



Journal Homepage: -www.journalijar.com
**INTERNATIONAL JOURNAL OF
 ADVANCED RESEARCH (IJAR)**

Article DOI:10.21474/IJAR01/1890
 DOI URL: <http://dx.doi.org/10.21474/IJAR01/1890>



RESEARCH ARTICLE

COMPARISON OF SOME CLASSICAL AND META-HEURISTIC OPTIMIZATION TECHNIQUES IN THE ESTIMATION OF THE LOGIT MODEL PARAMETERS.

Ozge Akkus¹ and Emre Demir².

1. Department of Statistics, Faculty of Science, Mugla Sıtkı Kocman University, Turkey.
2. Department of Biostatistics, Faculty of Medicine, Hitit University, Turkey.

Manuscript Info

Manuscript History

Received: 13 August 2016
 Final Accepted: 22 September 2016
 Published: October 2016

Key words:-

Genetic algorithm, logit model, Nelder-Mead, Newton-Raphson, parameter estimation.

Abstract

The aim of this study is to discuss the success of the Genetic Algorithm (GA) approach compared with the conventional Newton-Raphson (NR) and Nelder-Mead (NM) algorithms in the estimation of the Binary Logit Model (BLM) parameters over a simulated data and a real data set on Alopecia disease. NR algorithm requires restrictive assumptions such as the continuity of the objective function and determination of starting points for the model parameters in iterative process. As for the NM algorithm, it does not require differentiable objective function but still suffers from the starting point problem. In this study, the best set of parameters that maximize the likelihood function in BLM is found using both NR and NM algorithms. Then, considering the limitations of the conventional methods, the success of GA is investigated on condition that all the assumptions of the NR and NM methods are satisfied. The results show that when the assumptions of the classical techniques are valid, the GA approach can achieve to obtain very close result to NR and NM. This also implies that it is a good alternative to the NR and NM methods when the requirements of the classical methods cannot be satisfied. Model results of NR, NM and GA are compared in terms of the estimated values and the maximum likelihood function value.

Copy Right, IJAR, 2016., All rights reserved.

Introduction:-

When we study with a categorical dependent variable, regression model parameters can be estimated using the Maximum Likelihood Estimator (MLE). The aim is to find the best estimation result which maximizes the likelihood or log-likelihood functions. Because the likelihood equations obtained by taking the first order derivatives of the likelihood function with respect to the parameters are not linear, some iterative procedures such as Gauss-Newton, Newton-Raphson, Levenberg-Marquardt, Direct Search and Steepest Descent are developed within the category of the classic optimization techniques for the solution of an unconstrained nonlinear optimization problem.

These algorithms require restrictive assumptions such as the continuity of the likelihood function. They also have the risk of not finding the best solution due to the risk of getting stuck on local optimums relative to the selected starting points. Newton-Raphson (NR) algorithm is the most popular one among the classical techniques. Nelder-Mead (NM) algorithm can be seen as a good alternative to the NR algorithm. The most important advantage of NM algorithm is that it eliminates the requirement of the objective function (in our case, the likelihood function) to be differentiable.

Corresponding Author: Ozge Akkus.

Address:- Department of Statistics, Faculty of Science, Mugla Sıtkı Kocman University, Turkey

However, the problem of selecting the most suitable starting point for parameters still exists in the iteration process. Unlike the traditional optimization approach, Genetic Algorithms (GA) have the advantage of being able to find the global optimum points in a much easier manner and does not suffer from the starting point problems and the continuity of the functions to be optimized.

In this study, first a simulated data set is used for assessing the success of the algorithms. Binary Logit Model (BLM) is constructed using NR, NM and GA algorithms. Considering the difficulties mentioned above, the performance of GA is investigated when the classical optimization conditions are satisfied. The results suggest that GA gives very close results to the NR and NM results when the assumptions of the classical optimization techniques are provided. This also implies that it is a good alternative to the NR and NM methods when the requirements of the classical methods cannot be satisfied.

In the second part of the application, the performance of the studied optimization approaches is discussed over a real data set on Alopecia disease. The best model results are interpreted and the statistical significant factors influencing the disease are determined.

Literature review;-

Some of the researches in the current literature related to the GA and NM algorithms are presented below.

Lee and Kim (2015) compared the performance of the genetic algorithm and particle swarm optimization approaches for inverse surface radiation problems.

In their study, Stylianou *et al.* (2015) discussed the effectiveness of neural networks, genetic algorithms, support vector systems and many other heuristic optimization techniques to investigate the mortality rate due to burns.

Yuan and Lee (2015) compared the effectiveness of different methods for calculating a risk probability taking into account the financial indicators of 88 Taiwan companies in period of 2001-2010.

Pfeifer *et al.* (2015) focused on delays in the completion of project. Genetic algorithm approach was utilized to estimate the maximum risk of project delays will move.

In the study made by Hadji *et al.* (2015), the theoretical and experimental analyses of maximum power point tracking were presented over photovoltaic systems (PV) using genetic algorithm.

Aguilar-Rivera *et al.* (2015) discussed the use of many evolutionary computation methods such as genetic algorithms, genetic programming, multi-purpose evolutionary algorithm for the solution of the financial problems. It is concluded that as the interest of other methods varies depending of time, genetic algorithm maintains its popularity in each slice.

In his study on the bankruptcy probability of SMEs in Italy, Gordini (2014) compared the results obtained from the genetic algorithm, logistic regression and support vector machines. It is observed that genetic algorithm gives significantly successful results.

Johnson *et al.* (2013) used logistic regression and genetic algorithm in the diagnosis of Alzheimer disease.

Meng and Weng (2011) used the genetic algorithm approach and logistic regression analysis to assess the risk of the working area. The results demonstrated that genetic algorithm approach performs better than binary logistic regression model.

In the study of Babaoğlu *et al.* (2010), the efficiency of particle swarm optimization and genetic algorithm is compared over a data set on coronary artery disease.

Kohet *et al.* (2008) used genetic algorithm approach in constructing a probability scoring system for assessing the adverse drug reaction. That the new founded system provided excellent signals at %83 ratio is observed.

Liu and Ong (2008) used genetic algorithm approach to determine the most effective variables and number of clusters in order to get successful results from the cluster analysis for the marketing segmentation. Considering the number of

clusters and variables determined, it is observed that the genetic algorithm is extremely successful in finding the global optimum points.

In the study made by Hadi and Gonzalez-Andujar (2009), seedling time was estimated as a function of thermal time using both nonlinear regression and genetic algorithm. Given the flexibility and other advantages of assumptions conclude that genetic algorithm is a more effective method.

Methodology:-

Binary Logit Model (BLM):-

The purpose of the BLM is to establish an acceptable model which can identify the relationship between the dependent variable and explanatory variables using minimum number of variables in order to achieve the best model fit. BLM model is easy to apply and interpret thus it is commonly used in many areas, especially in social sciences, biology, medicine, economy, agriculture and veterinary sciences.

In BLM, dependent variable Y has two categories and coded as “0” and “1”. The probability expression of the model that uses the “logit link function” is expressed as the following.

$$P(Y = 1 / X) = \frac{\exp(\sum_{i=1}^N \hat{\beta}_k x_{ik})}{1 + \exp(\sum_{i=1}^N \hat{\beta}_k x_{ik})} \quad (1)$$

In Eq.(1), $P(Y = 1 / X)$ represents the probability of observing the category of “1” conditional on the explanatory variables X, $\hat{\beta}$ is the estimated model parameters and N is the number of observations. The relationship between the expected value of Y conditional on X is not in linear form. Considering that the model parameters are also nonlinear, MLE is commonly used for getting efficient estimations of the parameters (see Menard, 2002 for model assumptions, theoretical background of the model and other details).

Maximum Likelihood Estimation Method for BLM:-

Probability models are generally expressed as follows:

$$P_i = P(Y_i = 1 / X_i) \quad (2)$$

It is equal to $P(Y_i = 0 / X_i) = 1 - P_i$ for binary dependent variable models. Considering each observations has a Bernoulli distribution with P_i parameter, the probability of Y_i being equal to 0 or 1 is given in Eq.(3).

$$P(Y_i / X_i) = P_i^{Y_i} (1 - P_i)^{1 - Y_i} \quad (3)$$

The probability of observing N sample depending on the explanatory variables X and the dependent variable set Y is equal to the multiplication of N probability expression under the assumption that observations are independent.

$$P(Y / X) = \prod_{i=1}^N P_i^{Y_i} (1 - P_i)^{1 - Y_i} \quad (4)$$

Because P_i and $P(Y / X)$ are dependent on parameters and the aim is to estimate the unknown parameters, this dependency could be revealed by defining a likelihood function L(.).

$$L(Y / X, \beta) = P(Y / X) \quad (5)$$

L(.) denotes the probability (or likelihood) of observing a sample Y. The basis of the MLE is to find the optimal estimated $\hat{\beta}$ vector which maximizes the probability of observing Y.

$$L(Y / X, \hat{\beta}) = \max_{\beta} L(Y / X, \hat{\beta}) \quad (6)$$

The likelihood function for BLM is given below.

$$L(Y / X, \beta) = \prod_{i=1}^N \left[\frac{\exp(\sum \beta_k x_{ik})}{1 + \exp(\sum \beta_k x_{ik})} \right]^{Y_i} \left[\frac{1}{1 + \exp(\sum \beta_k x_{ik})} \right]^{1 - Y_i} \quad (7)$$

The logarithm of both sides of Eq.(7) is taken for much easier calculations.

$$\log L(Y / X, \beta) = \sum_{i=1}^N [Y_i \log P_i + (1 - Y_i) \log (1 - P_i)] \quad (8)$$

If $\hat{\beta}$ maximizes $L(Y / X, \beta)$, then it also maximizes $\log L(Y / X, \beta)$. Therefore, in order to obtain parameter estimations maximizing the probability of observing sample Y , first order derivatives are taken, equalized to 0 and a total of K equations for K parameters are obtained. Simultaneous solution of these K equations provides likelihood estimations. Likelihood equations for the BLM are derived from the expression given below.

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^N [Y_i - \frac{\exp(\sum \beta_k x_{ik})}{1 + \exp(\sum \beta_k x_{ik})}] x_{ij} = 0 \quad ; \quad \begin{matrix} i = 1, 2, \dots, N \\ j = 1, 2, \dots, K \end{matrix} \quad (9)$$

Because the likelihood equations are not linear in β_k , open forms of these equations could not be obtained. In this case, some traditional iterative optimization procedures like Newton-Raphson or Fisher-Scoring algorithms are applied (Agresti, 2002).

Newton-Raphson (NR) Algorithm:-

In NR algorithm, approximate parameter estimations can be determined depending on the starting point. Using Taylor series expansion and omitting high level terms, the first step of the iteration for NR algorithm can be expressed as follows.

$$\beta_1 = \beta_0 - H^{-1}(\beta_0) S(\beta_0) \quad (10)$$

In Eq.(10), H is the Hessian matrix consisting of the second order partial derivatives of the log-likelihood function. These algorithms based on unconstrained optimization of an objective function are commonly used in traditional methods for the maximization of the log-likelihood function. However, due to the restrictive assumptions mentioned above, the efficiency of the NM algorithm in the estimation of BLM model parameters is discussed, as well.

Nelder-Mead (NM) Algorithm:-

The classical NM method which is one of the multivariate, unconstrained, non-derivative optimization techniques is a simplex method developed by John Ashworth Nelder and Roger Mead in 1964 in order to find the local minimum of a multivariate function. Simplex, which has a N -dimensional geometrical form, contains $(n+1)$ points in n dimension. Thus, a simplex in two dimensions is constituted by a triangle. A simplex in three dimensions turns into a triangular prism with four surfaces. While finding the coordinates and minimum value of the point where the n -dimensional function is a minimum with this method, initial coordinates of $(n+1)$ point must be provided. Given these coordinates, the algorithm searches surface topography and moves to a minimum. Calculated value depending on the starting coordinates may be the local or global minimum.

In this method, $(n+1)$ point is mutually placed as a regular simplex (simple and regular) in n -dimensional Euclidean Space. As the simplex is a triangle in two dimensions, the process begins with taking three corner points of the first triangle. The point which has the biggest function value for the maximum problem and the smallest function value for the minimum problem is called ("Best Vertex: B) while the next one is called (Good Vertex: G). In maximization problems, the point with the smallest function value and in minimization problems, the point with the largest function value is called (Worst Vertex: W). This process is repeated for each iteration and moves to the optimum solution with one series of triangles in two dimensional problems. The point where the corner points of the triangle called B, G, and W is equal to each other is the optimal point.

In two dimensions, this method is a model search method comparing function values in corners of the triangle. Considering the minimization of a two-variable $Z = f(x, y)$ function, it takes the greatest value in the worst corner of the triangle. Therefore, this worst corner is changed place with a new point. Thus, a new triangle is obtained. Now the search process is conducted in this new triangle. So this process becomes a series of triangles where corner point gets smaller in each iteration step. The process ends with triangles approaching a single point, which is the desired minimum point. This method, developed for two-variable $Z = f(x, y)$ function can be extended to n -variable

$f(x) = f(x_1, x_2, \dots, x_n)$ function. In NM algorithm, movement of the simplex is provided by three basic operations, which are (Reflection:R), Expansion (Expansion:E) and Contraction (Contraction:C) (Mathews and Fink, 2004).

The Starting Triangle:-

For the function $f(x,y)$ to be minimized, $V_k = (x_k, y_k)$ ($k=1,2,3$) is created by providing values for three corner points of the triangle. Then, after function values for each of these points are calculated using $Z_k = f(x_k, y_k)$, they are sorted as $Z_1 \leq Z_2 \leq Z_3$. Corner points of the triangle are defined as $B(x_1, y_1)$, $G(x_2, y_2)$, $W(x_3, y_3)$.

Midpoint of Good Side:-

It is used for calculating the midpoint of the line drawn between points B and G. This point is found by calculating the mean value of the coordinates.

$$M = (B + G)/2 = [(x_1 + x_2)/2, (y_1 + y_2)/2] \quad (11)$$

Finding Point R Using the Reflection Procedure:-

In order to determine the point R, first the point M in the BG line of the triangle is calculated as shown below. Secondly, a line is formed between points W and M. The length of this line is called "d". Finally, a line with the length of "d" is drawn from point M, and the point where this line ends is called point R. This way, R point is found by using reflection operation. The vectorial notation of point R is given in Eq.(12) (Mathews and Fink, 2004).

$$R = M + (M - W) = 2M - W \quad (12)$$

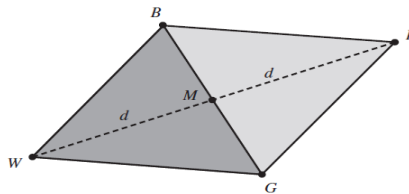


Fig.1:- Finding the Point R Using the Middle Point of M and Reflection Operator

Finding Point E Using the Expansion Procedure:-

If the function value in point R is smaller than the function value in point W (if $f(R) < f(w)$), then the iteration moves towards the minimum. The R point could be further from the minimum point. Therefore, when the line from point M to point R is expanded for the length of "d", end point E is obtained. Thus, a BGE triangle from BGW triangle is obtained by using the expansion process.

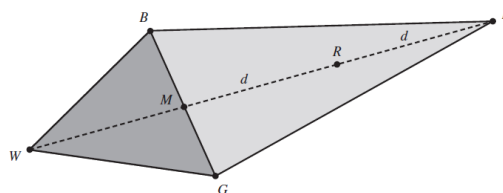


Fig2:- Producing the BGE Triangle and Point E by the Expansion Operator

If the function value in point E is smaller than the function value in point R ($f(E) < f(R)$), a better corner than R is found. Vectorial formulation for point E obtained by expansion process is as given in Eq.(13) (Mathews and Fink, 2004).

$$E = R + (R - M) = 2R - M \quad (13)$$

Finding Point C Using the Contraction Procedure:-

If the function value in point R is equal to the function value ($f(R) = f(W)$), another point must be tested. Function may be smaller in point M but M and W cannot be written again because there is a need to obtain a new triangle. When midpoints of lines WM and MR are called C1 and C2 respectively, our new triangle will be called BGC. In this case, BGC1 or BGC2 will not be different in two dimensions. However, there will be a difference in multiple dimensions (Mathews and Fink, 2004).

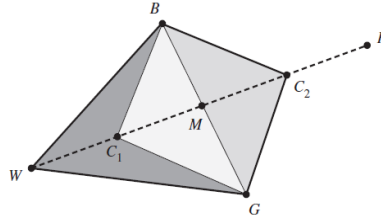


Fig3:-Producing the BGE Triangle with the Middle Points of C1 and C2 by the Reduction Operator

Contraction of Triangle Towards Point B:-

If the function value in point C is not smaller than the function value in point W ($f(c) \leq f(w)$), points G and W must be contracted towards point B (Fig.4). In the first step of this process, point M (which is midpoint of BG line) instead of point G and point S (which is midpoint of BW line) instead of point W are used. The next sequential operation steps are always conducted by using midpoints towards point B. The process ends when the point B is obtained (Mathews and Fink, 2004).

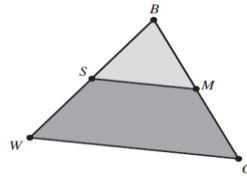


Fig4:-Transaction of Reduction to the Point B

Logical Decisions for Each Step:-

In each step described above, a new corner is found and this is replaced with W. We can explain it via an algorithm given below.

If $f(R) < f(G)$, Phase 1; If $f(R) \leq f(G)$, Phase 2 is regulated.

Phase 1: Reflection or Expansion

If $f(B) < f(R)$, R is used instead of W. If $f(B) \leq f(R)$, E and $f(E)$ are calculated.

If $f(E) < f(B)$, E is used instead of W. If $f(E) \leq f(R)$, R is used instead of W.

Phase 2: Shrinkage of Contraction

If $f(R) < f(W)$, R is used instead of W. $C = (W + M)/2$ and $f(C)$ value is calculated.

If $f(C) < f(W)$, C is used instead of W.

If $f(C) \leq f(W)$, $f(S)$ is calculated instead of S. S is used instead of W and M is used instead of G (Mathews and Fink, 2004; Tektaş, 2010).

We mainly focus on finding the roots of the log-likelihood equations which maximize the likelihood function of $(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)$ by using the most suitable method. NM algorithm seems to be a good alternative to the NR algorithm. However, considering restrictive assumptions of the iterative methods, GA is one of most commonly used meta-heuristic methods, especially in the optimization of non-linear functions.

Genetic Algorithm:-

GA is an intelligence technique that calculates the best solution for the problem using evolutionary processes. Influenced by Charles Darwin's evolution theory, GA is a meta-heuristic algorithm that was introduced in the study of Rechenberg (1973) called "evolution strategies" (He was a pioneer in evolutionary computation and artificial evolution); developed by John Holland (1975) and, popularized by a book published by Goldberg in 1989 (for details see; Genetic Algorithms in Search, Optimization and Machine Learning). GA, like other heuristic algorithms, does not guarantee the optimum point which minimizes or maximizes of an objective function. However, its solution quality is very high.

Unlike other optimization methods, GA conducts the search using a community of candidate solutions instead of just one candidate solution. Thus, the solution space is scanned in parallel with the multiple starting points. This

characteristic is the major factor enabling GA's to find the global optimum value without getting stuck on local optimums. GA works with coded formats of candidate solutions (according to certain rules) instead of candidate solutions themselves. This encoding is formed by using an alphabet consisting of 0's and 1's (BIT - Binary digit) in the classical genetic algorithms. While binary coding is common in GA's, symbols and real number encoding are also used in broad applications (Mitchell, 1999; Karr and Freeman, 1999).

In GA, the optimum solution of a problem is performed by some basic operations. These are the determination of the objective function, encoding, creating the initial population, selection, crossover and mutation operations. First, encoding is performed in accordance with the nature of the problem. Then, the starting population is created, often randomly. Suitable value for each array in the initial population is calculated. Finally, selection, crossover and mutation operators are used in order to change the population and create a new generation. GA operations proceed by assigning the adaptation value to the new population until the best solution is found (for details of schema theorem, see Holland, 1992; Goldberg, 1989; Reeves and Rowe, 2002).

Each individual in the population is called chromosome. In accordance with the fitness condition, chromosomes evolve and create new generations. Two chromosomes from current generation are obtained using Crossover or Mutation operators to create new generations. His new individual is called the 'Parent'. Parents matching the eligibility criteria (fitness value) are kept in the generation while parents who are not, are eliminated from the generation. This process is called the natural selection. Thus, GA's produce constantly improving solutions in accordance with the "survival of the fittest" rule found in nature. For this purpose, in order to determine what's best, GA's use a fitness function and for creating new solutions, they use operators like Recombination, Crossover and Mutation. During these processes, the change for the best chromosomes to be selected is maximized. A few generations later, the algorithm converges to the optimum solution where the best chromosomes exist.

Encoding:-

It is required to decide how solutions of individuals in the population are represented. It usually takes place in the form of chromosomes. Special structures in each chromosome are called 'Genes'. In binary encoding, every chromosome consists of character sequences like (0 or 1). Binary encoding is the most commonly used encoding method as well as the one most suitable for the structure of GA. However, there is no general coding technique that applies to all problems.

Starting Population:-

In this step, n chromosomes are chosen randomly as a starting group. Determination of the initial population is important in terms of possibilities like GA to fast converge or kept in local optimums. Population size is one of the parameters that display the performance of the solution (Pasia, Hermosilla and Ombao, 2005). Determining a small number of populations can cause the search space not to search effectively while determining large number of population can cause the processing load to increase and algorithm to be slowdown. Some studies suggest that population size should be around 20-30 while others argue that ideal population size should be between 50 and 100 (Reeves and Rowe, 2002).

Fitness Value:-

In each iteration step of GA, fitness value for each individual in the population is calculated and recorded. Individuals with higher fitness value are more likely to be transferred to the next generation. GA cannot perform the function of selection without fitness values. A new population is created in each iteration step. In this population, every individual has a fitness value and the chromosome with the best fitness value is recorded. After the iteration is completed, all populations are compared and the best of each population's best value are considered for the optimum solutions of the GA.

Selection Operator:-

The selection process is the process of selecting a chromosome in the old population to the new population which will be created according to fitness values, in order to have individuals with higher fitness value in new generations. The parent with high fitness value increases its chance for selection by taking the principle of the selection method used. For the selection process, Tournament, Roulette Wheel, Sorting, Steady State and Stochastic Uniform methods can be used. Goldberg and Deb (1991) stated that none of these methods have advantages of one another.

Crossover Operator:-

The basic function of this operator is to increase the probability of desired outcome by enabling new generations to produce different individuals than earlier generations and thus working in the solution space. Crossover employs the idea that new chromosomes become better by taking good genes from the old chromosomes. Crossover operator is expected to create better chromosomes by combining good characteristics of the current chromosomes. Probability of crossover specifies how often crossover operation will be performed. If there is not any crossover (if probability of crossover is 0%), the child will have the same chromosomes as the parents. If the probability of crossover is 100%, children are completely created by crossover.

Mutation Operator:-

Mutation prevents the searches to be stuck to local optimum points. How often crossover operation is to be performed to the individuals are determined by the probability of crossover (P_c) while it is determined by the probability of P_m for the Mutation operation. Mutation is often used within a smaller probability than Crossover. The reason for this is in order not to lose high adaptation levels achieved by Crossover. Crossover probability is often between 0.25 and 1 while Mutation probability is between 0.01 and 0.001. Mutation probability determines the frequency for chromosome parts to mutate. If there is no mutation, parents are directly copied without Crossover. If there is mutation, one or more parts of parent chromosomes change. If mutation probability is 100%, all chromosomes will change, if it is 0%, none of the chromosome will change.

Application:-

In this part of the study, performance comparisons of NR, NM and GA methods in the estimation of BLM model parameters are made over a simulated data and a real data set on Alopecia disease.

Simulated data Results:-**NR results for the estimation of BLM parameters:-**

Applicants generally use the results of this algorithm regardless of considering the iteration number, starting points and whether the function is differentiable or not. However, as mentioned earlier, this method may not achieve convergence in number of iterations if a starting point is chosen far away from the best solution. The NR algorithm results of the BLM model are given below.

Table 3- BLM Results Obtained by the Algorithm NR

	$\hat{\beta}$	Std.Error	Wald	df	Sig.	Exp($\hat{\beta}$)
Constant	0.744	0.817	0.829	1	0.363	2.103
X ₁	0.333	0.620	0.288	1	0.591	1.395
X ₂	-1.175	0.607	3.745	1	0.053	0.309
X ₃	-0.060	0.616	0.010	1	0.922	0.941
X ₄	-0.006	0.022	0.090	1	0.764	0.994
Min(-Log-LikelihoodFunction)	32.4960					

The minimum log-likelihood function value is 32.4960 according to the NR algorithm.

NM results for the estimation of BLM parameters:-

In this part of the study, parameter estimations for the same data set are obtained by using the NM algorithm. Since this algorithm is not included in the standard package programs, Matlab program is used for constructing the required commands for estimation. The general representation of the NM commands and explanations are as the following.

$$[x,y, \text{exitflag}, \text{output}] = \text{fminsearch}(@\text{likelihood}, [-0:0,-0:0,-0:0,-0:0,-0:0]) \quad (14)$$

x: represents the estimated parameter values;

y: represents the optimized value of the log-likelihood function;

exitflag: shows one of the integer values (-1, 0, 1) for specifying the reason for the algorithm to stop.

output: monitorize the results.

fminsearch: is added to the "command window" to obtain the best parameter estimations. Starting corner points for all parameters are taken "0". After the command is run, results given in the last row of Table 4 are obtained.

X = 0.647 0.313 -1.18 -0.037 -0.003; Y = 32.4829; exitflag = 1; output = iterations: 453; funcCount: 714; algorithm: 'Nelder-Mead simplex direct search'; message: [1x196 char]

Table 4:- LR Results Obtained by the Algorithm NM

Corner Points	[-3:3,-3:3,-3:3,-3:3,-3:3]	[-1:1,-1:1,-1:1,-1:1,-1:1]	[-0:0,-0:0,-0:0,-0:0,-0:0]
	$\hat{\beta}$	$\hat{\beta}$	$\hat{\beta}$
Constant	0.647	0.647	0.647
X ₁	0.313	0.313	0.313
X ₂	-1.18	-1.18	-1.18
X ₃	-0.037	-0.037	-0.037
X ₄	-0.003	-0.003	-0.003
Min Function Value	32.48285466470697	32.48285466470732	32.48285466876738
Iteration number	5146	1269	453

When Table 4 is examined, the optimum value of the log-likelihood function is found **32.48285466470697** in **5146** iteration. Three different starting corner points and different iteration numbers are selected.

GA results for the estimation of BLM parameters:-

In this section, the performance of GA is discussed in terms of the best starting population; selection, crossover and mutation operators; crossover and mutation probabilities over randomly generated data. Tables 5, 6 and 7 give the estimated LR results under different combinations of alternatives in GA such as “Population Size”, “Selection Function”, “Number of Generations”, “Stopping Criteria” etc.

Table 5:-GA Results According to the Different Population Size

Population Size	20	40	60	80	100
Selection Funct.	Roulette	Roulette	Roulette	Roulette	Roulette
Mutation Funct.	Gaussian	Gaussian	Gaussian	Gaussian	Gaussian
Mutation Ratio	0.2	0.2	0.2	0.2	0.2
Mutation Shrink	1.0	1.0	1.0	1.0	1.0
CrossoverFen	Intermediate	Intermediate	Intermediate	Intermediate	Intermediate
Crossover Ratio	1.0	1.0	1.0	1.0	1.0
Generations	5000	5000	5000	5000	5000
StallGenLimit	5000	5000	5000	5000	5000
TolFun	1e-6	1e-6	1e-6	1e-6	1e-6
TolCon	1e-6	1e-6	1e-6	1e-6	1e-6
Others	Default	Default	Default	Default	Default
$\hat{\beta}_0$ (constant)	0.647	0.647	0.647	0.647	0.647
$\hat{\beta}_1$	0.313	0.313	0.313	0.313	0.313
$\hat{\beta}_2$	-1.18	-1.18	-1.18	-1.18	-1.18
$\hat{\beta}_3$	-0.036	-0.037	-0.037	-0.037	-0.037
$\hat{\beta}_4$	-0.003	-0.003	-0.003	-0.003	-0.003
Min Function Value	32.48285466522573	32.48285466470753	32.482854664724	32.48285466470948	32.48285466470892

In order to achieve the optimum results, at least 100 different combinations are examined by individually trying alternatives. Four alternatives for the Mutation function (Use constraint dependent default, Gaussian, Uniform, Adaptive feasible); ten for Mutation rates; six for Crossover function (Scattered, Single point, Two point, Intermediate, Heuristic, Arithmetic); ten for Crossover rates; five for Selection preferences.

Table 5 contains the estimated parameters and the minimum values of the negative of the log-likelihood function according to the different population sizes while other factors are the same. It shows that the best result (the smallest function value of **32.48285466470753**) is obtained when the population size is 40. This table also emphasizes the selection of the optimal population size for getting better results.

Table 6 shows the results under different number of generations and stopping criteria for the iteration process. We can easily see that better results could be obtained as the value of the stopping criteria decreases. In this case, the iteration number will increase. Thus, it is an expected result that we could have much better results as the iteration number increases.

“Stopping criteria” and “number of generation” are taken as variable and population size is set to 60. The smallest function number (32.482854664706964) is obtained for the stopping criteria of “1e-6” for the number of generations “20000”. This means that the optimal generation number has to be truly determined for better model results. Additionally, when the results are examined according to three different stopping values of 1e-6, “1e-10” and “1e-15”, optimal function value is obtained for the “1e-6” stopping value in the scenario which the number of generation is set to 20000.

Table 6:- GA Results According to the Different Combinations

Population Size	60	60	60	60	60
Selection Function	Roulette	Roulette	Roulette	Roulette	Roulette
Mutation function	Gaussian	Gaussian	Gaussian	Gaussian	Gaussian
Mutation ratio	0.2	0.2	0.2	0.2	0.2
Mutation shrink	1.0	1.0	1.0	1.0	1.0
CrossoverFcn	Intermediate	Intermediate	Intermediate	Intermediate	Intermediate
Crossover ratio	1.0	1.0	1.0	1.0	1.0
Generations	10000	20000	30000	20000	20000
StallGenLimit	10000	20000	30000	20000	20000
TolFun	1e-6	1e-6	1e-6	1e-10	1e-15
TolCon	1e-6	1e-6	1e-6	1e-10	1e-15
Others	Default	Default	Default	Default	Default
$\hat{\beta}_0$ (constant)	0.647	0.647	0.647	0.647	0.647
$\hat{\beta}_1$	0.313	0.313	0.313	0.313	0.313
$\hat{\beta}_2$	-1.18	-1.18	-1.18	-1.18	-1.18
$\hat{\beta}_3$	-0.037	-0.037	-0.037	-0.037	-0.037
$\hat{\beta}_4$	-0.003	-0.003	-0.003	-0.003	-0.003
Min Function Value	32.48285466470744	32.482854664706964	32.48285466470703	32.48285466470710	32.48285466470703

In Table 6, the optimal value is obtained when the “Roulette” selection function is used. As a result of all the treatments made up till now, even they do not very much differ from each other; the smallest value found of the negative of the log-likelihood function is 32.4828546647066964.

Table 7:- GA Results According to the Different Selection Methods

PopulationSize	60	60	60	60	60
Selection Funct.	Stochastic Uniform	Remainder	Uniform	Tournament (size 4)	Roulette
Mutation function	Gaussian	Gaussian	Gaussian	Gaussian	Gaussian
Mutation ratio	0.2	0.2	0.2	0.2	0.2
Mutation shrink	1.0	1.0	1.0	1.0	1.0
CrossoverFcn	Intermediate	Intermediate	Intermediate	1.0	Intermediate
Crossover ratio	1.0	1.0	1.0	Intermediate	1.0
Generations	20000	20000	20000	1.0	20000
StallGenLimit	20000	20000	20000	20000	20000
TolFun	1e-6	1e-6	1e-6	20000	1e-6
TolCon	1e-6	1e-6	1e-6	1e-6	1e-6
Others	Default	Default	Default	1e-6	Default
	Default			Default	
$\hat{\beta}_0$ (constant)	0.647	0.647	0.647	0.646	0.647
$\hat{\beta}_1$	0.313	0.313	0.313	0.314	0.313
$\hat{\beta}_2$	-1.18	-1.18	-1.18	-1.18	-1.18
$\hat{\beta}_3$	-0.037	-0.037	-0.037	-0.036	-0.037
$\hat{\beta}_4$	-0.003	-0.003	-0.003	-0.003	-0.003
Min Function Value	32.4828546647070	32.4828546647071	32.48285478452	32.48285501285	32.482854664706964

In Table 7, results obtained according to the different selection functions are given. These are the “Stochastic Uniform”, “Roulette”, “Uniform” and “Tournament”. The best solution is obtained from the function of “Roulette” whereas the worst solution is for “Tournament” function.

The combination of the optimum results through all the trials is given by Table 8 and the graphical representation of the iterative process is presented by Figure 5.

Table 8:- Parameters Used in GA Coding

GA parametresi	Değer / Metod
PopulationType	'doubleVector'
PopulationSize	60
SelectionFunction	Roulette
Mutationfunction	Gaussian
Mutationratio	0.2
Mutationshrink	1.0
CrossoverFcn	Intermediate
Crossoverratio	1.0
Migration Direction	'forward'
Others (InitialPopulation, InitialScores <i>et al.</i>)	Usedefault
Stoppingcriteria	
Generations	20000
StallGenLimit	20000
TolFun	1e-6
TolCon	1e-6
Output	
PlotFcns	@gaplotbestf
Display	'iter','final','diagnose'
Others	Usedefault

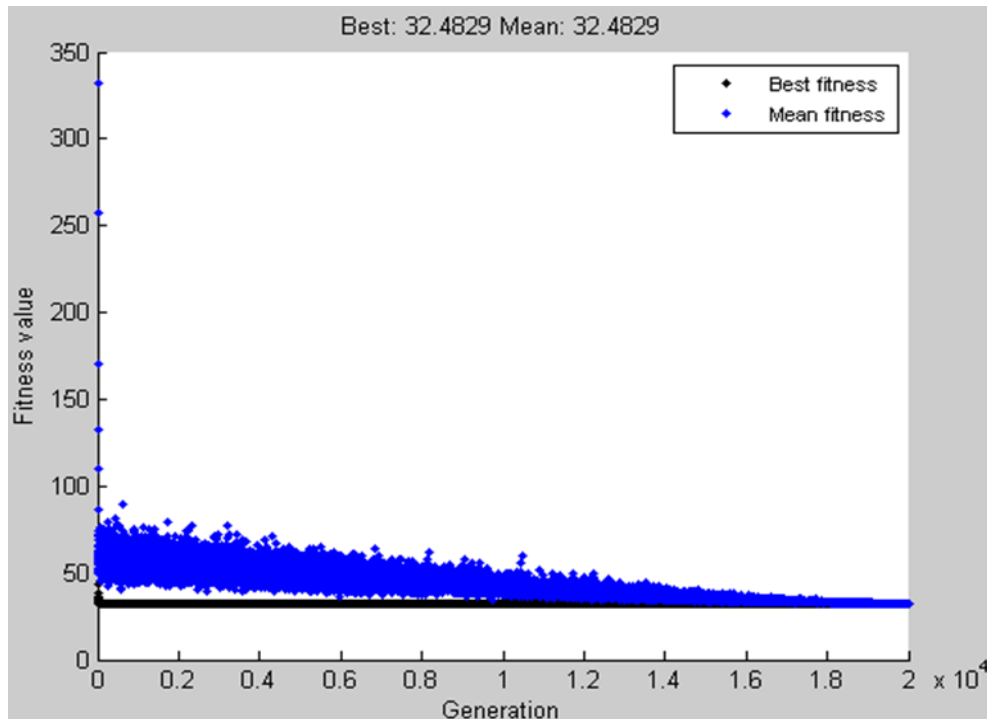


Figure 5:- Fitness Value (20000 iteration)

When Figure 5 is examined, the starting value of the fitness function, the resulting values and the spread of the search points at the end of the 20000 iterations can clearly be seen. The fitness value is approximately 50 and 100 in initial iterations. At the end of the 10000 iteration and after the application of the genetic operators, it falls into the interval of 30-35.

Table 9:- Iteration Numbers and Optimum Values in GA.

Generation	f-frequency	f(x)	Best	MeanStall		
			f(x)	Generation		
		19981	1198920	32.48	32.48	1927
		19982	1198980	32.48	32.48	1928
		19983	1199040	32.48	32.48	1929
		19984	1199100	32.48	32.48	1930
		19985	1199160	32.48	32.48	1931
		19986	1199220	32.48	32.48	1932
		19987	1199280	32.48	32.48	1933
		19988	1199340	32.48	32.48	1934
		19989	1199400	32.48	32.48	1935
		19990	1199460	32.48	32.48	1936
		19991	1199520	32.48	32.48	1937
		19992	1199580	32.48	32.48	1938
		19993	1199640	32.48	32.48	1939
		19994	1199700	32.48	32.48	1940
		19995	1199760	32.48	32.48	1941
		19996	1199820	32.48	32.48	1942
		19997	1199880	32.48	32.48	1943
		19998	1199940	32.48	32.48	1944
		19999	1200000	32.48	32.48	1945
		20000	1200060	32.48	32.48	1946

Comparison of NR, NM and GA results:-

The final comparison is on the performance of NR, NM and GA methods over the minimum value of the negative of the log-likelihood in Table 10.

Table 10:- Comparison of the BLM Parameter Estimates of NR, NM and GA.

ParameterEstimation	BLM Model					
	NR	p-value	NM	p-value	GA	p-value
$\hat{\beta}_0$ (Constant)	0.744	0.363	0.647	0.441	0.647	0.441
$\hat{\beta}_1$	0.333	0.591	0.313	0.610	0.313	0.610
$\hat{\beta}_2$	-1.175	0.053	-1.18	0.051	-1.18	0.051
$\hat{\beta}_3$	-0.060	0.922	-0.037	0.952	-0.037	0.952
$\hat{\beta}_4$	-0.006	0.764	-0.003	0.889	-0.003	0.889
MinFunction Value	32.4960		32.48285466470697		32.482854664706964	

When the results are compared, no significant variations are observed in terms of the parameter estimation. According to the minimum values of the negative of the log-likelihood function, GA is the most powerful method, followed by NM and NR.

This result implies the success of the GA in the parameter estimation of the categorical dependent variable modeling when the assumptions of the classic optimization techniques are provided. This also means that GA will probably give efficient estimates even though the classical optimization assumptions are not met.

Real data application:-

In this part of the study we have compared the performance of NR, NM and GA approaches in the estimation process over a real data set on Alopecia disease and interpreted the best model results obtained. Data set has been taken from the records of Dermatology Unit of Çorum Training and Research Hospital of Hitit University in 2013. The total number of patients applied to the hospital with hair loss problem is 95 during 2013.

The study variable (alopecia disease) is binary. The diagnosed patients are coded as “1” and the non-diagnosed patients are coded as “0”. Variables affecting the Alopecia diagnosis and the necessary abbreviations are given in Table 11.

Table 11:- Variables Used in the Analysis.

	Variable Definition	VariableType
Y	AlopeciaDisease (dependentvariable)	Qualitative 1 = 'Diseased' 0 = 'Nondiseased'
X ₁	Age	Quantitative
X ₂	Gender	Qualitative 1 = 'Male' 0 = 'Female'
X ₃	ALT (AlanineAminotransferase)	Quantitative
X ₄	AST (AspartateAminotransferase)	Quantitative
X ₅	Glucose	Quantitative
X ₆	HDL (High DensityLipoprotein)	Quantitative
X ₇	LDL (LowDensityLipoprotein)	Quantitative
X ₈ , X ₉ , X ₁₀	Cholesterol, Free t3 andFree t4 thyroids	Quantitative
X ₁₁	TSH (ThyroidStimulatingHormone)	Quantitative
X ₁₂	Trigliserit (Tg)	Quantitative

The explanations of some medical terms we used in the analysis is presented below.

ALT and AST in Table 9 are enzymes existent in liver cell. Cholesterol is a greasy element existent in circulation system. HDL is good cholesterol by blood, which is a special type of molecular proteins generated in liver transports. LDL is bad cholesterol by blood, which is a special type of molecular proteins generated in liver transports. Free t3 and Free t4 are thyroid gland hormone. TSH is a thyroid stimulating hormone and Trigliserit is a principle component of

grease. For the unique qualitative explanatory variable of “Gender” is coded as “1” for male patients and “0” for females.

MLE estimation results of BLM model parameters obtained via NR, NM and GA are given in Table 12.

The optimal values of the statistically significant parameters (i.e. the roots of the log-likelihood function) are marked bold.

In GA, we get the following log-likelihood function by substituting the independent variable values of patients in Eq.(5).

$$f(\beta_0, \beta_1, \dots, \beta_k) = 52*x(1)+1481*x(2)+37*x(3)+1167*x(4)+1189*x(5)+4899*x(6)+2255.4*x(7)+8430*x(8) +4826.07*x(9)+178.45*x(10)+55.28*x(11)+103.67*x(12)+6695*x(13)-\log(1+\exp(x(1)+18*x(2)+x(3)+20*x(4)+\dots+\log(1+\exp(x(1)+25*x(2)+10.6*x(4)+16.5*x(5)+88*x(6)+41.3*x(7)+140*x(8)+89.09*x(9)+3.19*x(10)+0.62*x(11)+2.2*x(12)+136.7*x(13)-\log(1+\exp(x(1)+25*x(2)+13*x(4)+17*x(5)+81*x(6)+69.7*x(7)+202*x(8)+118.7*x(9)+3.01*x(10)+0.86*x(11)+0.19*x(12)+68*x(13))) \quad (15)$$

Table 12:- NR, NM and GA Results of the Alopecia Data.

	NR		NM			GA	
	$\hat{\beta}$	p-value	$\hat{\beta}$	p-value	Exp ($\hat{\beta}$)	$\hat{\beta}$	p-value
Constant	2.985	0.346	2.930	0.418	18.731	2.55	0.418
Age	-0.025	0.310	-0.024	0.337	0.976	-0.023	0.337
Gender	1.618	0.009*	1.622	0.010	5.063	1.589	0.010
ALT	-0.051	0.185	-0.050	0.187	0.951	-0.049	0.187
AST	0.135	0.022*	0.134	0.022	1.143	0.135	0.021
Glucose	-0.004	0.687	-0.004	0.689	0.996	-0.004	0.682
HDL	-0.012	0.791	-0.010	0.818	0.990	-0.01	0.826
Cholesterol	0.011	0.774	0.009	0.757	1.009	0.012	0.772
LDL	-0.032	0.387	-0.030	0.362	0.970	-0.032	0.401
Free t3 thyroid	-1.611	0.038*	-1.611	0.048	0.199	-1.531	0.046
Free t4 thyroid	2.357	0.026*	2.372	0.028	10.72	2.325	0.002
TSH	-0.039	0.871	-0.016	0.889	0.984	-0.034	0.889
Triglycerit	0.006	0.515	0.006	0.529	1.006	0.005	0.555
Min (-log-likelihood func.)	52.8833		52.789756584169076			52.79292900514353	

*Coefficient is statistically significant at a significance level of 5%. ; Base category: Female; CI stands for Confidence Interval

GA optimization command in MATLAB is generally expressed as the following [Michalewicz, 1996].

$$[x, fval] = ga(@likelihood, 13, options) \quad (16)$$

In Eq.(16), ‘ga’ stands for the genetic algorithm and called by entering variable count (‘13’) with the definition of the operators (‘likelihood’); ‘x’ displays the independent variables that optimize the function and ‘fval’ displays the optimum value of the function. In order to find the value of 52.79292900514353, we set:

Table 13:- Set of Parameters Used in the Analysis.

GA Parameter	Value / Method
Population Type – Population Size	'double Vector' -50
Selection Function	Roulette
Reproduction Crossover fraction	0.95
Mutation function	Gaussian
Mutation ratio and Mutation shrink	0.15 and 0.88
Crossover Function and Crossover ratio	Crossover intermediate and 1.0
Migration Direction	'forward'
Others (Initial Population, Initial Scores <i>et al.</i>)	Use default
Stopping Criteria	
Generations and StallGenLimit	50000 and 5000

TolFun and TolCon	1e-6 and 1e-6
Output	
Plot Fcns	@gplotbestf
Display	'iter', 'final', 'diagnose'
Others	Use default

The graphical representation of the results after 50000 iterations is given in Figure 6.

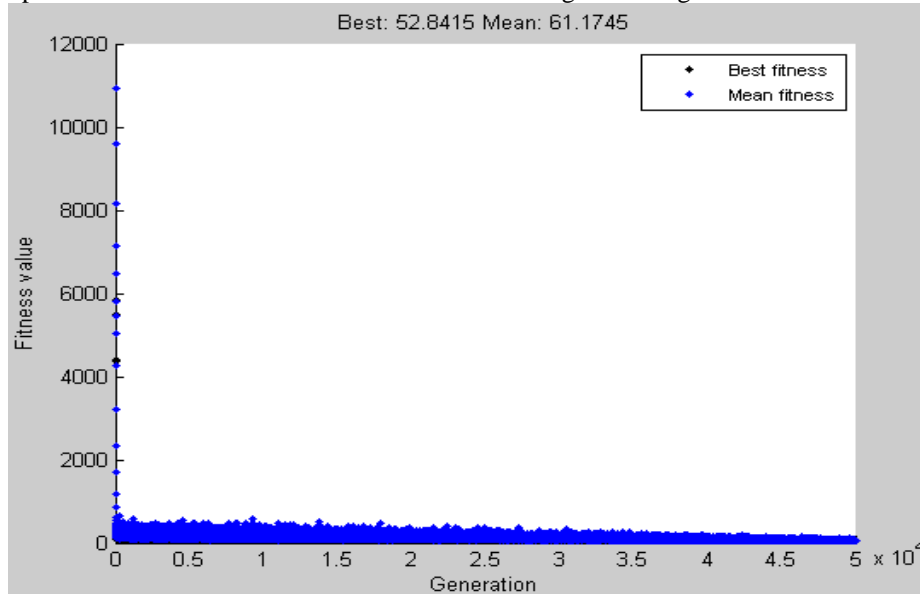


Figure 6:- Fitness Value (50000 iterations)

The value of the log-likelihood function is around 100 at the end of the first iteration. It slowly approaches to the 50-55 range, which includes the optimal value of 52.789756584169056 obtained from NM algorithm. As for GA, the optimum value is 52.79292900514353, which is not considerably far from the NM result.

The results show that we can obtain very close and sometimes even better results in BLM using GA with different operators and rates when the necessary conditions for applying the classical methods are satisfied. Additionally, these findings represent the success of GA because we can use GA method successfully in more flexible circumstances.

A Brief Interpretation of the Statistical Results:-

According to the NM algorithm results in Table 12, the statistically significant factors affecting the Alopecia disease at a %5 significance level are Gender ($p=0.010$), AST ($p=0.022$), Free t3 thyroid ($p=0.048$) and Free t4 thyroid ($p=0.028$).

Exp ($\hat{\beta}$) values give the odds ratios in LR analysis. This value is 5.063 for Males. This means that Males are nearly 5 times possible to exposure Alopecia disease compared with Females.

Since the estimated coefficients of AST and Free t4 thyroid are positive, they are all associated with higher probability of having the disease with a unit change in their values. Inversely, a unit change in Free t3 thyroid leads to the decrease in the probability of having the disease. As for the interpretation of their odds-ratio values, a unit increase in the AST and Free t4 thyroid values result in 1.143 and 10.72 times increases in the probability of having the disease, respectively. A unit decrease in the value of Free t3 thyroid leads to nearly 5.025 (1/0.199) times increases the probability of having the disease.

Conclusion:-

In this study, the success of two classical (NR and NM) and a meta-heuristic optimization method (GA) in the parameter estimation of LR model are examined over simulated and real data sets.

In LR model, the method used for finding parameter estimations is based on the maximization of the likelihood function via the method of MLE. In the classical optimization approach, the performance of the process depends on the validity of the required assumptions. That is, likelihood function has to be differentiable and true starting points related to the parameters have to be defined. In addition, iterative methods should be used to estimate approximate parameters in categorical dependent variable modeling due to the nonlinear structure of the likelihood equations obtained by taking the first order derivatives of the likelihood equations with respect to the each parameter. The classical NR algorithm is one of the best known and widely used root finding methods. NM algorithm is somewhat different from NR in terms of the required assumptions. Namely, it is in the search methods category in the classical optimization approach and unlike NR, it does not require differentiable objective function.

In NR and NM methods, search for the solution starts from a certain point and moves to another point according to some criteria. In order to find better parameter estimations in these methods, a starting point can be determined with a preliminary graphical drawing. However, incorrect selection of the starting point may cause the solution not to be found or solution time to be increased.

In GA, search process is performed on a set of potential solutions instead of a single point and solutions are evaluated until the best solutions are found. GA has certain advantages like not requiring derivatives or other complementary information and not getting stuck to local optimum points in order to find global optimum points for the solution of the problem. For this reason, in order to examine the performance of the GA method in the parameter estimation of the binary logit model when the assumptions of the classical optimization techniques are satisfied, its success on parameter estimation are demonstrated over an application.

In the application part of the study, first randomly generated data is used where the dependent variable is binary. LR model parameters are estimated using NR, NM and GA methods. In GA, many different combinations of the elements are applied and the best combination is tried to find that gives the best set of the estimated parameters. Results show that GA can successfully be used in categorical dependent variable modeling as well as the classical approaches on condition that the best component of the algorithm is defined correctly. Additionally, we could infer from the results that, if GA is powerful in the estimation even though the strict conditions are provided from the classical methods, it can be applied to the data, unhesitatingly under the worst conditions.

References:-

1. Agresti, A., 2002. Categorical Data Analysis, 2th edition, John Wiley&Sons Inc., Hoboken, 710p, New Jersey.
2. Aguilar-Rivera, R., Valenzuela-Rendón, M., Rodríguez-Ortiz, J.J., 2015. Geneticalgorithm and Darwinian approaches in financial applications: A survey. *Expert Systems with Applications*, 42(21): 7684-7697.
3. Babaoğlu, İ., Findik, O., Ülker, E., 2010. A comparison of features selection models utilizing binary particle swarm optimization and genetic algorithm in determining coronary artery disease using support vector machine. *Expert Systems with Applications*, 37(4): 3177-3183.
4. Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization and Machine Learning. Addison Wesley, Reading, MA, 507p.
5. Goldberg D.E., Deb, K., 1991. A Comparative Analysis of Selection Schemes Used in Genetic Algorithms, *Foundations of Genetic Algorithms*. CA: Morgan Kaufmann, 69-93, San Francisco.
6. Gordini, N., 2014. A geneticalgorithm approach for firms bankruptcy prediction: Empirical evidence from Italy. *Expert Systems with Applications*, 41(14): 6433-6445.
7. Hadi, H.S., J.L. Gonzalez-Andujar, 2009. Comparison of fitting weed seedling emergence models with nonlinear regression and genetic algorithm, *Computers and Electronics in Agriculture*, 65(1): 19-25.
8. Hadji, S., Gaubert, J.P., Krim, F., 2015. Theoretical and experimental analysis of geneticalgorithms based mppt for PV systems. *Energy Procedia*, 74: 772-787.
9. Holland, J.H., 1975. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, 183p. USA.
10. Holland, J.H., 1992. *Adaptation in Natural and Artificial Systems*. 2th edition, The MIT Press, Cambridge, London.
11. Karr C.L., Freeman M. L., 1999. *Industrial Applications of Genetic Algorithms*. CRC Press, 350p, USA.
12. Koh, Y., Yap, C.W., Li, S.C., 2008. A quantitative approach of using geneticalgorithm in designing a probability scoring system of an adverse drug reaction assessment system. *International Journal of Medical Informatics*, 77(6): 421-430.

13. Johnson, P., Graham, P., Wilson, P., Macaulay, L., Maruff, P., Savage, G., Ellis, K., Martins, R., Rowe, C., Masters, C., Ames, D., Zhang, P., 2013. Genetic algorithm with logistic regression for Alzheimer's disease diagnosis and prognosis, *Alzheimer's & Dementia*, 9(4): 455-456.
14. Lee, K.H., Kim, K.W., 2015. Performance comparisons of particle swarm optimization and genetic algorithm for inverse surface radiation problem. *International Journal of Heat and Mass Transfer*, 88: 330-337.
15. Liu, H.H., Ong, C.S., 2008. Variable selection in clustering for marketing segmentation using genetic algorithms. *Expert Systems with Applications*, 34(1): 502-510.
16. Mathews, J. H., Fink, K.K., 2004. *Numerical Methods Using Matlab*, 4th Edition, Prentice-Hall Inc., Upper Saddle River, New Jersey, USA, 696p.
17. Menard, S. 2002., *Applied Logistic Regression Analysis*, 2th Edition, Sage Publications, 128p., USA.
18. Meng, Q., Weng, J., 2011. A genetic algorithm approach to assessing work zone casualty risk. *Safety Science*, 49: 1283-1288.
19. Mitchell, M., 1999. *An Introduction to Genetic Algorithms*, 5rd Edition, The MIT Press, Cambridge, London, 143p.
20. Nelder, J.A. and Mead, R., 1964. A simplex method for function minimization, *Comput. J.*, 7: 308-313.
21. Pasia, J., Hermosilla, A., Ombao, H., 2005. A useful tool for statistical estimation: genetic algorithm. *Journal of Statistical Computation and Simulation*, 75(4): 237-251.
22. Pfeifer, J., Barker, K., Ramirez-Marquez, J.E., Morshedlou, N., 2015. Quantifying the risk of project delays with a genetic algorithm. *International Journal of Production Economics*, 170(A): 34-44.
23. Reeves, C.R., Rowe, J.E., 2002. *Genetic Algorithms Principles and Perspectives. A Guide to GA Theory*. Kluwer Academic Press, 332p., USA.
24. Rechenberg, I., 1973. *Evolutions Strategie—Optimierung technischer systemen nach Prinzipien Der Biologischen Evolution*. (PhD. Thesis). Fromman-Holzboog, Germany.
25. Stylianou, N., Akbarov, A., Kontopantelis, E., Buchan, I., W. Dunn, K., 2015. Mortality risk prediction in burn injury: comparison of logistic regression with machine learning approaches. *Burns*, 41(5): 925-934.
26. Yuan, F.C., Lee, C.H., 2015. Using least squares support vector regression with genetic algorithm to forecast beta systematic risk. *Journal of Computational Science*, 11: 26-33.