

Research Article

Estimates of Inequality Indices Based on Simple Random, Ranked Set, and Systematic Sampling

Pooja Bansal, Sangeeta Arora, and Kalpana K. Mahajan

Department of Statistics, Panjab University, Chandigarh 160014, India

Correspondence should be addressed to Sangeeta Arora; sarora131@gmail.com

Received 16 June 2013; Accepted 2 August 2013

Academic Editors: X. Dang and S. Lototsky

Copyright © 2013 Pooja Bansal et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Gini index, Bonferroni index, and Absolute Lorenz index are some popular indices of inequality showing different features of inequality measurement. In general simple random sampling procedure is commonly used to estimate the inequality indices and their related inference. The key condition that the samples must be drawn via simple random sampling procedure though makes calculations much simpler but this assumption is often violated in practice as the data does not always yield simple random sample. Nonsimple random samples like Ranked set sampling or stratified sampling are gaining popularity for estimating these indices. The purpose of the present paper is to compare the efficiency of simple random sample estimates of inequality indices with their nonsimple random counterparts. Monte Carlo simulation technique is applied to get the results for some specific distributions.

1. Introduction

Lorenz curve [1] and the associated Gini index [2] are one of the most popular and frequently used tools to measure income inequality. Certain other variants of Lorenz curve, namely, Generalized Lorenz curve [3], Absolute Lorenz curve [4], Bonferroni curve [5], and Comic curves [3, 6] and associated inequality indices, that is, Generalized Lorenz index, Absolute Lorenz index, Bonferroni index, and Comic index, are some of the popular alternatives to Gini index used to study certain specialized features of inequality. In practice, mostly the simple random sampling procedure is used to derive the statistical inference or obtaining the sample estimates of these inequality measures. The key condition that the samples must be drawn via simple random sampling procedure though makes calculations much simpler but this assumption is often violated in practice as the data does not always yield simple random sample. For example, in India the socioeconomic data collected by the National Sample Survey Organization (NSSO) is not drawn through simple random sampling but follows two-stage stratified sampling techniques. Also in the United States, commonly used income and earnings data, such as the Current Population Survey

(CPS) and the Panel Study of Income Dynamics (PSID), are all multistage random samples, where simple random sampling is not the only method applied at each stage. One may quote stratified sampling, cluster sampling, multistage cluster sampling, and so forth, as alternatives to simple random sampling while estimating inequality indices [7].

Besides, these available traditional sampling methods, the method of Ranked set sampling (RSS), have also gained popularity in the literature for estimating parameters like population mean and variance and its associated statistical inference. This RSS method of sampling is proved to be more efficient than simple random sampling [8]. Thus, the RSS technique may also be extended to obtain the sample estimates of these inequality measures, and it will be extended to compare the efficiency of RSS with simple random and other sampling estimates as applied to inequality measures. [9] have given the comparison of ranked set sample estimate of Gini index with simple random sampling for a non-standard Lorenz curve equation using simulation technique and are not true in general as no standard distributions are considered.

The purpose of the present work is to compare the nonsimple random estimates of the inequality measures

TABLE 1: Simple random estimates of inequality indices.

Inequality indices	Simple random sampling estimates
Gini index (G)	$\widehat{G} = 1 - \frac{2}{n} \sum_{i=1}^n L(p_i)$
Bonferroni index (B)	$\widehat{B} = 1 - \frac{1}{n} \sum_{i=1}^n B(p_i)$
Absolute Lorenz index (A)	$\widehat{A} = -\frac{1}{n} \sum_{i=1}^n AL(p_i)$

TABLE 2: Ranked set sample estimates of inequality indices.

Inequality indices	Ranked set sampling estimates
Gini index (G)	$\widehat{G} = 1 - 2 \frac{\sum_{j=1}^r \sum_{i=1}^m L(p_{(i)j})}{n}$
Bonferroni index (B)	$\widehat{B} = 1 - \frac{\sum_{j=1}^r \sum_{i=1}^m B(p_{(i)j})}{n}$
Absolute Lorenz index (A)	$\widehat{A} = -\frac{1}{n} \sum_{j=1}^r \sum_{i=1}^m AL(p_{(i)j})$

with that of simple random estimates. In particular, we have applied two nonsimple random sampling techniques—Ranked set sampling and systematic sampling methods and have obtained the sample estimates of some popular inequality indices—Gini, Bonferroni, and Absolute Lorenz index via method of Monte Carlo simulation [10] and have compared their efficiencies with those of simple random estimates. Simulated results showing the relative efficiencies have been obtained for Pareto, Exponential, and power distribution. It is interesting to note that the simple random estimates have much less efficiency as compared to their nonsimple counterparts, that is, ranked set or systematic sample estimates. We have also compared the ranked set estimates with systematic sample estimates and found that systematic sample estimates are more efficient than the ranked set sample estimates.

The present paper has been organized in the following manner. In Section 2, we have applied the estimates of some popular inequality indices—Gini Index, Bonferroni Index, and Absolute Lorenz Index for three sampling procedures, namely, simple random sampling, ranked set sampling, and systematic sampling—and applied the method of Monte Carlo method of simulation to obtain the estimates. A brief introduction of the inequality indices and their estimation in case of three sampling techniques is given in this section. Simulation study has been performed in Section 3, where we have computed the efficiencies of these estimates for some specific distributions, namely, Exponential, Pareto, and Power distributions. The table showing the various comparison of Mean Square Errors (MSE) under simple random sampling, ranked set sampling, and systematic sampling is given at the end of Section 3 along with the conclusion.

2. Monte Carlo Estimates for Different Sampling Techniques

In this section, we have discussed three sampling techniques, namely, simple random sampling, ranked set sampling, and

TABLE 3: Systematic sample estimates of inequality indices.

Inequality indices	Systematic random sampling estimates
Gini index (G)	$\widehat{G} = 1 - \frac{2}{N} \sum_{i=1}^m \sum_{k_j=1}^{N_i} g(p'_{K_i})$
Bonferroni index (B)	$\widehat{B} = 1 - \frac{1}{N} \sum_{i=1}^m \sum_{k_j=1}^{N_i} g(p'_{K_i})$
Absolute Lorenz index (A)	$\widehat{A} = -\frac{1}{N} \sum_{i=1}^m \sum_{k_j=1}^{N_i} g(p'_{K_i})$

systematic sampling procedures and obtained the Monte Carlo estimates of the three indices Gini, Bonferroni, and Absolute Lorenz.

2.1. Simple Random Sampling Procedure. The sample mean Monte Carlo procedure for the inequality indices using simple random sampling is explained below.

2.1.1. Gini Index. Gini index is defined as

$$G = 1 - 2 \int_0^1 L(p) dp \quad (1)$$

$$\begin{aligned} &= 1 - 2 \int_0^1 \frac{L(p)}{f(p)} \cdot f(p) dp \\ &= 1 - 2E \left[\frac{L(p)}{f(p)} \right] \\ &= 1 - 2I_1, \end{aligned} \quad (2)$$

where

$$L(p) = \frac{1}{\mu} \int_0^p F^{-1}(t) dt, \quad 0 \leq p \leq 1 \quad (3)$$

is the Lorenz curve at p ordinates [1] and

$$\begin{aligned} F^{-1}(p) &= \inf \{x : F(x) \geq p\}, \\ I_1 &= E \left[\frac{L(p)}{f(p)} \right]. \end{aligned} \quad (4)$$

Since p is distributed uniformly over $[0, 1]$, that is, $p \sim U(0, 1)$,

$$\begin{aligned} f(p) &= \begin{cases} 1 & \text{if } 0 < p < 1, \\ 0 & \text{o.w.} \end{cases} \\ \implies I_1 &= E[L(p)]. \end{aligned} \quad (5)$$

An unbiased estimator of I is sample mean

$$\widehat{I}_1 = \frac{1}{n} \sum_{i=1}^n L(p_i), \quad (\text{Reference [10]}), \quad (6)$$

where n is the sample size.

TABLE 4: Gini index (G) for given distributions.

Distribution	cdf	Lorenz curve $L(p)$	Gini index (G)
Exponential	$1 - e^{-\lambda x}, \lambda > 0, x \geq 0$	$p + (1 - p) \log(1 - p)$	$\frac{1}{2}$
Pareto	$1 - \left(\frac{x}{x_0}\right)^{-\alpha}, \alpha > 1, x \geq x_0 > 0$	$1 - (1 - p)^{(\alpha-1)/\alpha}$	$\frac{1}{(2\alpha - 1)}$
Power	$x^\alpha, \alpha > 1, 0 \leq x \leq 1$	$p^{(\alpha+1)/\alpha}$	$\frac{1}{(2\alpha + 1)}$

TABLE 5: Bonferroni index (B) for given distributions.

Distribution	cdf	Bonferroni curve $B(p)$	Bonferroni index (B)
Exponential	$1 - e^{-\lambda x}, \lambda > 0, x \geq 0$	$\frac{p + (1 - p) \log(1 - p)}{p}$	$\frac{\pi^2}{6} - 1$
Pareto	$1 - \left(\frac{x}{x_0}\right)^{-\alpha}, \alpha > 1, x \geq x_0 > 0$	$\frac{1 - (1 - p)^{(\alpha-1)/\alpha}}{p}$	$\psi(2) - \psi\left(\frac{(\alpha-1)}{\alpha}\right)$
Power	$x^\alpha, \alpha > 1, 0 \leq x \leq 1$	$p^{1/\alpha}$	$\frac{1}{(\alpha + 1)}$

Where $\psi(x)$ is the digamma function.

Therefore, unbiased estimator of G is

$$\widehat{G} = 1 - 2\widehat{I}_1. \quad (7)$$

Using (6)

$$\widehat{G} = 1 - \frac{2}{n} \sum_{i=1}^n L(p_i), \quad (8)$$

$$\begin{aligned} \text{Var}(\widehat{G}) &= \text{var}\left[1 - \frac{2}{n} \sum_{i=1}^n L(p_i)\right] \\ &= \frac{4}{n^2} \text{var}\left[\sum_{i=1}^n L(p_i)\right] \\ &= \frac{4}{n} \left\{E[L(p)]^2 - I_1^2\right\}, \quad \text{from (5)} \\ &= \frac{4}{n} \left[\int_0^1 (L(p))^2 dp - \left(\frac{1-G}{2}\right)^2 \right], \quad \text{from (1).} \end{aligned} \quad (9)$$

Algorithm for computing simple random estimate for Gini index is given as follows.

- (i) Generate a sequence $\{X_i\}_{i=1}^N$ of N random numbers from $U(0, 1)$.
- (ii) Compute $p_i = X_i$, $i = 1, \dots, N$.
- (iii) Compute $L(p_i)$, $i = 1, \dots, N$.
- (iv) Compute $\widehat{G} = 1 - (2/n) \sum_{i=1}^n L(p_i)$.

2.1.2. Bonferroni Curve. Bonferroni index is defined as

$$\begin{aligned} B &= 1 - \int_0^1 B(p) dp \\ &= 1 - \int_0^1 \frac{B(p)}{f(p)} \cdot f(p) dp \\ &= 1 - E\left[\frac{B(p)}{f(p)}\right] \\ &= 1 - I_2, \end{aligned} \quad (10)$$

where

$$B(p) = \frac{L(p)}{p} = \frac{1}{\mu \cdot p} \int_0^p F^{-1}(t) dt, \quad 0 \leq p \leq 1 \quad (11)$$

is the Bonferroni curve defined at p ordinates [5]

$$I_2 = E\left[\frac{B(p)}{f(p)}\right]. \quad (12)$$

Since p is distributed uniformly over $[0, 1]$, that is, $p \sim U(0, 1)$,

$$\implies I_2 = E[B(p)]. \quad (13)$$

An unbiased estimator of I is sample mean

$$\widehat{I}_2 = \frac{1}{n} \sum_{i=1}^n B(p_i), \quad (14)$$

where n is the sample size.

Therefore, unbiased estimator of B is

$$\widehat{B} = 1 - \widehat{I}_2. \quad (15)$$

TABLE 6: Absolute Lorenz index (A) for given distributions.

Distribution	cdf	Absolute Lorenz curve	Absolute Lorenz index (A)
Exponential	$1 - e^{-\lambda x}, \lambda > 0, x \geq 0$	$\frac{(1-p)\log(1-p)}{\lambda}$	$\frac{1}{4}\lambda$
Pareto	$1 - \left(\frac{x}{x_0}\right)^{-\alpha}, \alpha > 1, x \geq x_0 > 0$	$\frac{\alpha x_0 (1-p) \left[1 - (1-p)^{-1/\alpha}\right]}{(\alpha-1)}$	$\frac{\alpha x_0}{2(\alpha-1)(2\alpha-1)}$
Rectangular	$\frac{x-a}{\theta}, a \leq x \leq a+\theta, \theta > 0, a \geq 0$	$\frac{\theta p(p-1)}{2}$	$\frac{\theta}{12}$

TABLE 7: Relative efficiency of Gini index for exponential distribution.

r	m	MSE (RSS)	MSE (SRS)	MSE (SYS)	$E(1)$	$E(2)$	$E(3)$
2	3	0.02498093	0.04457217	0.00910646	1.784248	4.894566	2.743210
	4	0.01530442	0.03360818	0.004001166	2.195979	8.399597	3.824990
	5	0.01046073	0.02692506	0.002174010	2.573918	12.384975	4.811721
	6	0.007608975	0.02218654	0.001295014	2.915837	17.132278	5.875593
	3	0.01002222	0.01747601	0.003524347	1.743726	4.958652	2.843710
5	4	0.00611229	0.01333912	0.001584625	2.182345	8.417840	3.857247
	5	0.004159828	0.01100573	0.0008643665	2.645717	12.732712	4.812574
	6	0.003042587	0.00911744	0.0005193771	2.996608	17.554567	5.858146
	3	0.002467123	0.00450777	0.0009057722	1.827136	4.976715	2.723779
	4	0.001525546	0.003313244	0.0004123536	2.171842	8.034958	3.699606
20	5	0.001024757	0.002690909	0.0002184094	2.625900	12.320482	4.691909
	6	0.0007222134	0.002298709	0.0001300434	3.182866	17.676476	5.553634
	3	0.0005033155	0.0008831595	0.0001777265	1.754684	4.969205	2.831967
	4	0.0003078470	0.0006709811	8.14811e-05	2.179593	8.234807	3.778140
	5	0.0002099437	0.0005397776	4.331869e-05	2.571059	12.460617	4.846492
100	6	0.0001524544	0.0004400118	2.622736e-05	2.886186	16.776824	5.812800
	3	0.0002471209	0.0004459992	8.781071e-05	1.804781	5.079098	2.814246
	4	0.000153431	0.0003294056	4.053465e-05	2.14693	8.126519	3.785181
	5	0.0001031339	0.0002670593	2.214463e-05	2.589443	12.059777	4.657287
	6	7.600503e-05	0.0002178877	1.308801e-05	2.866754	16.647886	5.807226
200	3	0.0001002982	0.0001764347	3.693063e-05	1.759102	4.777463	2.715854
	4	5.943145e-05	0.0001320398	1.595697e-05	2.221716	8.274741	3.724482
	5	4.208938e-05	0.0001076813	8.8025e-06	2.558395	12.233036	4.781526
	6	3.117683e-05	9.065143e-05	5.230148e-06	2.907654	17.332479	5.960984

Using (14),

$$\hat{B} = 1 - \frac{1}{n} \sum_{i=1}^n B(p_i), \quad (16)$$

$$\begin{aligned} \text{Var}(\hat{B}) &= \text{var} \left[1 - \frac{1}{n} \sum_{i=1}^n B(p_i) \right] \\ &= \frac{1}{n^2} \text{var} \left[\sum_{i=1}^n B(p_i) \right] \\ &= \frac{1}{n} \left\{ E[B(p)]^2 - I_2^2 \right\}, \quad \text{from (13)} \\ &= \frac{1}{n} \left[\int_0^1 (B(p))^2 dp - (1-B)^2 \right], \quad \text{from (10).} \end{aligned} \quad (17)$$

Algorithm for computing simple random estimate for Bonferroni index is given as follows.

- (i) Generate a sequence $\{X_i\}_{i=1}^N$ of N random numbers from $U(0, 1)$.
- (ii) Compute $p_i = X_i, i = 1, \dots, N$.
- (iii) Compute $B(p_i), i = 1, \dots, N$.
- (iv) Compute $\hat{B} = 1 - (1/n) \sum_{i=1}^n B(p_i)$.

TABLE 8: Relative efficiency of Gini index for pareto ($\alpha = 3$) distribution.

r	m	MSE (RSS)	MSE (SRS)	MSE (SYS)	$E(1)$	$E(2)$	$E(3)$
2	3	0.02332234	0.04529824	0.00589787	1.942268	7.680441	3.954367
	4	0.01381014	0.03470658	0.002580012	2.513122	13.452100	5.352743
	5	0.009193211	0.02779215	0.001352259	3.023116	20.552387	6.798410
	6	0.00659181	0.02279101	0.0008011307	3.457473	28.448554	8.228133
5	3	0.009292055	0.01775528	0.002399531	1.910802	7.399479	3.872446
	4	0.005570443	0.01392200	0.001050541	2.499263	13.252220	5.302452
	5	0.003769394	0.01075287	0.0005584852	2.852678	19.253635	6.749318
	6	0.00268615	0.009290222	0.0003206307	3.458554	28.974836	8.377707
20	3	0.002275199	0.004430962	0.0006094044	1.947505	7.270971	3.733480
	4	0.001390079	0.003490160	0.0002649288	2.510764	13.173955	5.246991
	5	0.0009236206	0.002648793	0.0001380824	2.867836	19.182698	6.688909
	6	0.0006634941	0.002265676	8.109345e - 05	3.414765	27.939075	8.181846
100	3	0.000461317	0.00090545	0.0001197988	1.962750	7.558089	3.850765
	4	0.0002732137	0.0006699247	5.275955e - 05	2.452017	12.697695	5.178469
	5	0.0001887224	0.0005433437	2.791159e - 05	2.879063	19.466598	6.761435
	6	0.0001309059	0.0004580686	1.628860e - 05	3.499220	28.122036	8.036658
200	3	0.0002253251	0.0004546124	6.028802e - 05	2.017584	7.540676	3.737477
	4	0.0001392043	0.0003422813	2.601848e - 05	2.458842	13.155315	5.350209
	5	9.250598e - 05	0.0002770614	1.426628e - 05	2.995065	19.420718	6.484240
	6	6.674519e - 05	0.0002295553	7.895343e - 06	3.439278	29.074772	8.453742
500	3	9.473933e - 05	0.0001802806	2.443929e - 05	1.902912	7.376671	3.876517
	4	5.606348e - 05	0.0001383186	1.074361e - 05	2.467178	12.874499	5.218309
	5	3.699474e - 05	0.0001102943	5.493003e - 06	2.981350	20.079053	6.734884
	6	2.701999e - 05	9.018714e - 05	3.301841e - 06	3.337793	27.314198	8.183310

2.1.3. *Absolute Lorenz Index.* Absolute Lorenz index is defined as

$$\begin{aligned}
 A &= - \int_0^1 \text{AL}(p) dp \\
 &= - \int_0^1 \frac{\text{AL}(p)}{f(p)} \cdot f(p) dp \\
 &= - E \left[\frac{\text{AL}(p)}{f(p)} \right] \\
 &= - I_3,
 \end{aligned} \tag{18}$$

where

$$\text{AL}(p) = -\frac{\mu}{2} + \int_0^1 x F(x) dF(x) \tag{19}$$

is the Absolute Lorenz curve defined at p ordinates [6]

$$I_3 = E \left[\frac{\text{AL}(p)}{f(p)} \right]. \tag{20}$$

Since p is distributed uniformly over $[0, 1]$, that is, $p \sim U(0, 1)$,

$$\implies I_3 = E[\text{AL}(p)]. \tag{21}$$

An unbiased estimator of I is sample mean

$$\hat{I}_3 = \frac{1}{n} \sum_{i=1}^n \text{AL}(p_i), \tag{22}$$

where n is the sample size.

Therefore, unbiased estimator of A is

$$\hat{A} = -\hat{I}_3 \tag{23}$$

using (22)

$$\hat{A} = -\frac{1}{n} \sum_{i=1}^n \text{AL}(p_i), \tag{24}$$

$$\begin{aligned}
 \text{Var}(\hat{A}) &= \text{var} \left[-\frac{1}{n} \sum_{i=1}^n \text{AL}(p_i) \right] \\
 &= \frac{1}{n^2} \text{var} \left[\sum_{i=1}^n \text{AL}(p_i) \right] \\
 &= \frac{1}{n} \{E[\text{AL}(p)]^2 - I_3^2\}, \quad \text{from (21)} \\
 &= \frac{1}{n} \left[\int_0^1 (\text{AL}(p))^2 dp - A^2 \right], \quad \text{from (18).}
 \end{aligned} \tag{25}$$

Algorithm for computing simple random estimate for Absolute Lorenz index is given as follows.

TABLE 9: Relative efficiency of Gini index for power ($\alpha = 3$) distribution.

r	m	MSE (RSS)	MSE (SRS)	MSE (SYS)	$E(1)$	$E(2)$	$E(3)$
2	3	0.03002883	0.05904002	0.006465061	1.966111	9.132168	4.644787
	4	0.01837878	0.04518542	0.002769992	2.458565	16.312473	6.634958
	5	0.01202134	0.03688425	0.001390412	3.068232	26.527569	8.645883
	6	0.008600471	0.02950916	0.000838656	3.431109	35.186250	10.255064
5	3	0.01183113	0.02385245	0.00263158	2.016075	9.063927	4.495828
	4	0.007109205	0.01767874	0.001112017	2.486739	15.897904	6.393072
	5	0.004782672	0.01397249	0.0005790662	2.921482	24.129348	8.259284
	6	0.003415356	0.01160659	0.0003343554	3.398354	34.713332	10.214748
20	3	0.003007737	0.005933085	0.0006660064	1.972607	8.908450	4.516078
	4	0.001856484	0.004438954	0.0002774993	2.391054	15.996271	6.690049
	5	0.001176903	0.003597289	0.0001420534	3.056573	25.323498	8.284934
	6	0.0008649264	0.00293501	8.396861e - 05	3.393364	34.953657	10.300592
100	3	0.0006040928	0.001167984	0.0001285668	1.933451	9.084647	4.698669
	4	0.0003599259	0.0008980578	5.426289e - 05	2.495118	16.550128	6.633003
	5	0.0002358945	0.0007081798	2.842227e - 05	3.002103	24.916370	8.299636
	6	0.000173568	0.0005906876	1.640179e - 05	3.403206	36.013606	10.582260
200	3	0.0002970297	0.0005835807	6.760029e - 05	1.964722	8.632814	4.393912
	4	0.0001840207	0.0004393907	2.766382e - 05	2.387724	15.883226	6.652035
	5	0.0001199216	0.0003552514	1.421686e - 05	2.962364	24.988035	8.435168
	6	8.84773e - 05	0.0002953183	8.21482e - 06	3.337786	35.949455	10.770449
500	3	0.0001217598	0.0002407522	2.666487e - 05	1.977272	9.028816	4.566300
	4	7.261201e - 05	0.0001784493	1.124082e - 05	2.457572	15.875114	6.459672
	5	4.8425e - 05	0.0001436364	5.589331e - 06	2.966163	25.698317	8.663828
	6	3.399731e - 05	0.0001184066	3.256773e - 06	3.482823	36.357032	10.438956

- (i) Generate a sequence $\{X_i\}_{i=1}^N$ of N random numbers from $U(0, 1)$.
- (ii) Compute $p_i = X_i$, $i = 1, \dots, N$.
- (iii) Compute $\text{AL}(p_i)$, $i = 1, \dots, N$.
- (iv) Compute $\widehat{A} = -(1/n) \sum_{i=1}^n \text{AL}(p_i)$.

Simple random sampling estimates of Gini, Bonferroni, and Absolute Lorenz indices using Monte Carlo approach are given in the Table 1.

2.2. Ranked Set Sampling Procedure. General, ranked set sampling procedure involves the following steps [6].

(i) Select a simple random sample of m units from the population and subject it to ordering on attribute of interest via some ranking process, that is,

$$X_{(1)11} < \dots < X_{(m)11}. \quad (26)$$

This judgment ranking can result from a variety of mechanisms, including expert's opinion, visual comparisons, or the use of easy-to-obtain auxiliary variables, but it cannot involve actual measurements of the attribute of interest on sample units.

Once this judgement ranking of m units in our initial random sample has been accomplished, the item judged to be smallest is included as the first item in the ranked set sample and the attribute of interest is formally measured on this unit

and remaining $(m - 1)$ unmeasured units in the first random sample are not considered further.

$$X_{(1)11} < \dots < X_{(m)11} \implies X_{[1]11}. \quad (27)$$

This measurement is denoted by $X_{[1]11}$, the smallest orderstatistic or the smallest judgement order item. It may or may not actually have the smallest attribute measurement among the m sampled units. Note that $X_{(2)11} < \dots < X_{(m)11}$ are not considered further in the selection of our ranked set sample. The sole purpose of $(m - 1)$ units is to help in selecting an item for measurement that represents the smaller attribute values in the population.

(ii) Following selection of $X_{[1]11}$, a second independent random sample of size m is selected from the population and judgement ranking without formal measurement on the attribute of interest is made and the second smallest item of m units in second random sample is selected and is included in the ranked set sample for measurement of attribute of interest; that is,

$$X_{(1)21} < \dots < X_{(m)21} \implies X_{[2]21}. \quad (28)$$

The second measured observation is denoted by $X_{[2]21}$.

(iii) Similarly, from a third independent random sample, we select the unit judgement ranked to be the third smallest, $X_{[3]31}$, for measurement and inclusion in the ranked-set sample, that is,

$$X_{(1)31} < \dots < X_{(m)31} \implies X_{[3]31}. \quad (29)$$

TABLE 10: Relative efficiency of Bonferroni index for rectangular ($\theta = 3$) distribution.

r	m	MSE (RSS)	MSE (SRS)	MSE (SYS)	$E(1)$	$E(2)$	$E(3)$
2	3	0.06202481	0.1235527	0.01377850	1.991989	8.967065	4.501565
	4	0.0381469	0.09533465	0.005916627	2.499146	16.113007	6.447407
	5	0.02535770	0.07558111	0.002995199	2.980598	25.234086	8.466115
	6	0.01791203	0.06181676	0.001748279	3.451131	35.358636	10.245521
5	3	0.02490822	0.05036193	0.005549141	2.0219	9.075626	4.488662
	4	0.01489808	0.03685433	0.002358322	2.473765	15.627353	6.317237
	5	0.00982462	0.03053955	0.001206127	3.108471	25.320344	8.145593
	6	0.007220886	0.02487922	0.0006952421	3.445452	35.784973	10.386146
20	3	0.006252341	0.01245597	0.001364699	1.992209	9.127265	4.581480
	4	0.003685856	0.009230949	0.0005877362	2.504425	15.705939	6.271276
	5	0.002550649	0.007501943	0.0002866664	2.941190	26.169593	8.897621
	6	0.001808095	0.00619635	0.0001755058	3.427005	35.305671	10.302195
100	3	0.001249233	0.002446855	0.0002789078	1.958686	8.772989	4.479018
	4	0.0007403018	0.001914464	0.0001151496	2.586059	16.625885	6.429044
	5	0.000504336	0.001465721	6.01177e - 05	2.906238	24.380856	8.389143
	6	0.0003529966	0.001248915	3.414360e - 05	3.538037	36.578305	10.338588
200	3	0.0006307183	0.001248089	0.0001376957	1.978838	9.064110	4.580523
	4	0.0003722346	0.0009302265	5.962156e - 05	2.499033	15.602183	6.243289
	5	0.0002493340	0.0007560325	2.945604e - 05	3.032207	25.666468	8.464614
	6	0.0001767726	0.000618749	1.745216e - 05	3.500253	35.454007	10.128981
500	3	0.0002467436	0.0005050265	5.529099e - 05	2.046767	9.133975	4.462637
	4	0.0001511701	0.0003785248	2.356102e - 05	2.503966	16.065722	6.416110
	5	9.98555e - 05	0.0002998347	1.208355e - 05	3.002686	24.813461	8.263755
	6	7.085566e - 05	0.000252477	6.991167e - 06	3.563258	36.113713	10.135026

This process is continued until we have selected the unit judgement ranked to be largest of m units in the m th random sample, denoted by $X_{[m]m1}$; that is, $X_{(1)m1} < \dots < X_{(m)m1} \Rightarrow X_{[m]m1}$ for inclusion in our ranked set sample.

This entire process is referred to as a cycle and the number of observations in each random sample, m in our example, is called set size.

Thus, to complete a single ranked set cycle, we need to judge rank m independent random samples of size m involving a total of m^2 sample units in order to obtain m measured observations. These m observations represent a balanced ranked set sample with set size m .

In practice, the sample size m is kept small to ease the visual ranking, the RSS literature suggested $m = 2, 3, 4, 5$ or 6. Therefore, in order to obtain a ranked set sample with a desired total number of measured observation $n = r * m$, we repeat the entire cycle process r independent times, yielding the data

$$X_{[1]ij} < \dots < X_{[m]ij}, \quad i = 1, \dots, m; \quad j = 1, \dots, r. \quad (30)$$

Algorithm for obtaining the RSS estimate of Gini index via method of Monte Carlo simulation is given as follows.

- (a) Generate an RSS $U_{[m]ij}$, $i = 1, \dots, m; \quad j = 1, \dots, r$ of size $n = m \times r$ from $U(0, 1)$.
- (b) Assign $p_{(i)j} = U_{[m]ij}$, $i = 1, \dots, m; \quad j = 1, \dots, r$.
- (c) Compute $L(p_{(i)j})$.

- (d) Compute the ranked set sample estimate of Gini index as

$$\widehat{G} = 1 - 2 \frac{\sum_{j=1}^r \sum_{i=1}^m L(p_{(i)j})}{n}. \quad (31)$$

Algorithm for obtaining the RSS estimate of Bonferroni index via method of Monte Carlo simulation is given as following.

- (a) Generate a RSS $U_{[m]ij}$, $i = 1, \dots, m; \quad j = 1, \dots, r$ of size $n = m \times r$ from $U(0, 1)$.
- (b) Assign $p_{(i)j} = U_{[m]ij}$, $i = 1, \dots, m; \quad j = 1, \dots, r$.
- (c) Compute $B(p_{(i)j})$.
- (d) Compute the ranked set sample estimate of Bonferroni index as follows.

$$\widehat{B} = 1 - \frac{\sum_{j=1}^r \sum_{i=1}^m B(p_{(i)j})}{n}. \quad (32)$$

Algorithm for obtaining the RSS estimate of Absolute Lorenz index via method of Monte Carlo simulation is given as follows.

- (a) Generate a RSS $U_{[m]ij}$, $i = 1, \dots, m; \quad j = 1, \dots, r$ of size $n = m \times r$ from $U(0, 1)$.
- (b) Assign $p_{(i)j} = U_{[m]ij}$, $i = 1, \dots, m; \quad j = 1, \dots, r$.

TABLE 11: Relative efficiency of Bonferroni index for exponential distribution.

r	m	MSE (RSS)	MSE (SRS)	MSE (SYS)	$E(1)$	$E(2)$	$E(3)$
2	3	0.005826023	0.01085295	0.001687301	1.86284	6.432136	3.452865
	4	0.003477376	0.008182712	0.0007416163	2.353128	11.033619	4.688915
	5	0.002229820	0.006484168	0.0003876997	2.907933	16.724718	5.751410
	6	0.001690472	0.005454195	0.0002310937	3.226432	23.601660	7.315093
5	3	0.002319801	0.004423554	0.0006583515	1.906868	6.719137	3.523651
	4	0.001337874	0.003238717	0.0002952573	2.420794	10.969134	4.531214
	5	0.0009042652	0.002668140	0.0001607564	2.950617	16.597411	5.625065
	6	0.0006640723	0.002151227	9.39234e - 05	3.239447	22.904058	7.070361
20	3	0.0005728049	0.001073308	0.0001686894	1.873776	6.362629	3.395619
	4	0.0003477821	0.0008003468	7.638552e - 05	2.301288	10.477729	4.552985
	5	0.0002317084	0.0006640685	4.120043e - 05	2.865966	16.117999	5.623932
	6	0.0001682733	0.0005685107	2.490111e - 05	3.378496	22.830737	6.757663
100	3	0.0001145552	0.0002185333	3.457197e - 05	1.907668	6.321112	3.313528
	4	7.047152e - 05	0.0001665346	1.634645e - 05	2.363148	10.187814	4.311121
	5	4.76686e - 05	0.0001299972	9.897842e - 06	2.727103	13.133893	4.816060
	6	3.368466e - 05	0.0001080593	6.489773e - 06	3.20796	16.650706	5.190422
200	3	5.866765e - 05	0.0001113661	1.836847e - 05	1.898254	6.062895	3.193932
	4	3.637083e - 05	8.33706e - 05	9.291632e - 06	2.292238	8.972654	3.914364
	5	2.444185e - 05	6.776085e - 05	5.651263e - 06	2.772329	11.990390	4.325024
	6	1.846277e - 05	5.761897e - 05	4.05105e - 06	3.120819	14.223219	4.557527
500	3	2.478413e - 05	4.502048e - 05	8.473576e - 06	1.816505	5.313044	2.924873
	4	1.551794e - 05	3.491700e - 05	4.598167e - 06	2.250106	7.593678	3.374810
	5	1.109970e - 05	2.797701e - 05	3.230694e - 06	2.520520	8.659752	3.435701
	6	8.396832e - 06	2.362577e - 05	2.635701e - 06	2.813653	8.963752	3.185806

(c) Compute $AL(p_{(i)j})$.

(d) Compute the ranked set sample estimate of Absolute Lorenz index as

$$\widehat{A} = - \frac{\sum_{j=1}^r \sum_{i=1}^m AL(p_{(i)j})}{n}. \quad (33)$$

Ranked set sampling estimates of Gini, Bonferroni, and Absolute Lorenz indices using Monte Carlo approach are given in Table 2.

2.3. Systematic Sampling Procedure. Systematic sampling procedure is particular case of stratified sampling procedure where the stratas are of equal size.

2.3.1. Gini Index. Gini index is defined as

$$\begin{aligned} G &= 1 - 2 \int_0^1 L(p) dp \\ &= 1 - 2I_1. \end{aligned} \quad (34)$$

Its unbiased estimator in case of stratified sampling is given as

$$\implies \widehat{G} = 1 - 2\widehat{I}_1, \quad (35)$$

where

$$\begin{aligned} I_1 &= \int_0^1 L(p) dp = \int_D L(p) dp \\ &= \int_D \frac{L(p)}{f(p)} \cdot f(p) dp = \int_D g(p) \cdot f(p) dp, \end{aligned} \quad (36)$$

where

$$g(p) = \frac{L(p)}{f(p)}. \quad (37)$$

Let region D be divided into m disjoint subregions D_i , $i = 1, \dots, m$; that is, $D = \bigcup_{i=1}^m D_i$ and $D_k \cap D_j = \emptyset$, $k \neq j$ where \emptyset is null set.

Define

$$\begin{aligned} I_{1i} &= \int_{D_i} g(p) \cdot f(p) dp, \\ P_i &= \int_{D_i} f(p) dp, \quad \sum_{i=1}^m P_i = 1, \\ I_1 &= \int_D g(p) \cdot f(p) dp = \sum_{i=1}^m \int_{D_i} g(p) \cdot f(p) dp = \sum_{i=1}^m I_{1i}. \end{aligned} \quad (38)$$

TABLE 12: Relative efficiency of Bonferroni index for pareto ($\alpha = 3$) distribution.

r	m	MSE (RSS)	MSE (SRS)	MSE (SYS)	$E(1)$	$E(2)$	$E(3)$
2	3	0.0004945458	0.0009228851	0.0001830762	1.866127	5.040989	2.701311
	4	0.0003013004	0.0007029501	8.527298e - 05	2.333054	8.243527	3.533363
	5	0.0002011407	0.0005511385	4.790515e - 05	2.740064	11.504786	4.198728
	6	0.0001507043	0.0004636178	2.906871e - 05	3.076342	15.949032	5.184417
5	3	0.0001979686	0.0003752962	7.295216e - 05	1.895736	5.144415	2.713677
	4	0.0001211198	0.0002751505	3.327257e - 05	2.271723	8.269590	3.640230
	5	8.27896e - 05	0.0002221785	1.902978e - 05	2.683652	11.675306	4.350528
	6	5.93931e - 05	0.0001855068	1.188047e - 05	3.123373	15.614433	4.999221
20	3	4.902086e - 05	9.181063e - 05	1.821531e - 05	1.872889	5.040300	2.691190
	4	3.027866e - 05	7.056706e - 05	8.84171e - 06	2.330587	7.981155	3.424525
	5	2.076647e - 05	5.352124e - 05	4.825012e - 06	2.577291	11.092457	4.303921
	6	1.512937e - 05	4.587691e - 05	2.918024e - 06	3.032308	15.721910	5.184800
100	3	9.94706e - 06	1.860485e - 05	3.674845e - 06	1.870386	5.062758	2.706797
	4	6.097801e - 06	1.383726e - 05	1.745582e - 06	2.269221	7.927018	3.493277
	5	3.963538e - 06	1.106979e - 05	9.846638e - 07	2.792905	11.242203	4.025270
	6	2.950961e - 06	9.162954e - 06	6.031919e - 07	3.105075	15.190778	4.892242
200	3	4.966537e - 06	9.244004e - 06	1.813586e - 06	1.861257	5.097086	2.738518
	4	2.995594e - 06	6.901277e - 06	8.687236e - 07	2.303809	7.944157	3.448271
	5	2.069782e - 06	5.520984e - 06	4.823448e - 07	2.667423	11.446136	4.291084
	6	1.466736e - 06	4.654902e - 06	2.952955e - 07	3.173647	15.763539	4.967011
500	3	1.986694e - 06	3.73913e - 06	7.290523e - 07	1.882087	5.128754	2.725036
	4	1.206677e - 06	2.715055e - 06	3.486712e - 07	2.250026	7.786863	3.460788
	5	8.166757e - 07	2.197725e - 06	1.930648e - 07	2.691062	11.383354	4.230060
	6	5.903014e - 07	1.834876e - 06	1.190148e - 07	3.108371	15.417209	4.959899

Define

where p'_i is distributed according to $f(p)/P_i$ on D_i

$$g_i(p') = \begin{cases} g(p) & p \in D_i, \\ 0 & \text{o.w.}, \end{cases}$$

$$I_{1i} = \int_{D_i} g(p) \cdot f(p) dp = \int_{D_i} P_i \cdot g(p) \cdot \frac{f(p)}{P_i} dp \quad (39)$$

$$= P_i \int_{D_i} g(p') \cdot \frac{f(p)}{P_i} dp = P_i E[g_i(p')],$$

where

$$\int_{D_i} \frac{f(p)}{P_i} dp = 1. \quad (40)$$

Random variable p' is distributed according to $f(p)/P_i$ on D_i . I_{1i} can be estimated by

$$\hat{I}_{1i} = \frac{P_i}{N_i} \sum_{k_i=1}^{N_i} g(p'_{k_i}), \quad i = 1, \dots, m. \quad (41)$$

Therefore, integral I_1 is estimated by

$$\hat{I}_1 = \sum_{i=1}^m \hat{I}_{1i} = \sum_{i=1}^m \frac{P_i}{N_i} \sum_{k_i=1}^{N_i} g(p'_{k_i}), \quad (42)$$

$$\implies \widehat{G} = 1 - 2\hat{I}_1,$$

$$\widehat{G} = 1 - 2 \sum_{i=1}^m \frac{P_i}{N_i} \sum_{k_i=1}^{N_i} g(p'_{k_i}), \quad (43)$$

$$\text{Var}(\hat{I}_1) = \sum_{i=1}^m \frac{P_i^2}{N_i} \cdot \text{Var}[g(p'_i)]$$

$$= \sum_{i=1}^m \frac{P_i^2 \sigma_i^2}{N_i},$$

where

$$\sigma_i^2 = \text{Var}[g(p'_i)] = \frac{1}{P_i} \int_{D_i} g^2(p) \cdot f(p) dp - \frac{I_{1i}^2}{P_i^2}, \quad (44)$$

$$\text{Var}(\widehat{G}) = \text{Var}(1 - 2\hat{I}_1) = 4 \text{Var}(\hat{I}_1) = 4 \cdot \sum_{i=1}^m \frac{P_i^2 \sigma_i^2}{N_i}.$$

In particular, if $P_i = 1/m$ and $N_i = N/m$, we get the systematic sampling estimate of Gini index as

$$\widehat{G} = 1 - \frac{2}{N} \sum_{i=1}^m \sum_{k_i=1}^{N_i} g(p'_{k_i}). \quad (45)$$

TABLE 13: Relative efficiency of absolute Lorenz index for pareto ($\alpha = 3, x_0 = 200$) distribution.

r	m	MSE (RSS)	MSE (SRS)	MSE (SYS)	$E(1)$	$E(2)$	$E(3)$
2	3	20.70694	28.60494	10.57485	1.381418	2.704997	1.958131
	4	13.5872	21.77554	5.390407	1.602651	4.039684	2.520626
	5	9.815137	17.37329	3.06463	1.770051	5.668968	3.202715
	6	7.333079	14.3627	1.935963	1.958618	7.418892	3.787820
5	3	8.490195	11.32587	4.191259	1.333994	2.702260	2.025691
	4	5.523156	8.495127	2.138295	1.538093	3.972851	2.582972
	5	3.839579	6.901874	1.21959	1.79756	5.659176	3.148254
	6	2.848677	5.687296	0.7867597	1.996469	7.228759	3.620771
20	3	2.050789	2.857925	1.037954	1.393574	2.753422	1.975800
	4	1.377661	2.149993	0.521389	1.560612	4.123587	2.642290
	5	0.9706388	1.714577	0.301321	1.766442	5.690201	3.221278
	6	0.7303821	1.413525	0.1940954	1.935322	7.282630	3.763006
100	3	0.4120384	0.5679051	0.2109544	1.378282	2.692075	1.953211
	4	0.270179	0.4210431	0.1080168	1.558386	3.897941	2.501268
	5	0.1956007	0.3382184	0.06211348	1.729127	5.445169	3.149086
	6	0.1465505	0.2881033	0.03872126	1.965897	7.440442	3.784755
200	3	0.2098861	0.2912766	0.1067612	1.387785	2.728300	1.965940
	4	0.1389657	0.2102593	0.05198104	1.513031	4.044923	2.673392
	5	0.1003489	0.1699745	0.03067446	1.693835	5.541239	3.271415
	6	0.07082068	0.1436694	0.01940810	2.028637	7.402548	3.649027
500	3	0.08326179	0.1144187	0.04192018	1.374205	2.729442	1.986198
	4	0.0548825	0.08290751	0.02124204	1.510636	3.902992	2.583674
	5	0.03854217	0.0686116	0.01242014	1.780170	5.524221	3.103199
	6	0.02853415	0.05816632	0.007684862	2.038481	7.568948	3.713033

Algorithm for computing systematic random sampling estimate is given as follows.

- (i) Divide the range $[0, 1]$ of the cumulative distribution into m intervals each of width $1/m$.
- (ii) Generate $\{U_k, k = 1, \dots, N/m; i = 1, \dots, m\}$ from $U(0, 1)$.
- (iii) Compute $Y_{k_i} = (i - 1 + U_{k_i})/m; k_i = 1, \dots, N/m; i = 1, \dots, m$.
- (iv) Compute $p_{k_i} = F^{-1}(Y_{k_i})$.
- (v) Compute

$$\widehat{G} = 1 - \frac{2}{N} \sum_{i=1}^m \sum_{k_i=1}^{N_i} g(p_{k_i}). \quad (46)$$

2.3.2. Bonferroni Index. Similarly, Bonferroni index is defined as

$$\begin{aligned} B &= 1 - \int_0^1 B(p) dp \\ &= 1 - I_2. \end{aligned} \quad (47)$$

Its stratified estimate is given as

$$\implies \widehat{B} = 1 - \widehat{I}_2, \quad (48)$$

where

$$\begin{aligned} I_2 &= \int_0^1 B(p) dp = \int_D B(p) dp \\ &= \int_D \frac{B(p)}{f(p)} \cdot f(p) dp = \int_D g(p) \cdot f(p) dp, \end{aligned} \quad (49)$$

where

$$g(p) = \frac{B(p)}{f(p)}. \quad (50)$$

Let region D be divided into m disjoint subregions $D_i, i = 1, \dots, m$, that is, $D = \bigcup_{i=1}^m D_i$ and $D_k \cap D_j = \emptyset, k \neq j$, where \emptyset is null set.

Define

$$I_{2i} = \int_{D_i} g(p) \cdot f(p) dp,$$

$$P_i = \int_{D_i} f(p) dp, \quad \sum_{i=1}^m P_i = 1,$$

$$I_2 = \int_D g(p) \cdot f(p) dp = \sum_{i=1}^m \int_{D_i} g(p) \cdot f(p) dp = \sum_{i=1}^m I_{2i}. \quad (51)$$

TABLE 14: Relative efficiency of absolute Lorenz index for rectangular ($\theta = 3$) distribution.

r	m	MSE (RSS)	MSE (SRS)	MSE (SYS)	$E(1)$	$E(2)$	$E(3)$
2	3	0.0018881	0.002099699	0.00105751	1.11207	1.985512	1.785420
	4	0.001239755	0.001571081	0.0004786396	1.267251	3.282388	2.590164
	5	0.0008749366	0.001232462	0.0002404251	1.408630	5.126179	3.639123
	6	0.0006767988	0.001043956	0.0001412633	1.542491	7.390143	4.791045
5	3	0.0007426104	0.000812753	0.0004213969	1.094454	1.928711	1.762259
	4	0.0005023406	0.0006275718	0.0001826594	1.249295	3.435749	2.750149
	5	0.0003583338	0.0005023294	9.73136e - 05	1.401848	5.161965	3.682258
	6	0.0002722183	0.0004232076	5.552897e - 05	1.554662	7.621384	4.902275
20	3	0.0001900359	0.0002072712	0.0001066924	1.090695	1.942699	1.781157
	4	0.0001270530	0.0001569948	4.505728e - 05	1.235664	3.484338	2.819811
	5	9.21853e - 05	0.0001217164	2.413549e - 05	1.320345	5.043047	3.819492
	6	6.948416e - 05	0.0001039845	1.432389e - 05	1.496521	7.259515	4.850928
100	3	3.739423e - 05	4.054800e - 05	2.125268e - 05	1.084339	1.907901	1.759507
	4	2.485619e - 05	3.087114e - 05	9.368584e - 06	1.241990	3.295177	2.653143
	5	1.785318e - 05	2.452328e - 05	4.819807e - 06	1.373609	5.088021	3.704128
	6	1.317393e - 05	2.022185e - 05	2.820976e - 06	1.534990	7.168388	4.669990
200	3	1.865214e - 05	2.0668e - 05	1.051127e - 05	1.108077	1.966270	1.774490
	4	1.261365e - 05	1.541276e - 05	4.629849e - 06	1.221911	3.328998	2.724419
	5	9.145499e - 06	1.250770e - 05	2.469242e - 06	1.367635	5.065401	3.703768
	6	6.61361e - 06	1.047759e - 05	1.394760e - 06	1.584246	7.512110	4.741755
500	3	7.542962e - 06	8.347197e - 06	4.224176e - 06	1.106621	1.976053	1.785665
	4	5.027219e - 06	6.440567e - 06	1.884210e - 06	1.281139	3.418179	2.668078
	5	3.551602e - 06	4.918697e - 06	9.686214e - 07	1.384923	5.078039	3.666657
	6	2.65354e - 06	4.210024e - 06	5.679265e - 07	1.586569	7.412973	4.672330

Define

where p'_i is distributed according to $f(p)/P_i$ on D_i

$$g_i(p') = \begin{cases} g(p) & p \in D_i, \\ 0 & \text{o.w.,} \end{cases}$$

$$\implies \widehat{B} = 1 - \widehat{I}_2,$$

$$\begin{aligned} I_{2i} &= \int_{D_i} g(p) \cdot f(p) dp = \int_{D_i} P_i \cdot g(p) \cdot \frac{f(p)}{P_i} dp \quad (52) \\ &= P_i \int_{D_i} g(p') \cdot \frac{f(p)}{P_i} dp = P_i E[g_i(p')], \end{aligned}$$

$$\widehat{B} = 1 - \sum_{i=1}^m \frac{P_i}{N_i} \sum_{k_i=1}^{N_i} g(p'_{k_i}), \quad (56)$$

where

$$\int_{D_i} \frac{f(p)}{P_i} dp = 1. \quad (53)$$

$$\begin{aligned} \text{Var}(\widehat{I}_2) &= \sum_{i=1}^m \frac{P_i^2}{N_i} \cdot \text{Var}[g(p'_i)] \\ &= \sum_{i=1}^m \frac{P_i^2 \sigma_i^2}{N_i}, \end{aligned}$$

Random variable p' is distributed according to $f(p)/P_i$ on D_i . I_{2i} can be estimated by

$$\widehat{I}_{2i} = \frac{P_i}{N_i} \sum_{k_i=1}^{N_i} g(p'_{k_i}), \quad i = 1, \dots, m. \quad (54)$$

where

$$\sigma_i^2 = \text{Var}[g(p'_i)] = \frac{1}{P_i} \int_{D_i} g^2(p) \cdot f(p) dp - \frac{I_{2i}^2}{P_i^2}, \quad (57)$$

$$\text{Var}(\widehat{G}) = \text{Var}(1 - \widehat{I}_2) = \text{Var}(\widehat{I}_2) = \sum_{i=1}^m \frac{P_i^2 \sigma_i^2}{N_i}.$$

Therefore, integral I_2 is estimated by

$$\widehat{I}_2 = \sum_{i=1}^m \widehat{I}_{2i} = \sum_{i=1}^m \frac{P_i}{N_i} \sum_{k_i=1}^{N_i} g(p'_{k_i}), \quad (55)$$

In particular, if $P_i = 1/m$ and $N_i = N/m$, we get the systematic sampling estimate of Bonferroni index as

$$\widehat{B} = 1 - \frac{1}{N} \sum_{i=1}^m \sum_{k_i=1}^{N_i} g(p'_{k_i}). \quad (58)$$

TABLE 15: Relative efficiency of absolute Lorenz index for exponential ($\lambda = 3$) distribution.

r	m	MSE (RSS)	MSE (SRS)	MSE (SYS)	$E(1)$	$E(2)$	$E(3)$
2	3	0.0001772438	0.0002139371	9.98993e - 05	1.207022	2.141528	1.774225
	4	0.0001160937	0.001597166	4.811047e - 05	1.375756	3.319789	2.413065
	5	8.333168e - 05	0.0001291083	2.622109e - 05	1.549330	4.923834	3.178040
	6	6.31974e - 05	0.0001062572	1.622325e - 05	1.681354	6.549686	3.895483
5	3	7.01547e - 05	8.522752e - 05	3.934262e - 05	1.214851	2.166290	1.783173
	4	4.750068e - 05	6.440414e - 05	1.899583e - 05	1.355857	3.390436	2.500585
	5	3.310539e - 05	5.144982e - 05	1.046542e - 05	1.554122	4.916173	3.163312
	6	2.532269e - 05	4.341858e - 05	6.32867e - 06	1.714612	6.860617	4.001266
20	3	1.813884e - 05	2.141139e - 05	1.017992e - 05	1.180417	2.103296	1.781825
	4	1.161498e - 05	1.671353e - 05	4.712122e - 06	1.438963	3.546922	2.464915
	5	8.348938e - 06	1.282813e - 05	2.611685e - 06	1.536498	4.911821	3.196763
	6	6.300591e - 06	1.076669e - 05	1.601593e - 06	1.708839	6.722488	3.933953
100	3	3.576201e - 06	4.282342e - 06	1.949306e - 06	1.197455	2.196855	1.834602
	4	2.350533e - 06	3.236420e - 06	9.673106e - 07	1.376887	3.345792	2.429967
	5	1.667317e - 06	2.622885e - 06	5.363255e - 07	1.573117	4.890472	3.108778
	6	1.220823e - 06	2.203291e - 06	3.159623e - 07	1.804759	6.973272	3.863825
200	3	1.776957e - 06	2.109820e - 06	1.023774e - 06	1.187322	2.060826	1.735693
	4	1.175403e - 06	1.581642e - 06	4.78013e - 07	1.345617	3.308784	2.458935
	5	8.277293e - 07	1.306533e - 06	2.650969e - 07	1.578454	4.928511	3.122365
	6	6.383724e - 07	1.065240e - 06	1.576827e - 07	1.668682	6.755592	4.048462
500	3	7.048948e - 07	8.251703e - 07	4.060795e - 07	1.170629	2.032041	1.735854
	4	4.695757e - 07	6.461275e - 07	1.943934e - 07	1.375981	3.323814	2.415595
	5	3.350298e - 07	5.157602e - 07	1.047704e - 07	1.539446	4.922766	3.197752
	6	2.468869e - 07	4.345264e - 07	6.523285e - 08	1.760022	6.661159	3.784702

Algorithm for computing systematic random sampling estimate is given as follows.

- (i) Divide the range $[0, 1]$ of the cumulative distribution into m intervals each of width $1/m$.
- (ii) Generate $\{U_k, k = 1, \dots, N/m; i = 1, \dots, m\}$ from $U(0, 1)$.
- (iii) Compute $Y_{k_i} = (i - 1 + U_{k_i})/m; k_i = 1, \dots, N/m; i = 1, \dots, m$.
- (iv) Compute $p_{k_i} = F^{-1}(Y_{k_i})$.
- (v) Compute

$$\hat{B} = 1 - \frac{1}{N} \sum_{i=1}^m \sum_{k_i=1}^{N_i} g(p_{k_i}). \quad (59)$$

2.3.3. Absolute Lorenz Index. Absolute Lorenz index is defined as

$$\begin{aligned} A &= - \int_0^1 AL(p) dp \\ &= - I_3. \end{aligned} \quad (60)$$

Its stratified estimate is given as

$$\implies \hat{A} = -\hat{I}_3, \quad (61)$$

where

$$\begin{aligned} I_3 &= \int_0^1 AL(p) dp = \int_D AL(p) dp \\ &= \int_D \frac{AL(p)}{f(p)} \cdot f(p) dp = \int_D g(p) \cdot f(p) dp, \end{aligned} \quad (62)$$

where

$$g(p) = \frac{AL(p)}{f(p)}. \quad (63)$$

Let region D be divided into m disjoint subregions $D_i, i = 1, \dots, m$; that is, $D = \bigcup_{i=1}^m D_i$ and $D_k \cap D_j = \emptyset, k \neq j$ where \emptyset is null set.

Define

$$\begin{aligned} I_{3i} &= \int_{D_i} g(p) \cdot f(p) dp, \\ P_i &= \int_{D_i} f(p) dp, \quad \sum_{i=1}^m P_i = 1, \\ I_3 &= \int_D g(p) \cdot f(p) dp = \sum_{i=1}^m \int_{D_i} g(p) \cdot f(p) dp = \sum_{i=1}^m I_{3i}. \end{aligned} \quad (64)$$

Define

$$g_i(p') = \begin{cases} g(p) & p \in D_i, \\ 0 & \text{o.w.,} \end{cases}$$

$$I_{3i} = \int_{D_i} g(p) \cdot f(p) dp = \int_{D_i} P_i \cdot g(p) \cdot \frac{f(p)}{P_i} dp \quad (65)$$

$$= P_i \int_{D_i} g(p') \cdot \frac{f(p)}{P_i} dp = P_i E[g_i(p')],$$

where

$$\int_{D_i} \frac{f(p)}{P_i} dp = 1. \quad (66)$$

Random variable p' is distributed according to $f(p)/P_i$ on D_i .

I_{3i} can be estimated by

$$\hat{I}_{3i} = \frac{P_i}{N_i} \sum_{k_i=1}^{N_i} g(p'_{k_i}), \quad i = 1, \dots, m. \quad (67)$$

Therefore, integral I_3 is estimated by

$$\hat{I}_3 = \sum_{i=1}^m \hat{I}_{3i} = \sum_{i=1}^m \frac{P_i}{N_i} \sum_{k_i=1}^{N_i} g(p'_{k_i}), \quad (68)$$

where p'_i is distributed according to $f(p)/P_i$ on D_i

$$\implies \hat{A} = -\hat{I}_3,$$

$$\hat{A} = -\sum_{i=1}^m \frac{P_i}{N_i} \sum_{k_i=1}^{N_i} g(p'_{k_i}),$$

$$\text{Var}(\hat{I}_3) = \sum_{i=1}^m \frac{P_i^2}{N_i} \cdot \text{Var}[g(p'_i)] \quad (69)$$

$$= \sum_{i=1}^m \frac{P_i^2 \sigma_i^2}{N_i},$$

where

$$\sigma_i^2 = \text{Var}[g(p'_i)] = \frac{1}{P_i} \int_{D_i} g^2(p) \cdot f(p) dp - \frac{I_{3i}^2}{P_i^2}, \quad (70)$$

$$\text{Var}(\hat{A}) = \text{Var}(-\hat{I}_3) = \text{Var}(\hat{I}_3) = \sum_{i=1}^m \frac{P_i^2 \sigma_i^2}{N_i}.$$

In particular, if $P_i = 1/m$ and $N_i = N/m$, we get the systematic sampling estimate of Absolute Lorenz index as

$$\hat{A} = -\frac{1}{N} \sum_{i=1}^m \sum_{k_i=1}^{N_i} g(p'_{k_i}). \quad (71)$$

Algorithm for computing systematic random sampling estimate is given as follows.

- (i) Divide the range $[0, 1]$ of the cumulative distribution into m intervals each of width $1/m$.
- (ii) Generate $\{U_{k_i}, k_i = 1, \dots, N/m; i = 1, \dots, m\}$ from $U(0, 1)$.
- (iii) Compute $Y_{k_i} = (i - 1 + U_{k_i})/m$; $k_i = 1, \dots, N/m$; $i = 1, \dots, m$.
- (iv) Compute $p_{k_i} = F^{-1}(Y_{k_i})$.
- (v) Compute

$$\hat{A} = -\frac{1}{N} \sum_{i=1}^m \sum_{k_i=1}^{N_i} g(p_{k_i}). \quad (72)$$

Systematic random sampling estimates of Gini, Bonferroni, and Absolute Lorenz indices using Monte Carlo approach are given in Table 3.

3. Simulation Study

In this section, we illustrate the performance of the estimators for the three inequality indices based on the previously mentioned sampling procedures. 10000 random samples each of size 6, 8, 10, 12, 15, 20, 25, 30, 60, 80, 100, 120, 300, 400, 500, 600, 800, 1000, 1200, 1500, 2000, 2500, and 3000 were generated to compare the ranked set and systematic sampling estimates with simple random sampling estimates of the Gini index from exponential, Pareto ($\alpha = 3$), and Power ($\alpha = 3$); Bonferroni index from exponential, Pareto ($\alpha = 3$), Power ($\alpha = 3$), and Absolute Lorenz index from exponential ($\beta = 3$), Pareto ($\alpha = 3$), and rectangular ($\theta = 3$) distribution. The simulated ranked set samples with set size $m = 3, 4, 5, 6$ and number of repetitions, $r = 2, 5, 20, 200$, and 500 were chosen. The mean squared errors (MSE) and efficiencies were computed for the three procedures. Simulation results are shown in Tables 7, 8, 9, 10, 11, 12, 13, 14, and 15.

The expressions for Gini index, Bonferroni, and absolute Lorenz index are given in Tables 4, 5, and 6. The relation between Pareto, Power, and Rectangular distribution can be noted from the following remark.

Remark

- (1) If random variable $X \sim \text{Pow}(\alpha, \beta)$, that is, $f(x) = \alpha \beta^\alpha x^{\alpha-1}$, $0 \leq x \leq \beta$, then $Y = 1/X \sim \text{Par}(\alpha, 1/\beta)$; that is, $f(y) = (\alpha/\beta)(\beta/y)^{\alpha+1}$, $1/\beta \leq y \leq \infty$.
- (2) The Power function distribution $\text{Pow}(\alpha, \beta)$, with parameter $\alpha = 1$ and β , results in rectangular distribution defined on the interval $[0, 1/\beta]$.

For simple random sampling MSEs are defined as

$$\begin{aligned} \text{MSE}(\text{SRS}) &= \frac{\sum_{i=1}^{10000} (\widehat{G}_{\text{SRS}} - G)^2}{N}, \\ \text{MSE}(\text{SRS}) &= \frac{\sum_{i=1}^{10000} (\widehat{B}_{\text{SRS}} - B)^2}{N}, \\ \text{MSE}(\text{SRS}) &= \frac{\sum_{i=1}^{10000} (\widehat{A}_{\text{SRS}} - A)^2}{N}, \end{aligned} \quad (73)$$

where \widehat{G}_{SRS} , \widehat{B}_{SRS} , and \widehat{A}_{SRS} denote the simple random sample estimates of Gini, Bonferroni, and Absolute Lorenz indices (Table 1).

Similarly, for ranked set and systematic sampling the MSEs can be defined where \widehat{G}_{RSS} , \widehat{B}_{RSS} , and \widehat{A}_{RSS} denote the ranked set sample estimates and \widehat{G}_{SYS} , \widehat{B}_{SYS} , and \widehat{A}_{SYS} denote the systematic random sample (SYS) estimates of Gini, Bonferroni and Absolute Lorenz indices (Tables 2 and 3).

Gini index (G), Bonferroni index (B), and Absolute Lorenz index (A) for given distributions is given in the Tables 4–6.

Efficiencies are defined as

$$E(1) = \frac{\text{MSE}(\text{SRS})}{\text{MSE}(\text{RSS})}. \quad (74)$$

$E(1) > 1$ implies that Ranked set sample estimates are better than Simple random sample estimates.

Similarly,

$$E(2) = \frac{\text{MSE}(\text{SRS})}{\text{MSE}(\text{SYS})}. \quad (75)$$

$E(2) > 1$ implies that Systematic random sample (SYS) estimates are better than Simple random sample estimates.

$$E(3) = \frac{\text{MSE}(\text{RSS})}{\text{MSE}(\text{SYS})} \quad (76)$$

$E(3) > 1$ implies that Systematic random sample (SYS) estimates are better than Ranked set sample estimates.

The results have been presented in Tables 7–15 and it can be seen that the simple random estimates have much less efficiency as compared to their ranked set or systematic sample estimates. Systematic sample estimates are more efficient than the ranked set sample estimates. The same trend is prevalent through all the distributions irrespective of the inequality index being used. We also observe that relative efficiency increases as m increases irrespective of the sampling technique used.

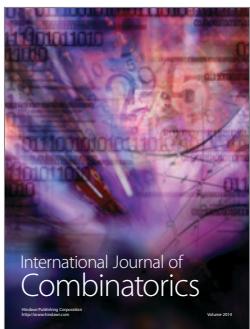
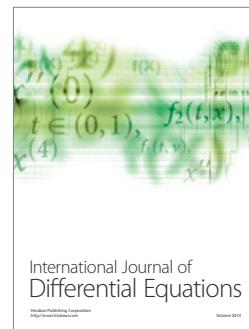
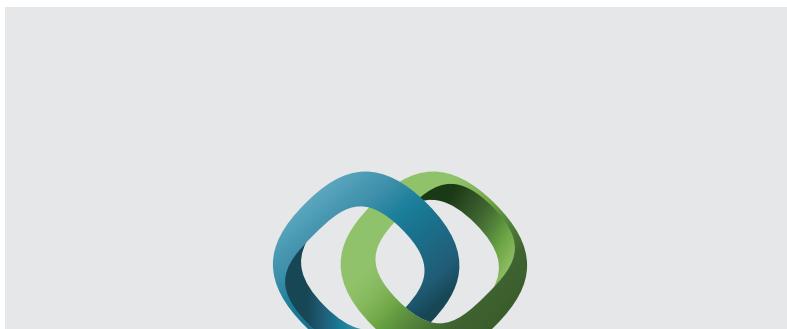
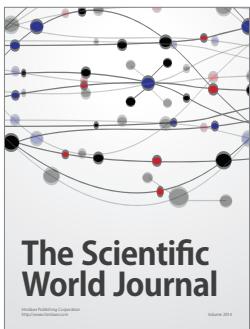
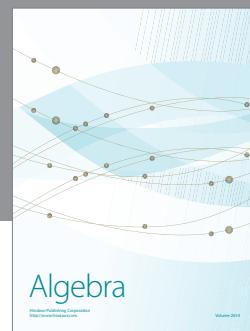
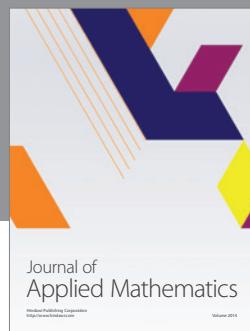
4. Conclusion

Ranked set sample estimates of inequality do improve the efficiency of all the inequality indices but compared to systematic sampling the ranked set sample estimates are less efficient. For example, in Table 7 while looking at the relative efficiencies for Gini index using different sample estimates we

see that the efficiency of RSS is almost two to three times more than SRS (ref. Col. E(1)) while efficiency of SYS is almost four to eight times more than RSS (ref. Col. E(3)). The efficiency of SYS versus SRS is noticeably high as is evident from col. E(2). The relative efficiency trend as given in the last three columns continues both for small as well as large sample sizes. One may conclude that SRS estimates though commonly used and simple to compute show less efficiency than their nonsimple random counterparts. One should use non-simple random estimate of inequality indices when the underlying data or practical situation so demands.

References

- [1] M. O. Lorenz, "Methods of measuring the concentration of wealth," *Publication of the American Statistical Association*, vol. 9, pp. 209–219, 1905.
- [2] C. Gini, "Measurement of inequality of incomes," *The Economic Journal*, vol. 31, pp. 124–126, 1921.
- [3] A. F. Shorrocks, "Ranking income distributions," *Economica*, vol. 50, pp. 3–17, 1983.
- [4] P. Moyes, "A new concept of Lorenz domination," *Economics Letters*, vol. 23, no. 2, pp. 203–207, 1987.
- [5] G. M. Giorgi, *Concentration Index, Bonferroni, Encyclopedia of Statistical Sciences*, vol. 2, John Wiley & Sons, New York, NY, USA, 1998.
- [6] S. Arora, K. Jain, and S. Pundir, "On cumulated mean income curve," *Model Assisted Statistics and Applications*, vol. 1, no. 2, pp. 107–114, 2006.
- [7] B. Zheng, "Testing Lorenz curves with non-simple random samples," *Econometrica*, vol. 70, no. 3, pp. 1235–1243, 2002.
- [8] G. A. McIntyre, "A method for unbiased selective sampling, using ranked sets," *Australian Journal of Agricultural Research*, vol. 2, pp. 385–390, 1952.
- [9] M. M. Al-Talib and A. D. Al-Nasser, "Estimation of Gini-index from continuous distribution based on ranked set sampling," *Electronic Journal of Applied Statistical Analysis*, vol. 1, pp. 33–41, 2008.
- [10] R. Rubinstein, *Simulation and the Monte Carlo Method*, John Wiley & Sons, New York, NY, USA, 1981.



Submit your manuscripts at
<http://www.hindawi.com>

