A Statistical Study on the Parameters of the Skew Normal Distribution Depending on the Use of the Genetic Algorithm Using the Simulation Method

Murtadha Alaa Alkhafaji, Rabab Abdulrida Saleh

Department of Statistics, College of Administration and Economics, University of Baghdad, Iraq

Email: mortada.alaa1101a@coadec.uobaghdad.edu.iq

Abstract. Recent statistical research has witnessed activity on the study of skew normal distribution (SND) due to the fact that the data set does not fit well with the normal distribution due to skewness and excessive kurtosis. For the purpose of estimating the parameters of the model (SND), the maximum likelihood method (ML) was used, but the probability equations of this method do not have clear solutions in the distribution (SND), and the problem was solved using the genetic algorithm (GA) and other repetitive techniques are Newton Raphson, Nelder Mead and Iteratively Reweighting Algorithm, using the simulation method with different sample sizes and comparing the preference of results methods used based on criteria (Mean, Mse and Def). It has been concluded that (ML) capabilities using the (GA) of parameters (SND) are best in the case of a small or medium sample size and the best (IR) algorithm at a large sample size.

1. Introduction

The symmetric normal distribution is considered the most prevalent theoretically due to its important properties, but in practice and analysing the data we find in many cases that the data is not distributed in a normal distribution. The reason is due to the presence of outliers that cause skewness in the data. Since the skew is responsible for the asymmetry in the shape of the distribution, and previous studies confirm that most of the phenomena of life contain a proportion of the skew. Symmetric Here lies the importance of studying a skew normal distribution (SND).

The (SND) distribution shares common features and characteristics of the normal distribution, adding the mechanism of the skewness parameter ($\omega$) that is responsible for regulating the skewness and is considered the turning point from the normal to the skew normal distribution and gives the distribution (SND) greater flexibility and more appropriate to the data under study.

In our research we studied experimental data with different sample sizes and different skewness levels using the Monte Carlo simulation method. The values of the simulation results express the accuracy and best methods and methods used in the research for the studied model and obtaining the best results from the use of statistical methods and methods of improvement appropriate to the model and relying on them according to scientific rules that result in taking the correct and scientific decision appropriate to the studied problem.

The main problem of the research is to overcome the problems of probability equations in the (ML) when estimating the parameters of the (SND), which are considered the most prominent statistical methods for estimating probability distributions and because of the emergence of nonlinear functions.
when the possibility function is derived and thus the results of (ML) have no solutions Evident in estimating the model parameters of a skew normal distribution modelpp.3.

The aim of our research is to study the (SND) of the characteristics and features, study the effect of the skewness parameter ($\omega$), and address the problem of nonlinear functions in the (ML) equations of the (SND) . For the purpose of estimating the parameters of the skew normal distribution model, the genetic algorithm (GA) was used in addition to iterative algorithms (NR, NM and IR), then comparison between the preference of these methods based on criteria (Mean, Mse and Def)

2. Skew Normal Distribution (SND)

The univariate skew normal distribution was introduced by the researcher Azzalini in 1985, the mathematical formula for the probability density function (PDF) for the variable $y$ distributed (SND) for the location parameter ($\epsilon$), the scale parameter ($\tau$) and the shape parameter ($\omega$) are written as follows[3] :

$$f(y, \epsilon, \tau, \omega) = \frac{2}{\tau} \phi\left(\frac{y-\epsilon}{\tau}\right) \psi\left(\omega \frac{y-\epsilon}{\tau}\right) \quad y \in R \ldots (1)$$

Where $\phi$ and $\psi$ are the (pdf) and (cdf) functions of the normal distribution, respectively. The skew normal distribution converts to the normal distribution when the shape parameter value is ($\omega =0$) , and when the location and scale parameters values is ($\epsilon= 0$ and $\tau = 1$) respectively , the (SND) becomes a standard skew normal distribution (SSND) With the parameter of skewness ($\omega$)[2], it is expressed mathematically as follows :

$$f(y, \omega) = 2 \phi(y) \varphi(\omega y) \quad y \in R \ldots (2)$$

The skewness increases (asymmetry of the distribution) with the increase of the skewness parameter ($\omega$) [8].

3. Method Maximum Likelihood (ML)

The ML method is considered one of the most prominent methods used to estimate the parameters of the probability model to characterize the estimated of maximum likelihood with high efficiency. The (ML) estimators of the parameters consider as points that lead to increase the logarithm of the likelihood function to its maximum limit [5]. The logarithm function of the (SND) is written as follows:

$$lnL = n\ln(2) - n\ln(\tau) - \frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^{n} \left(\frac{y_i-\epsilon}{\tau}\right)^2 + \sum_{i=1}^{n} \ln \psi\left(\omega \frac{y_i-\epsilon}{\tau}\right) \ldots (3)$$

Deriving the logarithm function of the skew normal distribution (SND) for the parameters ($\epsilon$, $\tau$, $\omega$), we obtain nonlinear equations as follows :

$$\frac{\partial lnL}{\partial \epsilon} = \sum_{i=1}^{n} \left(\frac{y_i-\epsilon}{\tau}\right) - \omega \sum_{i=1}^{n} \frac{\phi\left(\omega \frac{y_i-\epsilon}{\tau}\right)}{\psi\left(\omega \frac{y_i-\epsilon}{\tau}\right)} = 0 \ldots (4)$$

$$\frac{\partial lnL}{\partial \tau} = -n + \sum_{i=1}^{n} \left(\frac{y_i-\epsilon}{\tau}\right)^2 - \omega \sum_{i=1}^{n} \frac{\phi\left(\omega \frac{y_i-\epsilon}{\tau}\right)}{\psi\left(\omega \frac{y_i-\epsilon}{\tau}\right)} \left(\frac{y_i-\epsilon}{\tau}\right) \ldots (5)$$

$$\frac{\partial lnL}{\partial \omega} = \sum_{i=1}^{n} \frac{\phi\left(\omega \frac{y_i-\epsilon}{\tau}\right)}{\psi\left(\omega \frac{y_i-\epsilon}{\tau}\right)} \left(\frac{y_i-\epsilon}{\tau}\right) = 0 \ldots (6)$$
Where \( w(y_i) = \frac{\omega y_i - \epsilon}{\psi(y_i)} \) are nonlinear equations, and because of these equations are nonlinear, the calculation of parameters by the (ML) method is considered not has clear solutions in estimating a skew normal distribution [5].

4. Newton Raphson Algorithm (NR)
This method was introduced by Newton and Raphson in 1960, which is used to find a solution for nonlinear functions or equations that are difficult to solve using the traditional method.

In our study, we aim to solve nonlinear greatest possibility equations (4-6) and get the best estimate of the distribution parameters based on elementary estimates. An iterative method is used to find the solution (NR) algorithm [8]. The mathematical form of the (NR) algorithm is written in the following rule

\[
\hat{\theta}^{(p+1)} = \hat{\theta}^{(p)} - (W^{(p)})^{-1} \cdot T^{(p)} \quad \theta = (\epsilon, \tau, \omega) \quad \ldots (7)
\]

\( p \) The number of iterations \( (p = 1, 2, 3, \ldots) \)
\( \hat{\theta}^{(p)} \) and \( \hat{\theta}^{(p+1)} \) represent the estimation of the parameters of the maximum likelihood of iteration \( (p) \) and the subsequent iteration \( (p + 1) \), respectively.
\( T^{(p)} \) represents the vector of the first derivative of the logarithm of the likelihood

\[
T = \frac{\partial L_n L}{\partial \theta} = \left( \frac{\partial L_n L}{\partial \epsilon} \frac{\partial L_n L}{\partial \tau} \frac{\partial L_n L}{\partial \omega} \right) \quad \ldots (8)
\]

\( W^{(p)} \) represents the second matrix of the logarithm of the likelihood

\[
W = \frac{\partial^2 L_n L}{\partial \theta \partial \theta} = \left[ \begin{array}{ccc}
\frac{\partial^2 L_n L}{\partial \epsilon^2} & \frac{\partial^2 L_n L}{\partial \epsilon \partial \tau} & \frac{\partial^2 L_n L}{\partial \epsilon \partial \omega} \\
\frac{\partial^2 L_n L}{\partial \tau \partial \epsilon} & \frac{\partial^2 L_n L}{\partial \tau^2} & \frac{\partial^2 L_n L}{\partial \tau \partial \omega} \\
\frac{\partial^2 L_n L}{\partial \omega \partial \epsilon} & \frac{\partial^2 L_n L}{\partial \omega \partial \tau} & \frac{\partial^2 L_n L}{\partial \omega^2}
\end{array} \right] \quad \ldots (9)
\]

5. Nelder Mead Algorithm (NM)
It is a numerical algorithm in which geometric relations are used to reach the best minimum of the objective function, and it was formulated by researchers Nelder and Mead in 1965. In previous years, it was used in many practical applications and research for its effectiveness in finding results, its simplicity, as well as its use without needing to derive the function and in many mathematical problems similar to its ability to find solutions when faced nonlinear functions [4].

The basis of the work of the (NM) algorithm is the initial Simplex configuration, since its working mechanism depends on the order of Simplex points \( (\lambda_1, \lambda_2, \ldots, \lambda_{n+1}) \) in ascending order, and substituting the points in the objective function. The objective function is written as follows [7]:

\[
R(\lambda) = -ln L(\lambda) \quad \lambda = (\epsilon, \tau, \omega) \quad \ldots (10)
\]

\[
R(\lambda_1) < R(\lambda_2) < \ldots < R(\lambda_{n+1}) \quad \ldots (11)
\]

Where \( \lambda_1 \) and \( R(\lambda_1) \) refer to the best point and the best value of a objective function while \( \lambda_{n+1} \) and \( R(\lambda_{n+1}) \) refer to the worst point and worst value of the objective function. These points go through a set of processes iteratively (reflection, expansion, contraction and shrink) [7], and in these operations the worst point is replaced by a new point that is better than the previous one, and the process repeated to obtain the best minimum of the objective function that represents the best solution.
6. Iterative Re-Weighting Algorithm (IR)

It is one of the types of function maximization algorithms, and it is used in many previous studies in calculating the capabilities of the maximum potential parameters of the (SND). At the beginning of the work of the algorithm, initial values of the parameters are selected and the nonlinear weight function is updated:

\[ W_i^{(p)} = \frac{\psi(\omega_i^{(p)})}{\psi(\omega_i^{(p)})} \text{ Where } t_i = \frac{y_i - \bar{y}(p)}{\bar{y}(p)} \]  

Parameter estimates of the SND algorithm are calculated by the following laws [8]:

\[ \epsilon^{(P+1)} = \bar{y} - \omega^{(P)} \bar{W}(P) \tau^{(P)} \]  

\[ \tau^{(p+1)} = -\omega^{(p)} \sum_{i=1}^{n} W_i^{(p)} y_i - n \bar{W}(p) \bar{y} + \frac{\left(\omega^{(p)} \sum_{i=1}^{n} W_i^{(p)} y_i - n \bar{W}(p) \bar{y}\right)^2}{2n} + 4n \sum_{i=1}^{n} (y_i - \bar{y})^2 \]  

where \( \bar{W} \) and \( \bar{y} \) are the mean values of the \( W \) and \( y \), respectively.

7. Genetic Algorithm (GA)

It is a randomized research technique based on the mechanism of natural selection and genetics to find the best solution to the problem under study. It was invented and developed by John Holland in 1975. The GA algorithm begins with a randomly selected population of possible solutions and ends with the optimal solution. When applying the algorithm to a problem [1], an appropriate coding method for chromosomes must be determined and the fitness function must be determined.

The genetic algorithm goes through several stages and steps to reach the optimal solution, which will be mentioned as follows [6]:

1. Initialization: It is the process of generating many primary solutions (chromosomes) in a random manner.
2. Fitness function: evaluating the population under study by calculating the value of each chromosome with the evaluation function and the resulting value expresses the efficiency of the chromosome.
3. Selection: choosing the best chromosomes for the formation of the fathers in future generations, based on the evaluation function.
4. Crossover: the mating process that occurs between each parental chromosome that was selected in the selection process to produce offspring.
5. Mutation: random changes in the genes of a chromosome through a change in one or more genes in the son chromosome.
6. Stopping and Termination Criterion: (GA) continues to find the generations sequentially to find the best solution, and (GA) tests whether or not the stopping condition exists according to the nature of the problem studied.

8. Simulation

It is one of the important methods to study many of the approved problems that are difficult for the applied field in the absence of data or difficulty in obtaining [7]. The Monte Carlo simulation method was used to generate random data distributed in (SSND). The genetic algorithm (GA) and iterative algorithms (NR, NM and IR) were used in calculating the capabilities of (ML) parameters and discussing the preference of methods in estimating the model and comparing them.

To observe the effect of the skewness parameter \( \omega \), different values of the parameter \( \omega = 0, 0.5, 1.5, 2, 3, 4 \) were taken with different sample sizes \( n = 20, 50, 100, 200, 500 \), noting that when the parameter value is 0, the warp is equal to zero, the shape of the distribution is symmetrical, and when it increases, the warp increases.

The results were compared by extracting the statistical parameters, the mean, the mean square of error (Mse), and the sum of the mean squares of error for the parameters (Def). The smaller the value of (Mse and Def) is better. We note from the simulation results:
1- In Table (1), when ($\epsilon = 0, \tau = 1, \omega = 0$), we find that the mean squares of error (Mse) for the parameter ($\epsilon$) that using the sample size ($n = 20, 100, 200$), the genomic algorithm (GA) gives less (MSE) then followed by the algorithm (NM) for volumes ($n = 50, 500$). As for the parameter ($\tau$), (GA) products appear to be the best in small and medium samples and the (IR) algorithm is best in large samples.

The superiority of (IR) was observed by estimating the skewness parameter ($\omega$) in all of the samples used. The results of (Def) for the skew normal distribution model for small and medium samples ($n = 20, 50, 100$) show that (GA) is better because it has the lowest value for the (Def) criterion. Whereas in large samples ($n = 200, 500$) the (IR) algorithm was the best because it was the lowest (Def) value.

**Table (1) when ($\epsilon = 0 , \tau =1 , \omega = 0$)**

| n  | Method | $\hat{\epsilon}$ | Mse | $\hat{\tau}$ | Mse | $\hat{\omega}$ | Mse | Def  |
|----|--------|------------------|-----|--------------|-----|---------------|-----|------|
| 20 | NR (ML)| 0.000907 | 5   | 0.002031     | 1   | 0.388449      | 9   | 0.405361 |
|    | IR (ML)| 0.233060 | 8   | 0.229789     | 9   | 0.327560      | 2   | 0.334364 |
|    | NM(ML)| 0.849161 | 7   | 0.254647     | 7   | 0.375051      | 7   | 0.424142 |
|    | GA   | 0.184256 | 5   | 0.232137     | 7   | 0.065394      | 0   | 0.067472 |
| 50 | NR (ML)| 0.000241 | 5   | 0.001107     | 7   | 0.259506      | 1   | 0.263901 |
|    | IR (ML)| -0.090862| 5   | 0.290564     | 6   | 0.2194755     | 0   | 0.225738 |
|    | NM(ML)| 1.242986 | 9   | 0.118359     | 9   | 0.337342      | 4   | 0.337511 |
|    | GA   | 0.195996 | 6   | 0.259296     | 0   | 0.111822      | 5   | 0.113187 |
| 10 | NR (ML)| 0.146615 | 5   | 0.404685     | 5   | 0.279341      | 9   | 0.281737 |
|    | IR (ML)| -0.103749| 9   | 0.242716     | 6   | 0.254315      | 9   | 0.257835 |
|    | NM(ML)| 1.185520 | 7   | 0.392794     | 4   | 0.345234      | 7   | 0.346408 |
|    | GA   | 0.238913 | 2   | 0.307702     | 2   | 0.247803      | 9   | 0.249434 |
| 20 | NR (ML)| 0.072415 | 5   | 0.202752     | 2   | 0.246900      | 9   | 0.249756 |
|    | IR (ML)| -0.247208| 8   | 0.234964     | 4   | 0.230678      | 6   | 0.233177 |
|    | NM(ML)| 1.483016 | 4   | 0.707907     | 7   | 0.313142      | 7   | 0.318885 |
|    | GA   | 0.233834 | 0   | 0.304668     | 8   | 0.451937      | 5   | 0.453425 |
| 50 | NR (ML)| 0.318354 | 1   | 0.733658     | 8   | 0.257269      | 6   | 0.257710 |
|    | IR (ML)| -0.393921| 6   | 0.213396     | 6   | 0.247683      | 2   | 0.249174 |
In Table (2), when ($\varepsilon = 0, \tau = 1, \omega = 0.5$), we note that the (IR) algorithm is superior to other methods for all sample sizes in calculating the parameter estimation ($\varepsilon$) because it has less (Mse) than the other methods. As for estimating the parameter ($\tau$), the genetic algorithm was the best for small and medium sample sizes and showed a clear superiority to the (IR) algorithm in large samples. The (IR) algorithm was also the best in most sample sizes in estimating the torsion parameter ($\omega$).

Table (2) when ($\varepsilon = 0, \tau = 1, \omega = 0.5$)

| n   | Method  | Mean   | Mse    | Mean   | Mse    | Mean   | Mse    | Def  |
|-----|---------|--------|--------|--------|--------|--------|--------|------|
| 20  | NR (ML) | 0.07013| 0.01918| 0.16306| 0.38897| 0.00082| 0.0000001| 0.40815|
|     | IR (ML) | 0.48713| 0.00087| 0.26039| 0.31977| 0.00056| 0.00000000| 0.32064|
|     | NM(ML)  | 0.82472| 0.01725| 0.25464| 0.37505| 1.78511| 0.034470| 0.42677|
|     | GA(ML)  | 0.18455| 0.00189| 0.23213| 0.06539| 0.24330| 0.000588| 0.06787|
| 50  | NR (ML) | 0.05415| 0.00517| 0.25455| 0.25966| 0.00017| 0.0000009| 0.26483|
|     | IR (ML) | 0.40052| 0.00023| 0.32213| 0.21531| 0.00004| 0.0000004| 0.21554|
|     | NM(ML)  | 1.20986| 0.00000| 0.11835| 0.33734| 0.91117| 0.00057| 0.33740|
|     | GA(ML)  | 0.19631| 0.00097| 0.25929| 0.11182| 0.26339| 0.000697| 0.11349|
| 10  | NR (ML) | 0.05911| 0.00277| 0.16689| 0.27938| 0.00036| 0.0000008| 0.28215|
|     | IR (ML) | 0.37505| 0.00022| 0.27083| 0.25148| 0.00010| 0.0000001| 0.25171|
|     | NM(ML)  | 1.16554| 0.00049| 0.39279| 0.34523| 1.17777| 0.000537| 0.34626|
|     | GA(ML)  | 0.23917| 0.00082| 0.30770| 0.24780| 0.32660| 0.001065| 0.24969|
| 20  | NR (ML) | 0.10551| 0.00123| 0.30043| 0.24691| 0.00023| 0.0000002| 0.24814|
|     | IR (ML) | 0.38948| 0.00005| 0.26069| 0.22893| 0.00001| 0.0000004| 0.22899|
|     | NM(ML)  | 1.45815| 0.00223| 0.70790| 0.31314| 2.22215| 0.003293| 0.31867|
|     | GA(ML)  | 0.23402| 0.00068| 0.30466| 0.45193| 0.31997| 0.001023| 0.45364|
| 50  | NR (ML) | 0.09397| 0.00051| 0.21850| 0.25727| 0.00043| 0.0000003| 0.25778|

2- In Table (2), when ($\varepsilon = 0, \tau = 1, \omega = 0.5$), we note that the (IR) algorithm is superior to other methods for all sample sizes in calculating the parameter estimation ($\varepsilon$) because it has less (Mse) than the other methods. As for estimating the parameter ($\tau$), the genetic algorithm was the best for small and medium sample sizes and a clear superior to the (IR) algorithm in large samples. The (IR) algorithm was also the best in most sample sizes in estimating the torsion parameter ($\omega$).
In Table (3), when \( \varepsilon = 0, \tau = 1, \omega = 1.5 \), results of (Mse) appear for the parameter \( \varepsilon \) when using the sample sizes \( n = 20, 100, 200 \) given (GA) The lowest value and the better (NM) algorithm for samples \( n = 50, 500 \). In parameter \( \tau \), the (GA) algorithm outperformed all other methods. While the (IR) algorithm got the least (Mse) in large sample size, and the (NR) algorithm in large samples in the calculation of parameter estimation \( \omega \).

As can be seen, (GA) is superior in small and medium samples of the sum of the mean square error (Def) of the distribution parameters, and the (IR) algorithm is the best when using large samples.

### Table (3) when \( \varepsilon = 0, \tau = 1, \omega = 1.5 \)

| N   | Method | Mean | Mse  | Mean | Mse  | Mean   | Mse  | Def   |
|-----|--------|------|------|------|------|--------|------|-------|
| 20  | NR(ML) | 0.003041 | 0.014081 | 0.006488 | 0.387208 | 0.001867 | 0.000285 | 0.401575 |
|     | IR (ML)| 0.093807 | 0.009676 | 0.189816 | 0.337871 | 0.127456 | 0.000125 | 0.347673 |
|     | NM (ML)| 0.882136 | 0.011528 | 0.254647 | 0.375051 | 1.667478 | 0.025463 | 0.412043 |
|     | GA (ML)| 0.183524 | 0.001229 | 0.232137 | 0.065394 | 0.246020 | 0.000290 | 0.066914 |
| 50  | NR (ML)| 0.000821 | 0.003441 | 0.003418 | 0.259173 | 0.000528 | 0.000059 | 0.262674 |
|     | IR (ML)| -0.334597 | 0.011257 | 0.246270 | 0.225384 | 0.061245 | 0.000000 | 0.236642 |
|     | NM (ML)| 1.243161 | 0.000335 | 0.118359 | 0.337342 | 0.866680 | 0.00290 | 0.337969 |
|     | GA (ML)| 0.195269 | 0.000488 | 0.259290 | 0.111822 | 0.265653 | 0.000449 | 0.112761 |
| 100 | NR (ML)| 0.000479 | 0.001943 | 0.001286 | 0.279248 | 0.000302 | 0.000012 | 0.281204 |
|     | IR (ML)| -0.35874 | 0.006400 | 0.203107 | 0.258325 | 0.066639 | 0.000009 | 0.264736 |
|     | NM (ML)| 1.212432 | 0.001020 | 0.392794 | 0.345234 | 1.148462 | 0.00346 | 0.346601 |
|     | GA (ML)| 0.238312 | 0.000412 | 0.307702 | 0.247803 | 0.328794 | 0.000860 | 0.249077 |
| 200 | NR (ML)| 0.240729 | 0.000841 | 0.659226 | 0.246868 | 0.158672 | 0.000003 | 0.247712 |
|     | IR (ML)| -0.57393 | 0.004844 | 0.198027 | 0.233194 | 0.055950 | 0.000004 | 0.238043 |
|     | NM (ML)| 1.517825 | 0.002992 | 0.707907 | 0.313142 | 2.090984 | 0.002981 | 0.319116 |
In Table (4), when ($\varepsilon = 0$, $\tau = 1$, $\omega = 2$) with respect to (Mse), (GA) gives the best estimate for the parameter ($\varepsilon$) in most sample sizes and the average sample size for the parameter ($\tau$). We note also that in the calculation of the parameter estimation ($\omega$), (IR) gave the least output in terms of (Mse) for small and medium sample size and (NR) the best in using a large sample.

The results of (Def) for the total (Mse) of the model parameters showed that (GA) was the best in the small and medium sample sizes, while (IR) was the best in the large-sized samples.

Table (4) when ($\varepsilon=0$ , $\tau=1$ , $\omega=2$)

| N  | Method | $\hat{\varepsilon}$ Mean | $\hat{\varepsilon}$ Mse | $\hat{\tau}$ Mean | $\hat{\tau}$ Mse | $\hat{\omega}$ Mean | $\hat{\omega}$ Mse | Def  |
|----|--------|-------------------------|------------------------|------------------|-----------------|------------------|-----------------|------|
| 20 | NR (ML) | 0.003316 | 0.013959 | 0.007061 | 0.007061 | 0.002231 | 0.002231 | 0.401344 |
|    | IR (ML) | 0.091815 | 0.009674 | 0.184248 | 0.339320 | 0.143655 | 0.000175 | 0.349170 |
|    | NM (ML) | 0.883569 | 0.011405 | 0.254647 | 0.375051 | 1.657697 | 0.024708 | 0.411165 |
|    | GA (ML) | 0.183431 | 0.001215 | 0.232137 | 0.065394 | 0.246268 | 0.000267 | 0.066878 |
| 50 | NR (ML) | 0.000909 | 0.003373 | 0.003766 | 0.259123 | 0.000655 | 0.000072 | 0.262569 |
|    | IR (ML) | -0.34539 | 0.011461 | 0.239420 | 0.226305 | 0.070619 | 0.000001 | 0.237768 |
|    | NM (ML) | 1.246473 | 0.000365 | 0.118359 | 0.337342 | 0.862116 | 0.000326 | 0.338035 |
|    | GA (ML) | 0.195159 | 0.000471 | 0.259290 | 0.111822 | 0.265894 | 0.000426 | 0.112720 |
| 100| NR (ML) | 0.000528 | 0.001916 | 0.001414 | 0.279234 | 0.000370 | 0.000015 | 0.281166 |
|    | IR (ML) | -0.36777 | 0.006497 | 0.197035 | 0.258943 | 0.076428 | 0.000132 | 0.265454 |
|    | NM (ML) | 1.214141 | 0.001042 | 0.392794 | 0.345234 | 1.145518 | 0.003280 | 0.346605 |
|    | GA (ML) | 0.238225 | 0.000401 | 0.307702 | 0.247803 | 0.329018 | 0.000841 | 0.249045 |
| 200| NR (ML) | 0.267190 | 0.000825 | 0.730450 | 0.246863 | 0.198158 | 0.000004 | 0.247692 |
5. In Table (5), when ($\epsilon = 0$, $\tau = 1$, $\omega = 5$) the results of (Mse) appear for the parameter ($\epsilon$) that the algorithm (GA) and (NM) are the best, and it was noticed that in the parameter ($\tau$) the (GA) was the best in medium sample and the (IR) algorithm was best in large samples for the parameter ($\tau$), also in calculating the shape parameter estimator ($\omega$) in small and medium samples, and the (NR) algorithm was the best in large samples.

As usual, (GA) was the best at sample size ($n = 20, 50, 100$) to estimate all model parameters because it gave less (Def) and (IR) was the best in large samples.

Table (5) when ($\epsilon = 0$, $\tau = 1$, $\omega = 5$)

| n  | Method | $\hat{\epsilon}$ Mean | Mse | $\hat{\tau}$ Mean | Mse | $\hat{\omega}$ Mean | Mse | Def |
|----|--------|------------------------|-----|-------------------|-----|----------------------|-----|-----|
| 20 | NR (ML)| 0.003316               | 0.013959 | 0.007061       | 0.007061 | 0.002231           | 0.002231 | 0.401344 |
|    | IR (ML)| 0.091815               | 0.009674 | 0.184248       | 0.339320 | 0.143655           | 0.000175 | 0.349170 |
|    | NM (ML)| 0.883569               | 0.011405 | 0.254647       | 0.375051 | 1.657697           | 0.024708 | 0.411165 |
|    | GA (ML)| 0.183431               | 0.001215 | 0.232137       | 0.065394 | 0.246268           | 0.000267 | 0.066878 |
| 50 | NR (ML)| 0.000909               | 0.003373 | 0.003766       | 0.259123 | 0.000655           | 0.000072 | 0.262569 |
|    | IR (ML)| -0.34539               | 0.011461 | 0.239420       | 0.226305 | 0.070619           | 0.000001 | 0.237768 |
|    | NM (ML)| 1.246473               | 0.000365 | 0.118359       | 0.337342 | 0.862116           | 0.000326 | 0.338035 |
|    | GA (ML)| 0.195159               | 0.000471 | 0.259290       | 0.111822 | 0.265894           | 0.000426 | 0.112720 |
|100 | NR (ML)| 0.000528               | 0.001916 | 0.001414       | 0.279234 | 0.000370           | 0.000015 | 0.281166 |
|    | IR (ML)| -0.36777               | 0.006497 | 0.197035       | 0.258943 | 0.076428           | 0.000013 | 0.265454 |
|    | NM (ML)| 1.214141               | 0.001042 | 0.392794       | 0.345234 | 1.145518           | 0.000328 | 0.346605 |
|    | GA (ML)| 0.238225               | 0.000401 | 0.307702       | 0.247803 | 0.329018           | 0.000841 | 0.249045 |
In Table (6), when \((\varepsilon = 0, \tau = 1, \omega = 4)\), we note the superiority of (GA) in terms of (Mse) in most sample sizes of parameter \((\varepsilon)\) and it was also the best in small and medium-sized samples in Calculation of parameter estimation \((\tau)\) and \((\text{IR})\) is better in large samples of parameter \((\tau)\) and in small and medium sample sizes of skewness parameter \((\omega)\). While (NR) got less (Mse) when the sample size \((n = 200,500)\) for parameter \((\omega)\).

The (Def) results like the previous ones showed that (GA) is the best among other methods for model estimation at sample size \((n = 20,50,100)\), while (IR) is better when using large data.

Table (6) when \((\varepsilon=0, \tau=1, \omega=4)\)
|       | NM (ML)          | GA (ML)          | 200 NR (ML)       | IR (ML)          | 200 NM (ML)       | GA (ML)          | 500 NR (ML)       | IR (ML)          | 500 NM (ML)       | GA (ML)          |
|-------|------------------|------------------|-------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
|       | 1.210116         | 0.238188         | 0.285316          | -0.55586         | 1.517109         | 0.233281         | 0.121373         | -0.75710         | 1.01269          | 0.211844         |
|       | 0.000991         | 0.000430         | 0.000844          | 0.004677         | 0.002982         | 0.003171         | 0.00360          | 0.002793         | 0.002020         | 0.000453         |
|       | 0.392794         | 0.307702         | 0.781672          | 0.181014         | 0.707907         | 0.304668         | 0.276677         | 0.162985         | 1.164084         | 0.274592         |
|       | 0.345234         | 0.247803         | 0.246859          | 0.234357         | 0.313142         | 0.451937         | 0.257260         | 0.249932         | 0.322534         | 1.224832         |
|       | 1.144254         | 0.329114         | 0.222111          | 0.071326         | 2.071133         | 0.321786         | 0.094852         | 0.06615          | 1.473151         | 1.224832         |
|       | 0.000321         | 0.000832         | 0.000004          | 0.000008         | 0.002925         | 0.000849         | 0.000006         | 0.000004         | 0.000004         | 0.000072         |
|       | 0.346546         | 0.249066         | 0.247708          | 0.239043         | 0.319050         | 0.453103         | 0.257621         | 0.252729         | 0.323616         | 1.226005         |

9. Conclusions
The ML results of the (SND) parameters were improved by employing the genetic algorithm (GA) and iterative methods (NR, NM and IR) by extracting the results from the simulation experiment and then analysing the results so, the following conclusions were reached:

1- The results of the genetic algorithm (GA) in estimating the model is superior to the skew normal distribution (SND) at the sample sizes (n = 20, 50, 100) because it has the lowest (Def) value among the methods used regardless of the skewness value.

2- The results of the iterative re-weighting algorithm (IR) of the model were superior to the sample sizes (n = 200, 500) for having the lowest value of the (Def) criterion among the methods used.

3- We note that the (GA) genetic algorithm was the best in most cases when calculating the estimation of the location parameter (𝜖) at the sample size (n = 20, 100, 200) and the (NR) algorithm was better when the sample size (n = 20, 100, 200) and The (IR) algorithm has a clear advantage when the torsion size is very slight.

4- We found that when calculating the estimator of the measurement parameter (τ), it exceeds (GA) at sample sizes (n = 20, 50, 100), and the higher the sample size, the better the (IR) algorithm because it has less (Mse).

5- When calculating the estimator of the shape parameter (𝜔), we concluded that at small and medium sample sizes (n = 20, 50, 100) the (IR) algorithm was the best, but for large sample sizes the (NR) algorithm was better in calculating the capabilities of the parameter (𝜔).

References
[1] Al-Douri Y, Hamodi H and Lundberg J 2020 Risk-based life cycle cost analysis using a two-level multi-objective genetic algorithm International Journal Of Computer Integrated Manufacturing
[2] Ashour S, Abdel-hameed M 2010 Approximate skew normal distribution Journal of Advanced Research 1 341–350
[3] AZZALINI A 1985 A Class of Distributions which Includes the Normal Ones Scand J Statist 12 171-178
[4] Barati R 2011 Parameter Estimation of Nonlinear Muskingum Models Using Nelder-Mead Simplex Algorithm *Journal Of Hydrologic Engineering*

[5] Celik N, Senoglu B and Arslan O 2015 Estimation and Testing in One-Way ANOVA when the Errors are Skew-Normal *Revista Colombiana de Estadística* Vol 38 Issue 1 pp 75-91

[6] Elvira O, Jaen C, Morinigo S, Morales Velazquez, Osornio R and Romero T 2020 Genetic Algorithm Methodology for the Estimation of Generated Power and Harmonic Content in Photovoltaic Generation *Appl. Sci* 10 542

[7] Kılıç M 2020 Using Genetic Algorithms for Parameter Estimation of a Two-Component Circular Mixture Model *H. Dutta et al AISC IIII* pp 99–110

[8] Yalçınkaya A, Şenoglu B And Yolcu U 2017 Maximum Likelihood Estimation For The Parameters Of Skew Normal Distribution Using Genetic Algorithm *Swarm and Evolutionary Computation*