BEST: International Journal of Humanities, Arts, Medicine and Sciences (BEST: IJHAMS)
ISSN (P): 2348-0521, ISSN (E): 2454-4728
Vol. 3, Issue 12, Dec 2015, 227-238

USING ARTIFICIAL BEE COLONY ALGORITHMTO DETERMINE THE OPTIMAL STRATA BOUNDARIES

MOWAFAQ MUHAMMED AL-KASSAB & MUHAMMED ABDULJABBAR AL-HASAWY
Department of Statistics & Informatics, College of Mathematics and Computer Sciences, University of Mosul, Mosul, Iraq

ABSTRACT

Stratified random sampling is used when the researcher wants to highlight a specific subgroup within the population. This technique is useful in such researches because it ensures the presence of the key subgroup within the sample. A few numerical and computational techniques have been created for this reason.Some apply to highly skewed populations and some apply to any kind of populations. This paper proposes an ABC algorithm to solve the problem of stratum boundary while distributing the sample size according to Proportional Allocation method. The ABC algorithm is tested on two groups of populations and a comparative study with Genetic Algorithm (GA) of Keskintürk and Er (2007), Kozak’s (2004), Lavallée and Hidiroglou’s (1988) and Dalenius and Hodges(1959) methods have been implemented. The numerical results show the ability of the proposed algorithm to find the optimal stratified boundaries for a set of standard populations and various standard test functions compared with other algorithms.

KEYWORDS: Stratified random sampling, Artificial Bee Colony, Optimal Strata Boundaries, Proportional Allocation

1. INTRODUCTION

Stratified random sampling is a generally utilized sampling procedures particularly for heterogeneous population. Stratified examining is ideally utilized because of its ability of yielding so as to enhance measurable exactness a smaller variance of the estimator, compared with simple random sampling. With a specific end goal to decrease the variance of the estimator in stratified sampling the problems of stratum limit determination and sample allocation must be resolvedinitially.

A principal use of stratification, in order to obtain a better precision, is in defining what percentage of the sample must be taken from each stratum once we have chosen a non-uniform allocation scheme, that is, a non-trivial functional relation between the size of each stratum and the number of sample units to be collected in it. Thus, it is important to consider the allocation scheme itself in order to do a suitable stratification [6].

Several numerical and computational methods have been developed for obtaining the optimum boundaries in stratified sampling. Some apply to highly skewed populations and some apply to any kind of populations. An early and very simple method is the cumulative square root of the frequency method (cum√f) of Dalenius& Hodges in 1959 [8]. More recently Lavallée & Hidiroglou algorithm [16] and Gunning & Horgan's (2004) geometric method[9] have been proposed for highly skewed populations whereas Kozak's (2004) random search method [15] and Keskinturk&Er's (2007) genetic algorithm (GA) method [14] have been proposed for even non-skewed populations.

This study presents the ABC algorithm for the determination of stratum boundaries. In order to explore the efficiency of ABC algorithm, we compare its efficiency with GA, Kozak, LH and Delanius and Hodges methods

The rest of the paper is organized as follows. Section 2 describes stratified random sampling. In section 3, Background of ABC are summarized. ABC algorithm for optimal stratum boundaries is discussed in section 4. In order to test the efficiency of the proposed ABC, a comparative study with GA, Kozak, LH and Dalenius and Hodges methods is performed in Section 5. Conclusions are drawn in Section 6.

2. STRATIFIED RANDOM SAMPLING

There are a few alternative methods such as equal, proportional [7], and Neyman allocation [17]. The equal allocation method is the most straightforward technique where every stratum sample size is the same. With the proportional allocation method, the sample size in each stratum is proportional to the size of that stratum. These two methods are efficient and suitable if the variances within the stratum are similar.

In this paper, each character expresses the value as follows. *Y:*stratification variable; *N:*population size; *n:* sample size; *L:* number of strata; N_h: number of elements in stratum $h(h = 1,.. ., L)$;n_h: sample size in stratum *h; Y_h:* mean of elements in stratum *h;* $\bar{y}_{st}$ estimated mean in stratified sampling; $V(\bar{y}_{st})$:variance of the estimated mean in stratified sampling .

In stratified sampling [6], a population with N units is divided into L groups with N_1,N_2, ...,N_i, ...,N_L units respectively. These groups are called strata. There is no overlap among them and together they exhaust the population. Thus, we have

$$N_1 + N_2 + ... + N_h + ... + N_L = N \tag{1}$$

After the strata definition, which is based on characteristics of the population, sampling units are selected in each stratum, independently, according to specific criteria of selection. The sample sizes of the strata are denoted by n_1, n_2, ...,n_h, ..., n_L, respectively. The size of the sample taken from the population and symbolized by the n. Thus

$$\sum_{h=1}^{L} n_h = n \tag{2}$$

The population mean of the stratum h, denoted byμ_h.

$$\mu_h = \frac{1}{N_h}\sum_{i=1}^{N_h} Y_{hi} \tag{3}$$

The mean of the sample taken from the stratum h, denoted by $\bar{y}_h$.

$$\bar{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi} \ldots\ldots \quad (4)$$

The population variance of the stratum h, denoted by $\sigma^2{}_h$.

$$\sigma^2{}_h = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(Y_{hi} - \mu_h)^2 \tag{5}$$

The variance of the sample taken from stratum h, denoted by $S^2{}_h$.

$$S^2{}_h = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}(y_{hi} - \bar{y}_h)^2 \tag{6}$$

The weight of stratum h denoted by W_h is:

$$W_h = \frac{N_h}{N}. \tag{7}$$

It also can be obtained the population mean denoted by μ,by multiplyingmean ofstratum h by weight of stratum h:

$$\mu = \sum_{h=1}^{L} W_h \mu_h \quad (8)$$

If we multiply the mean of the sample for stratum h by its weight, we getthe stratified sample mean denoted by$\bar{y}_{st}$:

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h \quad (9)$$

Moreover, the variance of the stratified sample mean is:

$$V(\bar{y}_{st}) = \sum_{h=1}^{L} W^2{}_h \frac{\sigma^2{}_h}{n_h} \quad (10)$$

When the total sample size n is allocated using Proportional allocation. The variance of the stratified mean becomes:

$$V_{Prop(\bar{y}_{st})} = \frac{1}{n}\left(\sum_{h=1}^{L} W_h {\sigma_h}^2\right) \ldots \quad (11)$$

3. OVERVIEW OF THE ARTIFICIAL BEE COLONY ALGORITHM

By simulating the forging behavior of bee colonies, artificial bee colony (ABC) algorithm, which is a swarm intelligence-based optimization algorithm, was proposed by Karaboga in (2005) for numerical function optimization [10]. The main steps of ABC algorithm can be described as follows.

Initialization

Repeat

Employed bee stage: Place the employed bees on the food sources in the memory.

Onlooker bee stage: Place the onlooker bees on the food sources in the memory.

Scout bee stage: Send the scout bees to the search area for discovering new food sources.

Until (conditions are satisfied)

In this algorithm, the colony comprises of three sorts of honey bees: employed bees, onlooker bees and scout bees. Half of the colony is employed bees, and the other half is onlooker bees. The employed bees investigate the food source and send the information of the food source to the onlooker bees. The onlooker bees pick a food source to exploit based on the information shared by the employed bees. The scout bee, which is one of the employed bees whose food source are deserted, finds a new food source randomly. The position of a food source is a possible solution to the optimization problem. Denote the food source number as *SN*, the position of the i^{th} food source as x_i(i=1,…, *SN*), which is a *D* dimensional vector [11,12].

In ABC algorithm, the i^{th} fitness value *i fit* for a minimization problem is defined as [13]:

$$\text{fit}_i = \begin{cases} 1/(1+f_i) & if f_i \geq 0 \\ 1+abs(f_i) & if f_i < 0 \end{cases} \ldots\ldots \quad (12)$$

Where$f_i$is the cost value of the i^{th}solution. The probability that food source being selected by an onlooker bee is given by:

$$p_i = \frac{fit_i}{\sum_{i=1}^{SN} fit_i} \cdots \tag{13}$$

A candidate solution from the old one can be generated as:

$$v_{ij} = x_{ij} + \phi_{ij}\left(x_{ij} - x_{kj}\right). \tag{14}$$

Where $k \in \{1,2,\dots,SN\}$, $k \neq i \cdots$ and $j \in \{1,2,\dots,D\}$ are randomly selected indices, ϕ_{ij} ∈[-1, 1] is a uniformly distributed random number. The candidate solution is compared with the old one, and the better one should be remained [11].

If the abandoned food source is x_i, the scout bee exploits a new food source according to:

$$x_{ij} = x_{min,j} + rand(0,1)\left(x_{max,j} - x_{min,j}\right) \tag{15}$$

Where $x_{max,j}$ and $x_{min,j}$ are the upper and lower bounds of the j^{th} dimension of the problem's search space[12].

4. ABC ALGORITHM FOR STRATIFIED SAMPLING

In this section, we describe the ABC algorithm for determining stratum boundaries in stratified sampling.

4.1Input Information

For stratum boundary determination using Proportional allocation, the software implemented takes into consideration the following parameters:

- Number of strata L.
- Population data D that represents the study population, or population function f(x) in the period [0,1].

4.2 Fitness Function

Fitness function is a critical factor in the ABC Algorithm. Every Source of food in the ABC's population has a fitness value, and it moves in solution space with respect to its previous position where it has met the best fitness value. In this paper, the fitness value is the variance of Proportional allocation in stratified sampling denoted as Eq. (11) that must be minimized through the iteration process.

4.3Food source Structure

The composition and shape of the particle in stratified sampling differs in the way of representation from the most representations found in the literature, which represented by a single vector structure. The range of ascending values subject to stratification must be divided into L parts by points $Y_1 < Y_2 < \cdots < Y_{L-1}$. Each such part corresponds to a stratum boundary. The length of particle equal to the number of the strata L. The first gene in particle refers to the sequence of last observation in the first stratum, so it refers to the size of first stratum N_1.The second gene refers to the sequence of the last observation in second stratum. The difference between the values of the first gene and the second gene refers to the size of second stratum N_2 and so on. Therefore, the gene value refers to the stratified boundaries for each stratum, and the difference between gene and previous gene refers to the size of stratum. Figure 1 illustrate the food source representation of six strata boundaries for a population contains 30 observations.

4	8	14	18	27	30

Figure 1: Food Source Representation for Six Strata Observations

4.4 Initial Population Creation

Population initialization is a crucial step in swarm intelligence algorithms, because it can affect the quality of the solutions and the convergence speed. Therefore, the size of the population (number of food sources) and the way the initial population is created have a significant influence in the performance of the algorithm and to the quality of the result. Since each food source must contain a number of genes equal to the number of strata L, The last gene must have a value of "N" because it represents the upper limit of the last stratum. In general, the ideal situation would be to have the greatest possible diversity of food sources through the search space.

4.5 Search Mechanism

It is well known that both the exploration and exploitation abilities are necessary for the population based algorithms. How to balance these two abilities to achieve good optimization performance is very important.

In ABC algorithm, the employed bee stage represents the exploration ability of the algorithm, and the onlooker bee stage represents the exploitation ability of the algorithm. The search equation proposed in ABC algorithmis good at exploration but poor at exploitation, so that it will affect the convergence speed of the algorithm.

Inspired of PSO [5], in order to improve the exploitation ability of ABC algorithm, take the advantages of the searchequation in PSO, the global best solution will be considered in the new search equation in the onlooker beestage. The modified search equation in onlooker bee stageis described as follows:

$$v_{ij} = x_{ij} + \phi_{ij}\left(x_{ij} - x_{kj}\right) + \vartheta_{ij}\left(y_j - x_{ij}\right) \ldots (16)$$

Where$k$ ∈{1,2,..., SN} is a random selected index which is different from i,$j \in \{1,2,\ldots,D\}$ is a randomselected index,$y_j$is the j^{th} element of the global best solution, $\phi_{ij} = (-1,1), \vartheta_{ij} \in (0,1.5)$, are both uniformlydistributed random numbers.

Differential evolution (DE) [4] is a population based algorithm to function optimization, whose main strategy isto generate a new position for an individual by calculating vector differences between other randomly selected members in the population. “DE/current-to-rand/1” is a variant DE mutation strategy, which can effectivelymaintain population diversity according to randomness ofthe search equation. Motivated by “mutation strategy of Deferential Evolution and based on the property of ABCalgorithm, a new search equation in employed bee stage isproposed as follows:

$$v_{ij} = x_{ij} + \phi_{ij}\left(x_{ij} - x_{kj}\right) + c_{ij}\left(x_{r1j} - x_{r2,j}\right) \ldots \qquad (17)$$

Where$\phi_{ij}, c_{ij} \in (-1,1)andi \in \{1,2,\ldots,SN\}, j \in \{1,2,\ldots,D\}, r1, r2 \in \{1,2,\ldots,SN\} and r1, r2 \neq i$, $\phi_{ij}, c_{ij}$are both negativeor both positive, which can keep the search direction thesame.

In general, inspired by DE and PSO, the new searchequation and search mechanism are proposed to balancethe exploration ability and exploitation ability in ABCalgorithm. In the employed bee stage, search equation (17)is used to keep the exploration ability of ABC algorithm;while, in the onlooker bee stage, search equation (16) isemployed to increase the exploitation ability of thealgorithm.

Based on the above analysis, the main steps of the proposed artificial bee colony are as follows.

Algorithm: Hybrid Artificial bee Colony Algorithm

Initialize the food sources by using the initialization method in subsection 4.4, and evaluate thepopulation, $trail_i$= 0, (i=1, 2,..,SN). $Cycle$ =1.

Repeat

Step 1: Find the new food source for employed beeaccording to (17) and evaluate its quality.

Step 2: Use a greedy selection process and select thebetter solution between the new food source and the oldone.

Step 3: If solution does not improve $trail_i$= $trail_i$+1,otherwise$trail_i$=0.

Step 4: Calculate the probability according to (13) andapply roulette wheel selection scheme to choose a foodsource for onlooker bees.

Step 5: Find the new food source for onlooker bees according to (16) and evaluate its quality.

Step 6: Use a greedy selection process and select thebetter solution between the new food source and the oldone.

Step 7: If solution does not improve $trail_i$= $trail_i$+1, else$trail_i$=0.

Step 8: If max($trail_i$) >limit, replace this solution with a new one produced by the initializationmethod in subsection 4.4.Memorize the best solution achieved so far.

$Cycle$ = $Cycle$ +1

Until ($Cycle$ = Maximum Cycle Number)

5. NUMERICAL EXPERIMENTS

The ABC experiments for the stratification sampling has been performed on two groups of populations: data and functions, to find optimal strata boundaries based on variance of Proportional allocation. All experiments are implemented using Matlab 8.2.0 (R2013b).

5.1 Testing ABC Algorithm to find the Stratified Boundaries for Populations of Data

Some populations are used for stratification with different skewness, kurtosis, mean, standard deviation and size properties. Those populations that are available in the R stratification[18] and GA4Stratification[19] packages are used for stratification. Each of the populations are divided into 3, 4, 5 and 6 strata. The total sample size is 100 and the boundaries are obtained with Kozak and GA methods with random initial boundaries. These populations are:

Pop1: An accounting population of debtors in an Irish firm (Debtors).

Pop2: Number of municipal employees of 284 municipalities in Sweden in 1984 (ME84).

Pop3: Simulated Data from the Monthly Retail Trade Survey of Statistics Canada (MRTS).

Pop4: Population in thousands of 284 municipalities in Sweden in 1975 (P75).

Pop5: Real estate values in millions of kronor according to 1984 assessment of 284 municipalities in Sweden in 1984 (REV84).

Pop6: The resources in millions of dollars of large commercial US banks (USbanks).

Pop7: The population in thousands of US cities in 1940 (UScities).

Pop8: The number of students in four-year US colleges in 1952-1953 (UScolleges).

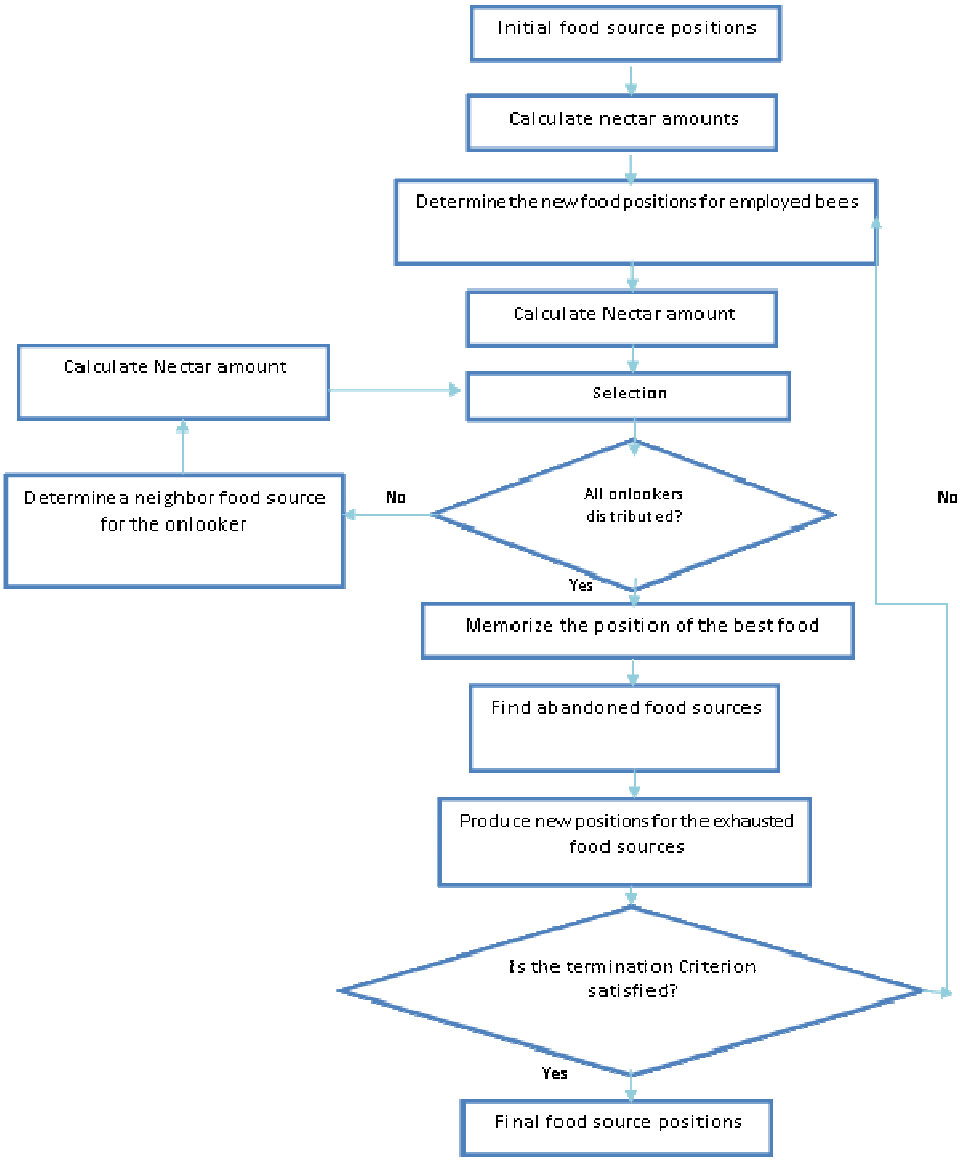

Figure 2: Artificial Bee Colony Algorithm Flowchart

Table 1: Summary Statistics of the Populations

Pop	Name	N	Range	Skewness	Kurtosis	Mean	StdDev.
Pop1	Debtors	3369	27960	6.44	59.00	838.64	1873.99
Pop2	ME84	284	46901	8.64	84.04	1779.07	4253.13
Pop3	MRTS	2000	486225	8.61	136.20	16882.8	21574.88
Pop4	P75	284	667	8.43	88.56	28.81	52.87
Pop5	REV84	284	59530	7.83	81.33	3088.09	4746.16
Pop6	Usbanks	357	907	2.07	4.06	225.62	190.46
Pop7	Uscities	1038	188	2.87	9.12	32.57	30.4
Pop8	Uscolleges	677	9423	2.45	5.80	1563	1799.06

In order to compare the efficiency of four methods (ABC algorithm, GA, LH and Kozak), the variance of the estimator given in Eq. (12) is calculated. We implement our proposed algorithm using MATLAB programming language on a PC (CPU 2.27 GHz,3GB RAM). ABC parameter settings used for stratifying these examples are shown in Table 2. Table 3 summarizes the variances of the estimators obtained with ABC, GA, Kozak's and LH methods. The computational time of ABC and GA algorithms are also shown in Table 3.

Table 2: ABC Parameter

ABC parameters	H =2,3	H=5,6
Colony size	100	100
Max iteration	500	600
Limit	30	35

Whereas the sample sizes given in Table 4 in Appendix. The results obtained by the ABC algorithm are better than the ones observed by using GA, LH and Kozak's methods.

5.2 Testing ABC Algorithm to Find the Stratified Boundaries for some Probability Density Functions

The proposed ABC is tested using three probability density functions. Table 5 shows the details of thesefunctions.Table 6 list the comparison results of the ABC algorithm, GA andDelanius and Hodges [8] methods for the threeprobability density functionsusing four different strata.

Table 3: Variances of the Estimators for Stratification Examples Obtained with ABC, GA,Kozak and LH Methods

H	LH	Kozak	GA	Time	ABC	Time
Pop1 : Debtors						
3	7810.8	10155	6703.7	73	**6616.4**	**20**
4	5598.5	7293.6	4135.8	73	**3902.1**	**20**
5	3837.4	4892.6	2610.8	88	**2460.6**	**29**
6	3725.2	4550	1868.5	92	**1808.4**	**31**
3	144310	144310	19525	16	**19525**	**10**
4	144530	144110	14934	17	**14760**	**10**
5	145310	145310	12488	21	**12275**	**13**
6	145280	145680	11394	21	**11351**	**13**
3	1469500	1991000	1397000	43	**1395392**	**14**
4	1181600	1229000	909640	44	**906217**	**23**
5	1094400	1133600	621460	53	**615481**	**19**
6	1096500	1094400	510150	54	**500851**	**28**
3	20.747	20.287	4.476515	16	**4.47555**	**10**
4	20.689	20.213	3.270845	17	**3.26949**	**10**
5	20.656	20.856	2.822907	21	**2.8209**	**12**
6	20.267	20.915	2.672325	21	**2.6502**	**13**
3	155080	152900	35702	16	**35701.8**	**9**
4	154510	155070	25989	17	**25989**	**10**
5	154210	154180	21600	21	**21532**	**13**
6	144820	154060	19885	21	**19877**	**12**
3	71.382	72.54	39.302	17	**39.302**	**9**
4	71.276	68.339	20.955	17	**20.955**	**10**
5	70.317	62.686	14.378	22	**14.375**	**12**

Table 3 – Cond.,

6	62.233	62.233	11.413	22	**10.293**	**13**
3	1.24158	1.757414	1.114905	25	**1.11427**	**11**
4	0.88556	1.506219	0.650937	26	**0.65046**	**21**
5	0.85368	1.463	0.424536	32	**0.4226**	**28**
6	0.8146	0.447133	0.305801	32	**0.3039**	**29**
3	3670.4	4225.9	3573.3	22	**3570.3**	**20**
4	3216.2	3686.6	2015.5	20	**2000.1**	**17**
5	3344.3	3506.7	1334.2	25	**1315.4**	**16**
6	3454.5	3147.9	953.98	25	**940.5**	**22**

Table 5: Probability Density Functions (f1-f3)

Function	Range
$f_1(x) = xe^{-x}$	$0 \le x < \infty$
$f_2(x) = e^{-x}$	$0 \le x < \infty$
$f_3(x) = 2(1-x)$	$0 \le x \le 1$

Table 6: Comparison Results off1, f2 and f3

F		D & H		GA		ABC	
		Vprop ($\bar{y}_{st}$)	strataboundaries	Vprop ($\bar{y}_{st}$)	strataboundaries	Vprop ($\bar{y}_{st}$)	strataboundaries
f1	2	0.3667	1.27	0.34913440	1.5911	0.34913440	1.5911
	3	0.1958	0.73	0.17720829	1.0152	0.17720827	1.0152
			2.04		2.6027		2.6028
	4	0.1227	0.52	0.10691183	0.7531	0.10689992	0.7519
			1.27		1.7674		1.7655
			2.61		3.3508		3.3475
	5	0.0858	0.39	0.07254241	0.6202	0.07138297	0.5953
			0.92		1.3847		1.3450
			1.68		2.4173		2.3508
			3.02		4.0114		3.9103
f2	2	0.6950	2.36	0.68632146	2.5717	0.68632146	2.5717
	3	0.3658	1.54	0.34791562	1.7886	0.34791561	1.7885
			3.26		3.7616		3.7613
	4	0.2257	1.20	0.20943699	1.4125	0.20943132	1.4063
			2.27		2.7276		2.7236
			3.94		4.5874		4.5843
	5	0.1519	1.01	0.13958637	1.1944	0.13952445	1.1732
			1.82		2.2140		2.1850
			2.86		3.4510		3.4109
			4.49		5.2345		5.1993
f3	2	0.0157	0.35	0.01548003	0.3820	0.01548003	0.3820
	3	0.0073	0.23	0.00716084	0.2513	0.00716084	0.2514
			0.50		0.5372		0.5373
	4	0.0042	0.18	0.00411284	0.1882	0.00411281	0.1879
			0.37		0.3927		0.3920
			0.62		0.6247		0.6242
	5	0.0030	0.12	0.00266640	0.1499	0.00266630	0.1502
			0.25		0.3095		0.3086
			0.40		0.4828		0.4836
			0.64		0.6802		0.6804

6. CONCLUSIONS

Stratified sampling is a sampling methodology used for heterogeneous populations as a part of request to acquire accuracy than different methods of sampling. This paper proposes an ABC algorithm for finding the optimal stratified boundaries with Proportional allocation and its performance, is evaluated using different test problems. The numerical results show the efficiency and capabilities of ABC algorithm in finding the Optimal Strata Boundaries. Amazingly, its performance better than other methods such as Kozak, GA, LH and Delanius and Hodges methods. This confirms that ABC can be efficiently utilized in the stratification of heterogeneous populations. Future research might use ABC algorithm where factors such as sample cost, the number of strata, and the sample size vary.

REFERENCES

1. Al-Saffawi. S.Y.(1976), Applying Statistical Techniques Analysis and Estimating a certain Agricultural Production ", M.SC. Thesis, University of Baghdad, Iraq.
2. Al-KassabMMT and Al-Hasow A., (1996) "*Method to find Stratum Boundaries Using Neyman Allocation*" Master Thesis, University of Mosul, Iraq.
3. Al-Kassab, MMT & Al-Taay, H, (1994)"*Approximately Optimal Stratification Using Neyman Allocation* ", J, of Tanmiat Al-Rafidain.
4. Bäck, T. (1996), "Evolutionary algorithms in theory and practice". New York: Oxford Univ. Press.
5. Bogdan Korel, (1990)," *Software Test Data Generation*", IEEE, Computer Society and Association for Computing Machinery.
6. Cochran, W. G. 1977" *Sampling Techniques, 3rd ed*." John Wiley & Sons, Inc. USA.
7. Daghistani, THN, (1995) " *An Approximately Optimal Stratification Using Proportional Allocation* " Master Thesis, University of Mosul, Iraq.
8. Dalenius,T.,Hodges, J.L.Jr. (1959)." *Minimum Variance Stratification*", Journal of the American StatisticalAssociation, 54, 285, pp.88-101.
9. Gunning, P., Horgan, J.M. (2004), "A New Algorithm for the Construction of Stratum Boundaries in Skewed Populations". Survey Methodology, 30,2.
10. Karboga, D. and Basturg B., (2007), "*A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm"*, J. Glob. Optim. Vol. 39, pp.459-471.
11. Karaboga, D., Basturk, B. (2008). *"On the performance of artificial bee colony (ABC) algorithm"*, Applied Soft Computing, Vol. 8, pp. 687-697.
12. Kaur, A., Goyal, S., (2011), *"A survey on the Applications of bee colony optimization Techniques"*, International Journal on Computer Science and Engineering (IJCSE), ISSN: 0975-3397 Vol. 3 No. 8.
13. Khaze, S., Maleki, I., Hojjatkhah, S. and Bagherinia, (2013). "*Evaluation the efficiency of Artificial Bee Colony and the Firefly Algorithm in Solving The Continuous Optimization Problem* ", International journal on Computational Science and Applications, Vol. 3, No. 4, pp.23-35.

14. Keskintürk, T., Er, Ş., (2007) "*A Genetic Algorithm Approach to Determine Stratum Boundaries and Sample Sizes of Each Stratum in Stratified Sampling* ". Computational Statistics &Data Analysis, 52, 1, pp.53-67.
15. Kozak, M., (2004),*"Optimal Stratification Using Random Search Method in Agricultural Surveys"*. Statistics in Transition, 6, 5, pp.797-806, (2004).
16. Lavallée, P., Hidiroglou, M., (1988),"On the Stratification of Skewed Populations", Survey Methodology, 14, 1, pp.33-43, (1988).
17. Neyman, Jerzy. (1934), “On the Two Different Aspects of the Representative Methods:The Method of Stratified Sampling and the Methodof Purposive Selection”, Journal of the Royal Statistical Society, 97 (4), pp.558-625.
18. R: stratification. http://CRAN.R-project.org/package=stratification
19. R:GA4Stratification.http://CRAN.Rproject.org/package=GA4stratification

Table 4: Size of the Strata (N_h) obtained from ABC, GA and Kozak’s Methods

HKozak GAABC				
		Pop1 : Debtors		
3	Nh	2673561135	2690545134	312122226
4	Nh	207191430381	208590130281	29852957316
5	Nh	189295433513949	189295533913647	2847361119348
6	Nh	153390549326512647	160495642622111844	241459322896308
		Pop2 : ME84		
3	Nh	1457861	1457861	254282
4	Nh	115644461	115644461	2473142
5	Nh	5469564164	5469564164	183732332
6	Nh	546133343765	546956411945	17668241222
		Pop3 : MRTS		
3	Nh	1204688108	1227671102	176721320
4	Nh	101774830332	102374220332	1682290262
5	Nh	77467536910532	74969837115032	1191654134192
6	Nh	51358045828113632	52157345528313632	108469016546132
		Pop4 : P75		
3	Nh	1507757	1507757	250322
4	Nh	111734357	111734357	18077243
5	Nh	6468523466	12361331948	1558134113
6	Nh	456639343367	458752331849	111735228173
		Pop5 : REV84		
3	Nh	1388165	1388165	215672
4	Nh	64816970	64816970	15888362
5	Nh	6169513469	6474533954	1458737133
6	Nh	575137422869	616042432652	130764023132
		Pop6 : Usbanks		
3	Nh	2128461	2128461	2587524
4	Nh	1111127361	1111127361	212847318
5	Nh	110101543260	110101543260	111112744218
6	Nh	516397543260	546890533260	11010154363818
		Pop7 : Uscities		
3	Nh	74919396	74919396	79519251
4	Nh	43435615494	43440915540	43441215339
5	Nh	22627129814994	3933671508939	3933821359137
6	Nh	2262712851288939	2742632451288939	2262712851289137
		Pop8 : Uscolleges		
3	Nh	47813069	47813069	48113561
4	Nh	25623411869	25623411869	27223111361
5	Nh	19216614510569	253221826061	2722251113435
6	Nh	133179166775369	132180166785269	25521981533435