

The Distribution of the Concentration Ratio for Samples from a Uniform Population

Giovanni Girone, Antonella Nannavecchia

Faculty of Economics, University of Bari, Bari, Italy

Email: giovanni.girone@uniba.it, nannavecchia@lum.it

Received 24 October 2014; revised 20 November 2014; accepted 16 December 2014

Copyright © 2015 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In the present paper we derived, with direct method, the exact expressions for the sampling probability density function of the Gini concentration ratio for samples from a uniform population of size $n = 6, 7, 8, 9$ and 10 . Moreover, we found some regularities of such distributions valid for any sample size.

Keywords

Gini Concentration Ratio, Uniform Distribution, Order Statistics, Probability Density Function

1. Introduction

In 1914 Corrado Gini [1] introduced the concentration ratio R for the measure of inequality among values of a frequency distribution. The Gini index is widely used in fields as diverse as sociology, health science, engineering, and in particular, economics to measure the inequality of income distribution.

Various aspects of the Gini index have been taken into account. One of the most interesting topics regards the estimation of the concentration ratio (Hoeffding, 1948 [2]; Glasser, 1962 [3]; Cucconi, 1965 [4]; Dall'Aglio, 1965 [5]). More recently, Deltas (2003) [6] discussed the sources of bias of the Gini coefficient for small samples. This has implications for the comparison of inequality among subsamples, some of which may be small, and the use of the Gini index in measuring firm size inequality in markets with a small number of firms. Barret and Donald (2009) [7] considered statistical inference for consistent estimators of generalized Gini indices. The empirical indices are shown to be asymptotically normally distributed using functional limit theory. Moreover, asymptotic variance expressions are obtained using influence functions. Davidson (2009) [8] derived an approximation for the estimator of the Gini index by which it is expressed as a sum of IID random variables. This approximation allows developing a reliable standard error that is simple to compute. Fakoor, Ghalibaf and Azarnoosh (2011) [9] considered nonparametric estimators of the Gini index based on a sample from length-bi-

ased distributions. They showed that these estimators are strongly consistent for the Gini index. Also, they obtained an asymptotic normality for the corresponding Gini index.

Girone (1968) [10] focused on the study of the sampling distribution of the Gini index and in 1971 [11] derived the exact expression for samples drawn from an exponential population. In 1971 Girone [12] obtained, with direct method, the sampling distribution function of the Gini ratio for samples of size $n \leq 5$ drawn from a uniform population.

In the present note (Section 2), we calculate the joint probability density function (p.d.f.) of the random sample of size n and, then, the joint p.d.f. of the n order statistics. Hence, we transform one of the order statistics in their average and the remaining $n - 1$ order statistics are divided by the same average. We calculate the joint p.d.f. of the new n variables and integrating with respect to the average we obtain the joint p.d.f. of the other $n - 1$ variables. One of these variables is transformed in the concentration ratio. We calculate the joint p.d.f. of the concentration ratio and of the other $n - 2$ variables and at last we integrate this p.d.f. with respect to the $n - 2$ variables obtaining the marginal p.d.f. of the concentration ratio. The main difficulty of this procedure consists in the identification of the region of integration of the $n - 2$ variables, for two reasons: firstly the need to decompose this region into subregions which allow identifying directly the limits of integration and secondly the growing number of such subregions that makes the derivation heavy.

In Sections 3-7, using the software Mathematica, we derive the exact distributions of the concentration ratio for samples from a uniform distribution of size $n = 6, 7, 8, 9$ and 10. Moreover (Section 8), we find some regularities of such distributions valid for any sample size.

2. The Procedure to Derive the Distribution of the Concentration Ratio

Let random variables X_1, X_2, \dots, X_n from a uniform population have p.d.f.

$$f(x) = \begin{cases} 1, & 0 < x < 1, \\ 0, & \text{elsewhere.} \end{cases} \quad (1)$$

The joint p.d.f. of the variables is

$$h(x_1, x_2, \dots, x_n) = \begin{cases} 1, & 0 < x_i < 1, \text{ for } i = 1, 2, \dots, n, \\ 0, & \text{elsewhere.} \end{cases} \quad (2)$$

The joint p.d.f. of the order statistics $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ is

$$h(x_{(1)}, x_{(2)}, \dots, x_{(n)}) = \begin{cases} n!, & 0 < x_{(1)} < x_{(2)} < \dots < x_{(n)} < 1, \\ 0, & \text{elsewhere.} \end{cases} \quad (3)$$

By transforming the variables

$$S = X_{(1)} + X_{(2)} + \dots + X_{(n)},$$

$$D_{(i)} = \frac{X_{(i)}}{S}, \quad \text{for } i = 1, 2, \dots, n-1,$$

whose Jacobian is

$$J = S^{n-1},$$

we obtain the joint p.d.f. of the variables S and $D_{(1)}, D_{(2)}, \dots, D_{(n-1)}$ that can be written as

$$g(s, d_{(1)}, d_{(2)}, \dots, d_{(n-1)}) = (n-1)! s^{n-1}, \quad (4)$$

$$\text{for } 0 < sd_{(1)} < sd_{(2)} < \dots < sd_{(n-1)} < s(1 - d_{(1)} - d_{(2)} - \dots - d_{(n-1)}) < 1.$$

We integrate expression [4] with respect to the variable S and obtain the joint p.d.f. of the variables $D_{(1)}, D_{(2)}, \dots, D_{(n-1)}$ that can be written as

$$f(d_{(1)}, d_{(2)}, \dots, d_{(n-1)}) = \frac{(n-1)!}{(1-d_{(1)}-d_{(2)}-\dots-d_{(n-1)})^n}, \quad (5)$$

for $0 < d_{(1)} < d_{(2)} < \dots < d_{(n-1)} < 1-d_{(1)}-d_{(2)}-\dots-d_{(n-1)}$.

By transforming the variable $D_{(n-1)}$ in the variable R i.e. the concentration ratio

$$R = 1 - \frac{2}{n-1} \sum_{i=1}^{n-1} (n-i) D_{(i)},$$

from which we get

$$D_{(n-1)} = \frac{(n-1)(1-R)}{2} - \sum_{i=1}^{n-2} (n-i) D_{(i)},$$

the Jacobian of the transformation is

$$J = \frac{n-1}{2}$$

and the joint p.d.f. of the variable R and $D_{(1)}, D_{(2)}, \dots, D_{(n-2)}$ is

$$h(d_{(1)}, d_{(2)}, \dots, d_{(n-2)}, R) = \frac{2^n (n-1)!}{\left(2 - (n-1)(1-R) + 2 \sum_{i=1}^{n-2} (n-i-1) d_{(i)}\right)^n}, \quad (6)$$

for

$$0 < d_{(1)} < d_{(2)} < \dots < d_{(n-2)} < \frac{(n-1)(1-R)}{2} - \sum_{i=1}^{n-2} (n-i) d_{(i)} < 1 - \frac{(n-1)(1-R)}{2} + \sum_{i=1}^{n-2} (n-i-1) d_{(i)}. \quad (7)$$

By integrating expression [6] with respect to the variables $D_{(1)}, D_{(2)}, \dots, D_{(n-2)}$ over the regions determined by inequalities [7], we get the marginal p.d.f. of the concentration ratio R .

3. The Distribution of the Concentration Ratio for $n = 6$

The procedure indicated in Section 2 is used to obtain the following p.d.f. (Figure 1) of the concentration ratio R for random samples of size $n = 6$:

$$g(R) = \frac{27}{50\left(\frac{1}{5}+R\right)^2} - \frac{1728}{25\left(\frac{2}{5}+R\right)^2} + \frac{19683}{25\left(\frac{3}{5}+R\right)^2} - \frac{55296}{25\left(\frac{4}{5}+R\right)^2} + \frac{3375}{2(1+R)^2}, \text{ for } 0 < R < \frac{1}{5},$$

$$g(R) = -\frac{125}{144R^2} + \frac{7853}{75\left(\frac{1}{5}+R\right)^2} - \frac{449523}{400\left(\frac{2}{5}+R\right)^2} + \frac{677147}{225\left(\frac{3}{5}+R\right)^2} - \frac{55296}{25\left(\frac{4}{5}+R\right)^2}, \text{ for } \frac{1}{5} < R < \frac{2}{5},$$

$$g(R) = \frac{32}{75\left(-\frac{1}{5}+R\right)^2} - \frac{11131}{240R^2} + \frac{33773}{75\left(\frac{1}{5}+R\right)^2} - \frac{449523}{400\left(\frac{2}{5}+R\right)^2} + \frac{19683}{25\left(\frac{3}{5}+R\right)^2}, \text{ for } \frac{2}{5} < R < \frac{3}{5},$$

$$g(R) = -\frac{27}{400\left(-\frac{2}{5}+R\right)^2} + \frac{437}{75\left(-\frac{1}{5}+R\right)^2} - \frac{11131}{240R^2} + \frac{7853}{75\left(\frac{1}{5}+R\right)^2} - \frac{1728}{25\left(\frac{2}{5}+R\right)^2}, \text{ for } \frac{3}{5} < R < \frac{4}{5},$$

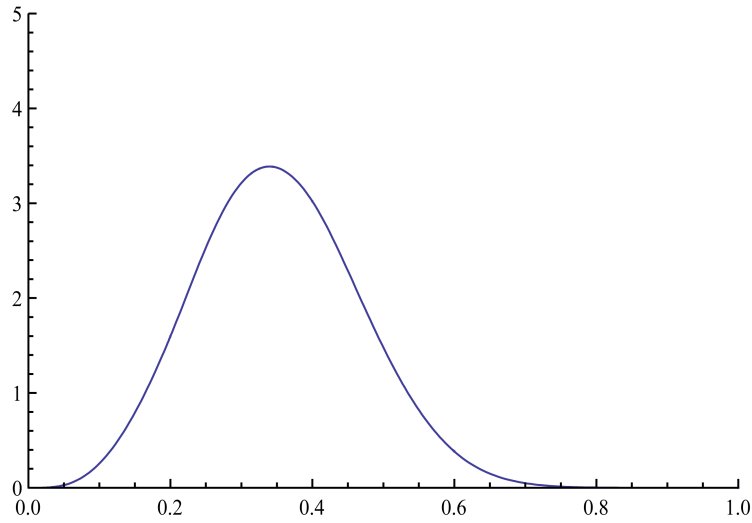


Figure 1. Probability density function of the concentration ratio R for random samples of size $n = 6$ from a uniform population.

$$g(R) = \frac{1}{450\left(-\frac{3}{5} + R\right)^2} - \frac{27}{400\left(-\frac{2}{5} + R\right)^2} + \frac{32}{75\left(-\frac{1}{5} + R\right)^2} - \frac{125}{144R^2} + \frac{27}{50\left(\frac{1}{5} + R\right)^2}, \text{ for } \frac{4}{5} < R < 1.$$

Characteristic values of the distribution are:

$$\text{mean } E(R) = -3 + \frac{11696 \log 2}{15} + \frac{243 \log 3}{2} - \frac{1250 \log 5}{3} = 0.35222,$$

$$\text{second moment } E(R^2) = 16 - \frac{69056 \log 2}{25} - \frac{24219 \log 3}{50} + \frac{18125 \log 5}{12} = 0.13716,$$

$$\text{third moment } E(R^3) = -\frac{1368}{25} + \frac{678688 \log 2}{125} + \frac{156006 \log 3}{125} - \frac{12625 \log 5}{4} = 0.05785,$$

$$\text{fourth moment } E(R^4) = \frac{89237}{625} - \frac{25434496 \log 2}{3125} - \frac{7691436 \log 3}{3125} + \frac{15290 \log 5}{3} = 0.02606,$$

$$\text{standard deviation } \sigma(R) = 0.11444,$$

$$\text{index of skewness } \gamma_1(R) = 0.20793,$$

$$\text{index of kurtosis } \gamma_2(R) = -0.16767.$$

The distribution of the concentration ratio R for samples of size $n = 6$ from a uniform population shows a slight positive skewness and platykurtosis.

4. The Distribution of the Concentration Ratio for $n = 7$

The procedure indicated in Section 2 is used to obtain the following p.d.f. (Figure 2) of the concentration ratio R for random samples of size $n = 7$:

$$g(R) = -\frac{117649}{518400\left(\frac{1}{6} + R\right)^2} + \frac{117649}{1620\left(\frac{2}{6} + R\right)^2} - \frac{1058841}{640\left(\frac{3}{6} + R\right)^2} + \frac{3764768}{405\left(\frac{4}{6} + R\right)^2} - \frac{367653125}{20736\left(\frac{5}{6} + R\right)^2} + \frac{1058841}{100(1+R)^2}, \text{ for } 0 < R < \frac{1}{6},$$

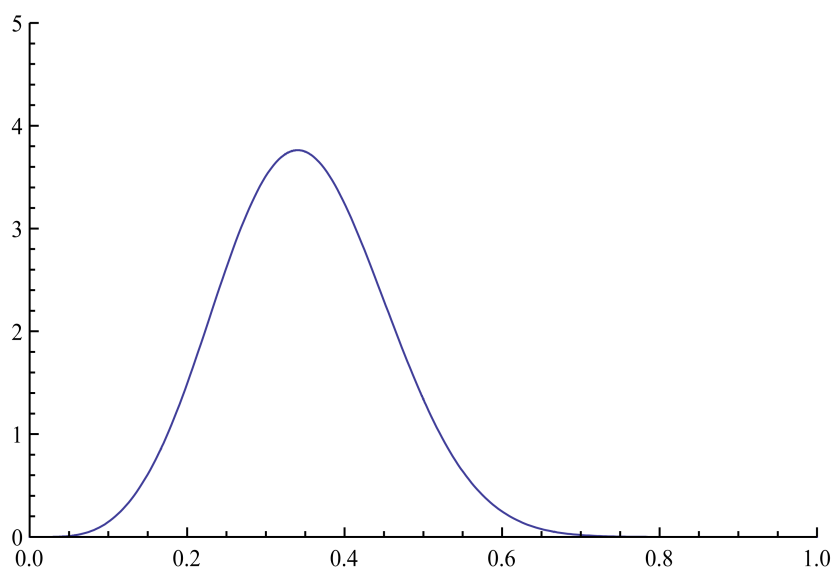


Figure 2. Probability density function of the concentration ratio R for random samples of size $n = 7$ from a uniform population.

$$g(R) = \frac{9}{20R^2} - \frac{14356253}{103680\left(\frac{1}{6} + R\right)^2} + \frac{2450309}{810\left(\frac{2}{6} + R\right)^2} - \frac{2099205}{128\left(\frac{3}{6} + R\right)^2} \\ + \frac{49230947}{1620\left(\frac{4}{6} + R\right)^2} - \frac{367653125}{20736\left(\frac{5}{6} + R\right)^2}, \quad \text{for } \frac{1}{6} < R < \frac{2}{6},$$

$$g(R) = -\frac{3125}{10368\left(-\frac{1}{6} + R\right)^2} + \frac{7853}{90R^2} - \frac{46303907}{25920\left(\frac{1}{6} + R\right)^2} + \frac{7450309}{810\left(\frac{2}{6} + R\right)^2} \\ - \frac{2099205}{128\left(\frac{3}{6} + R\right)^2} + \frac{3764768}{405\left(\frac{4}{6} + R\right)^2}, \quad \text{for } \frac{2}{6} < R < \frac{3}{6},$$

$$g(R) = \frac{32}{405\left(-\frac{2}{6} + R\right)^2} - \frac{1064201}{51840\left(-\frac{1}{6} + R\right)^2} + \frac{33773}{90R^2} - \frac{46303907}{25920\left(\frac{1}{6} + R\right)^2} \\ + \frac{2450309}{810\left(\frac{2}{6} + R\right)^2} - \frac{1058841}{640\left(\frac{3}{6} + R\right)^2}, \quad \text{for } \frac{3}{6} < R < \frac{4}{6},$$

$$g(R) = -\frac{9}{1280\left(-\frac{3}{6} + R\right)^2} + \frac{463}{324\left(-\frac{2}{6} + R\right)^2} - \frac{1064201}{51840\left(-\frac{1}{6} + R\right)^2} + \frac{7853}{90R^2} \\ - \frac{14356253}{103680\left(\frac{1}{6} + R\right)^2} + \frac{117649}{1620\left(\frac{2}{6} + R\right)^2}, \quad \text{for } \frac{4}{6} < R < \frac{5}{6},$$

$$g(R) = \frac{1}{8100\left(-\frac{4}{6} + R\right)^2} - \frac{9}{1280\left(-\frac{3}{6} + R\right)^2} + \frac{32}{405\left(-\frac{2}{6} + R\right)^2} - \frac{3125}{10368\left(-\frac{1}{6} + R\right)^2} + \frac{9}{20R^2} - \frac{117649}{518400\left(\frac{1}{6} + R\right)^2}, \quad \text{for } \frac{5}{6} < R < 1.$$

Characteristic values of the distribution are:

$$\text{mean } E(R) = -\frac{7}{2} + \frac{35072 \log 2}{15} - \frac{21797 \log 3}{160} + \frac{359375 \log 5}{288} + \frac{823543 \log 7}{1440} = 0.34951,$$

$$\text{second moment } E(R^2) = \frac{763}{36} + \frac{3806128 \log 2}{405} + \frac{451251 \log 3}{80} - \frac{359375 \log 5}{72} - \frac{15647317 \log 7}{6480} = 0.13291,$$

third moment

$$E(R^3) = -\frac{18179}{216} - \frac{2771144 \log 2}{135} - \frac{4266351 \log 3}{320} + \frac{18546875 \log 5}{1728} + \frac{5764801 \log 7}{960} = 0.05417,$$

fourth moment

$$E(R^4) = \frac{165193}{648} + \frac{40601588 \log 2}{1215} + \frac{7638867 \log 3}{320} + \frac{263234375 \log 5}{15552} - \frac{895191241 \log 7}{77760} = 0.02342,$$

$$\text{standard deviation } \sigma(R) = 0.10367,$$

$$\text{index of skewness } \gamma_1(R) = 0.18545,$$

$$\text{index of kurtosis } \gamma_2(R) = -0.14535.$$

The distribution of the concentration ratio R for samples of size $n = 7$ from a uniform population shows slight positive skewness and platykurtosis, both lower than those obtained for samples of size $n = 6$.

5. The Distribution of the Concentration Ratio for $n = 8$

The procedure indicated in Section 2 is used to obtain the following p.d.f. (Figure 3) of the concentration ratio R for random samples of size $n = 8$:

$$g(R) = \frac{8192}{99225\left(\frac{1}{7} + R\right)^2} - \frac{2097152}{33075\left(\frac{2}{7} + R\right)^2} + \frac{663552}{245\left(\frac{3}{7} + R\right)^2} - \frac{536870912}{19845\left(\frac{4}{7} + R\right)^2} + \frac{128000000}{1323\left(\frac{5}{7} + R\right)^2} - \frac{169869312}{1225\left(\frac{6}{7} + R\right)^2} + \frac{137682944}{2025(1+R)^2}, \quad \text{for } 0 < R < \frac{1}{7},$$

$$g(R) = -\frac{16807}{86400R^2} + \frac{9613463}{66150\left(\frac{1}{7} + R\right)^2} - \frac{2548367503}{423360\left(\frac{2}{7} + R\right)^2} + \frac{386863168}{6615\left(\frac{3}{7} + R\right)^2} - \frac{173031518287}{846720\left(\frac{4}{7} + R\right)^2} + \frac{19007619787}{66150\left(\frac{5}{7} + R\right)^2} - \frac{169869312}{1225\left(\frac{6}{7} + R\right)^2}, \quad \text{for } \frac{1}{7} < R < \frac{2}{7},$$

$$g(R) = \frac{81}{490\left(-\frac{1}{7} + R\right)^2} - \frac{14356253}{120960R^2} + \frac{31094051}{6615\left(\frac{1}{7} + R\right)^2} - \frac{18601017487}{423360\left(\frac{2}{7} + R\right)^2} + \frac{1969741961}{13230\left(\frac{3}{7} + R\right)^2} - \frac{173031518287}{846720\left(\frac{4}{7} + R\right)^2} + \frac{128000000}{1323\left(\frac{5}{7} + R\right)^2}, \quad \text{for } \frac{2}{7} < R < \frac{3}{7},$$

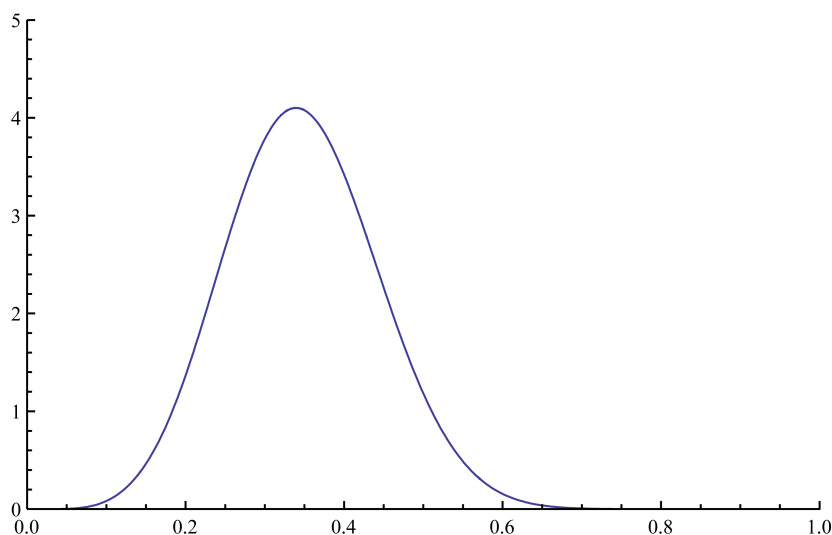


Figure 3. Probability density function of the concentration ratio R for random samples of size $n = 8$ from a uniform population.

$$g(R) = -\frac{15625}{254016\left(-\frac{2}{7}+R\right)^2} + \frac{274531}{6615\left(-\frac{1}{7}+R\right)^2} - \frac{46303907}{30240R^2} + \frac{268282153}{19845\left(\frac{1}{7}+R\right)^2} \\ - \frac{18601017487}{423360\left(\frac{2}{7}+R\right)^2} + \frac{386863168}{6615\left(\frac{3}{7}+R\right)^2} - \frac{536870912}{19845\left(\frac{4}{7}+R\right)^2}, \quad \text{for } \frac{3}{7} < R < \frac{4}{7},$$

$$g(R) = \frac{64}{6615\left(-\frac{3}{7}+R\right)^2} - \frac{2472719}{423360\left(-\frac{2}{7}+R\right)^2} + \frac{1254307}{6615\left(-\frac{1}{7}+R\right)^2} - \frac{46303907}{30240R^2} \\ + \frac{31094051}{6615\left(\frac{1}{7}+R\right)^2} - \frac{2548367503}{423360\left(\frac{2}{7}+R\right)^2} + \frac{663552}{245\left(\frac{3}{7}+R\right)^2}, \quad \text{for } \frac{4}{7} < R < \frac{5}{7},$$

$$g(R) = -\frac{81}{156800\left(-\frac{4}{7}+R\right)^2} + \frac{15949}{66150\left(-\frac{3}{7}+R\right)^2} - \frac{2472719}{423360\left(-\frac{2}{7}+R\right)^2} + \frac{274531}{6615\left(-\frac{1}{7}+R\right)^2} \\ - \frac{14356253}{120960R^2} + \frac{9613463}{66150\left(\frac{1}{7}+R\right)^2} - \frac{2097152}{33075\left(\frac{2}{7}+R\right)^2}, \quad \text{for } \frac{5}{7} < R < \frac{6}{7},$$

$$g(R) = \frac{1}{198450\left(-\frac{5}{7}+R\right)^2} - \frac{81}{156800\left(-\frac{4}{7}+R\right)^2} + \frac{64}{6615\left(-\frac{3}{7}+R\right)^2} - \frac{15625}{254016\left(-\frac{2}{7}+R\right)^2} \\ + \frac{81}{490\left(-\frac{1}{7}+R\right)^2} - \frac{16807}{86400R^2} + \frac{8192}{99225\left(\frac{1}{7}+R\right)^2}, \quad \text{for } \frac{6}{7} < R < 1.$$

Characteristic values of the distribution are:

$$\text{mean } E(R) = -4 + \frac{3475456 \log 2}{315} + \frac{2775303 \log 3}{560} - \frac{2421875 \log 5}{1008} - \frac{3411821 \log 7}{720} = 0.34747,$$

$$\text{second moment } E(R^2) = \frac{190}{7} - \frac{3184576 \log 2}{63} - \frac{8776431 \log 3}{392} + \frac{4140625 \log 5}{392} + \frac{7882483 \log 7}{360} = 0.12985,$$

third moment

$$E(R^3) = -\frac{6016}{49} + \frac{658405376 \log 2}{5145} + \frac{1492752159 \log 3}{27440} - \frac{129921875 \log 5}{5488} - \frac{13596863 \log 7}{240} = 0.05160,$$

fourth moment

$$E(R^4) = \frac{145475}{343} - \frac{3729880384 \log 2}{15435} - \frac{2307400911 \log 3}{24010} + \frac{533984375 \log 5}{14406} + \frac{9851303 \log 7}{90} = 0.02162,$$

$$\text{standard deviation } \sigma(R) = 0.09544,$$

$$\text{index of skewness } \gamma_1(R) = 0.16867,$$

$$\text{index of kurtosis } \gamma_2(R) = -0.12824.$$

The distribution of the concentration ratio R for samples of size $n = 8$ from a uniform population shows slight positive skewness and platykurtosis, both lower than those obtained for samples of size $n = 6$ and 7 .

6. The Distribution of the Concentration Ratio for $n = 9$

The procedure indicated in Section 2 is used to obtain the following p.d.f. (Figure 4) of the concentration ratio R for random samples of size $n = 9$:

$$g(R) = -\frac{531441}{20070400 \left(\frac{1}{8} + R\right)^2} + \frac{531441}{11200 \left(\frac{2}{8} + R\right)^2} - \frac{10460353203}{2867200 \left(\frac{3}{8} + R\right)^2} + \frac{2125764}{35 \left(\frac{4}{8} + R\right)^2} - \frac{41518828125}{114688 \left(\frac{5}{8} + R\right)^2} + \frac{10460353203}{11200 \left(\frac{6}{8} + R\right)^2} - \frac{437664515463}{409600 \left(\frac{7}{8} + R\right)^2} + \frac{544195584}{1225(1+R)^2}, \quad \text{for } 0 < R < \frac{1}{8},$$

$$g(R) = \frac{1024}{14175R^2} - \frac{9819117949}{77414400 \left(\frac{1}{8} + R\right)^2} + \frac{15242823}{1600 \left(\frac{2}{8} + R\right)^2} - \frac{1441266427931}{9289728 \left(\frac{3}{8} + R\right)^2} + \frac{857395628}{945 \left(\frac{4}{8} + R\right)^2} - \frac{6604248318741}{2867200 \left(\frac{5}{8} + R\right)^2} + \frac{2358496602787}{907200 \left(\frac{6}{8} + R\right)^2} - \frac{437664515463}{409600 \left(\frac{7}{8} + R\right)^2}, \quad \text{for } \frac{1}{8} < R < \frac{2}{8},$$

$$g(R) = -\frac{823543}{11059200 \left(-\frac{1}{8} + R\right)^2} + \frac{9613463}{75600R^2} - \frac{715845826021}{77414400 \left(\frac{1}{8} + R\right)^2} + \frac{1262942489}{8640 \left(\frac{2}{8} + R\right)^2} - \frac{2582172507227}{3096576 \left(\frac{3}{8} + R\right)^2} + \frac{156844988749}{75600 \left(\frac{4}{8} + R\right)^2} - \frac{6604248318741}{2867200 \left(\frac{5}{8} + R\right)^2} + \frac{10460353203}{11200 \left(\frac{6}{8} + R\right)^2}, \quad \text{for } \frac{2}{8} < R < \frac{3}{8},$$

$$g(R) = \frac{81}{2240 \left(-\frac{1}{8} + R\right)^2} - \frac{183689449}{3096576 \left(-\frac{2}{8} + R\right)^2} + \frac{31094051}{7560R^2} - \frac{193012968877}{3096576 \left(\frac{1}{8} + R\right)^2} + \frac{2971536239}{8640 \left(\frac{2}{8} + R\right)^2} - \frac{2582172507227}{3096576 \left(\frac{3}{8} + R\right)^2} + \frac{857395628}{945 \left(\frac{4}{8} + R\right)^2} - \frac{41518828125}{114688 \left(\frac{5}{8} + R\right)^2}, \quad \text{for } \frac{3}{8} < R < \frac{4}{8},$$

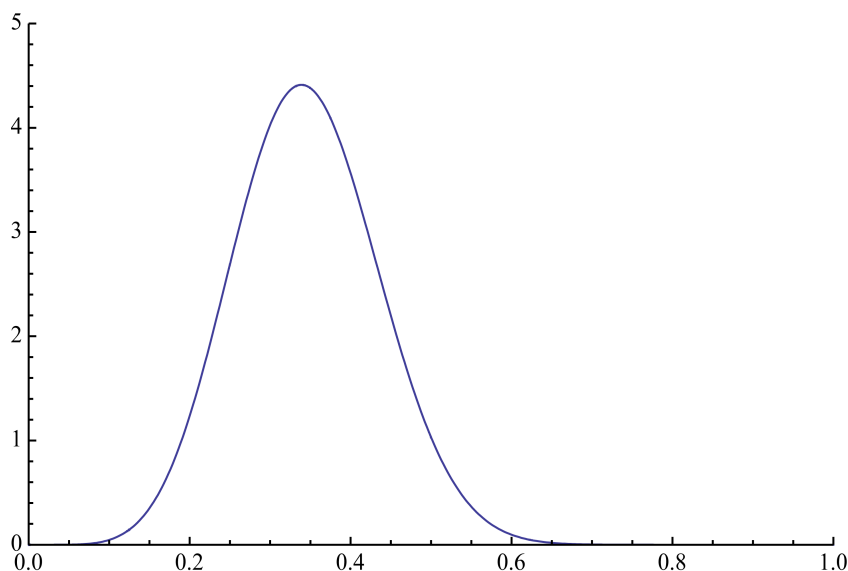


Figure 4. Probability density function of the concentration ratio R for random samples of size $n = 9$ from a uniform population.

$$g(R) = \frac{78125}{9289728 \left(-\frac{3}{8} + R\right)^2} + \frac{783437}{60480 \left(-\frac{2}{8} + R\right)^2} - \frac{2575720699}{3096576 \left(-\frac{1}{8} + R\right)^2} + \frac{268282153}{22680R^2} \\ - \frac{193012968877}{3096576 \left(\frac{1}{8} + R\right)^2} + \frac{1262942489}{8640 \left(\frac{2}{8} + R\right)^2} - \frac{1441266427931}{9289728 \left(\frac{3}{8} + R\right)^2} + \frac{2125764}{35 \left(\frac{4}{8} + R\right)^2}, \quad \text{for } \frac{4}{8} < R < \frac{5}{8},$$

$$g(R) = \frac{4}{4725 \left(-\frac{4}{8} + R\right)^2} - \frac{90129527}{77414400 \left(-\frac{3}{8} + R\right)^2} + \frac{19593601}{302400 \left(-\frac{2}{8} + R\right)^2} - \frac{2575720699}{3096576 \left(-\frac{1}{8} + R\right)^2} \\ + \frac{31094051}{7560R^2} - \frac{715845826021}{77414400 \left(\frac{1}{8} + R\right)^2} + \frac{15242823}{1600 \left(\frac{2}{8} + R\right)^2} - \frac{10460353203}{2867200 \left(\frac{3}{8} + R\right)^2}, \quad \text{for } \frac{5}{8} < R < \frac{6}{8},$$

$$g(R) = \frac{81}{2867200 \left(-\frac{5}{8} + R\right)^2} + \frac{2251}{75600 \left(-\frac{4}{8} + R\right)^2} - \frac{90129527}{77414400 \left(-\frac{3}{8} + R\right)^2} + \frac{783437}{60480 \left(-\frac{2}{8} + R\right)^2} \\ - \frac{183689449}{3096576 \left(-\frac{1}{8} + R\right)^2} + \frac{9613463}{75600R^2} - \frac{9819117949}{77414400 \left(\frac{1}{8} + R\right)^2} + \frac{531441}{11200 \left(\frac{2}{8} + R\right)^2}, \quad \text{for } \frac{6}{8} < R < \frac{7}{8},$$

$$g(R) = \frac{1}{6350400 \left(-\frac{6}{8} + R\right)^2} - \frac{81}{2867200 \left(-\frac{5}{8} + R\right)^2} + \frac{4}{4725 \left(-\frac{4}{8} + R\right)^2} - \frac{78125}{9289728 \left(-\frac{3}{8} + R\right)^2} \\ + \frac{81}{2240 \left(-\frac{2}{8} + R\right)^2} - \frac{823543}{11059200 \left(-\frac{1}{8} + R\right)^2} + \frac{1024}{14175R^2} - \frac{531441}{20070400 \left(\frac{1}{8} + R\right)^2}, \quad \text{for } \frac{7}{8} < R < 1.$$

Characteristic values of the distribution are:

$$\text{mean } E(R) = -\frac{9}{2} - \frac{18425272 \log 2}{315} - \frac{1948617 \log 3}{1120} + \frac{1953125 \log 5}{576} + \frac{109531219 \log 7}{5760} = 0.34589,$$

second moment

$$E(R^2) = \frac{1083}{32} + \frac{31176874 \log 2}{105} + \frac{209048769 \log 3}{35840} - \frac{150390625 \log 5}{9216} - \frac{1101076991 \log 7}{11520} = 0.12754,$$

third moment

$$E(R^3) = -\frac{43983}{256} - \frac{69161579 \log 2}{84} + \frac{394860663 \log 3}{57344} + \frac{947265625 \log 5}{24576} + \frac{31654522291 \log 7}{122880} = 0.04969,$$

fourth moment

$$E(R^4) = \frac{2732815}{4096} + \frac{58613649 \log 2}{35} - \frac{36660941253 \log 3}{458752} - \frac{36018359375 \log 5}{589824} - \frac{739122430613 \log 7}{1474560} = 0.02032,$$

$$\text{standard deviation } \sigma(R) = 0.08889,$$

$$\text{index of skewness } \gamma_1(R) = 0.15559,$$

$$\text{index of kurtosis } \gamma_2(R) = -0.11467.$$

The distribution of the concentration ratio R for samples of size $n = 9$ from a uniform population shows slight positive skewness and platykurtosis, both lower than those obtained for samples of size $n = 6, 7$ and 8 .

7. The Distribution of the Concentration Ratio for $n = 10$

The procedure indicated in Section 2 is used to obtain the following p.d.f. (Figure 5) of the concentration ratio R for random samples of size $n = 10$:

$$g(R) = \frac{78125}{10287648 \left(\frac{1}{9} + R\right)^2} - \frac{10000000}{321489 \left(\frac{2}{9} + R\right)^2} + \frac{234375}{56 \left(\frac{3}{9} + R\right)^2} - \frac{5120000000}{45927 \left(\frac{4}{9} + R\right)^2} + \frac{762939453125}{734832 \left(\frac{5}{9} + R\right)^2} \\ - \frac{30000000}{7 \left(\frac{6}{9} + R\right)^2} + \frac{450375078125}{52488 \left(\frac{7}{9} + R\right)^2} - \frac{2621440000000}{321489 \left(\frac{8}{9} + R\right)^2} + \frac{4613203125}{1568(1+R)^2}, \quad \text{for } 0 < R < \frac{1}{9},$$

$$g(R) = -\frac{59049}{2508800R^2} + \frac{6102360983}{64297800 \left(\frac{1}{9} + R\right)^2} - \frac{29486161727807}{2351462400 \left(\frac{2}{9} + R\right)^2} + \frac{91871139}{280 \left(\frac{3}{9} + R\right)^2} - \frac{282890791328125}{94058496 \left(\frac{4}{9} + R\right)^2} \\ + \frac{112482309211387}{9185400 \left(\frac{5}{9} + R\right)^2} - \frac{78160683773061}{793800 \left(\frac{6}{9} + R\right)^2} + \frac{1465749080718869}{64297800 \left(\frac{7}{9} + R\right)^2} - \frac{2621440000000}{321489 \left(\frac{8}{9} + R\right)^2}, \quad \text{for } \frac{1}{9} < R < \frac{2}{9},$$

$$g(R) = \frac{32768}{1148175 \left(-\frac{1}{9} + R\right)^2} - \frac{9819117949}{87091200R^2} + \frac{133551832049}{9185400 \left(\frac{1}{9} + R\right)^2} \\ - \frac{34956443674391}{1161216 \left(\frac{2}{9} + R\right)^2} + \frac{227124686777}{68040 \left(\frac{3}{9} + R\right)^2} - \frac{4488452581986787}{335923200 \left(\frac{4}{9} + R\right)^2} \\ + \frac{239423780652283}{9185400 \left(\frac{5}{9} + R\right)^2} - \frac{8684520419229}{358400 \left(\frac{6}{9} + R\right)^2} + \frac{450375078125}{52488 \left(\frac{7}{9} + R\right)^2}, \quad \text{for } \frac{2}{9} < R < \frac{3}{9},$$

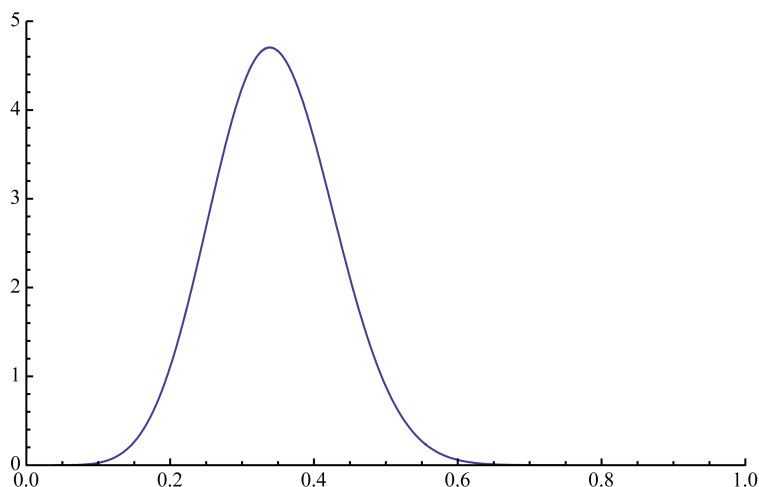


Figure 5. Probability density function of the concentration ratio R for random samples of size $n = 10$ from a uniform population.

$$g(R) = -\frac{5764801}{335923200\left(-\frac{2}{9}+R\right)^2} + \frac{605566249}{9185400\left(-\frac{1}{9}+R\right)^2} - \frac{715845826021}{87091200R^2} \\ + \frac{373815952381}{1837080\left(\frac{1}{9}+R\right)^2} - \frac{167366716643141}{94058496\left(\frac{2}{9}+R\right)^2} + \frac{2371170173011}{340200\left(\frac{3}{9}+R\right)^2} \\ - \frac{4488452581986787}{335923200\left(\frac{4}{9}+R\right)^2} + \frac{112482309211387}{9185400\left(\frac{5}{9}+R\right)^2} - \frac{30000000}{7\left(\frac{6}{9}+R\right)^2}, \quad \text{for } \frac{3}{9} < R < \frac{4}{9},$$

$$g(R) = \frac{3}{560\left(-\frac{3}{9}+R\right)^2} - \frac{1859135071}{94058496\left(-\frac{2}{9}+R\right)^2} + \frac{4305254531}{1837080\left(-\frac{1}{9}+R\right)^2} \\ - \frac{193012968877}{3483648R^2} + \frac{429101140253}{918540\left(\frac{1}{9}+R\right)^2} - \frac{167366716643141}{94058496\left(\frac{2}{9}+R\right)^2} \\ + \frac{227124686777}{68040\left(\frac{3}{9}+R\right)^2} - \frac{282890791328125}{94058496\left(\frac{4}{9}+R\right)^2} + \frac{762939453125}{734832\left(\frac{5}{9}+R\right)^2}, \quad \text{for } \frac{4}{9} < R < \frac{5}{9},$$

$$g(R) = -\frac{78125}{94058496\left(-\frac{4}{9}+R\right)^2} + \frac{195677}{68040\left(-\frac{3}{9}+R\right)^2} - \frac{29538353821}{94058496\left(-\frac{2}{9}+R\right)^2} \\ + \frac{12705254531}{1837080\left(-\frac{1}{9}+R\right)^2} - \frac{193012968877}{3483648R^2} + \frac{373815952381}{1837080\left(\frac{1}{9}+R\right)^2} \\ - \frac{34956443674391}{94058496\left(\frac{2}{9}+R\right)^2} + \frac{91871139}{280\left(\frac{3}{9}+R\right)^2} - \frac{512000000}{45927\left(\frac{4}{9}+R\right)^2}, \quad \text{for } \frac{5}{9} < R < \frac{6}{9},$$

$$g(R) = \frac{64}{1148175\left(-\frac{5}{9} + R\right)^2} - \frac{404606309}{2351462400\left(-\frac{4}{9} + R\right)^2} + \frac{779623}{48600\left(-\frac{3}{9} + R\right)^2}$$

$$- \frac{29538353821}{94058496\left(-\frac{2}{9} + R\right)^2} + \frac{4305254531}{1837080\left(-\frac{1}{9} + R\right)^2} - \frac{715845826021}{87091200R^2}$$

$$+ \frac{133551832049}{9185400\left(\frac{1}{9} + R\right)^2} - \frac{29486161727807}{2351462400\left(\frac{2}{9} + R\right)^2} + \frac{234375}{56\left(\frac{3}{9} + R\right)^2}, \quad \text{for } \frac{6}{9} < R < \frac{7}{9},$$

$$g(R) = -\frac{3}{2508800\left(-\frac{6}{9} + R\right)^2} + \frac{180731}{64297800\left(-\frac{5}{9} + R\right)^2} - \frac{404606309}{2351462400\left(-\frac{4}{9} + R\right)^2}$$

$$+ \frac{195677}{68040\left(-\frac{3}{9} + R\right)^2} - \frac{1859135071}{94058496\left(-\frac{2}{9} + R\right)^2} + \frac{605566249}{9185400\left(-\frac{1}{9} + R\right)^2}$$

$$- \frac{9819117949}{87091200R^2} + \frac{6102360983}{64297800\left(\frac{1}{9} + R\right)^2} - \frac{10000000}{321489\left(\frac{2}{9} + R\right)^2}, \quad \text{for } \frac{7}{9} < R < \frac{8}{9},$$

$$g(R) = \frac{1}{257191200\left(-\frac{7}{9} + R\right)^2} - \frac{3}{2508800\left(-\frac{6}{9} + R\right)^2} + \frac{64}{1148175\left(-\frac{5}{9} + R\right)^2}$$

$$- \frac{78125}{94058496\left(-\frac{4}{9} + R\right)^2} + \frac{3}{560\left(-\frac{3}{9} + R\right)^2} - \frac{5764801}{335923200\left(-\frac{2}{9} + R\right)^2}$$

$$+ \frac{32768}{1148175\left(-\frac{1}{9} + R\right)^2} - \frac{59049}{2508800R^2} + \frac{78125}{10287648\left(\frac{1}{9} + R\right)^2}, \quad \text{for } \frac{8}{9} < R < 1.$$

Characteristic values of the distribution are:

$$\text{mean } E(R) = -5 + \frac{686419424 \log 2}{2835} - \frac{88683579 \log 3}{1120} + \frac{60546875 \log 5}{6048} - \frac{40353607 \log 7}{810} = 0.34462,$$

second moment

$$E(R^2) = -\frac{4977475650505}{1666598976} + \frac{5506009449897293 \log 2}{57868020} - \frac{3870789417061723 \log 3}{115736040}$$

$$+ \frac{23520914453125 \log 5}{5143824} - \frac{155338805614063 \log 7}{8266860} = 0.12574,$$

third moment

$$E(R^3) = -\frac{56500}{243} + \frac{302665066912 \log 2}{76545} - \frac{1609375311 \log 3}{1120}$$

$$+ \frac{99482421875 \log 5}{489888} - \frac{133852914419 \log 7}{174960} = 0.04823,$$

fourth moment

$$E(R^4) = \frac{6581554}{6561} - \frac{10398297681152 \log 2}{1240029} + \frac{1892901777 \log 3}{560} - \frac{10382748828125 \log 5}{19840464} + \frac{1189583980753 \log 7}{787320} = 0.01933,$$

standard deviation $\sigma(R) = 0.08352$,

index of skewness $\gamma_1(R) = 0.14505$,

index of kurtosis $\gamma_2(R) = -0.10366$.

The distribution of the concentration ratio R for samples of size $n = 10$ from a uniform population shows slight positive skewness and platykurtosis, both lower than those obtained for samples of size $n = 6, 7, 8$ and 9 .

8. Some Regularities of the Distributions

The analysis of the p.d.f. for $n = 2, 3, \dots, 10$ shows some regularities:

- The p.d.f. of the concentration ratio R , for $0 < R < 1/n$ and for samples of size n , can be expressed by

$$g(R) = \sum_{i=1}^{n-1} \frac{(-1)^{1+i} [n(n-1)]^{n-1} \binom{n-2}{i-1}}{[(n-1)!]^2 \left(\frac{n-i}{n-1} + R\right)^2};$$

- Furthermore, the p.d.f. of the concentration ratio R , for $(n-1)/n < R < 1$ and for samples of size n , can be expressed by

$$g(R) = \sum_{i=1}^{n-1} \frac{(-1)^{n+1-i} (n+1-i)^{n-1} \binom{n-2}{i-1}}{[(n-1)!]^2 \left(\frac{2-i}{n-1} + R\right)^2};$$

- The density of the concentration ratio R , for $0 < R < 1/n$ and for samples of size n , is given by

$$\int_0^{1/n} g(R) dR = \frac{n^{n-2}}{[(n-1)!]^2};$$

- The density of the concentration ratio R , for $(n-1)/n < R < 1$ and for samples of size n , is given by

$$\int_{1/n}^1 g(R) dR = \frac{1}{[(n-1)!]^2};$$

- The j th term of the density of the concentration ratio R , denoted as $a_{i,j}$, verifies the following symmetry

$$a_{i,j} = a_{j,i}.$$

The coefficients of the $a_{i,i}$ terms of the p.d.f. of the concentration ratio R for samples of size $n-1$ multiplied by $(n-1)/n$ become the coefficients of the $a_{i+1,i+1}$ terms of the same p.d.f. for sample of size n .

These results are valid for every sample size and may allow reducing the heavy calculation to determine the p.d.f. of the concentration ratio R .

9. Concluding Remarks

In the present paper we obtain the distributions of the Gini concentration ratio R for samples of size $n = 6, 7, 8, 9$ and 10 drawn from a uniform population. We use the same method used by Girone [12] to derive the same distributions for samples of size $n \leq 5$. We obtain the p.d.f. of the concentration ratio R calculating a multiple integral in $n-1$ dimensions for each region from $(k-1)/(n-1)$ to $k/(n-1)$ for $k = 1, 2, \dots, n-1$. The limits of integration are defined by solving the inequalities of the order statistics divided by the sample

mean and expressed in terms of the concentration ratio R for the values assumed in each of such regions. The calculation of the limits of integration is particularly heavy and requires a very long processing time.

The obtained results show that the p.d.f. of the concentration ratio R is given by hyperbolic splines with degree 2 and with nodes in $k/(n-1)$ for $k=1,2,\dots,n-1$. Such distributions are unimodal with mean tending to $1/3$, which is the value of the concentration ratio R for the population, and have decreasing standard deviation. Moreover, the distributions show a slight positive skewness and platykurtosis that tend to decrease as n increases.

Beyond the possibility to obtain similar results for samples of larger size, open problems are the derivation of the exact expression for the mean and the other features of the distribution of the concentration ratio R for random samples of size n drawn from a uniform population.

References

- [1] Gini, C. (1914) L'ammontare e la composizione della ricchezza delle nazioni. Bocca, Torino.
- [2] Hoeffding, W. (1948) A Class of Statistics with Asymptotically Normal Distribution. *Annals of Mathematical Statistics*, **19**, 293-325.
- [3] Glasser, G.J. (1962) Variance Formulas for the Mean Difference and the Coefficient of Concentration. *Journal of the American Statistical Association*, **57**, 648-654. <http://dx.doi.org/10.1080/01621459.1962.10500553>
- [4] Cucconi, O. (1965) Sulla distribuzione campionaria del rapporto R di concentrazione. *Statistica*, **25**, 119.
- [5] Dall'Aglio, G. (1965) Comportamento asintotico delle stime della differenza media e del rapporto di concentrazione. *Metron*, **24**, 379-414.
- [6] Deltas, G. (2003) The Small-Sample Bias of the Gini Coefficient: Results and Implications for Empirical Research. *Review of Economics and Statistics*, **85**, 226-234. <http://dx.doi.org/10.1162/rest.2003.85.1.226>
- [7] Barrett, G.F. and Donald, S.G. (2009) Statistical Inference with Generalized Gini Indices of Inequality, Poverty, and Welfare. *Journal of Business & Economic Statistics*, **27**, 1-17. <http://dx.doi.org/10.1198/jbes.2009.0001>
- [8] Davidson, R. (2009) Reliable Inference for the Gini Index. *Journal of Econometrics*, **150**, 30-40. <http://dx.doi.org/10.1016/j.jeconom.2008.11.004>
- [9] Fakoor, V., Ghalibaf, M.B. and Azarnoosh, H.A. (2011) Asymptotic Behaviors of the Lorenz Curve and Gini Index in Sampling from a Length-Biased Distribution. *Statistics and Probability Letters*, **81**, 1425-1435. <http://dx.doi.org/10.1016/j.spl.2011.04.013>
- [10] Girone, G. (1968) Sul comportamento campionario simulato del rapporto di concentrazione. *Annali della Facoltà di Economia e Commercio dell'Università degli Studi di Bari*, **23**, 5-11.
- [11] Girone, G. (1971) La distribuzione del rapporto di concentrazione per campionamenti casuali di variabili esponenziali. *Studi di Probabilità, Statistica e Ricerca operativa in onore di Giuseppe Pompilj*, Oderisi, Gubbio.
- [12] Girone, G. (1971) La distribuzione del rapporto di concentrazione per piccoli campioni estratti da una popolazione uniforme. *Annali dell'Istituto di Statistica dell'Università degli Studi di Bari*, **36**, 31-52.

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or **Online Submission Portal**.

