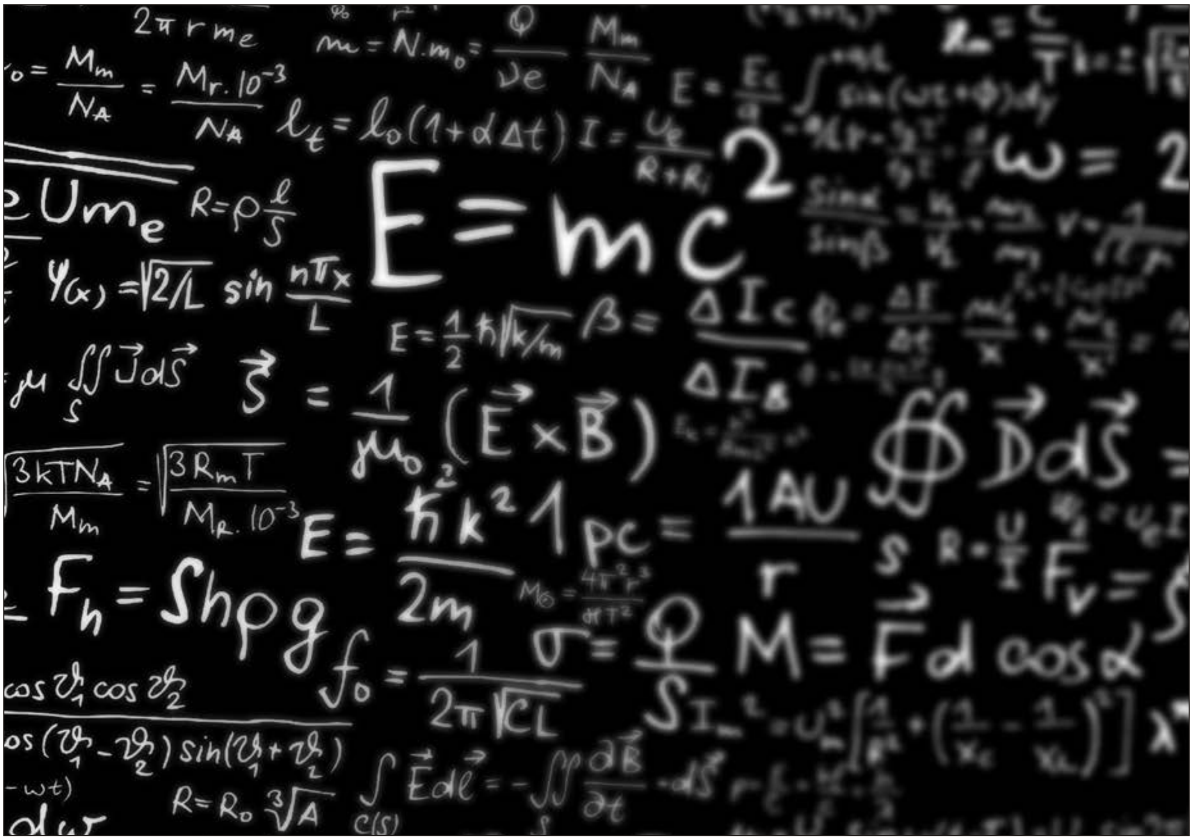

Asian Journal of Mathematics & Statistics

ISSN 1994-5418

science alert

ANSInet
an open access publisher
http://ansinet.com

Asian Journal of Mathematics and Statistics 3 (3): 130-138, 2010
ISSN 1994-5418

The Ranked Sample-Mean Monte Carlo Method for Unidimensional Integral Estimation

[1]A.D. Al-Nasser and [2]Mohammed Al-Talib
[1]Division of Area Planning and Residents Relations,
WRM, Emirate of Abu Dhabi, UAE
[2]Department of Mathematics, Southern Illinois University Carbondale, USA

Abstract: This study introduced the idea of using the novel ranked set sampling scheme for the Monte Carlo integral estimation problem. We proposed and discussed the unidimensional integral problem. It is demonstrated that this approach provides an unbiased and more efficient estimators than the traditional estimators based on simple random sampling. The method is illustrated by examples for estimating π and $\{f(x) = e^{-x^2}, 0 \le x \le 1\}$. An application to estimate the Gini index is proposed.

Key words: Ranked set sampling, gaussian integral, the number π, gini index

INTRODUCTION

A defined integral, such as I, which cannot be explicitly evaluated, can be obtained by a variety of numerical methods. Therefore, the importance of good Monte Carlo integration scheme is evident. Some of these methods were given by Rubinstein (1981) and Morgan (1984) for univariate integration problem. In this study our concern is in the sample mean Monte Carlo method for integral estimation. Consider the one dimensional integral:

$$I = \int_a^b g(x)dx \tag{1}$$

This integral can be represented as expected value of some random variable. Indeed, let us rewrite the integral as:

$$I = \int_a^b \frac{g(x)}{f(x)} f(x)dx$$

Assuming that f(x) is any pdf such that f(x) > 0 and a< x< b, when g (x)≠0; then,

$$I = E\left[\frac{g(x)}{f(x)}\right]$$

For simplicity, suppose that *X* is distributed uniformly over [a,b]; i.e., X~ U(a,b), then:

Corresponding Author: A.D. Al-Nasser, Division of Area Planning and Residents Relations, WRM, Emirate of Abu Dhabi, UAE

$$\frac{I}{b-a} = E[g(x)]$$

Therefore an unbiased estimator of *I* based on SRS is given by:

$$\hat{\theta} = (b-a)\frac{\sum_{i=1}^{n} g(x_i)}{n}$$

This is an unbiased estimator with variance:

$$Var(\hat{\theta}) = \frac{1}{n}\left((b-a)\int_a^b g^2(x)dx - I^2\right)$$

It is in any case interesting to see how simple random numbers may be used to evaluate deterministic integral. However, one can utilize the idea of Rank Set Sampling (RSS) of McIntyre (1952) for integrals approximation. The majority of research of RSS has been concerned with estimating the population mean. Few works in the literature were considering the RSS in Monte Carlo methods; Samawi (1999) used the random Beta sampler to evaluate non-stochastic integrals. In a similar fashion, Al-Saleh and Samawi (2000) investigated the use of the Steady State RSS for integrals approximation. It turned out that this procedure improve the efficiency of Monte Carlo methods much further. Later, Samawi and Al-Saleh (2007) used the importance sampling technique with RSS on the multiple integrals approximation.

In this study, we used the simulated RSS for univariate integral estimation based on the sample mean Monte Carlo method.

RANKED SET SAMPLING

The balanced RSS scheme involves of drawing m sets of SRS each of size m from a population and ranking each set with respect to the variable of interest. Then, from the first set the element with the smallest rank is chosen for the actual measurement. From the second set the element with the second smallest rank is chosen. The process is continued by keep selecting the i^{th} order statistics of the i^{th} random sample until the element with the largest rank from the m^{th} set is chosen. The scheme yields the following data:

$$\begin{array}{cccccc}
1 & X_{(1:m)} & X_{(2:m)} & X_{(3:m)} & \cdots & X_{(m:m)} \\
2 & X_{(1:m)} & X_{(2:m)} & X_{(3:m)} & \cdots & X_{(m:m)} \\
3 & X_{(1:m)} & X_{(2:m)} & X_{(3:m)} & \cdots & X_{(m:m)} \\
 & \vdots & \vdots & \vdots & \cdots & \vdots \\
m & X_{(1:m)} & X_{(2:m)} & X_{(3:m)} & \cdots & X_{(m:m)}
\end{array}$$

Hence, the selected RSS will be denoted by:

$$\{X_{[i:m]}, i = 1, 2, ..., m\}$$

where $X_{[I: m]}$ is the ith order statistics of the ith random sample of size m and it is denoted by the ith judgment order statistics. It can be noted that the selected elements are independent order statistics but not identically distributed.

In practice, the sample size m is kept small to ease the visual ranking, RSS literature suggested that m = 2, 3, 4, 5 or 6. Therefore, if a sample of larger size is needed, then the entire cycle may be repeated several times; say r times, to produce a RSS sample of size n = rm. Then the element of the desired sample will be in the form:

$$\begin{matrix} X_{[1:m]1} & X_{[1:m]2} & X_{[1:m]3} & \cdots & X_{[1:m]r} \\ X_{[2:m]1} & X_{[2:m]2} & X_{[2:m]3} & \cdots & X_{[2:m]r} \\ X_{[3:m]1} & X_{[3:m]2} & X_{[3:m]3} & \cdots & X_{[3:m]r} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ X_{[m:m]1} & X_{[m:m]2} & X_{[m:m]3} & \cdots & X_{[m:m]r} \end{matrix} \tag{2}$$

which can be represented as:

$$\{X_{[i:m]j}, i = 1, 2, ..., m, j = 1, 2, ..., r\}$$

where $X_{[I:m]j}$ is the ith judgment order statistics in the jth cycle, which is the ith order statistics of the ith random sample of size m in the jth cycle. It should be noted that all of $X_{[I:m]j}$'s are mutually independent, in addition, the $X_{[I:m]j}$ in the same row of (2) are identically distributed; More details can be found by Al-Nasser and Al-Rawwash (2007).

ONE DIMENSIONAL INTEGRAL USING RSS

In order to plan sample mean Monte Carlo RSS design for the problem in Eq. 1, n RSS should be selected. Then the integral estimation has the following steps:

Step1: Generate a RSS of size n= m x r from U(a, b)

$$\{U_{[i:m]j}, i = 1, 2, ..., m, j = 1, 2, ..., r\}$$

Step 2: Compute $X_{(i)j} = a + (b - a)\, U_{[i:m]j}$
Step 3: Compute $g(X_{(i)j})$
Step 4: Find the ranked sample-mean estimator

$$\hat{\theta}_{RSS} = (b-a)\frac{\sum_{j=1}^{m}\sum_{i=1}^{r} g(x_{(i)j})}{mr} \tag{3}$$

Lemma 1

$\hat{\theta}_{RSS}$ is unbiased estimator for I given in Eq. 1.

Proof

In order to proof this lemma, just take the expected value of Eq. 3; for both sides:

$$E(\hat{\theta}_{RSS}) = \frac{b-a}{mr}\sum_{j=1}^{m}\sum_{i=1}^{r} E\left(g(x_{(i)j})\right)$$

Since, RSS are independent order statistics; this expectation can be rewritten as:

$$E(\hat{\theta}_{RSS}) = \frac{b-a}{mr} m.\sum_{i=1}^{r} E\left(g(x_{(i)j}\right)$$
$$= \frac{b-a}{r}\sum_{i=1}^{r} E\left(g(x_{(i)}\right)$$
$$= (b-a)\int_a^b g(x)f(x)dx$$

But f (x) is U (a, b), which means:

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

and this complete the prove of the lemma.

EMPIRICAL STUDY

In this section we carry out some experiments to compare the efficiency of SRS and RSS, in three different areas, Mathematics, Statistics and Economics, by considering the problem of estimating Pi, Gaussian integral and the Gini index. For these comparisons, we generate 50,000 random samples, each of size n = m×r, where m takes the values 2, 6, 20 and 500; r takes the values 3, 4, 5 and 6. The simulated MSE, Bias and the EFF of the parameter were used as a criterion in the comparison, as follows:

$$\text{MSE}=\frac{\sum_{i=1}^{NOI}(\theta\text{-}\hat{\theta}_i)^2}{NOI},\ \text{Bias} = \frac{\sum_{i=1}^{NOI}(\theta\text{-}\hat{\theta}_i)}{NOI}$$

and the efficiency:

$$\text{EFF}=\frac{\text{MSE (SRS)}}{\text{MSE (RSS)}}$$

where θ is the exact value, $\hat{\theta}_i$ is the estimate of the parameter and NOI represent the number of generated sample (50,000).

The Number π

The constant π is an irrational number; that is, it cannot be written as the ratio of two integers. By using the equivalent of 96-sided polygons, Archimedes (287-212 BC) proved that $223/71<\pi < 22/7$. Taking the average of these values yields to 3.1419. However, π can be empirically estimated by drawing a large circle, then measuring its diameter and circumference and dividing the circumference by the diameter.

For any circle with radius r and diameter d = 2r, the circumference is πd and the area is πr^2. Further, π appears in formulas for areas and volumes of many other geometrical shapes based on circles, such as ellipses, spheres, cones and tori. Accordingly, ð appears in definite integrals that describe circumference, area or volume of shapes generated by circles. In the basic case, the area of a quadrant of a circle of radius unity is given by:

Table 1: Comparison between SRS and RSS in estimating π

m	r	Method	Mean	Bias	MSE	Efficiency
2	3	SRS	0.78527	-0.00043	0.00835	1.81697
		RSS	0.78523	-0.00047	0.00459	
	4	SRS	0.78518	-0.00054	0.00621	2.13453
		RSS	0.78527	-0.00043	0.00291	
	5	SRS	0.78573	0.00002	0.00496	2.50256
		RSS	0.78559	-0.00011	0.00198	
	6	SRS	0.78546	-0.00025	0.00417	2.92495
		RSS	0.78525	-0.00046	0.00142	
6	3	SRS	0.78539	-0.00031	0.00273	1.76121
		RSS	0.78526	-0.00044	0.00155	
	4	SRS	0.78515	-0.00055	0.00205	2.15098
		RSS	0.78540	-0.00031	0.00095	
	5	SRS	0.78541	-0.00029	0.00165	2.51765
		RSS	0.78540	-0.00031	0.00065	
	6	SRS	0.78516	-0.00055	0.00138	2.90075
		RSS	0.78535	-0.00035	0.00047	
20	3	SRS	0.78531	-0.00040	0.00083	1.78116
		RSS	0.78535	-0.00036	0.00046	
	4	SRS	0.78548	-0.00023	0.00061	2.13117
		RSS	0.78537	-0.00034	0.00028	
	5	SRS	0.78547	-0.00024	0.00049	2.51573
		RSS	0.78546	-0.00025	0.00019	
	6	SRS	0.78534	-0.00037	0.00041	2.92768
		RSS	0.78533	-0.00037	0.00014	
500	3	SRS	0.78543	-0.00028	0.00003	1.79253
		RSS	0.78538	-0.00033	0.00002	
	4	SRS	0.78538	-0.00033	0.00002	2.17451
		RSS	0.78543	-0.00027	0.00001	
	5	SRS	0.78540	-0.00031	0.00002	2.50562
		RSS	0.78540	-0.00031	0.00001	
	6	SRS	0.78538	-0.00033	0.00002	2.8517
		RSS	0.78538	-0.00032	0.00001	

$$\frac{\pi}{4} = \int_0^1 \sqrt{1 - x^2}\,dx \tag{4}$$

The results for estimating (Eq. 4) are shown in Table 1 showing the MSE, Bias as well as the EFF of the parameter estimates using SRS and RSS.

The Gaussian Integral

The Gaussian integral, or probability integral, is the improper integral of the Gaussian function over the entire real line. It is named after the German mathematician and physicist Carl Friedrich Gauss and the equation is:

$$\sqrt{\pi} = \int_{-\infty}^{\infty} e^{-x^2}\,dx \tag{5}$$

This integral has wide applications including normalization in probability theory and continuous Fourier transform. It also appears in the definition of the error function. The Gaussian integral can be solved analytically through the tools of calculus. That is, there is no elementary indefinite integral for

$$\int e^{-x^2}\,dx$$

but the definite integral given in Eq. 5 can be evaluated. The gaussian integral is also can be used to evaluate the exact value of π.

Table 2: Comparison between SRS and RSS in estimating normal probailities

m	r	Method	Mean	Bias	MSE	Efficiency
2	3	SRS	0.74719	-0.03851	0.00822	1.65648
		RSS	0.74678	-0.03892	0.00496	
	4	SRS	0.74723	-0.03848	0.00653	1.81814
		RSS	0.74673	-0.03898	0.00359	
	5	SRS	0.74685	-0.03885	0.00553	1.91966
		RSS	0.74703	-0.03867	0.00288	
	6	SRS	0.74680	-0.03890	0.00487	1.95495
		RSS	0.74696	-0.03875	0.00249	
6	3	SRS	0.74693	-0.03878	0.00374	1.41662
		RSS	0.74701	-0.03870	0.00264	
	4	SRS	0.74697	-0.03874	0.00217	1.17573
		RSS	0.74686	-0.03884	0.00185	
	5	SRS	0.74684	-0.03887	0.00285	1.44932
		RSS	0.74687	-0.03884	0.00197	
	6	SRS	0.74691	-0.03880	0.00262	1.42243
		RSS	0.74675	-0.03896	0.00184	
20	3	SRS	0.74705	-0.03866	0.00217	1.16970
		RSS	0.74683	-0.03887	0.00185	
	4	SRS	0.74669	-0.03901	0.00202	1.18514
		RSS	0.74691	-0.03879	0.00171	
	5	SRS	0.74678	-0.03893	0.00191	1.16450
		RSS	0.74687	-0.03884	0.00164	
	6	SRS	0.74682	-0.03889	0.00185	1.15031
		RSS	0.74685	-0.03885	0.00161	
500	3	SRS	0.74679	-0.03892	0.00154	1.01139
		RSS	0.74684	-0.03886	0.00152	
	4	SRS	0.74684	-0.03887	0.00153	1.00837
		RSS	0.74685	-0.03886	0.00152	
	5	SRS	0.74681	-0.03889	0.00152	1.00684
		RSS	0.74682	-0.03890	0.00151	
	6	SRS	0.74681	-0.03890	0.00152	1.00693
		RSS	0.74682	-0.03888	0.00151	

Therefore, to have different results; we shrinkage the integral interval to be evaluated on a finite limits; say [a,b]. such formulation allow us to evaluate the area under normal curve. Without loss of generality, we consider the integral limits to be [0,1]. Under the simulation assumptions, the results for evaluating the integral

$$\int_0^1 e^{-x^2} dx$$

By using both sampling schemes; are shown in Table 2 showing the MSE, Bias as well as the EFF of the parameter estimates using SRS and RSS.

The Gini-Index

Economists use a cumulative distribution called Lorenz Curve to measure the distribution of income among households in a given country. Typically, a Lorenz Curve (Fig. 1) is defined on [0, 1], continuous, increasing and concave up and passes through (0, 0) and (1, 1).

For example, the point (a, b) on the curve represents the fact that the bottom a% of the households receive less than or equal to b% of the total income. The Gini Index (coefficient of inequality), is the ratio of the area of the region between y = x and the Lorenz Curve to the area under y = x. The Gini index G for an income distribution of a certain country which is represented by the Lorenz curve, for example:

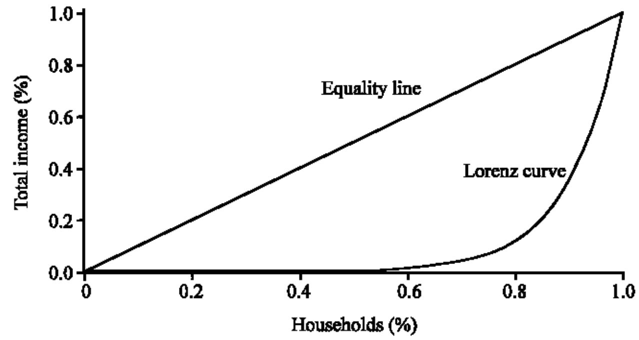

Fig. 1: The Lorenz curve

Table 3: Comparison between SRS and RSS in estimating gini index: p = 0.3

m	r	Method	Mean	Bias	MSE	Efficiency
2	3	SRS	0.17634	0.13434	0.01917	1.91670
		RSS	0.13986	0.09786	0.01000	
	4	SRS	0.18083	0.13883	0.02013	2.04137
		RSS	0.14004	0.09804	0.00986	
	5	SRS	0.18350	0.14150	0.02072	2.12170
		RSS	0.13996	0.09796	0.00976	
	6	SRS	0.18542	0.14342	0.02116	2.17311
		RSS	0.14005	0.09805	0.00973	
6	3	SRS	0.17636	0.13436	0.01842	1.89277
		RSS	0.13994	0.09794	0.00973	
	4	SRS	0.18102	0.13902	0.01961	2.02428
		RSS	0.14001	0.09801	0.00969	
	5	SRS	0.18356	0.14156	0.02027	2.10057
		RSS	0.13995	0.09795	0.00965	
	6	SRS	0.18535	0.14335	0.02074	2.15214
		RSS	0.13998	0.09798	0.00964	
20	3	SRS	0.17626	0.13425	0.01813	1.88100
		RSS	0.13998	0.09798	0.00964	
	4	SRS	0.18082	0.13882	0.01936	2.00833
		RSS	0.14005	0.09805	0.00963	
	5	SRS	0.18355	0.14156	0.02010	2.08988
		RSS	0.14000	0.09801	0.00962	
	6	SRS	0.18536	0.14336	0.02061	2.14286
		RSS	0.14001	0.09801	0.00961	
500	3	SRS	0.17630	0.13430	0.01804	1.87809
		RSS	0.14000	0.09800	0.00960	
	4	SRS	0.18082	0.13882	0.00002	2.00686
		RSS	0.14000	0.09800	0.00960	
	5	SRS	0.18354	0.14154	0.02003	2.08634
		RSS	0.13999	0.09799	0.00960	
	6	SRS	0.18536	0.14336	0.02055	2.14033
		RSS	0.13999	0.09800	0.00960	

$$L(x) = \frac{7}{12}x^2 + \frac{5}{12}x$$

depending on the fact that The Gini Index is

$$G = 2\int_0^1 [x - L(x)]dx.$$

The performance of the coefficient of inequality over the intervals [0, 0.3] is presented in Table 3, the results indicate that the estimator based on RSS is superior to the estimators

Table 4: Comparison between SRS and RSS in estimating Gini Index: p = 0.5

m	r	Method	Mean	Bias	MSE	Efficiency
2	3	SRS	0.19436	0.09713	0.01069	1.05666
		RSS	0.19444	0.09722	0.01011	
	4	SRS	0.19441	0.09718	0.01039	1.05490
		RSS	0.19446	0.09723	0.00985	
	5	SRS	0.19435	0.09713	0.01018	1.04767
		RSS	0.19441	0.09719	0.00971	
	6	SRS	0.19454	0.09732	0.01009	1.04721
		RSS	0.19443	0.09721	0.00964	
6	3	SRS	0.19459	0.09737	0.00990	1.02419
		RSS	0.19441	0.09719	0.00966	
	4	SRS	0.19443	0.09721	0.00976	1.01874
		RSS	0.19444	0.09722	0.00958	
	5	SRS	0.19445	0.09723	0.00970	1.01553
		RSS	0.19452	0.09730	0.00955	
	6	SRS	0.19439	0.09717	0.00965	1.01378
		RSS	0.19447	0.09725	0.00952	
20	3	SRS	0.19441	0.09718	0.00957	1.00629
		RSS	0.19440	0.09718	0.00951	
	4	SRS	0.19442	0.09720	0.00954	1.00571
		RSS	0.19443	0.09721	0.00948	
	5	SRS	0.19450	0.09728	0.00953	1.00688
		RSS	0.19442	0.09719	0.00947	
	6	SRS	0.19439	0.09717	0.00951	1.00322
		RSS	0.19446	0.09724	0.00948	
500	3	SRS	0.19445	0.09723	0.00946	1.00017
		RSS	0.19446	0.09723	0.00945	
	4	SRS	0.19445	0.09723	0.00946	1.00039
		RSS	0.19444	0.09722	0.00945	
	5	SRS	0.19445	0.09723	0.00946	1.00026
		RSS	0.19445	0.09723	0.00945	
	6	SRS	0.19444	0.09722	0.00945	1.00005
		RSS	0.19445	0.09723	0.00945	

Table 5: Comparison between SRS and RSS in estimating gini index: p = 1.0

m	r	Method	Mean	Bias	MSE	Efficiency
2	3	SRS	0.19479	0.00034	0.00124	1.10714
		RSS	0.19435	-0.00008	0.00112	
	4	SRS	0.19426	-0.00017	0.00094	1.24922
		RSS	0.19450	0.00005	0.00075	
	5	SRS	0.19438	-0.00006	0.00074	1.37719
		RSS	0.19451	0.00007	0.00054	
	6	SRS	0.19449	0.00005	0.00063	1.55606
		RSS	0.19439	-0.00005	0.00041	
6	3	SRS	0.19450	0.00006	0.00042	1.11369
		RSS	0.19442	-0.00002	0.00038	
	4	SRS	0.19458	0.00013	0.00031	1.25023
		RSS	0.19459	0.00014	0.00025	
	5	SRS	0.19433	-0.00011	0.00025	1.40257
		RSS	0.19444	-0.000004	0.00018	
	6	SRS	0.19448	0.00004	0.00021	1.58108
		RSS	0.19445	0.00001	0.00013	
20	3	SRS	0.19444	0.000004	0.00013	1.10382
		RSS	0.19447	0.00002	0.00011	
	4	SRS	0.19436	-0.00008	0.00009	1.23885
		RSS	0.19446	0.00002	0.00007	
	5	SRS	0.19448	0.00004	0.00008	1.40063
		RSS	0.19446	0.00002	0.00005	
	6	SRS	0.19449	0.00004	0.00006	1.55864
		RSS	0.19448	0.00004	0.00004	
500	3	SRS	0.19445	0.00001	0.000005	1.09834
		RSS	0.19445	0.00001	0.000004	
	4	SRS	0.19444	0.000003	0.000004	1.27188
		RSS	0.19445	0.00001	0.000003	
	5	SRS	0.19445	0.000006	0.000003	1.40112
		RSS	0.19444	-0.0000002	0.000002	
	6	SRS	0.19444	0.000002	0.000003	1.55179
		RSS	0.19444	0.0000008	0.000002	

based on SRS in estimating GINI index. Extending the area of coefficient of inequality to be in the interval [0, 0.5] we observed the same inference about the proposed technique as given in Table 4. Also, the similar were obtained in Table 5 as the inequality area is extended to be in the interval [0, 1].

CONCLUSION

In this study, we consider using simulated ranked set sampling for estimation the unidimensional integral. The method is illustrated by Monte Carlo experiments for estimating π and normal probabilities and the Gini index. All the simulation experiments indicated that estimators based on ranked set sample is more superior than the estimators based on simple random samples of the same size.

REFERENCES

Al-Nasser, A.D. and M. Al-Rawwash, 2007. A control chart based on ranked data. J. Applied Sci., 7: 1936-1941.

Al-Saleh, M.F. and H.M. Samawi, 2000. On the efficiency of monte carlo methods using steady state ranked simulated samples. Commun. Stat. Simulat. Comput., 29: 941-954.

McIntyre, G.A., 1952. A method of unbiased selective sampling using ranked sets. Aust. J. Agric. Res., 3: 385-390.

Morgan, B.J., 1984. Elements of Simulation. Chapman and Hall, London, ISBN: 0412-24590-6.

Rubenstein, R.Y., 1981. Simulation and the Monte Carlo Method. 1st Edn., John Wiley and Sons, New York.

Samawi, H.M., 1999. More efficient monte carlo methods obtained by using ranked set simulated samples. Commun. Stat. Simulat. Comput., 28: 699-713.

Samawi, H.M. and M.F. Al-Saleh, 2007. On the approximation of multiple integrals using multivariate ranked simulated sampling. Applied Math. Comput., 188: 345-352.