STATISTICA, anno XLVII, n. 2, 1987

ABOUT A GENERAL METHOD FOR THE LOWER AND UPPER DISTRIBUTION-FREE BOUNDS ON GINI'S CONCENTRATION RATIO FROM GROUPED DATA (*)

Giovanni Maria Giorgi and Andrea Pallini (**)

1. INTRODUCTION

Official statistical agencies collecting information on the distribution of individual or family income of a country generally publish the data grouped in intervals. In this case a method of estimating Gini's concentration ratio R without fitting curves to the data consists in determining the lower ($\hat{R}L$) and upper ($\hat{R}U$) bounds on $\hat{R}$.

In an interesting article, Gastwirth [6] proposes distribution-free bounds on the Gini index from grouped data when limits and mean income of each interval are known.

Careful examination of the material published on the subject shows that Pizzetti [23] was a precursor of Gastwirth [6], as we shall demonstrate in Section 2. We also wish to point out that the algorithms considered by both authors for estimating the upper bound on the concentration ratio are not rigorously correct from the mathematical point of view and this may be deduced from some of Benedetti's results [2, p. 178, 189 and 206-210].

Gastwirth and Smith subsequently propose the above bounds for testing the fit of a distribution to grouped data [7].

In our opinion, Gastwirth has the merit of initiating international discussion on the subject and motivating other scholars such as Mehran [18, 19] and Nygard and Sandström [21, p. 297-298] to take up the question and propose other solutions.

The debate has also recently broached certain inferential aspects, for example in the papers of McDonald and Ransom [17] and Gastwirth, Nayak and Krieger [8]. The former authors show that the bounds on the Gini index from sample

---

(*) Research partially supported by M.P.I.

The authors thank professor Carlo Benedetti of the University "La Sapienza" of Rome for his constructive criticism of the first draft of this paper and his suggestions on certain points dealt with in Section 2.

(**) Sections 1 and 2 are by G.M. Giorgi and Sections 3 and 4 by A. Pallini.

data cannot be used inferentially without taking into account the sampling variation which depends both on the number and type of class grouping used. The latter derive the asymptotic distribution of the bounds for random samples grouped in intervals.

The present paper contributes to the subject by demonstrating the relationship between the methods of Pizzetti and Gastwirth (Section 2) and proposing a new algorithm for estimating the upper bound on the concentration ratio partly based on Benedetti's results [2, p. 178, 189 and 206-210].

In Section 3 we give new formulas for RL and RU deduced from the relations between R and population parameters, some of which were considered in a recent papers of ours [12].

In the last section we give a numerical illustration of the procedures treated, based on data from the Italian Household Sample Survey (1984) by the Bank of Italy.

2. SOME THEORETICAL CONSIDERATIONS ON THE METHODS OF PIZZETTI AND GASTWIRTH

Let us suppose that income X is a random variable defined on $[0, \infty)$, with the continuous and differentiable cumulative distribution function $F(x)$

$$F(x) = \int_0^x dF(t) \tag{1}$$

Assuming that the first moment μ about zero exists, is finite and different from zero, then the first moment distribution function ${}_1F(x)$ is

$${}_1F(x) = \frac{1}{\mu} \int_0^x t\, dF(t) \tag{2}$$

Equations (1) and (2) define the Lorenz curve in the orthogonal plane $\{F(x); {}_1F(x)\}$.

Gini's concentration ratio $R \in [0,1]$ is

$$R = \frac{\Delta}{2\mu} \tag{3}$$

where

$$\Delta = \int_0^\infty \int_0^\infty | x - y |\, dF(x)\, dF(y) \tag{4}$$

is the mean difference.

If the interval in which the Lorenz curve is defined is partitioned into k subintervals $[F(a_{t-1}), F(a_t)]$ with $t = 1, 2, \ldots, k$, a numerical approximation of (3) is given by the trapezoidal rule:

$$R' = 1 - \sum_{t=1}^{k} [F(a_t) - F(a_{t-1})]\, [{}_1F(a_t) + {}_1F(a_{t-1})] \tag{5}$$

Obviously (5) gives an underestimate of (3) wich augments with the width of the intervals $[F(a_{t-1}), F(a_t)]$. The error incurred considering (5) instead of (3) is shown by 2 [shaded area] in Fig. 1 (1), according to the geometric approach of Kakwani [15, p. 98-100] to the bounds on the Gini index.

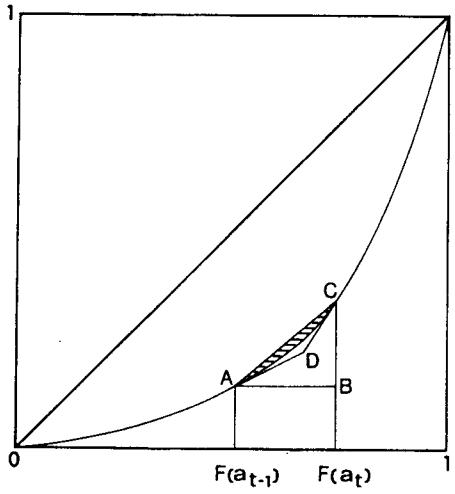

Figure 1

$$2\,[\text{shaded area}] = \frac{\Delta_t}{2\,\mu_t}\; 2\,[\text{area of triangle } ABC]$$

$$= \frac{\Delta_t}{2\,\mu_t}\,[F(a_t) - F(a_{t-1})]\,[{}_1F(a_t) - {}_1F(a_{t-1})] \tag{6}$$

where Δ_t and μ_t are the mean difference and mean income respectively in the interval t ($t = 1, 2, \ldots, k$).

Also since

$$[F(a_t) - F(a_{t-1})]\,\frac{\mu_t}{\mu} = {}_1F(a_t) - {}_1F(a_{t-1}) \tag{7}$$

we can write (3) as follows:

$$R = R' + \frac{1}{2\mu}\sum_{t=1}^{k}[F(a_t) - F(a_{t-1})]^2\,\Delta_t \tag{8}$$

Gastwirth [6, p. 310] proposes the following upper bound on the concentration ratio (2)

---

(1) See for example Pietra [22, p. 780], Kendall and Stuart [16, p. 49].

(2) See also Gastwirth, Nayak and Krieger [8, p. 269].

$$_1RU = R' + \frac{1}{\mu} \sum_{t=1}^{k} [F(a_t) - F(a_{t-1})]^2 \frac{(a_t - \mu_t)(\mu_t - a_{t-1})}{(a_t - a_{t-1})} \tag{9}$$

obtained by substituting the following statistic (3) for Δ_t in (8)

$$_1\Delta U_t = 2 \frac{(a_t - \mu_t)(\mu_t - a_{t-1})}{(a_t - a_{t-1})} \tag{10}$$

which is geometrically (see Fig. 1) equal to 2 [area of triangle ADC] $2\mu_t$ in which segments $\overline{AD}$ and $\overline{CD}$ have the same slope as the Lorenz curve at points A and C, i.e. a_{t-1}/μ and a_t/μ respectively.

In practice the problem is in estimating (9) from the incomes x_i of n recipients (individuals or households) grouped in k classes.

Following Gastwirth's approach [6, p. 306; 8, p. 269] let us suppose that the n incomes x_i constitute the elements of a simple random sample of size n drawn from a cumulative distribution function $F(x)$. Let us also assume that these incomes are arranged in non-decreasing order such that $x_{i-1} \leqslant x_i$ $(i = 1, 2, \ldots, n)$ and that the data are grouped in k classes $[a_{t-1}, a_t)$ with $t = 1, 2, \ldots, k$ where the last class is open and $a_0 = 0$.

It is well known that in the discrete case the mean difference (without repetition) is

$$\hat{\Delta} = \frac{1}{n(n-1)} \sum_{i \neq j} \sum |x_i - x_j| \tag{11}$$

and for data grouped in classes, Gini [9, p. 1239] proposes the following formula:

$$\hat{\Delta} = \hat{\Delta}' + \frac{1}{n(n-1)} \sum_{t=1}^{k} n_t (n_t - 1) \hat{\Delta}_t \tag{12}$$

where $\hat{\Delta}'$ is the index calculated assuming that each recipient earns an income equal to the mean of the class to which he belongs, while $\hat{\Delta}_t$ and n_t are the mean difference and the size of class t $(t = 1, 2, \ldots, k)$.

Dividing (12) by $2m$, with $m = (1/n) \Sigma x_i$, Gini [9, p. 1239] obtains the concentration ratio $R \in [0,1]$ from grouped data

$$\hat{R} = \hat{R}' + \frac{1}{2n(n-1)m} \sum_{t=1}^{k} n_t (n_t - 1) \hat{\Delta}_t \tag{13}$$

where $\hat{R}'$ is the Gini index calculated by $\hat{\Delta}'$.

Now to estimate (9) using (13) we need to substitute for $\hat{\Delta}_t$ in (13) an estimate of (10) obtained in the discrete case, calculating the maximum of $\hat{\Delta}_t$ compatible with the limits and the mean of class t.

If we consider the class t $(t = 2, \ldots, k-1)$, we have the maximum concentration when the x_j $(j = 1, 2, \ldots, n_t)$ it contains constitute a succession of

(3) The statistic (10) is deducible from Gini [11, p. 33] and Zanardi [26, p. 360]. The latter obtains it by another path i.e. by mean deviation curves.

h_t terms all equal to a_{t-1} and of $n_t - h_t$ terms all equal to $a_t - \epsilon_t$ where ϵ_t is any small positive number, since we have assumed that the class is of the type $[a_{t-1}, a_t)$. Thus, in agreement with the conditions:

$$0 < a_{t-1} \leqslant x_j \leqslant (a_t - \epsilon_t); \quad \sum_{j=1}^{n_t} x_j = n_t \; m_t = \text{constant}; \quad n_t = \text{constant} \quad (14)$$

the number h_t and $n_t - h_t$ of incomes equal to a_{t-1} and $a_t - \epsilon_t$ respectively may be obtained from the relation

$$h_t \, a_{t-1} + (n_t - h_t)(a_t - \epsilon_t) = n_t \, m_t \quad (15)$$

from which

$$h_t = n_t \frac{(a_t - \epsilon_t) - m_t}{(a_t - \epsilon_t) - a_{t-1}} \; ; \quad n_t - h_t = n_t \frac{m_t - a_{t-1}}{(a_t - \epsilon_t) - a_{t-1}} \quad (16)$$

However, as deduced from Benedetti [2, p. 178], h_t and $n_t - h_t$ may not be integers and in fact the site of the maximum in question requires us to consider the first $z_t = [h_t]$ incomes equal to a_{t-1} (where $[h_t]$ is the maximum integer contained in h_t), the income

$$x_{z_t+1} = \theta_t \, a_{t-1} + (1 - \theta_t)(a_t - \epsilon_t) \; ; \qquad \theta_t = h_t - [h_t]$$

and the remaining incomes equal to $a_t - \epsilon_t$.

As deduced again from Benedetti [2, p. 189, 206-210], for $\epsilon_t \to 0$ ($t = 2, \ldots, k-1$) we can therefore write the maximum of $\hat{\Delta}_t$ as follows (4):

$${}_1\hat{\Delta}U_t = \frac{2}{n_t (n_t - 1)} (a_t - a_{t-1}) [z_t (n_t - z_t)(1 - \theta_t) + (z_t + 1)(n_t - z_t - 1)\theta_t] \quad (17)$$

which for $\theta_t = 0$ becomes (5)

$${}_1\hat{\Delta}U_t = 2 \frac{n_t}{n_t - 1} \frac{(a_t - m_t)(m_t - a_{t-1})}{(a_t - a_{t-1})} \quad (18)$$

In particular if we consider the first class $[0, a_1)$, the maximizing distribution is composed of the first $z_1 = [h_1]$ incomes equal to zero, an income

$$x_{z_1+1} = (1 - \theta_1)(a_1 - \epsilon_1)$$

and the remaining $n_1 - z_1 - 1$ incomes equal to $a_1 - \epsilon_1$.

---

(4) Benedetti demonstrates a formula analogous to (17) for ungrouped data.

(5) See Gini [11, p. 33] regarding (18). It can also be demonstrated that for very large values of n_t it makes no difference whether we consider (18) or (17).

Thus for $\epsilon_1 \to 0$ the maximum of $\hat{\Delta}_1$ is

$$_1\hat{\Delta}U_1 = \frac{2}{n_1(n_1-1)}\, a_1 [z_1(n_1-z_1)(1-\theta_1) + (z_1+1)(n_1-z_1-1)\theta_1] \quad (19)$$

which for $\theta_1 = 0$ becomes ([6])

$$_1\hat{\Delta}U_1 = 2\,\frac{n_1}{n_1-1}\,\frac{m_1(a_1-m_1)}{a_1} \quad (20)$$

For the last class which we took to be open the maximizing distribution is obtained by considering the first $n_k - 1$ incomes equal to a_{k-1} and the last equal to $n_k\, m_k - (n_k - 1)\, a_{k-1}$, giving a maximum ([7]) of $\hat{\Delta}_k$

$$_1\hat{\Delta}U_k = 2\,(m_k - a_{k-1}) \quad (21)$$

Inserting the expressions (17), (19) and (21) in (13) we obtain the following estimate of (9):

$$\begin{aligned} _1\hat{R}U = \hat{R}' &+ \frac{a_1 [z_1(n_1-z_1)(1-\theta_1)+(z_1+1)(n_1-z_1-1)\theta_1]}{n(n-1)m} + \\ &+ \frac{1}{n(n-1)m}\sum_{t=2}^{k-1}(a_t-a_{t-1})[z_t(n_t-z_t)(1-\theta_t)+(z_t+1)(n_t-z_t-1)\theta_t] + \\ &+ \frac{n_k(n_k-1)(m_k-a_{k-1})}{n(n-1)m} \end{aligned} \quad (22)$$

Now we shall analyse the contributions of Pizzetti [23] and Gastwirth [6] for estimating (9).

Pizzetti [23, p. 582-583], in order to express the maximum concentration inside class t compatible with the limits and mean income of the class, uses several of Gini's results [11, p. 33] and obtains an algorithm ([8]) similar to (18) which, as we have seen, is a special case of (17) with $\theta_t = 0$. So inserting (18) in (13) he obtains ([9]) the upper bound $\hat{R}U^*$ on the concentration ratio from grouped data

$$\hat{R}U^* = \hat{R}' + \frac{1}{n(n-1)m}\sum_{t=1}^{k} n_t^2\,\frac{(a_t-m_t)(m_t-a_{t-1})}{(a_t-a_{t-1})} \quad (23)$$

([6]) See Gini [10, p. 6] regarding (20). It can be demonstrated here too that for very large values of n_1 we can use (20) instead of (19).

([7]) See Gini [11, p. 30].

([8]) Actually Pizzetti does not use (18) but a similar formula without the term $n_t/n_t - 1$ which we consider to be necessary as we are dealing in terms of mean difference without repetition.

([9]) For the reasons explained in footnote (8) Pizzetti does not obtain (23) but a similar formula with $n_t\,(n_t - 1)$ in the place of n_t^2.

Remembering that we assumed $a_0 = 0$ and the last income class open, by the above procedure we can rewrite (23) as follows:

$$\hat{R}U^* = \hat{R}' + \frac{n_1^2\, m_1}{n\,(n-1)\,m}\,\frac{a_1 - m_1}{a_1} + \\ + \frac{1}{n\,(n-1)\,m}\sum_{t=2}^{k-1} n_t^2\,\frac{(a_t - m_t)\,(m_t - a_{t-1})}{(a_t - a_{t-1})} + \\ + \frac{n_k\,(n_k - 1)}{n\,(n-1)\,m}\,(m_k - a_{k-1}) \tag{24}$$

The same result may be obtained, reasoning in terms of mean difference with repetition $\hat{\Delta}_r = [(n-1)/n]\,\hat{\Delta}$. In fact in this case (12) and (18) become

$$\hat{\Delta}_r = \hat{\Delta}_r' + \frac{1}{n^2}\sum_{t=1}^{k} n_t^2\,\hat{\Delta}_{rt} \tag{25}$$

$${}_1\hat{\Delta}U_{rt} = 2\,\frac{(a_t - m_t)\,(m_t - a_{t-1})}{(a_t - a_{t-1})} \tag{26}$$

by inserting (26) in (25) and dividing (25) by $[(n-1)/n]\,2m$ we obtain (23).

Gastwirth [6, p. 309-310] also begins with the mean difference with repetition, namely (25) in order to estimate the upper bound (9) of the Gini index and obtains the statistic

$$\hat{G}U^* = \hat{G}' + \frac{1}{n^2 m}\sum_{t=1}^{k} n_t^2\,\frac{(a_t - m_t)\,(m_t - a_{t-1})}{(a_t - a_{t-1})} \tag{27}$$

by substituting (26) for $\hat{\Delta}_{rt}$ in (25) and dividing the latter by $2m$. As we did before we can rewrite (27) as follows:

$$\hat{G}U^* = \hat{G}' + \frac{n_1^2\, m_1}{n^2\, m}\,\frac{a_1 - m_1}{a_1} + \\ + \frac{1}{n^2 m}\sum_{t=2}^{k-1} n_t^2\,\frac{(a_t - m_t)\,(m_t - a_{t-1})}{(a_t - a_{t-1})} + \\ + \frac{n_k\,(n_k - 1)}{n^2\, m}\,(m_k - a_{k-1}) \tag{28}$$

where $\hat{G}U^*$ is the upper bound on $\hat{G} = \hat{\Delta}_r/2m$ with $\hat{G} \in [0, 1 - (1/n)]$ while $\hat{G}'$ is the value assumed by this index ([10]) when all recipients have an income equal to the mean income of the class to which they belong:

$$\hat{G} = \frac{n-1}{n}\,\hat{R} \tag{29}$$

([10]) Geometrically, the index $\hat{G}$ is 2 [non-normalized area of concentration] (see for example Nygard and Sandström [21, p. 267]).

From this it is clear that Pizzetti [23] was a precursor of Gastwirth's [6] contribution. However neither realized that (23) and (27) are not rigorously correct from a mathematical point of view, as can be deduced from Benedetti [2, p. 189, 206-210]. The use of (24) and (28) without regard to the values of θ_1 and θ_t $(t = 2, \ldots, k-1)$ is justified only if the size of the classes is very large, which is not always the case in practice.

Finally it is interesting to recall some of the results of Hoeffding [14, p. 297, 313] who shows that the sampling estimate of the mean difference without repetition belongs to the class of the "*U*-statistics" which is a correct estimate [14, p. 310] of (4) and more generally that $\sqrt{n}\,(\hat{\Delta} - \Delta)$ is asymptotically normal $(0, \sigma^2\,(\hat{\Delta}))$. He also shows for R, expressed as a relation between two "*U*-statistics", that $\sqrt{n}\,(\hat{R} - R)$ is asymptotically normal ([11]) $(0, \sigma^2\,(\hat{R}))$.

This is why we prefer to start from the mean difference without repetition i.e. from (13), taking (17) and (19) into account for the correct calculation of ${}_1\hat{R}U$. In so doing we remain coherent with our hypothesis of estimating (9) on the basis of a simple random sample of size n.

We conclude that the most important aspect of the contributions of Pizzetti and Gastwirth is that although they did not consider (17) and (19) they succeeded in expressing the upper bound on the Gini index as a function of the maximum concentration in every class compatibly with the limits and the mean in the class itself. This was accomplished by Pizzetti using an algorithm similar to (23) and by Gastwirth with (27) and suggests the possibility of a general method for the construction of lower RL and upper RU bounds on the concentration ratio from grouped data.

3. LOWER *RL* AND UPPER *RU* BOUNDS ON THE GINI INDEX BASED ON SOME RL_t AND RU_t

Let us continue to consider the income X as a random variable as defined above and from (8) we can deduce the lower RL and upper RU bounds on the Gini index by substituting for the mean difference Δ_t $(t = 1, 2, \ldots, k)$ its bounds ΔL_t and ΔU_t

$$RL = R' + \frac{1}{2\mu} \sum_{t=1}^{k} [F(a_t) - F(a_{t-1})]^2\, \Delta L_t \tag{30}$$

$$RU = R' + \frac{1}{2\mu} \sum_{t=1}^{k} [F(a_t) - F(a_{t-1})]^2\, \Delta U_t \tag{31}$$

On the basis of the results of Stuart [25, p. 40], Michetti and Dall'Aglio [20, p. 185], Zanardi [26, p. 343] and Rigo [24, p. 629] it is possible to deduce

([11]) There are also some interesting results on the properties of the concentration ratio and its asymptotic behaviour in Michetti and Dall'Aglio [20], Dall'Aglio [5], Cucconi [4].

$$_2\Delta L_t = 0 \tag{32}$$

$$_2\Delta U_t = \frac{2}{\sqrt{3}}\,\sigma_t \tag{33}$$

where σ_t is the standard deviation in the interval t ($t = 1, 2, \ldots, k$). Inserting (32) into (30) and (33) into (31) we have

$$_2RL = R' \tag{34}$$

$$_2RU = R' + \frac{1}{\mu}\sum_{t=1}^{k}[F(a_t) - F(a_{t-1})]^2\,\frac{\sigma_t}{\sqrt{3}} \tag{35}$$

If we also consider some of the results of our recent paper [12, p. 381] we obtain

$$_3\Delta L_t = S_{\mu_t} \tag{36}$$

$$_3\Delta U_t = 2S_{\mu_t} \tag{37}$$

$$_4\Delta L_t = S_{me_t} \tag{38}$$

$$_4\Delta U_t = 2S_{me_t} \tag{39}$$

where S_{μ_t} and S_{me_t} are the mean deviations about the mean value μ_t and the median me_t in the interval t ($t = 1, 2, \ldots, k$).

Inserting (36) into (30) and (37) into (31) we obtain the following bounds:

$$_3RL = R' + \frac{1}{2\mu}\sum_{t=1}^{k}[F(a_t) - F(a_{t-1})]^2\,S_{\mu_t} \tag{40}$$

$$_3RU = R' + \frac{1}{\mu}\sum_{t=1}^{k}[F(a_t) - F(a_{t-1})]^2\,S_{\mu_t} \tag{41}$$

Furthermore if we insert (38) into (30) and (39) into (31) we have

$$_4RL = R' + \frac{1}{2\mu}\sum_{t=1}^{k}[F(a_t) - F(a_{t-1})]^2\,S_{me_t} \tag{42}$$

$$_4RU = R' + \frac{1}{\mu}\sum_{t=1}^{k}[F(a_t) - F(a_{t-1})]^2\,S_{me_t} \tag{43}$$

For estimating (33) we refer to Glasser [13, p. 177] and Bhandari and Mukerjee [3, p. 260] from which we deduce

$$_2\hat{\Delta}U_t = 2\,\frac{\hat{\sigma}_t}{\sqrt{3}}\,\sqrt{\frac{n_t + 1}{n_t - 1}} \tag{44}$$

and thus (13) becomes [12]

$$_2\hat{R}L = \hat{R}' \tag{45}$$

$$_2\hat{R}U = \hat{R}' + \frac{1}{2n\,(n-1)\,m} \sum_{t=1}^{k} n_t\,(n_t - 1)\,{}_2\hat{\Delta}U_t \tag{46}$$

For estimating (36) and (37) we refer to Glasser [13, p. 178] and obtain

$$_3\hat{\Delta}L_t = \frac{n_t}{n_t - 1} \cdot \hat{S}_{\mu_t} \tag{47}$$

$$_3\hat{\Delta}U_t = 2\hat{S}_{\mu_t} \tag{48}$$

which when inserted in (13) for Δ_t give

$$_3\hat{R}L = \hat{R}' + \frac{1}{2n\,(n-1)\,m} \sum_{t=1}^{k} n_t\,(n_t - 1)\,{}_3\hat{\Delta}L_t \tag{49}$$

$$_3\hat{R}U = \hat{R}' + \frac{1}{2n\,(n-1)\,m} \sum_{t=1}^{k} n_t\,(n_t - 1)\,{}_3\hat{\Delta}U_t \tag{50}$$

Finally to estimate (38) and (39) we refer to Benedetti [1, p. 182] from which we deduce

$$_4\hat{\Delta}L_t = \frac{n_t}{n_t - 1}\,\hat{S}_{me_t} \qquad (\text{even } n_t) \tag{51}$$

$$_4\hat{\Delta}L_t = \frac{n_t + 1}{n_t - 1}\,\hat{S}_{me_t} \qquad (\text{odd } n_t) \tag{52}$$

$$_4\hat{\Delta}U_t = 2\hat{S}_{me_t} \tag{53}$$

thus (13) can be written

$$_4\hat{R}L = \hat{R}' + \frac{1}{2n\,(n-1)\,m} \sum_{t=1}^{k} n_t\,(n_t - 1)\,{}_4\hat{\Delta}L_t \tag{54}$$

$$_4\hat{R}U = \hat{R}' + \frac{1}{2n\,(n-1)\,m} \sum_{t=1}^{k} n_t\,(n_t - 1)\,{}_4\hat{\Delta}U_t \tag{55}$$

([12]) By $\hat{R}'$ we intend the estimate of (5)

$$\hat{R}' = \frac{1}{n\,(n-1)\,m} \sum_{t=1}^{k} (N_t + N_{t-1} - 1)\,m_t\,n_t - 1\ ; \qquad N_t = \sum_{j=1}^{t} n_j$$

also deducible from Gini [9, p. 1213].

where we substitute (51) or (52) for ${}_4\hat{\Delta}L_t$ in (54) according to whether the number of recipients in each class is even or odd.

4. A NUMERICAL ILLUSTRATION

In order to empirically compare the intervals $[{}_s\hat{R}L, {}_s\hat{R}U]$ with $s = 1,2,3,4$ we have calculated the bounds from Italian Household Sample Survey data (1984) collected by the Bank of Italy ([13]). For the sake of simplicity and clarity in the application of the formulas we have used unweighted data. Considering the purely illustrative purpose of this exercise we have not taken sampling variations into account as suggested by McDonald and Ransom [17] as this would go beyond the purpose of this section and this paper in general. Our aim was to examine certain aspects of Pizzetti's and Gastwirth's methods for estimating the upper bound on R, to propose a modification in the light of Benedetti's findings [2, p. 189, 206-210] and to obtain new bounds on the Gini index from grouped data.

TABLE 1

Income Distribution Data in 10 Income Classes

i	Income Classes (Lire · 10^3)	(*) $\sum_{t=1}^{i} \frac{n_t}{n}$	(**) $\sum_{t=1}^{i} \frac{n_t m_t}{nm}$	m_t	$\hat{me}_t$	$\hat{S}_{\mu_t}$	$\hat{\sigma}_t$	$\hat{S}_{me_t}$
1	- 4000	.0115	.0018	3346.167	3464.5	411.715	591.523	402.833
2	4000- 5500	.0376	.0074	4749.183	4740.0	356.479	416.414	356.395
3	5500- 7000	.0710	.0168	6162.683	6160.0	382.314	439.506	382.295
4	7000- 9000	.1297	.0383	8027.233	8060.0	457.938	550.208	457.192
5	9000-11000	.1980	.0693	9961.344	10000.0	482.719	578.603	480.979
6	11000-15000	.3689	.1711	13055.466	13040.0	972.553	1123.490	972.532
7	15000-25000	.6903	.4568	19488.332	19350.0	2544.351	2936.384	2542.735
8	25000-35000	.8658	.6915	29333.817	29021.0	2462.260	2854.456	2451.899
9	35000-50000	.9581	.8634	40830.065	40286.0	3488.202	4170.939	3458.070
10	50000-	1.0000	1.0000	71422.297	62100.0	18000.658	27967.387	16495.577

(*) Cumulative share of families with income $< a_t$.
(**) Cumulative income share of families with income $< a_t$.

In order to remain coherent with the theoretical approach adopted, we assumed that the incomes x_i $(i = 1, 2, \ldots, n)$ were a simple random sample drawn from $F(x)$. Making use of the random sample observations we estimate the parameters entering in the Gini ratio's upper and lower bounds. These values may be found in Table 1 for data grouped in 10 classes and the bounds ${}_s\hat{R}L$

([13]) The authors wish to thank Dr. Rocco Pirrotta and Dr. Luigi Cannari of the Bank of Italy for kindly supplying the data of the 20th Sample Survey of Italian Household Budgets.

TABLE 2

Lower and Upper Bounds on the Gini index

	10 Income Classes				20 Income Classes		
s	${}_s\hat{R}L$	$\hat{R}$	${}_s\hat{R}U$	s	${}_s\hat{R}L$	$\hat{R}$	${}_s\hat{R}U$
1	.3224137	.3355915	.3421663	1	.3315671	.3355915	.3377711
2	.3224137	.3355915	.3359579	2	.3315671	.3355915	.3359179
3	.3322905	.3355915	.3421392	3	.3345806	.3355915	.3375762
4	.3322243	.3355915	.3419847	4	.3345223	.3355915	.3374488

and ${}_s\hat{R}U$ ($s = 1,2,3,4$) for grouping of 10 and 20 classes are shown in Table 2 where $\hat{R}$ indicates the exact sampling value of the concentration ratio.

Examination of the lower bounds ${}_s\hat{R}L$ in Table 2 reveals that ${}_1\hat{R}L = {}_2\hat{R}L$ are the smallest values. This is due to the fact that they were obtained under the assumption that every recipient has an income equal to the mean income of the class to which he belongs.

For the upper bounds ${}_s\hat{R}U$ we see that ${}_4\hat{R}U < {}_3\hat{R}U$ which is due to their different geometric significance in terms of the Lorenz curve [12, p. 381].

Finally from the data grouped into 20 classes it becomes evident that the lower and upper bounds enclose a narrower interval than that obtained from the consideration of only 10 classes. This is because more information is afforded by grouping into a greater number of classes.

Istituto di Statistica
Università di Siena

GIOVANNI MARIA GIORGI
ANDREA PALLINI

REFERENCES

[1] C. BENEDETTI (1961), *A proposito dei rapporti tra differenza media e scostamenti medio quadratico, semplice medio e semplice medio dalla mediana*, "Metron", vol. XXI, n. 1-4, pp. 181-185.

[2] C. BENEDETTI (1981), *Istituzioni di Statistica* (II edizione), Veschi, Roma.

[3] S.K. BHANDARI, R. MUKERJEE (1986), *Some Relations among Inequality Measures*, "Sankhya", vol. XLVIII, series B, pt. 2, pp. 258-261.

[4] O. CUCCONI (1965), *Sulla distribuzione campionaria del rapporto R di concentrazione*, "Statistica", anno XXV, n. 1, pp. 119-138.

[5] G. DALL'AGLIO (1965), *Comportamento asintotico delle stime della differenza media e del rapporto di concentrazione*, "Metron", vol. XXIV, n. 1-4, pp. 379-414.

[6] J.L. GASTWIRTH (1972), *The Estimation of the Lorenz Curve and Gini Index*, "Review of Economics and Statistics", vol. LIV, n. 3, pp. 306-316.

[7] J.L. GASTWIRTH, J.T. SMITH (1972), *A New Goodness of Fit Test*, "Proceedings of the American Statistical Association", pp. 320-321.

[8] J.L. GASTWIRTH, T.K. NAYAK, A.M. KRIEGER (1986), *Large Sample Theory for the*

Bounds on the Gini and Related Indices of Inequality Estimated from Grouped Data, "J. Business and Economic Statistics", vol. IV, n. 2, pp. 269-273.

[9] C. GINI (1914), *Sulla misura della concentrazione e della variabilità dei caratteri*, "Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti", A.A. 1913-14, tomo LXXIII, parte II, pp. 1203-1248.

[10] C. GINI (1930), *Sul massimo degli indici di variabilità assoluta e sulle sue applicazioni agli indici di variabilità relativa e al rapporto di concentrazione*, "Metron", vol. VIII, n. 3, pp. 3-65.

[11] C. GINI (1932), *Intorno alle curve di concentrazione*, "Metron", vol. IX, n. 3-4, pp. 3-76.

[12] G.M. GIORGI, A. PALLINI (1986), *Di talune soglie inferiori e superiori del rapporto di concentrazione*, "Metron", vol. XLIV, n. 1-4, pp. 377-390.

[13] G.J. GLASSER (1961), *Relationships between the Mean Difference and Other Measures of Variation*, "Metron", vol. XXI, n. 1-4, pp. 176-180.

[14] W. HOEFFDING (1948), *A Class of Statistics with Asymptotically Normal Distribution*, "Annals of Mathematical Statistics", vol. XIX, pp. 293-325.

[15] N.C. KAKWANI (1980), *Income Inequality and Poverty*, Oxford University Press, Oxford.

[16] M.G. KENDALL, A. STUART (1969), *The Advanced Theory of Statistics* (3rd Edition), vol. 1, Griffin, London.

[17] J.B. McDONALD, M.R. RANSOM (1981), *An Analysis of the Bounds for the Gini Coefficient*, "J. Econometrics", vol. 17, n. 2, pp. 177-188.

[18] F. MEHRAN (1975), *Bounds on the Gini Index Based on Observed Points of the Lorenz Curve*, "J.A.S.A.", vol. LXX, n. 349, pp. 64-66.

[19] F. MEHRAN (1975), *Bounds on the Gini Index of Income Inequality Based on Grouped Data*, in "Dealing with Grouped Income Distribution Data", World Employment Programme Research: Income Distribution and Employment, Working Paper n. 20, International Labour Office, Geneva, pp. 27-36.

[20] B. MICHETTI, G. DALL'AGLIO (1957), *La differenza semplice media*, "Statistica", anno XVII, n. 2, pp. 159-255.

[21] F. NYGARD, A. SANDSTRÖM (1981), *Measuring Income Inequality*, Almqvist & Wiksell International, Stockholm.

[22] G. PIETRA (1915), *Delle relazioni tra gli indici di variabilità* (I, II), "Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti", A.A. 1914-15, tomo LXXIV, parte II, pp. 775-804.

[23] E. PIZZETTI (1955), *Osservazioni sul calcolo aritmetico del rapporto di concentrazione*, Studi in Onore di Gaetano Pietra, Biblioteca di Statistica, n. 1, Cappelli, Bologna.

[24] P. RIGO (1985), *Lower and Upper Distribution Free Bounds for Gini's Concentration Ratio*, "Proceedings International Statistical Institute", 45th Session, Amsterdam, Contributed Papers, Book 2, pp. 629-630.

[25] A. STUART (1954), *The Correlation between Variate-Values and Ranks in Samples from a Continuous Distribution*, "British J. Statistical Psychology", vol. VII, part I, pp. 37-44.

[26] G. ZANARDI (1963), *Le curve degli scostamenti semplici medi*, "Memorie della Accademia Patavina di SS.LL.AA.", Classe di Scienze Matematiche e Naturali, vol. LXXV, A.A. 1962-63, pp. 335-360.

Su di un metodo generale per la determinazione delle soglie inferiori e superiori (distribution-free) del rapporto di concentrazione del Gini per dati raggruppati

Gli Autori propongono un nuovo algoritmo per la determinazione della soglia superiore (distribution-free) del rapporto di concentrazione R del Gini per dati raggruppati.

Rilevano in particolare che il procedimento utilizzato a tal fine da Pizzetti e da Gastwirth non è rigorosamente corretto da un punto di vista matematico come si desume da alcuni risultati dovuti a Benedetti.

Vengono inoltre fornite altre soglie inferiori e superiori (distribution-free) di R.

RÉSUMÉ

Sur une méthode générale pour les bornes inférieures et supérieures de la mesure de concentratione du Gini pour des données groupées

Les Auteurs proposent un nouvel algorithme pour la détermination de la borne supérieure (distribution-free) de la mesure de concentratione R du Gini pour des données groupées.

Il trouvent en particulier que l'approche utilisée dans ce but par Pizzetti et Gastwirth n'est pas rigoureusement correct d'un point de vue mathématique comme l'on peut déduire de quelques résultats dus à Benedetti.

Quelques bonnes inférieures et supérieures (distribution-free) de R sont en outre fournies.