14. ABOUT THE ACCURACY OF GINI INDEX FOR MEASURING THE POVERTY

Ştefan V. ŞTEFĂNESCU[1]

Abstract

The Gini index is often used to measure the income inequality presented inside a specified group of individuals and sometimes also for evaluating the "poverty" degree of this population. In this article we show that Gini's index is not always so adequate to measure the poverty level of an analyzed population. More precisely, we'll get two very different income distributions which finally give the same value for Gini's index.

Having in mind this aspect, we recommend with priority to use the Gini index only together with other similar or complementary poverty indicators to evaluate the poverty level of a given population.

Keyword: measurement and analysis of poverty, Gini index, Lorenz curve, income distribution, inequality index
JEL Classification: I32, D63, D31

1. Gini's measure

Let V , $0 \le V \le b$, be a random variable (r.v.) which defines the "income" distribution of the individuals from a given population P. The income v of an arbitrary person from P is nonnegative and, more, it is upper limited by the threshold b , $b > 0$, that is $0 \le v \le b$.

We denote by $f(v)$ and $F(v)$, $0 \le v \le b$, the probability distribution function (p.d.f.), respectively the cumulative distribution function (c.d.f.) of the r.v. V (Papoulis, 1990).

Taking into consideration the wages v for every person of the population P, the classical Lorenz curve visualizes graphically the variation in the proportion of individuals who have their income below an arbitrary threshold v, with $0 \le v \le b$ (Gastwirth, 1971). In this manner, we obtain a suggestive image about the fluctuation in the average income for the groups having individuals with their wages below the threshold v (Gastwirth, 1971).

[1] The Research Institute for Quality of Life - Romanian Academy,
E-mail: stefanst@fmi.unibuc.ro

Concretely, the Lorenz curve $L(u)$, $0 \le u \le 1$, attached to the r.v. V has the form (Gastwirth, 1971)

$$L(u) = \frac{1}{\mu}\int_0^u F^{-1}(t)\, dt = \frac{S(u)}{\mu}$$

where: $F^{-1}(t)$, $0 \le t \le 1$, is the inverse of the c.d.f. $F(v)$, $\mu = Mean(V)$ and

$$S(u) = \int_0^u F^{-1}(t)\, dt \quad , \quad 0 \le u \le 1 .$$

The use of the Lorenz curve is a very suitable graphical method to establish how large is the income inequality between the persons of the population P . In fact, the Lorenz curve permits us to compare and to visualize the income discrepancies, providing also information about the distribution of income depending on the mean wage of P (Gastwirth, 1971). The paper of Gastwirth suggests us some possible extensions regarding the concept of the “Lorenz curve”. Many other methods for comparing multidimensional indices of inequality are discussed in Lugo, 2007.

Comparatively with other indicators used to evaluate the poverty degree of the population P , the Gini’s index measures the income inequality on the basis of the classical Lorenz curve (Gastwirth, 1971). We must mention here that the Gini index γ is certainly the most popular indicator for measuring the inequality level of a given population P (Chakravarty, 1983; Maasoumi, Lugo, 2008; Sen, 1976; Zhou, Ang, 2010; Donaldson, Weymark, 1980; Foster, Greer, Thorbecker, 1984; Foster, McGillivray, Seth, 2009; Gastwirth, 1971; Kakwani, 1980; Lugo, 2007; Somarriba, Pena, 2009; Ştefănescu, 2009; Yitzhaki, 1983).

Concretely, respecting the previous notations, the Gini’s index $\gamma(V)$ associated to the income distribution V has the expression

$$\gamma(V) = 2\int_0^1 (u - L(u))\, du = 1 - \frac{2}{\mu}\int_0^1 S(u)\, du$$

Finally, we remark that the Gini’s index $\gamma(V)$ satisfies four essential conditions which are often imposed to any good poverty index:

A1.Continuity (Gini values are not so different for closed distributions).

A2.Anonymity(the invariance of the Gini index to a permutation of the income values).

A3.Invariance when we change the measure scale for the income.

A4.Dalton-Pigou transfer principle (the Gini index decreases when “money” is transfered from “rich” to “poor” people).

More details could be found in the following papers: Chakravarty, 1983; Foster, Greer, Thorbecker, 1984; Donaldson, Weymark, 1980.

Nevertheless, all these necessary axioms ***A1-A4*** are not sufficient for selecting an acceptable poverty indicator. More precisely, in the following we will show that the Gini index can take the same value for very different income distributions V .

2. A case study

2.1. Hypotheses

Roughly speaking, the population P could be partially characterized by the pair of probabilities (p,q) with $p \geq 0$, $q \geq 0$, $p+q \leq 1$ having the following meaning :

- p is the proportion of the "poor" people in the population P ;
- q is the proportion of "rich" persons inside P .

Hence, $1-p-q$ represents the probability of the individuals in the population P who belong to the "middle class".
To simplify the next computations, we will impose the additional restrictions:

H1. The income variable V varies in the interval $[0\,,1]$, that is we take the upper income bound $b=1$;

H2. Respecting the previous notations and the condition ***H1*** we consider $p = Pr(0 \leq V \leq 1/4)$, $q = Pr(3/4 \leq V \leq 1)$;

H3. The r.v. V is uniformly distributed for the income categories $[0.0\,, 0.25]$, $(0.25\,, 0.75)$, and $[0.75\,, 1.0]$, which define a partition of the domain $[0\,,1]$ (the support for the income variable V).

Remark 1. The hypothesis ***H1*** is not restrictive, since the Gini's index remains invariant if the unit for measuring the income V is modified (the previous axiom ***A3***; see also Chakravarty, 1983; Foster, Greer,Thorbecker, 1984; Donaldson, Weymark, 1980). Having in mind this property, to simplify the next computations, we accept in the subsequent that $0 \leq V \leq 1$.

Remark 2. The condition ***H3*** results naturally by applying the statistical principle regarding the insufficient reason (additional details in Papoulis, 1990, p. 17).

For more accuracy, we denote by $V_{(p,q)}$ a r.v. V , which satisfies the mentioned conditions ***H1-H3*** with specified values for the probabilities p and q .

2.2. Theoretical results

Respecting all the hypotheses ***H1-H3*** the p.d.f. $f_{p,q}(v)$ of the r.v. $V_{(p,q)}$ has the form

$$f_{p,q}(v) = \begin{cases} 4p & ;\ for\ 0 \leq v \leq 1/4 \\ 2-2p-2q & ;\ for\ 1/4 < v \leq 3/4 \\ 4q & ;\ for\ 3/4 \leq v \leq 1 \end{cases}$$

Proposition 1. For any $p \geq 0$, $q \geq 0$, $p+q \leq 1$ we get

$$\mu_{p,q} = Mean(V_{(p,q)}) = \frac{4-3p+3q}{8}$$

Proof. Indeed

$$\mu_{p,q} = Mean(_{(p,q)}) = \int_0^1 v\, f_{p,q}(v)\, dv =$$

$$= \int_0^{1/4} 4p\, v\, dv + \int_{1/4}^{3/4} (2-2p-2q)\, v\, dv + \int_{3/4}^{1} 4q\, v\, dv = \frac{4-3p+3q}{8}$$

Since the c.d.f. $F_{p,q}(v)$ of the r.v. $V_{(p,q)}$ is given by the formula

$$F_{p,q}(v) = \int_{-\infty}^{v} f_{p,q}(t)\, dt$$

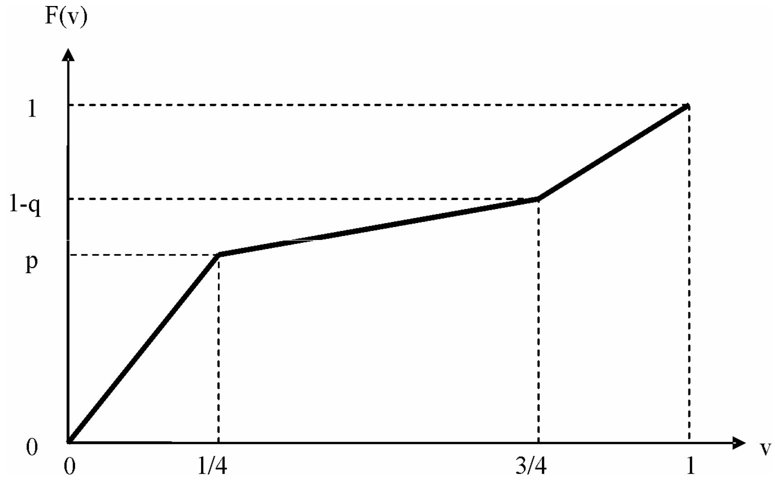
we finally deduce

$$F_{p,q}(v) = \begin{cases} 4p\,v & ;\ for\ 0 \leq v \leq 1/4 \\ (3p+q-1)/2 + (2-2p-2q)\,v & ;\ for\ 1/4 < v \leq 3/4 \\ 1-4q+4q\,v & ;\ for\ 3/4 \leq v \leq 1 \end{cases}$$

The linear form, on the subintervals $[0\,,1/4]$, $(1/4\,,3/4)$, and $[3/4\,,1]$ respectively, for the c.d.f. $F_{p,q}(v)$ is emphasized in Figure 1. Interpreting the graph in Figure 1 we also deduce the concrete significance of the probabilities p and q when $p \geq 0$, $q \geq 0$, $p+q \leq 1$.

Figure 1

The c.d.f. $F_{p,q}(v)$ of the income variable $V_{(p,q)}$



After a direct calculus, the inverse function $F_{p,q}^{-1}(t)$ of the c.d.f. $F_{p,q}(v)$ has the expression

$$F_{p,q}^{-1}(t)=\begin{cases} t/(4p) & ;\ for\ 0\le t\le p \\ (t-(3p+q-1)/2))/(2-2p-2q) & ;\ for\ p<t<1-q \\ (t-1+4q)/(4q) & ;\ for\ 1-q\le t\le 1 \end{cases}$$

Proposition 2. The application

$$S_{p,q}(u)=\int_0^u F_{p,q}^{-1}(t)\,dt \quad , \quad 0\le u\le 1 .$$

is given by the formula

$$S_{p,q}(u)=\begin{cases} c_{10}+c_{11}u+c_{12}u^2 & ;\ for\ 0\le u\le p \\ c_{20}+c_{21}u+c_{22}u^2 & ;\ for\ p<u<1-q \\ c_{30}+c_{31}u+c_{32}u^2 & ;\ for\ 1-q\le u\le 1 \end{cases}$$

with

$$c_{10}=0 \qquad c_{11}=0 \qquad c_{12}=\frac{1}{8p}$$

$$c_{20}=\frac{3p^2+pq-p}{8(1-p-q)} \qquad c_{21}=\frac{1-3p-q}{4(1-p-q)} \qquad c_{22}=\frac{1}{4(1-p-q)}$$

$$c_{30}=\frac{3q^2-3pq-4q+1}{8q} \qquad c_{31}=\frac{4q-1}{4q} \qquad c_{32}=\frac{1}{8q}$$

Proof. For any $0\le u\le p$ we deduce

$$S_{p,q}(u)=\int_0^u F_{p,q}^{-1}(t)\,dt=\int_0^u \frac{t}{4p}\,dt=\frac{1}{8p}u^2$$

When $p<u<1-q$ it results

$$S_{p,q}(u)=\int_0^u F_{p,q}^{-1}(t)\,dt=\int_0^p F_{p,q}^{-1}(t)\,dt+\int_p^u F_{p,q}^{-1}(t)\,dt=S_{p,q}(p)+\int_p^u \frac{t-(3p+q-1)/2}{2-2p-2q}\,dt=$$

$$=\frac{p}{8}+\frac{u-p}{4(1-p-q)}(u-2p-q+1)=$$

$$=\frac{3p^2+pq-p}{8(1-p-q)}+\frac{1-3p-q}{4(1-p-q)}u+\frac{1}{4(1-p-q)}u^2$$

When $1-q\le u\le 1$ we finally obtain

$$S_{p,q}(u)=\int_0^u F_{p,q}^{-1}(t)\,dt=\int_0^{1-q} F_{p,q}^{-1}(t)\,dt+\int_{1-q}^u F_{p,q}^{-1}(t)\,dt=S_{p,q}(1-q)+\int_{1-q}^u \frac{t-1+4q}{4q}\,dt=$$

$$=\frac{4-3p-4q}{8}+\frac{u^2}{8q}+\frac{(4q-1)\,u}{4q}+\frac{7q^2-8q+1}{8q}=$$

$$=\frac{3q^2-3pq-4q+1}{8q}+\frac{4q-1}{4q}u+\frac{1}{8q}u^2$$

Proposition 3. For any r.v. $V_{(p,q)}$ with $p\geq 0$, $q\geq 0$ and $p+q\leq 1$ the value of Gini index $\gamma(V_{(p,q)})$ has the expression

Proof. Using *Proposition 2* we get successively

$$I_1=\int_0^p S_{p,q}(u)\,du=\int_0^p (c_{10}+c_{11}u+c_{12}u^2)\,du=c_{10}\,p+\frac{c_{11}}{2}p^2+\frac{c_{12}}{3}p^3=\frac{1}{24p}p^3=\frac{p^2}{24}$$

$$I_2=\int_p^{1-q} S_{p,q}(u)\,du=\int_p^{1-q} (c_{20}+c_{21}u+c_{22}u^2)\,du=$$

$$=c_{20}\,((1-q)-p)+\frac{c_{21}}{2}((1-q)^2-p^2)+\frac{c_{12}}{3}((1-q)^3-p^3)=$$

$$=\frac{3p^2+pq-p}{8(1-p-q)}((1-q)-p)+\frac{1-3p-q}{8(1-p-q)}((1-q)^2-p^2)+$$

$$+\frac{1}{12(1-p-q)}((1-q)^3-p^3)=\frac{2p^2+5q^2+7pq-7p-10q+5}{24}$$

$$I_3=\int_{1-q}^1 S_{p,q}(u)\,du=\int_{1-q}^1 (c_{30}+c_{31}u+c_{32}u^2)\,du=$$

$$=c_{30}\,(1-(1-q))+\frac{c_{31}}{2}(1-(1-q)^2)+\frac{c_{32}}{3}(1-(1-q)^3)=$$

$$=\frac{3q^2-3pq-4q+1}{8q}(1-(1-q))+\frac{4q-1}{8q}(1-(1-q)^2)+\frac{1}{24q}(1-(1-q)^3)=$$

$$=\frac{-2q^2-9pq+12q}{24}$$

Hence

$$\int_0^1 S_{p,q}(u)\,du=\int_0^p S_{p,q}(u)\,du+\int_p^{1-q} S_{p,q}(u)\,du+\int_{1-q}^1 S_{p,q}(u)\,du=$$

$$= I_1 + I_2 + I_3 = \frac{3p^2 + 3q^2 - 2pq - 7p + 2q + 5}{24}$$

Applying *Proposition 1* and using the last result we deduce the Gini's index $\gamma(V_{(p,q)})$ of the random variable $V_{(p,q)}$, that is

$$h(p,q) = \gamma(V_{(p,q)}) = 1 - \frac{2}{\mu_{p,q}} \int_0^1 S_{p,q}(u)\, du = \frac{2 + 5p + 5q + 4pq - 6p^2 - 6q^2}{3(4 - 3p + 3q)}$$

3. Concrete results

3.1. The variation in Gini's index

The variation in Gini's indexes $h(p,q) = \gamma(V_{(p,q)})$ attached to the income variables $V_{(p,q)}$ with $p \geq 0$, $q \geq 0$, $p + q \leq 1$ is represented graphically in Figure 2.

The concrete form of the surface $z = h(p,q)$, $p \geq 0$, $q \geq 0$, $p + q \leq 1$, specified by *Proposition 3* suggests us the possibility to choose different parameters (p,q), which give the same value for the Gini index $\gamma(V_{(p,q)})$ (see also Figure 2).

Figure 2

The surface $z = h(p,q)$ for the Gini index $\gamma(V_{(p,q)})$

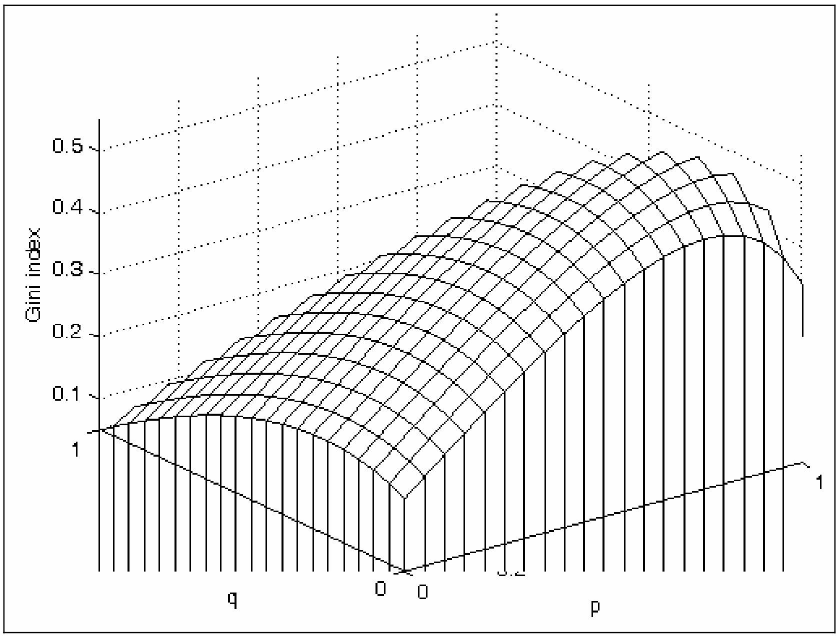


Table 1

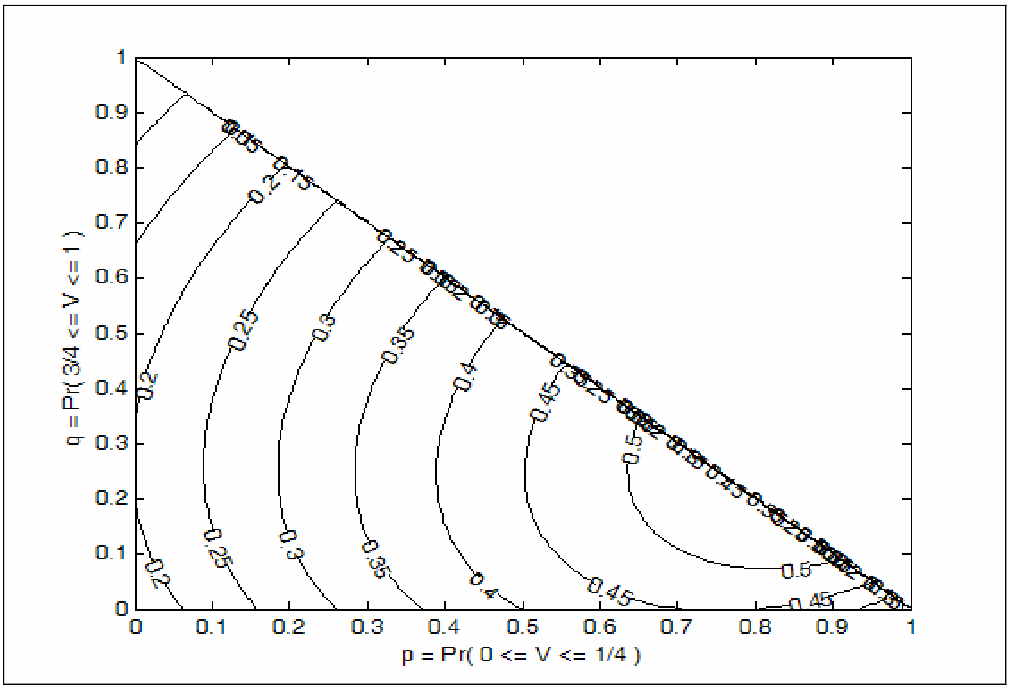
The values of the Gini index $\gamma(V_{(p,q)})$ for different $0 \le p,q \le 1$

P	q=0.1	q=0.2	q=0.3	q=0.4	q=0.5	q=0.6	q=0.7	q=0.8	q=0.9
0.1	0.243	0.254	0.255	0.248	0.233	0.213	0.189	0.160	0.127
0.2	0.296	0.307	0.307	0.299	0.283	0.262	0.235	0.205	---
0.3	0.345	0.357	0.357	0.347	0.330	0.308	0.280	---	---
0.4	0.391	0.404	0.404	0.393	0.375	0.351	---	---	---
0.5	0.433	0.447	0.447	0.436	0.417	---	---	---	---
0.6	0.469	0.486	0.486	0.475	---	---	---	---	---
0.7	0.497	0.517	0.519	---	---	---	---	---	---
0.8	0.512	0.539	---	---	---	---	---	---	---
0.9	0.508	---	---	---	---	---	---	---	---

This idea is clearly sustained when we represent the contour curves of the surface $z = h(p,q)$ for imposed levels z of the Gini index (Figure 3).

Figure 3

The contour curves obtained for different levels of the Gini index $\gamma(V_{(p,q)})$



More details are presented in Table 1, where the Gini index $\gamma(V_{(p,q)})$ is computed for an arbitrary income distribution $V_{(p,q)}$ with $p \geq 0$, $q \geq 0$, $p+q \leq 1$.

3.2. Examples and comments

Example 1. Applying *Proposition 3* we deduce the concrete Gini index values $\gamma(V)$ for many possible distributions (p, r, q) of the income variable $V_{(p,q)}$, where $r = 1 - p - q$.

The analyzed income distributions V_k , $1 \leq k \leq 26$, are presented in Table 2.

Interpreting the values from Table 2 we remark :

- There are many different income distributions (p, r, q) with the same value for the Gini index $\gamma(V)$.
- If we intend to use the Gini's index for quantifying the intensity of the poverty phenomenon, then we could obtain the same γ value for the populations P_k with very distinct proportions p_k of poor people. One should compare, for example, the income variables V_5 and V_{10}, where the proportions of the poor group are very different (Table 2). The contrast between the populations P_5 and P_{10} is pointed out more clearly in Figure 4.

Table 2

The Gini index $\gamma(V_k)$ for the income distribution (p_k, r_k, q_k)

k	p_k	r_k	q_k	$\gamma(V_k)$	k	p_k	r_k	q_k	$\gamma(V_k)$
1	0.300	0.322	0.378	0.35	2	0.300	0.570	0.130	0.35
3	0.350	0.135	0.515	0.35	4	0.350	0.623	0.027	0.35
5	0.400	0.254	0.346	0.40	6	0.400	0.446	0.154	0.40
7	0.450	0.067	0.483	0.40	8	0.450	0.500	0.050	0.40
9	0.500	0.500	0.000	0.40	10	0.950	0.041	0.009	0.40
11	0.550	0.016	0.434	0.45	12	0.550	0.359	0.091	0.45
13	0.600	0.358	0.042	0.45	14	0.650	0.335	0.015	0.45
15	0.700	0.299	0.001	0.45	16	0.800	0.200	0.000	0.45
17	0.850	0.140	0.010	0.45	18	0.900	0.074	0.026	0.45
19	0.950	0.001	0.049	0.45	20	0.650	0.010	0.340	0.50
21	0.650	0.174	0.176	0.50	22	0.700	0.190	0.110	0.50
23	0.750	0.167	0.083	0.50	24	0.800	0.126	0.074	0.50
25	0.850	0.075	0.075	0.50	26	0.900	0.013	0.087	0.50

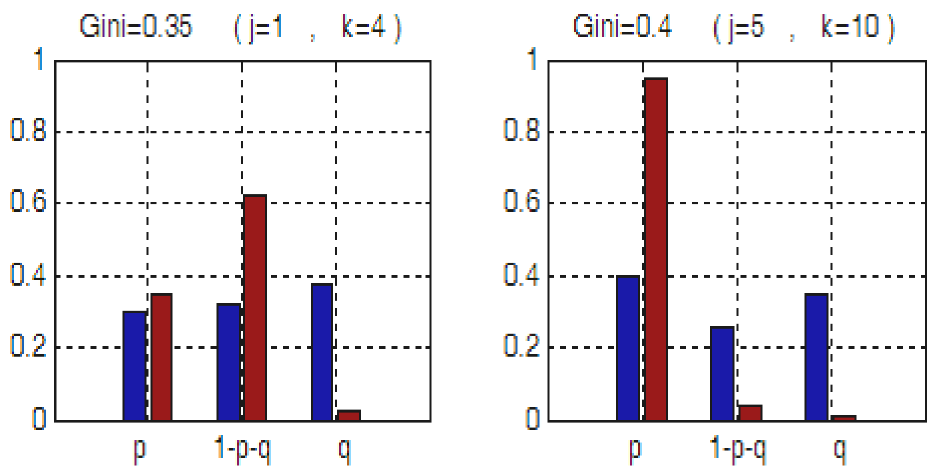
Some selected income distributions V_j, V_k from Table 2 were compared graphically in Figure 4 by using the bar-type diagrams. Moreover, the variables V_j, V_k verify the equality $\gamma = \gamma(V_j) = \gamma(V_k)$.

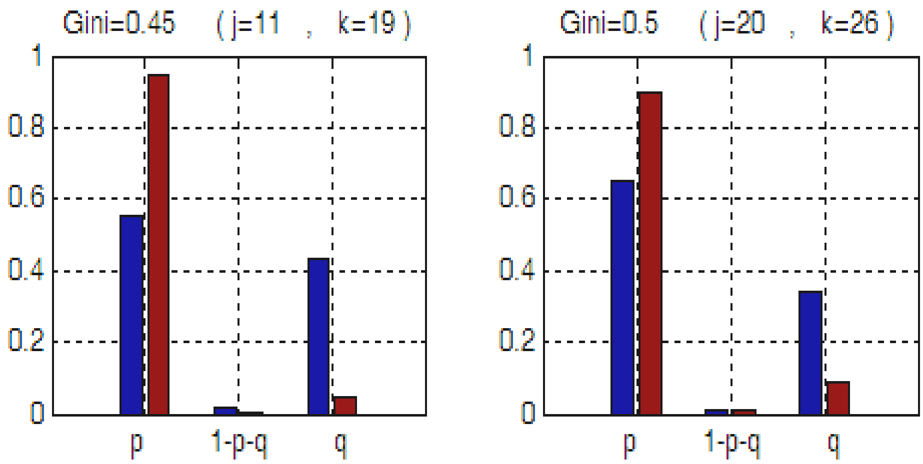
In the following, we shall comment shortly the four bar diagrams of Figure 4:

- The proportions of the poor in the populations P_1 and P_4 are close (between 0.30 and 0.35 in Table 2). But these two populations are clearly very distinct if we take into consideration the middle class or the rich people. Finally, the relation $\gamma(V_1) = \gamma(V_4) = 0.35$ is obtained.
- The populations P_5 and P_{10} are clearly dissimilar. Compare, for example, the proportions of individuals who belong to an arbitrary income category (the poor, middle class or rich people) in these populations (Table 2, Figure 4). Nevertheless, we get the equalities $\gamma(V_5) = \gamma(V_{10}) = 0.40$ (Table 2).
- The both populations P_{11} and P_{19} are not middle-class (Table 2, Figure 4) but they are enough distinct related to their proportions of the poor or rich people (see the concrete values in Table 2). However, the equality $\gamma(V_{11}) = \gamma(V_{19}) = 0.45$ is obtained (Table 2).
- An analogous discussion is related to the groups P_{20} and P_{26}, where we get the relations $\gamma(V_{20}) = \gamma(V_{26}) = 0.50$ (see Table 2 and Figure 4).

Figure 4

The distributions (p, r, q) of the random variables V_j and V_k





4. Final remarks

In the previous sections we proved that Gini's index is not always quite adequate to measure the poverty phenomenon. For this reason, we must improve the classical Gini index by proposing new extensions (Chakravarty, 1983; Foster, Greer, Thorbecker, 1984; Kakwani, 1980; Yitzhaki, 1983; Donaldson, Weymark, 1980).

Moreover, we recommend using comparatively many other complementary indices for measuring the social inequality. We mention here the indicators based on the entropy principle (see, for example, the Sen or Theill indices; Kakwani, 1980; Somarriba, Pena, 2009), the application of the minimum informational energy for choosing an appropriate measure for the poverty (the Onicescu coefficient, see Onicescu, Ştefănescu, 1979), to select the robust indicators on the basis of an ordinal approach (Sen, 1976; Foster, McGillivray, Seth, 2009), to establish a natural correlation with other associated processes (for example, the polarization phenomenon treated by Ştefănescu, 2009) or to combine and weight, too, the known indices to obtain new composite coefficients (Zhou, Ang, 2010).

Nevertheless, we could not neglect the multidimensional approach of the poverty phenomenon, where, beside the income component, many other discriminate variables between individuals (Lugo, 2007, Maasoumi, Lugo, 2008) are used frequently.

In fact, any indicator like $\gamma(F)$, which is associated to a c.d.f. $F(v)$ approximates the global behaviour of the v data. For this reason, the index $\gamma(F)$ is more adequate to measure only some characteristics of the distribution F (for example, the inequality, the polarization or the asymmetry properties).

Finally, we remark here that all the newly proposed inequality indicators must verify at least the four mentioned axioms ***A1-A4***, which are imposed by the specificity of the poverty phenomenon.

References

Chakravarty, S.R., 1983. A new index of poverty. *Mathematical Social Sciences*, 6, pp.307–313.

Donaldson, D. and Weymark, J.A., 1980. A single parameter generalization of the Gini indices of inequality. *Journal of Economic Theory*, 22, pp.67–86.

Foster, J., Greer, J. and Thorbecker, E., 1984. A class of decomposable poverty measures. *Econometrica*, 52(3), pp.761–776.

Foster, J., McGillivray, M. and Seth, S., 2009. Rank robustness of multidimensional well-being measures. *OPHI Working Paper* 26-2009.

Gastwirth, J.L., 1971. A general definition of a Lorenz curve. *Econometrica*, 39. pp.1037-1039.

Kakwani, N.C., 1980. On a class of poverty measures. *Econometrica,* 43, pp. 37-446.

Lugo, M.A., 2007. Comparing multidimensional indices of inequality: Methods and application. *Research on Economic Inequality*, 14, pp. 213-236.

Maasoumi, E. and Lugo, M.A., 2008. The information basis of multivariate poverty assessments. In: N. Kakwani, J. Silber eds. *Quantitative Approaches to Multidimensional Poverty Measurement*. Palgrave-MacMillan.

Onicescu, O. and Stefanescu, V., 1979. *Elements of informational statistics with applications*. Bucharest: Technical Publishing House (in Romanian).

Papoulis, A., 1990. *Probability and statistics*. New Jersey: Prentice Hall.

Sen, A.K., 1976. Poverty : An ordinal approach to measurement. *Econometrica*, 44(2), pp. 219-231.

Somarriba, N. and Pena, B., 2009. Synthetic indicators of quality of life in Europe. *Social Indicators Research*, 94, pp.115-133.

Stefanescu, St., 2009. Measurement of the bipolarization events. International Journal of Applied Mathematics and Computer Sciences, 5(2),pp. 74-81.

Yitzhaki, S., 1983. On an extension of the Gini index. *International Economic Review*, 24, pp.617–628.

Zhou, P., Ang, B.W. and Zhou, D.Q., 2010. Weighting and aggregation in composite indicator construction: A multiplicative optimization approach. *Social Indicators Research*, 96(1), pp.169-181.