Computation of Parameters Using Genetic Algorithm and Sequential Quadratic Programming in Sampling

Nursel Koyuncu

Abstract—**The sampling literature contains many examples of estimators of population parameter. To deal with this problem many authors have suggested family of estimators of population parameter. But in the case of generalization of these estimators, estimation of optimum values is a problem. Some authors can define estimator replacing the unknown parameters by their sample estimates. To get the optimum estimator, one need to solve complex mean square error equation with many parameters and nonlinear constraints. In this study we have tried to get these optimum parameter in stratified random sampling using genetic algorithms and sequential quadratic programming. A numerical example is also done to compare these algorithms. The results show that genetic algorithm is more efficient than sequential quadratic programming to solve the complex model with more parameters under non-linear constraints.**

Index Terms—**Efficiency, genetic algorithm, mean square error, stratified random sampling.**

I. INTRODUCTION

Genetic Algorithms (GAs) are stochastic optimization methods based on concepts of natural selection and genetics ([1], [2]). They work with a population of individuals, each representing a possible solution to a given problem. GAs typically works by iteratively generating and evaluating individuals using an evaluation function. The simplest of GAs work according to scheme given by [3]. Evolutionary Process consists following steps: 1. Create initial population (k=0), 2. Evaluate fitness, 3. Selection, 4. Crossover, 5. Mutation, 6. Select next generation, 7. Check stopping criteria (if yes: go next step; if no: k=k+1), 8. Get the best solution.

A. Selection Mechanism

Selection is the process which guides the evolutionary algorithm to the optimal solution by preferring choromosomes with high fitness. The choromosomes evolve trough successive iterations, called generations. Each generation, the choromosemes are evaluated using same measure of fitness.

B. Crossover

Crossover is the main genetic operator. It provides the main search operator while bit mutation simply serves as a background operator to ensure that all possible solutions can enter the population.

C. Mutation

Mutation is a background operator which produces some random changes in various choromosomes. In GAs, mutation play the role of replacing the genes lost during the evalutionary process in a new form or produce new genes which were nor explored before in the initial population. One way to do mutation would be to alter one or more genes.
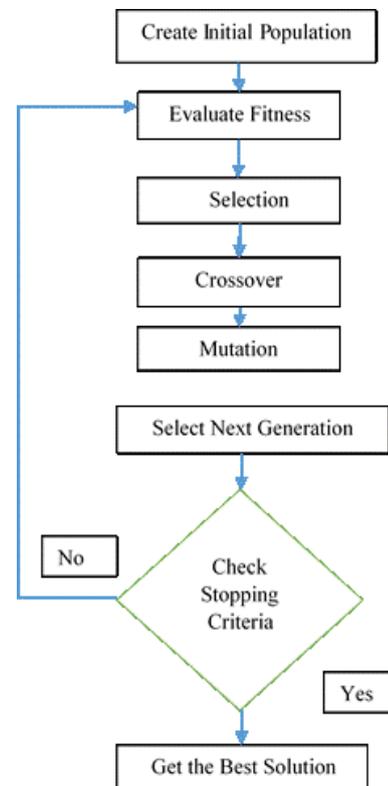

Fig. 1. Simple GAs process.

D. Fitness Function

GAs evaluates the individuals in the population using a selected fitness function. This function indicates should indicate how good or bad a candidate solution is.

GAs gain a great popularity due to their known attributes. These are:
1) GAs can handle both continuous and discrete optimization problems
2) They require no derivative information about the fitness criterion.
3) GAs have the advantageous over others each algorithms since it is less likely to be trapped by local minimum
4) GAs provide a more optimal and global solution.
5) In the case studies they are less sensitive to the presence of noise and uncertainty in measurements.

6) GAs use probabilistic operators not deterministic ones.

For these reasons they have been successful used in numerous applications in engineering and computer science.

They have been applied to a wide range of problems in diverse fields such as engineering, mathematics, operations research etc. Most of the problems in these fields are stated as constrained optimization problems. In survey sampling many authors have suggested many estimators of population mean. To deal with this problem family of estimators is proposed to collect all estimators in a form ([4]-[7]). But in this case parameter estimation is a problem. In this paper we have solved this problem using genetic algorithm and sequential quadratic programming.

II. NONLINEAR PROGRAMMING PROBLEM FORMULATION

A constraint nonlinear programming problem can be described as follows:

$$Min\ f(x), x \in F \subseteq S \subseteq R^n$$

Subject to

$$h_i(x) = 0,\ i = 1, 2, \dots, p$$

$$g_j(x) \le 0,\ j = p+1, \dots, q$$

Given that $a_k \le x_k \le b_k$, $k = 1, \dots, n$. $x = [x_1, \dots, x_n]$ is a vector of n variables.

$f(x)$ is the objective function.

$h_i(x)$ *i*. equality constraint $(i = 1, 2, \dots, p)$

$g_j(x)$: $(j = p+1, \dots, q)$ $q < n$ *j*. inequality constraint

S: whole search space

F: feasible search space

a_k and b_k present the lower and upper bounds of variable x_k respectively.

When you submit your final version, after your paper has been accepted, prepare it in two-column format, including figures and tables.

III. SEQUENTIAL QUADRATIC PROGRAMMING (SQP)

The sequential quadratic programming is a powerful technique for solving nonlinear constrained optimization problems. SQP allows you to closely mimic Newton's method for constrained optimization just is done for unconstraint optimization. At each iteration an approximation is made of the Hessian of the Lagrangian function using a quasi Newton updating method. This is then used to generate a Quadratic subproblem. Then its solution is used to form a search direction for a line search procedure.

SQP is an iterative method which solves at the k iteration a QP of the following form:

$$Minimize \frac{1}{2} d^t H_k d + \nabla f(x_k)^t d \qquad (1)$$

Subject to

$$\nabla h_i(x_k)^t d + h_i(x_k) = 0,\ i = 1, 2, \dots, p \qquad (2)$$

$$\nabla g_j(x_k)^t d + g_j(x_k) \le 0,\ j = p+1, \dots, q \qquad (3)$$

d is defined as the search direction and h_k is a positive definite approximation to the Hessian matrix of Lagrangian function can be described as:

$$L(x, \gamma, \beta) = f(x) + \sum_{i=1}^{p} \gamma_i h_i(x) + \sum_{j=p+1}^{q} \beta_j g_j(x) \qquad (4)$$

γ and β are the Lagrangian multipliers. The developed quadratic sub problems can then be solved using the active set strategy. The solution x_k at each iteration is updated according to

$$x_{k+1} = x_k + \alpha_k d_k \qquad (5)$$

α is defined as the step size and takes values in interval [0, 1]. After each iteration the matrix H_k is updated based on the Newton Method.

$$H_{k+1} = H_k + \frac{y_k y_k^t}{s_k y_k^t} - \frac{H_k s_k s_k^t H_k}{s_k^t H_k s_k} \qquad (6)$$

where

$$s_k = x_{k+1} - x_k \qquad (7)$$

$$y_k = \nabla L(x_{k+1}, \gamma_{k+1}, \beta_{k+1}) - \nabla L(x_{k+1}, \gamma_k, \beta_k) \qquad (8)$$

IV. STRATIFIED RANDOM SAMPLING

A stratified random sample is obtained by partitioning the population then taking a simple random sampling without replacement of specified size from each strata ([5], [6], [8]). With this method, we get more efficient estimates compared to simple random sampling. Consider a finite population $U = (u_1, u_2, \dots, u_N)$ of size *N* and let *y* and *x*, respectively, be the study and auxiliary variables associated with each unit u_j $(j = 1, 2, \dots, N)$ of the population. Let the population of size, *N*, is stratified into *L* strata with *h*-th stratum containing N_h units, where $h = 1, 2, \dots, L$ such that $\sum_{h=1}^{L} N_h = N$. A simple random sample of size n_h is drawn without replacement from the *h*-th stratum such that $\sum_{h=1}^{L} n_h = n$. Let (y_{hi}, x_{hi}) denote the observed values of *y* and *x* on the *i*-th unit of the *h*-th stratum, where $i = 1, 2, \dots, N_h$ and $h = 1, 2, \dots, L$. Moreover, let $\bar{y}_h = \sum_{i=1}^{n_h} \frac{y_{hi}}{n_h}$, $\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h$,

and $\bar{Y}_h = \sum_{i=1}^{N_h} \frac{y_{hi}}{N_h}$, $\bar{Y} = \bar{Y}_{st} = \sum_{h=1}^{L} W_h \bar{Y}_h$ be the sample and

population means of y, respectively, where $W_h = \frac{N_h}{N}$ is the

stratum weight. Similar expressions for x can also be defined.

To obtain the bias and the *MSE*, let us define $e_0 = (\bar{y}_{st} - \bar{Y})/\bar{Y}$ and $e_1 = (\bar{x}_{st} - \bar{X})/\bar{X}$. Using these notations,

$$E(e_0) = E(e_1) = 0,$$

$$V_{r,s} = \sum_{h=1}^{L} W_h^{r+s} \frac{E\left[(\bar{x}_h - \bar{X}_h)^r (\bar{y}_h - \bar{Y}_h)^s\right]}{\bar{X}^r \bar{Y}^s}.$$

We can write

$$E(e_0^2) = \frac{\sum_{h=1}^{L} W_h^2 \gamma_h S_{yh}^2}{\bar{Y}^2} = V_{0,2}$$

$$E(e_1^2) = \frac{\sum_{h=1}^{L} W_h^2 \gamma_h S_{xh}^2}{\bar{X}^2} = V_{2,0}$$

$$E(e_0 e_1) = \frac{\sum_{h=1}^{L} W_h^2 \gamma_h S_{xyh}}{\bar{X}\bar{Y}} = V_{1,1}$$

where

$$S_{yh}^2 = \frac{\sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2}{N_h - 1} \quad S_{xh}^2 = \frac{\sum_{i=1}^{N_h} (x_{hi} - \bar{X}_h)^2}{N_h - 1}$$

$$S_{xyh} = \frac{\sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)(x_{hi} - \bar{X}_h)}{N_h - 1}, \quad \gamma_h = \frac{1 - f_h}{n_h}$$

and

$$f_h = \frac{n_h}{N_h}.$$

Reference [4] suggested a family of estimators for the population mean in the stratified random sampling as

$$\bar{y}_{CST} = \bar{y}_{ST} \left(\frac{\bar{x}_{ST}}{\bar{X}}\right)^\delta \left[w + (1-w)\left(\frac{\bar{x}_{ST}}{\bar{X}}\right)^\varepsilon\right]^\eta \quad (9)$$

where δ, ε, η, and w can take finite values. When these four parameters are conveniently chosen, many estimators are obtainable. Also, when one parameter is considered as "free parameter", it is possible to obtain some subclasses of

estimators. Here, free parameter means that one of the four parameters is such a scalar that the mean square error of $\bar{y}_{CST}$ gets the minimum value. In other words, we minimize the MSE equation according to the free parameter. Rewriting class in terms of e's, we have

$$\bar{y}_{CST} = \bar{y}_{ST}(1 + e_1)^\delta \left[w + (1-w)(1 + e_1)^\varepsilon\right]^\eta \quad (10)$$

To obtain the bias and the *MSE*, the term $(1 + e_1)^\delta$ is expandable as

$$(1 + e_1)^\delta = a_0 + a_1 e_1 + a_2 e_1^2 + a_3 e_1^3 + a_4 e_1^4 + \dots \quad (11)$$

where

$$a_i = \begin{cases} \dfrac{\delta(\delta - 1)\dots(\delta - i + 1)}{i!} & i = 1,2,3,\dots \\ \\ 1 & i = 0 \end{cases} \quad (12)$$

and similarly the term $\left[w + (1-w)(1 + e_1)^\varepsilon\right]^\eta$ in (10) is expandable as

$$\left[w + (1-w)(1 + e_1)^\varepsilon\right]^\eta = b_0 + b_1 e_1 + b_2 e_1^2 + \dots$$

where

$$b_i = \begin{cases} \dfrac{1}{i!} \dfrac{\partial^i}{\partial(e_1)^i} \left[w + (1-w)(1 + e_1)^\varepsilon\right]^\eta \Big|_{e_1=0} & i = 1,2,3,\dots \\ \\ 1 & i = 0 \end{cases}$$

Up to the first order of approximation, the bias and the *MSE* of the estimator $\bar{y}_{CST}$ are respectively given by

$$Bias_I(\bar{y}_{CST}) = \bar{Y}(c_1 V_{1,1} + c_2 V_{2,0}), \quad (13)$$

$$MSE_I(\bar{y}_{CST}) = \bar{Y}^2 \left(V_{0,2} + c_1^2 V_{2,0} + 2c_1 V_{1,1}\right) \quad (14)$$

where

$$c_i = \sum_{j=0}^{i} \alpha_j b_{i-j}.$$

Up to the second order of approximation, the bias and the *MSE* of the estimator $\bar{y}_{CST}$ are respectively given by

$$Bias_{II}(\bar{y}_{CST}) = Bias_I(\bar{y}_{CST}) + \bar{Y}(c_2 V_{2,1} + c_4 V_{4,0} + c_3(V_{3,0} + V_{3,1})) \quad (15)$$

$$MSE_{II}\left(\bar{y}_{CST}\right) = MSE_{1}\left(\bar{y}_{CST}\right) + \bar{Y}^{2}\Big[\left(c_{2}^{2} + 2c_{1}c_{3}\right)V_{4,0}$$

$$+\, 2c_{1}c_{2}V_{3,0} + 2\left(2c_{1}c_{2} + c_{3}\right)V_{3,1}$$

$$+\, 2\left(c_{1}^{2} + c_{2}\right)V_{2,1} + \left(c_{1}^{2} + 2c_{2}\right)V_{2,2} + 2c_{1}V_{1,2}\Big]. \quad (16)$$

where

$$c_{1} = \varepsilon\eta\left(1-w\right) + \delta$$

$$c_{2} = \Big[\varepsilon^{2}\eta_{1}\left(1-w\right) + \varepsilon_{1}\eta\Big]\left(1-w\right)/2 + \delta\varepsilon\eta\left(1-w\right) + \frac{\delta\left(\delta-1\right)}{2}$$

$$c_{3} = \Big[\varepsilon^{2}\eta_{2}\left(1-w\right)^{2} + 3\varepsilon\varepsilon_{1}\eta_{1}\left(1-w\right) + \varepsilon_{2}\eta\Big]\left(1-w\right)/6$$

$$+\, \delta\Big[\varepsilon^{2}\eta_{1}\left(1-w\right) + \varepsilon_{1}\eta\Big]\left(1-w\right)/2$$

$$+\, \frac{\delta\left(\delta-1\right)}{2}\varepsilon\eta\left(1-w\right) + \frac{\delta\left(\delta-1\right)\left(\delta-2\right)}{3!}$$

V. NUMERICAL EXAMPLE

In this section, we use the data concerning the number of teachers as study variable and the number of students as auxiliary variable in both primary and secondary schools for 923 districts at 6 regions as: Marmara, Agean, Mediterranean, Central Anatolia, Black Sea, East and Southeast Anatolia in Turkey in 2007. We used the Neyman allocation for allocating the samples to different strata. Using this data set we try to minimize equation (16) both genetic algorithm and sequential quadratic algorithm. The optimum parameters and MSE values are given in Table II. According to Table II we can say that genetic algorithm gives more efficient result than sequential quadratic programming.

TABLE I: SOME MEMBERS OF DIANA (1993) ESTIMATOR

	δ	ε	η	w	$c_i \ \left(i = 1, ..., k\right)$
$\bar{y}_{RW(st)}$	-1	-1	-1		$\left(-w\right)^{i}$
$\bar{y}_{Gu(st)}$	-1	-1	1		$\left(-1\right)^{i}\left[i(1-w)+1\right]$
$\bar{y}_{RS(st)}$	-1	1	1		$\left(-1\right)^{i}w$

$\bar{y}_{SR(st)}$	0	1	2	$\dfrac{\varepsilon(\varepsilon-1)..(\varepsilon-i+1)}{(-1)i!}$
$\bar{y}_{TR}$	1	-1	1	$w \quad (i=1)$ $0 \quad (i=2,...)$
$\bar{y}_{sr(st)}$	0	-1	1	$\left(-1\right)^{i}\left(1-w\right)$
$\bar{y}_{sp(st)}$	0	1	1	$1-w \quad (i=1)$ $0 \quad (i=2,...)$
$\bar{y}_{rp(st)}$	-1	2	1	$1-2w \quad (i=1)$ $\left(-1\right)^{i}w \quad (i=2,...)$
$\bar{y}_{MS(st)}$	1	1	-1	$w\left(w-1\right)^{i-1}$
$\bar{y}_{SM(st)}$	-1	1	-1	$\left(-1\right)^{i}\left[1+\sum_{j=1}^{i}\left(1-w\right)^{j}\right]$

TABLE II: OPTIMUM PARAMETERS AND MSE VALUES USING GA AND SQP

δ	w	ε	η	MSE	Algorithm
0.15	1.732	1.827	-0.796	211.79	SQP
4.98	-0.212	-0.543	0.328	0.000131	GA

REFERENCES

[1] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison Wesley, Reading, MA, 1989.
[2] J. H. Holland, *Adaptation in Naturel and Artificial Systems*, MIT Press, Cambridge, 1975.
[3] A. Sheta and H. Turabieh, "A comparison between genetic algorithms and sequential quadratic programming in solving constrained optimization problems," *AIML Journal*, vol. 6, issue 1, pp. 67-74, 2006.
[4] G. Diana, "A class of estimators of the population mean in stratified random sampling," *Statistica*, vol. 53, no. 1, pp. 59-66, 1993.
[5] N. Koyuncu and C. Kadilar, "On the family of estimators of population mean in stratified sampling," *Pakistan Journal of Statistics*, vol. 26, issue 2, pp. 427-443, 2010.
[6] N. Koyuncu and C. Kadilar, "Family of estimators of population mean using two auxiliary variables in stratified random sampling," *Communications in Statistics: Theory and Methods*, vol. 38, issue 14, pp. 2398-2417, 2009.
[7] N. Koyuncu, "Efficient estimators of population mean using auxiliary attributes," *Applied Mathematics and Computation*, vol. 218, pp. 10900-10905, 2012.
[8] N. Koyuncu and C. Kadilar, "On improvement in estimating population mean in stratified random sampling," *Journal of Applied Statistics*, vol. 37, issue 6, pp. 999-1013, 2010.

Nursel Koyuncu was born in Ankara in 1984. She has completed her BSc, MSc and PhD degrees at Hacettepe University, Department of Statistics in Turkey. She has studied for one year at Erasmus MC University Medical Center, Department of Biostatistics in Netherlands. She is working as an assistant professor in Hacettepe University, Department of Statistics now.

She has papers about survey sampling, biostatistics and application of statistics.