Sankhyā: The Indian Journal of Statistics
1995, Volume 57, Series B, Pt. 1, pp. 10-18

SAMPLING DISTRIBUTION OF THE BONFERRONI INEQUALITY INDEX FROM EXPONENTIAL POPULATION

By G. M. GIORGI and R. MONDANI
University of Rome

SUMMARY. After having expressed the Bonferroni index as a ratio of two linear combinations of order statistics, its exact sampling distribution from exponential population is deduced.

1. Introduction

Over the last twenty years or so, the study of the income inequality has become more and more important and the international scientific community has contributed considerably to this topic. The Gini concentration ratio (1914) and the Lorenz curve (1905) are undoubtedly the inequality measures which have attracted the most interest. For an examination of the vast literature on this subject, see Giorgi (1990, 1992) and Moothathu (1991). However, apart from these and other well known indices, there are some measures which have not received due attention in spite of their having interesting characteristics.

One of these is the Bonferroni (1930) inequality index (B) which, unlike the Gini ratio, is more sensitive at lower levels of income distribution in as much as it gives "more weights to transfer among poor" as already shown by Nygard and Sandström (1981, p. 276). This makes the Bonferroni index particularly suitable for the study of an important aspect of income distribution, viz., the measurement of the intensity of poverty.

We shall here study some of the sampling aspects of the index B. This is important because if valid information is to be derived from an estimate of this index, it is necessary to know the sampling characteristics of the index. To be precise, here we shall first express the index B as a ratio of two linear combinations of order statistics, and then deduce the exact sampling distribution of the index for the case of exponential population.

Paper received. May 1993; revised. March 1994.

AMS (1985) *subject classifications.* Primary 62E15, 62G30; secondary 62P99.

Key words and phrases. Bonferroni inequality index, order statistics, exact sampling distribution, exponential distribution.

2. BONFERRONI INDEX AS A RATIO OF TWO LINEAR COMBINATIONS OF ORDER STATISTICS

Consider a random variable $X \in [0, \infty)$ with continuous and differentiable cumulative distribution function (c.d.f.)

$$F(x) = \int_0^x f(t)dt. \qquad \ldots(1)$$

If the first moment about zero

$$\mu = \int_0^\infty x f(x)dx \qquad \ldots(2)$$

exists and is finite and nonzero, the first moment distribution function (1st m.d.f.) is

$${}_1F(x) = \frac{1}{\mu}\int_0^x t f(t)dt. \qquad \ldots(3)$$

Thus, the partial mean value, for $X \leq x$, is

$$\mu_x = \frac{\int_0^x t f(t)dt}{\int_0^x f(t)dt} = \mu \frac{{}_1F(x)}{F(x)}, \qquad \ldots(4)$$

so that we may write

$$L(F(x)) = \frac{1}{\mu}\mu_x = \frac{{}_1F(x)}{F(x)}. \qquad \ldots(5)$$

Now, the index B (1930, p. 55, 85) may be written in the continuous case as

$$\begin{aligned} B &= \int_0^1 \frac{\mu - \mu_x}{\mu} dF = 1 - \frac{1}{\mu}\int_0^1 \mu_x dF \\ &= 1 - \frac{1}{\mu}\int_0^\infty \frac{\int_0^x t f(t)dt}{\int_0^x f(t)dt} f(x)dx \end{aligned} \qquad \ldots(6)$$

with $0 \leq B \leq 1$.

When the random variable X is income, $F(x)$ represents the cumulative share of recipents with income $\leq x$, while $L(F(x))$ is the mean density of their income.

The parametric equations (1) and (5) define the Bonferroni curve in the orthogonal plane $[F(x); L(F(x))]$ as shown in Figure 1. Such a curve may be represented by the equation

$$L = \phi(F) \qquad \ldots(7)$$

which is continuously increasing in as much as

$$\frac{dL}{dF} > 0. \qquad \ldots(8)$$

In the Bonferroni diagram (cf. Fig. 1), unlike the Lorenz curve (1905), the so-called egalitarian line is the one which connects the point (0, 1) to (1, 1), while

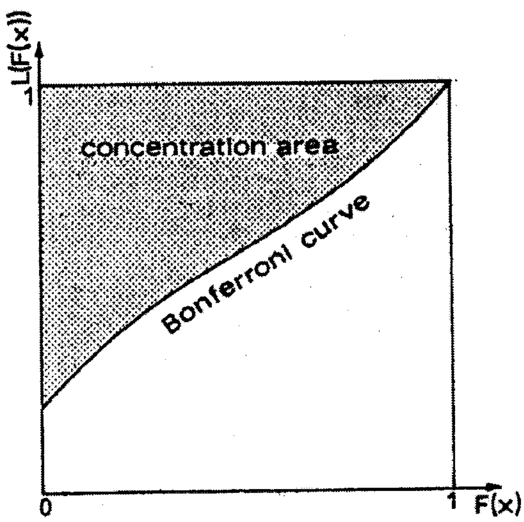

Fig.1. An example of the Banferroni diagram when $X \in (0, \infty)$.

the area contained between the latter and Bonferroni curve is the concentration area. It may be noted that with the increase of the difference

$$1 - L = 1 - \frac{{}_1F}{F} \qquad \ldots (9)$$

the ineqality grows. Theorefore the index B may be written (D'Addario, 1936, p. 11) as

$$\begin{aligned} B &= \int_0^1 (1-L)dF = 1 - \int_0^1 L\,dF \\ &= 1 - \int_0^1 \frac{{}_1F}{F}\,dF. \end{aligned} \qquad \ldots (10)$$

Denoting the i-th rank-ordered observation in a sample of n observations drawn from $F(x)$ by $x_{(i)}$, where $x_{(i-1)} \leq x_{(i)}$ for $i = 1, 2, .., n$, and the sample mean and the sample partial mean of the first i elements of the ordered sample by

$$m = \frac{1}{n}\sum_{i=1}^{n} x_{(i)} \qquad \ldots (11)$$

$$m_i = \frac{1}{i}\sum_{j=1}^{i} x_{(j)}; \qquad (i = 1, 2, ..., n) \qquad \ldots (12)$$

respectively, the sampling estimator of B may be written as (Bonferroni, 1930, p. 55)

$$\begin{aligned} B_n &= (1/(n-1)) \sum_{i=1}^{n-1} \frac{m - m_i}{m} \\ &= 1 - (1/(n-1)) \sum_{i=1}^{n-1} \frac{m_i}{m}, \qquad (0 \leq B_n \leq 1) \end{aligned} \qquad \ldots (13)$$

Therefore the index B is given by the complement to 1 of the arithmetic mean of the ratios between the sample partial means and the sample mean.

Using (11), (13) may be written as

$$B_n = 1 - \frac{n}{(n-1)\sum_{i=1}^{n} x_{(i)}} \sum_{i=1}^{n-1} m_i. \qquad \ldots (14)$$

Given that

$$\sum_{i=1}^{n-1} m_i = \sum_{i=1}^{n-1} \left(x_{(i)} \sum_{j=i}^{n-1} \frac{1}{j} \right) = \sum_{i=1}^{n-1} x_{(i)} \xi_i \qquad \ldots (15)$$

where

$$\xi_i = \sum_{j=i}^{n-1} \frac{1}{j}; \qquad (i = 1, 2, ..., n-1) \qquad \ldots (16)$$

by substituting (15) in (14) we obtain

$$\begin{aligned} B_n &= 1 - \frac{n}{(n-1)\sum_{t=1}^{n} x_{(i)}} \sum_{i=1}^{n-1} x_{(i)} \xi_i \\ &= \frac{\sum_{i=1}^{n-1} (n - 1 - n\xi_i) x_{(i)} + (n-1) x_{(n)}}{(n-1)\sum_{i=1}^{n} x_{(i)}}. \end{aligned} \qquad \ldots (17)$$

Putting $a_i = n - 1 - n\xi_i$ and $a_n = n - 1$, (17) becomes

$$B_n = \frac{\sum_{i=1}^{n-1} a_i x_{(i)} + a_n x_{(n)}}{(n-1)\sum_{i=1}^{n} x_{(i)}}. \qquad \ldots(18)$$

Thus we obtain the index B as a ratio of two linear combinations of order statistics which will enable us to link up with some results obtained by other authors. In particular, Cicchitelli (1976a) considers a random variable analogous to (18) of the type

$$Q = \frac{\sum_{i=1}^{n} a_i x_{(i)}}{\sum_{i=1}^{n} c_i x_{(i)}} \qquad \ldots(19)$$

with c.d.f.

$$G(Q) = Pr\left\{\frac{\sum_{i=1}^{n} a_i x_{(i)}}{\sum_{i=1}^{n} c_i x_{(i)}} \leq Q\right\} \qquad \ldots(20)$$

and a random variable

$$T = \sum_{i=1}^{n}(a_i - Qc_i)x_{(i)} \qquad \ldots(21)$$

with continuous c.d.f. $U(T)$. He shows that, if $\sum_{i=1}^{n} c_i x_{(i)} > 0$, then

$$G(Q) = Pr\left\{\sum_{i=1}^{n}(a_i - Qc_i)x_{(i)} \leq 0\right\} \qquad \ldots(22)$$

while, if $\sum_{i=1}^{n} c_i x_{(i)} \leq 0$, then

$$G(Q) = Pr\left\{\sum_{i=1}^{n}(a_i - Qc_i)x_{(i)} \geq 0\right\}. \qquad \ldots(23)$$

Cicchitelli (1976a, p. 222) also shows that, if $\sum_{i=1}^{n} c_i x_{(i)} > 0$ is satisfied, then

$$G(Q) = U(0). \qquad \ldots(24)$$

While if $\sum_{i=1}^{n} c_i x_{(i)} \leq 0$ is satisfied, then

$$G(Q) = 1 - U(0). \qquad \ldots (25)$$

Now, considering the results of Ali (1968, 1969), who determines $U(T)$ for the random variable T, on the basis of (24) it is possible to deduce the sampling distribution of the index B for the exponential population.

3. Sampling distribution of the index B

Given a random variable X with probability density function (p.d.f.)

$$f(x) = \lambda e^{-\lambda x}, \qquad x > 0, \qquad \ldots (26)$$

if

$$\frac{\sum_{i=h}^{n}(a_i - Qc_i)}{(n-h+1)} \neq \frac{\sum_{i=r}^{n}(a_i - Qc_i)}{(n-r+1)}; \quad h \neq r; \quad (h, r = 1, 2, ..., n) \qquad \ldots (27)$$

then following Ali (1969; p. 18) and Cicchitelli (1976a, p. 223), the c.d.f. of the random variable T defined by (21) results

$$U(T) = \sum_{i} \left\{ \frac{[(A_i - QC_i)/\lambda(n-i+1)]^{n-1} e^{-\lambda(n-i+1)T/(A_i - QC_i)}}{\prod_{\substack{j=1 \\ j \neq i}}^{n} \left[\frac{A_i - QC_i}{\lambda(n-i+1)} - \frac{A_j - QC_j}{\lambda(n-j+1)} \right]} \right\} \qquad \ldots (28)$$

where

$$A_i = \sum_{k=i}^{n} a_k; \quad C_i = \sum_{k=i}^{n} c_k; \quad (i = 1, 2, ..., n); \quad A_{n+1} = C_{n+1} = 0,$$

and the summation is extended to all the i, so that

$$A_i - QC_i < 0; \qquad (i = 1, 2, ..., n). \qquad \ldots (29)$$

At this point we have the necessary information for deducing the exact sampling distribution of the index B. In fact, on the basis of the hypotheses made, from (18) it follows

$$(n-1)\sum_{i=1}^{n} x_{(i)} > 0 \qquad \ldots (30)$$

so that it is possible to use (24).

From (28), putting $T = 0$ and for (24), we obtain (Cicchitelli, 1976a, p.223)

$$H(Q) = \sum_i \left\{ \frac{[(A_i - QC_i)/(n-i+1)]^{n-1}}{\prod_{\substack{j=1\\ j\neq i}}^{n} \left[\frac{A_i - QC_i}{(n-i+1)} - \frac{A_j - QC_j}{(n-j+1)} \right]} \right\}. \qquad \dots(31)$$

For the index B we have

$$A_i = \sum_{k=i}^{n} a_k = \sum_{k=i}^{n-1} a_k + a_n = (n-1)(n-i+1) - n\sum_{k=i}^{n-1} \xi_k; \ (i = 1, 2, ..., n-1);$$

$$A_n = n - 1 \qquad \dots(32)$$

$$C_i = \sum_{k=i}^{n} c_k = (n-1)(n-i+1); \qquad (i = 1, 2, .., n). \qquad \dots(33)$$

Thus (27) becomes

$$\frac{A_n - B_n C_h}{(n-h+1)} \neq \frac{A_r - B_n C_r}{(n-r+1)}; \ \ h \neq r; \quad (h, r = 1, 2, ..., n). \qquad \dots(34)$$

In order to satisfy (34) we must have

$$\frac{n}{(n-r+1)} \sum_{k=r}^{n-1} \xi_k \neq \frac{n}{(n-h+1)} \sum_{k=h}^{n-1} \xi_k \qquad \dots(35)$$

Therefore, on the basis of (31) the c.d.f. of B_n is given by

$$H(B_n) = \sum_i \left\{ \frac{[(\theta_i(1-B_n) - \tau_i)/(n-i+1)]^{n-1}}{\left[\frac{-\tau_i}{(n-i+1)}\right] \prod_{j=1_{j\neq i}}^{n-1} \left[\frac{\tau_j}{(n-j+1)} - \frac{\tau_i}{(n-i+1)} \right]} \right\} \qquad \dots(36)$$

where

$$\theta_i = (n-1)(n-i+1); \qquad (i = 1, 2, ..., n)$$

$$\tau_i = n \sum_{k=i}^{n-1} \xi_k; \quad \xi_k = \sum_{h=k}^{n-1} \frac{1}{h}; \quad (i = 1, 2, .., n-1); \ \ \tau_n = 0$$

Using (32) and (33), for (29) the summation of (36) is extended to all i so that

$$\theta_i(1 - B_n) - \tau_i < 0, \qquad \ldots(37)$$

from which we obtain

$$B_n > 1 - \frac{n}{(n-1)(n-i+1)} \sum_{k=i}^{n-1} \xi_k. \qquad \ldots(38)$$

In particular, for $i = n$, we must have from (37)

$$(n-1)(1 - B_n) < 0. \qquad \ldots(39)$$

Thus

$$B_n > 1 \qquad \ldots(40)$$

but, by definition $B_n \in [0, 1]$, so that the term with $i = n$ is not part of the summation.

From (36) the p.d.f. is deduced as

$$h(B_n) = \frac{\partial H(B_n)}{\partial B_n} = \sum_i \left\{ \frac{(n-1)^2 \left[(\theta_i(1-B_n) - \tau_i)/(n-i+1)\right]^{n-2}}{\left[\frac{\tau_i}{(n-i+1)}\right] \prod_{\substack{j=1 \\ j\neq i}}^{n-1} \left[\frac{\tau_j}{(n-j+1)} - \frac{\tau_i}{(n-i+1)}\right]} \right\}. \qquad \ldots(41)$$

Acknowledgements. The authors are indebted to an anonymous referee for some valuable remarks which enabled them to improve the paper.

REFERENCES

ALI, M. M. (1968). Distribution of linear combination of order statistics from the rectangular population (Abstract). *Ann. Math. Statist.*, **39**, 1086.

– – – – (1969). Distribution of linear combination of order statistics from the rectangular population. *Bull. Inst. Statist. Research and Training*, University of Dacca, **3**, 1-21.

BENEDETTI, C. (1982). Carlo Emilio Bonferroni (1892-1960). *Metron*, **40**, 3-36.

– – – – (1986). Sulla interpretazione benesseriale di noti indici di concentrazione e di altri. *Metron*, **44**, 421-429.

BONFERRONI, C. E. (1930). *Elementi di statistica generale.* Seeber, Firenze.

– – – – (1940). Un indice quadratico di concentrazione. *Atti del II convegno dell'Unione Matematici Italiani*, Bologna, 700-706.

CICCHITELLI, G. (1976a). The sampling distribution of the ratio of two linear combinations of order statistics. *Metron*, **34**, 219-228.

– – – – (1976b). The sampling distribution of linear combinations of order statistics. *Metron*, **34**, 269-302.

D'ADDARIO, R. (1936). *Curve di concentrazione, elasticitá, flessibilitá, densitá media e densitá marginale dei redditi*, Cressati, Bari.

GINI, C. (1914). Sulla misura della concentrazione e della variabilitá dei caratteri. *Atti del Regio Istituto Veneto di Scienze, Lettere ed Arti*, **73**, 1203-1248.

GIORGI, G. M. (1990). Bibliographic portrait of the Gini concentration ratio. *Metron*, **48**, 183-221.

– – –– (1992). *Il rapporto di concentrazione di Gini: genesi evoluzione ed una bibliografia commentata*. Ticci, Siena.

LORENZ, M. O. (1905). Methods of measuring the concentration of wealth. *Quart. Pubbl. American Statist. Assoc.*, **9**, 209-219.

MOOTHATHU, T. S. K. (1991). Lorenz curve and Gini index. *Calcutta Statist. Assoc. Bul.*, **40**, 307-324.

NYGARD, F. and SANDSTRÖM, A. (1981). *Measuring Income Inequality*, Almqvist and Wiksell International, Stockholm.

PIESCH, W. (1975). *Statistische Konzentrazionsmasse*. J.B.C. Mohr (Paul Siebeck), Tübingen.

DIPARTIMENTO DI STATISTICA
PROBAB. E. STATIST. APPL.
UNIVERSITÃ "LA SAPIENZA"
P.LE ALDO MORO N. 5
00185 ROMA
ITALY.