $\widetilde{A}_g(x_j)$ is a continuous function describing overall inequality in the selection process for access to the various percentile ranks of the population that underlies opportunity of access to the different goods $\underset{0\leq k\leq n}{G_k}$.

Suppose we have, for instance, two populations Q1 and Q2. In each population, n discrete goods G_k as defined above are considered. We dispose of n overall measures of inequality in the selection process for access to the n goods G_k . The various cutting points in the percentile ranks of the population underlying access from the less selective to the most selective good G_k usually differ from one population to another. Nevertheless, simple extrapolations may suffice to compare overall inequality in the selection process for access to each specific percentile rank of the population, i.e. intrinsic inequality in the selection process.

Discussion. Intensity of inequality in the selection process for access to a good G_k may depend on the overall rate of access x_k to G_k . This would not be the case if the $\widetilde{a}_g^k$ coefficients were equal. In the reverse case, comparisons of inequality in the selection process for access to only one of the goods G_k do not tell the whole story as far as explanation is called into question.

Note that the approach developed here can be applied to inequality in the selection process for access to a continuous good G, income for instance. In this case, the variable g described in 2.1, underlying a ranking of the population, could be directly inferred from the quantity of G one possesses. Analysis of inequality in the selection process would rely on comparisons of joint densities $\widetilde{f}(x, C_g^+)$ characterizing, in each population, opportunity distribution of individuals from disadvantaged subgroups as opposed to opportunity distribution of individuals from advantaged ones.

6. Conclusion

The issue at stake develops when interpreting divergence in opportunities of access to a discrete good G between populations. When measuring such discrepancies, alternative methods of analysis usually provide different and even contradictory results. These contradictions are supposed to translate ultimate indecisiveness of standpoints about inequalities. However, each measure isolates a different feature of inequality and may be subject to criteria of validity relative to its aim.

In an explanatory analysis of changes in opportunity of access to good G, one may aim at comparing inequality of opportunity in the selection process. The general $\widetilde{a}_g$ coefficient of opportunity inequality proposed here, which opposes advantaged subgroups (with a rate of access to G superior to average) to disadvantaged ones, permits such comparisons and is not sensitive to marginal distributions.

Provided that general case (1) applies, the overall level of inequality in the selection process $\widetilde{a}_g$ is calculated as a familiar statistics of association. The coefficient $\widetilde{a}_g$ represents twice the value of the difference of proportions comparing columns of the 2×2 contingency table. Note that the necessity to refer to insensitivity to margins in a broad sense has been claimed. Such a sense may apply to concepts of inequality involving interrelations between rows and columns of the contingency table. According to this broad meaning, the defined general level of inequality $\widetilde{a}_g$ is insensitive to margins with regard to its proper object of measure, while the difference of proportions comparing columns of the contingency table is not.

It should be noted that, as long as the problem of inequality of opportunity is at play, equality is not absurd as a reference for measuring discrepancies between subgroups, i.e., a sufficient level of heterogeneity within subgroups is likely to be achieved-thus, general case (1) is expected to apply. In other cases, provided that at least one member of the disadvantaged group accesses to G (and that at least one member of the advantaged group does not access to G), the coefficient of inequality in the selection process can also be calculated. In such cases its value differs from the difference of proportions comparing columns of the contingency table.

Finally, the coefficient of inequality of opportunity developed here exploits the fact that, with respect to the selection process for access to G, one can infer continuous distributions of relative opportunity within population, even if G is a discrete good. In doing so, the defined coefficient of inequality in the selection process aims to compare the results of microsociological processes generating inequalities observed at macrosocial level, whereas usual measures allow only for hypotheses on these matters. The process of access to G is still a black box, but comparisons of overall inequality relative to the selection process represent a positive tool for explaining the statistical relations observed.

Annexure. Calculation of $\widetilde{a}_g$ specific cases

In the following, m_g is the fraction of the whole population in C_g^+, x_j is the fraction of the whole population accessing to G and r_g the access rate to G of individuals from C_g^+. The equation of $\widetilde{d}(x, C_g^+)$ is $y = \widetilde{a}_g x + \widetilde{b}_g$. The equation of $d(x, C_g^+)$ is $y = a_g x + b_g$.

By construction (see 4.1),

$$\int_0^{x_j} d(x, C_g^+) = \mathrm{S1} \quad \text{and}$$

$$\int_{x_j}^1 d(x, C_g^+) = \mathrm{S2}$$

$$a_g = 2 \times \left[\frac{S_2}{1 - x_j} - \frac{S_1}{x_j}\right] = \frac{2 \times m_g \times (x_j - r_j)}{(1 - x_j) \times x_j} \quad \text{(see 4.2.1)}$$

(I) Condition for $\widetilde{d}(x, C_g^+)$ to intersect the basis of the square where $\widetilde{f}(x, C_g^+)$ is traced.

Let us specify the conditions of the case described by Figure 3.

For a given fraction m_g of the whole population in C_g^+, the maximal value of $\widetilde{a}_g$ for $\widetilde{d}(x, C_g^+)$ not to intersect the basis of the square where $\widetilde{f}(x, C_g^+)$ is traced is $\frac{\widetilde{a}_g}{2} = m_g$ (see Figure 3). In this limit case, we have $S1 = r_g m_g = m_g x_j^2$ so that $r_g = x_j^2$. Conversely, $S1 < m_g x_j^2 \Rightarrow \frac{\widetilde{a}_g}{2} > m_g$ (see Figure 3).

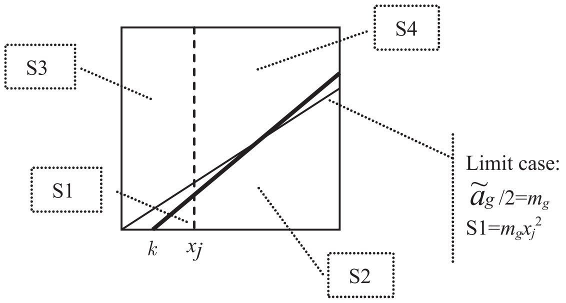

Figure 3

$\widetilde{d}(x, C_g^+)$ intersects the basis of the square where $\widetilde{f}(x, C_g^+)$ is traced

Hence $\frac{\widetilde{a}_g}{2} > m_g \Leftrightarrow \widetilde{d}(x, C_g^+)$ intersects the basis of the square and $\widetilde{d}(x, C_g^+)$ intersects the basis of the square

$\Leftrightarrow \quad r_g < x_j^2$

$\Leftrightarrow \quad \frac{a_g}{2} > m_g$ (the demonstration is analytically simple knowing $\frac{a_g}{2} = \frac{m_g \times (x_j - r_g)}{(1 - x_j) \times x_j}$)

$\Leftrightarrow \quad d(x, C_g^+)$ intersects the basis of the square

(II) Condition for $\widetilde{d}(x, C_g^+)$ to intersect the top of the square where $\widetilde{f}(x, C_g^+)$ is traced.

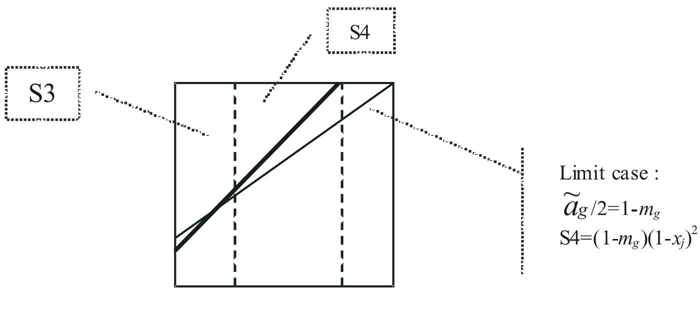

Figure 4

$\widetilde{d}(x, C_g^+)$ intersects the top of the square where $\widetilde{f}(x, C_g^+)$ is traced

Let us specify the conditions of the case described by Figure 4.

For a given fraction m_g of the whole population in C_g^+, the maximal value of $\widetilde{a}_g$ for $\widetilde{d}(x, C_g^+)$ not to intersect the top of the square where $\widetilde{f}(x, C_g^+)$ is traced is $\frac{\widetilde{a}_g}{2} = 1 - m_g$ (see Figure 4). In this limit case, we have $S4 = (1 - m_g)(1 - x_j)^2$. We have $S4 = 1 - x_j - m_g + r_g m_g$, we obtain: $r_g = \frac{1}{m}(x_j^2 + 2m_g x_j - m_g x_j^2 - x_j)$. Conversely $S4 < (1 - m_g)(1 - x_j)^2 \Rightarrow \frac{\widetilde{a}_g}{2} > 1 - m_g$ (see Figure 1).

Hence $\frac{\widetilde{a}_g}{2} > 1 - m_g$

$\Leftrightarrow$ $\widetilde{d}(x, C_g^+)$ intersects the top of the square where $\widetilde{f}(x, C_g^+)$ is traced and

$\widetilde{d}(x, C_g^+)$ intersects the top of the square

$\Leftrightarrow$ $r_g < \frac{1}{m}(x_j^2 + 2m_g x_j - m_g x_j^2 - x_j)$

$\Leftrightarrow$ $\frac{a_g}{2} > 1 - m_g$ (the demonstration is analytically trivial)

$\Leftrightarrow$ $d(x, C_g^+)$ intersects the top of the square

(III) Condition for $\widetilde{d}(x, C_g^+)$ to intersect the basis and the top of the square where $\widetilde{f}(x, C_g^+)$ is traced.

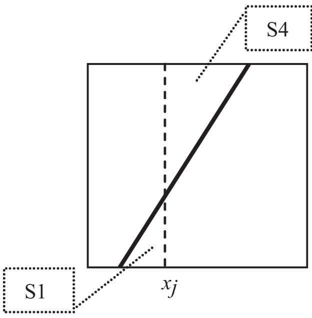

Figure 5

$\widetilde{d}(x, C_g^+)$ intersects the basis and the top of the square where $\widetilde{f}(x, C_g^+)$ is traced

Let us specify the conditions of the case described by Figure 5. According to I and II:

$1 - \frac{\widetilde{a}_g}{2} < m_g < \frac{\widetilde{a}_g}{2}$

$\Leftrightarrow$ $\widetilde{d}(x, C_g^+)$ intersects the basis and the top of the square where $\widetilde{f}(x, C_g^+)$ is traced and $\widetilde{d}(x, C_g^+)$ intersects the basis and the top of the square

$\Leftrightarrow \quad r_g < x_j^2$ and $r_g = \frac{1}{m_g}(x_j^2 + 2m_g x_j - m_g x_j^2 - x_j)$

$\Leftrightarrow \quad 1 - \frac{a_g}{2} < m_g < \frac{a_g}{2}$

Note that:

$$x_j^2 < \frac{1}{m_g}(x_j^2 + 2m_g x_j - m_g x_j^2 - x_j) \Leftrightarrow m_g > \frac{1}{2} \tag{5}$$

Proof. We have $0 < x_j < 1$.

(IV) Synthesis of the results

- The condition for $\widetilde{d}(x, C_g^+)$ to intersect the basis and not the top of the square (which supposes $m_g < \frac{1}{2}$) is:

$$\begin{aligned} & m_g < \inf\left(\frac{a_g}{2}, 1 - \frac{a_g}{2}\right) \\ \Leftrightarrow \qquad & \frac{1}{m_g}(x_j^2 + 2m_g x_j - m_g x_j^2 - x_j) < r_g < x_j^2 \end{aligned}$$

- The condition for $\widetilde{d}(x, C_g^+)$ to intersect the top and not the basis of the square (which supposes $m_g > \frac{1}{2}$) is:

$$\begin{aligned} & m_g > \sup\left(\frac{a_g}{2}, 1 - \frac{a_g}{2}\right) \\ \Leftrightarrow \qquad & x_j^2 < r_g < \frac{1}{m}(x_j^2 + 2m_g x_j - m_g x_j^2 - x_j) \end{aligned}$$

- The condition for $\widetilde{d}(x, C_g^+)$ to intersect the basis and the top of the square is:

$$\begin{aligned} & 1 - \frac{a_g}{2} < m_g < \frac{a_g}{2} \\ \Leftrightarrow \qquad & \begin{cases} r_g < x_j^2 & \text{if } m_g > \frac{1}{2} \\ r_g < \frac{1}{m_g}(x_j^2 + 2m_g x_j - m_g x_j^2 - x_j) & \text{if } m_g < \frac{1}{2} \end{cases} \end{aligned}$$

(V) Calculation of $\widetilde{a}_g$

- *First case.* $\widetilde{d}(x, C_g^+)$ intersects the basis and not the top of the square where $\widetilde{f}(x, C_g^+)$ is traced (thus, according to (3), $m_g < \frac{1}{2}$):

$$\begin{aligned} & m_g < \inf\left(\frac{a_g}{2}, 1 - \frac{a_g}{2}\right) \\ \Leftrightarrow \qquad & \frac{1}{m_g}(x_j^2 + 2m_g x_j - m_g x_j^2 - x_j) < r_g < x_j^2 \end{aligned}$$

  k is the abscissa of the intersection point with the basis of the square:

Calculating $S_1 = r_g m_g$ and $(\mathrm{S1} + \mathrm{S2}) = m_g$ we have:

$$\begin{cases} \widetilde{a}_g (x_j - k)^2 = 2 r_g m_g \\ \widetilde{a}_g (1-k)^2 = 2 m_g \end{cases}$$

We deduce $k = \frac{x_j - \sqrt{r_g}}{1 - \sqrt{r_g}}$

Substituting this expression for k in the second equation we find:

$$\widetilde{a} = 2 m_g \left[\frac{1 - \sqrt{r_g}}{1 - x_j} \right]^2$$

- *Second case.* $\widetilde{d}(x, C_g^+)$ intersects the top and not the basis of the square where $\widetilde{f}(x, C_g^+)$ is traced (thus, according to (3), $m_g > \frac{1}{2}$):

$$\begin{aligned} & m_g > \sup\left(\frac{a_g}{2}, 1 - \frac{a_g}{2}\right) \\ \Leftrightarrow \quad & x_j^2 < r_g < \frac{1}{m_g}(x_j^2 + 2 m_g x_j - m_g x_j^2 - x_j) \end{aligned}$$

  k' is the abscissa of the intersection point with the top of the square:

  Calculating $\mathrm{S4} = 1 - x_j - m_g + r_g m_g$ and $(\mathrm{S3} + \mathrm{S4}) = 1 - m_g$ we have:

$$\begin{cases} \widetilde{a}_g (k' - x_j)^2 = 2\mathrm{S4} \\ \widetilde{a}_g (k')^2 = 2(1 - m_g) \end{cases}$$

  We deduce $k' = \frac{x_j}{1 - \sqrt{\frac{\mathrm{S4}}{1 - m_g}}}$

  Substituting this expression for k' in the second equation $\left[\sqrt{\widetilde{a}_g} = \frac{\sqrt{2(1-m_g)}}{k'} \right]$ we find:

$$\widetilde{a}_g = 2 \left[\frac{\sqrt{1 - m_g} - \sqrt{1 - x_j - m_g + r_g m_g}}{x_j} \right]^2$$

- *Third case.* $\widetilde{d}(x, C_g^+)$ intersects the top and not the basis of the square where $\widetilde{f}(x, C_g^+)$ is traced:

$$\begin{aligned} & 1 - \frac{a_g}{2} < m_g < \frac{a_g}{2} \\ \Leftrightarrow \quad & \begin{cases} 1 - r_g < x_j^2 & \text{if } m_g > \frac{1}{2} \\ 2 - r_g < \frac{1}{m}(x_j^2 + 2 m_g x_j - m_g x_j^2 - x_j) & \text{if } m_g < \frac{1}{2} \end{cases} \end{aligned}$$

Calculating $S1 = r_g m_g$ and $S4 = 1 - x_j - m_g + r_g m_g$ we have:

$$\begin{cases} \widetilde{a}_g(k' - x_j)^2 = 2S4 \\ \widetilde{a}_g(x_j - k)^2 = 2S1 \end{cases}$$

By definition, $k = \frac{-\widetilde{b}_g}{\widetilde{a}_g}$ and $k' = \frac{1-\widetilde{b}_g}{\widetilde{a}_g}$ so that we have $k' = \frac{1}{\widetilde{a}_g} + k$.

From the two equations above we deduce $\frac{\frac{1}{\widetilde{a}_g} + k - x_j}{x_j - k} = \sqrt{\frac{S4}{S1}}$ so that

$$\frac{1}{\widetilde{a}_g} = (x_j - k)\left[\sqrt{\frac{S4}{S1}} + 1\right] \text{ and } \frac{1}{\widetilde{a}_g} = \frac{1}{2S1}(x_j - k)^2.$$

We deduce $(x_j - k) = 2S1\left[\frac{\sqrt{S1}+\sqrt{S4}}{\sqrt{S1}}\right]$ and with the second equation we have:

$$\widetilde{a}_g = \frac{1}{2}\left[\frac{1}{\sqrt{S1} + \sqrt{S4}}\right]^2 = \frac{1}{2}\left[\frac{1}{\sqrt{r_g m_g} + \sqrt{1 - x_j - m_g + r_g m_g}}\right]^2.$$

References

[1] A. B. Atkinson, On the measurement of inequality, *Journal of Economic Theory*, Vol. 2 (1970), pp. 244–263.

[2] R. M. Blackburn, J. Siltanen and J. Jarman, The measurement of occupational gender segregation: current problems and a new approach, *Journal of the Royal Statistical Society* (A), Vol. 158 (Part2) (1995), pp. 319–331.

[3] O. D. Duncan and B. Duncan, A methodological analysis of segregation indices, *American Sociological Review*, Vol. 20 (1955), pp. 210–217.

[4] R. J. Harrison, Do odds ratios really control for the availability of occupational positions in status contingency tables? *European Sociological Review*, Vol. 4 (1988), pp. 65–79.

[5] R. M. Hutchens, Segregation curves, Lorenz curves, and inequality in the distribution of people across occupation, *Mathematical Social Science*, Vol. 21 (1991), pp. 31–51.

[6] S. C. Kolm, Unequal inequalities, *Journal of Economic Theory*, Vol. 12 (1976), pp. 416–442.

[7] M. E. Sobel, Structural mobility, circulation mobility, and the analysis of occupational mobility: a conceptual mismatch, *American Sociological Review*, Vol. 29 (1983), pp. 16–23.

[8] A. B. Sorensen, The Structure of inequality and the Process of Attainment, *American Sociological Review*, Vol. 42 (1977), pp. 965–978.

[9] M. Watts, Occupational gender segregation: index measurement and econometric modelling, *Demography*, Vol. 35 (1998), pp. 489–496.

Received March, 2009