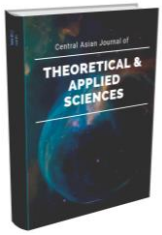

CENTRAL ASIAN JOURNAL OF THEORETICAL AND APPLIED SCIENCE

https://cajotas.centralasianstudies.org/index.php/CAJOTAS
Volume: 05 Issue: 04 | July 2024 ISSN: 2660-5317

Article

Comparison Of Data Collection Algorithms Using Some Mixed Distributions

Hassanien Adel Salih

General Directorate of Education in Najaf Governorate Najaf, Iraq
* Correspondence: hasaneen.adel24@qu.edu.iq

Abstract: The current research aims to compare different data clustering algorithms, focusing on algorithms that use Mixture distributions. We will discuss how these algorithms work, their advantages and disadvantages, and their efficiency in clustering diverse data in size and structure. A comprehensive analysis will also be conducted by applying these algorithms to multiple data sets to evaluate the performance, efficiency, and accuracy of clustering by using them in Mixture distributions. (Mixture Exponential Distribution, Mixture Weibull Distribution, Mixture Pareto distribution) were chosen as applications to study clustering algorithms. Comparing different data clustering algorithms when using Mixture distributions, which are a type of statistical models that depend on merging several probability distributions to represent data, as well as the widespread use of these algorithms in data analysis and extracting patterns of that data, which makes them a powerful tool in many practical applications such as classification, pattern recognition, and statistical predictions. After a detailed presentation of the different clustering algorithms, the algorithm evaluation mechanism, and some Mixture distributions, the researcher concluded that each algorithm has a work that cannot be dispensed with or replaced and that all algorithms are highly efficient in their field of work if the conditions and specifications of each algorithm are adhered to. Therefore, the researcher recommended dealing with these algorithms, each according to its work, to obtain the best results.

Keywords: Data Clustering Algorithms, Mixture Distributions, K-Means Algorithm, Mean Shift Algorithm, DBSCAN Algorithm, Agglomerative Hierarchical Clustering Algorithm, Expectation And Maximization Algorithm EM.

Citation:. Hassanien Adel Salih. Comparison Of Data Collection Algorithms Using Some Mixed Distributions Central Asian Journal of Theoretical and Applied Science 2024, 5(4), 400-416.

Received: 10th Apr 2024
Revised: 11th Mei 2024
Accepted: 24th Jun 2024
Published: 27th Jul 2024

1. Introduction

The scientific revolution and the development of modern technological capabilities during the past years have led to new methods for dealing with data, including collecting and mining data and other methods accompanying this tremendous scientific development in information technology.

A number of the most widely used and famous clustering algorithms among researchers, programmers and statisticians have been studied, including, but not limited to ((K-Means algorithm, MeanShift algorithm, DBSCAN algorithm, Agglomerative Hierarchical Clustering algorithm, EM expectation and maximisation algorithm)) and other standard algorithms in this field and research on comparing performance through several measures such as accuracy, efficiency and stability.

Mixture distributions are characterised by the ability to model complex data, as the algorithms used in applying these models vary in accuracy, flexibility, and efficiency. The choice of the appropriate algorithm depends on the nature of the data and the purpose of the analysis. This research is considered a step towards a deeper understanding of how to choose and apply appropriate algorithms in different contexts, as the researcher provides an overview of comparing clustering algorithms using Mixture distributions, and the research can be developed more deeply by studying specific cases and practical applications to analyse the performance of each algorithm in multiple scenarios.

This study will provide a scientific and objective comparison that helps researchers and practitioners determine the most appropriate algorithm for use in various data collection applications. Therefore, this research is an essential contribution to understanding the performance of clustering algorithms and guiding researchers towards the most appropriate algorithm for each data type.

2. Materials and Methods

Research problem:

Data clustering is one of the most prominent fields of data science and machine learning, as it aims to divide a large set of data into smaller groups, each of which has common characteristics. These techniques are widely used in data analysis and pattern discovery, and data clustering algorithms are among the most essential tools that help simplify and understand complex structures in data.

With the availability of many different algorithms for data clustering, such as K-Means algorithm, MeanShift algorithm, DBSCAN algorithm, Agglomerative Hierarchical Clustering algorithm, EM algorithm through Mixture distributions such as Mixture exponential distribution, Mixture Weibull distribution and Mixture Pareto distribution, the need arises to compare these algorithms in terms of performance, accuracy, and effectiveness when dealing with different data sets and Mixture distributions.

Importance of the research:

Clustering algorithms in light of Mixture distributions are essential in many fields, including data analysis, artificial intelligence and applied sciences. The importance of the research can be summarised in the following points:

1. Complex data analysis, as clustering algorithms are used to understand and analyse complex data through Mixture distributions that are more powerful in identifying hidden and overlapping patterns in the data.
2. Comparing algorithms is essential in revealing the most effective algorithms in certain circumstances, which improves the accuracy and speed of the performance of the clustering algorithms.
3. In most cases, the data is heterogeneous; therefore, Mixture distributions provide an effective way to interpret this type of data.
4. By comparing clustering algorithms, it is possible to determine which one provides more accurate and reliable results based on the data's nature, which helps avoid errors resulting from the misuse of a particular algorithm.
5. This type of research contributes to enhancing analytical capabilities to deal with modern and complex data, which helps make informed decisions and improve models' efficiency in various fields.

Research objective:

The current research aims to study and analyse the performance of several data collection algorithms when using Mixture distributions and try to understand which of these algorithms is more effective or accurate in collecting unclassified data, in addition to providing recommendations on the use of specific collection algorithms in particular

contexts based on their performance in analysing Mixture distributions. The research objectives included the following:

1. Determine the efficiency and accuracy by comparing algorithms in data collection and dealing with data that contain complex or asymmetric distributions.
2. Test the algorithms on accurate data and observe their performance and how the algorithms work with data that may not necessarily follow ideal probability distributions.
3. Compare the work of each algorithm in terms of stability and adaptation, in addition to dealing with challenges in the data if it is irregular or contains abnormal values.
4. Identify the most effective areas of use for each algorithm.
5. Study the impact of algorithms on different factors such as data size, distribution, and the extent to which performance is affected by changes in these factors.

3. Results and Discussion

The concept of clustering algorithms:

Algorithms can be considered a series of instructions and commands arranged by the programmer, statistician, or others to obtain sequential results. They are also a recipe for professionally processing data types. Many algorithms are adopted in programming processes, but we will deal with clustering algorithms in this research.

Data clustering processes are the process of dividing data and placing it in similar groups. It is one of the branches of data mining processes. The clustering algorithm divides data into several groups, as the similarity between points within a specific group is more excellent than between two points within two groups. The idea of clustering data is simple and very close to the human way of thinking, as the more we deal with a large amount of data, the more we tend to summarise the vast amount of data into a small number of groups or categories, and this is from To facilitate the analysis process (Al-Dhafri, 2024: 38).

The clustering algorithm can also be defined as creating groups of similar or related data. This is done by analysing the data, identifying similar elements, and grouping them into groups. Clustering algorithms are used in various fields, such as machine learning, data science, image processing, deep learning, biological learning, and others. The clustering algorithm aims to organise data, reduce complexity, improve performance, and better understand the data.

Clustering algorithms are widely used to organise, classify, compress, and build a data arrangement model. If we find a data cluster, it is possible to create a model of the problem based on those clusters through which similar or related data can be grouped. This is done by analysing the data, identifying similar elements, and grouping them into coherent groups.

Advantages of clustering algorithms:

Clustering algorithms are of great importance in classifying data into subgroups that can be controlled, and these algorithms have been used in several fields to provide data descriptions, including (Allyn and Robert, 2001:41):

1. Clustering algorithms effectively simplify data sets and create a classification for data sets.
2. Identifying data patterns within groups.
3. Detecting hidden structures in data.
4. Identifying outliers.
5. Compressing the prominent information of those groups (to store and retrieve information efficiently).
6. Analysing their structure in some statistical analyses through automatically clustering the data set.

7. It is possible to conclude the behaviour of the data set without the need for a detailed or even superficial understanding of that data, and this is of particular importance for work in modelling and simulation.
8. Clusters allow for a general prediction of where model parameters tend to be located within the state space and their compatibility with specific behaviour patterns within the simulation.
9. Assisting in decision-making processes.

The most important disadvantages and obstacles of clustering algorithms:

Clustering algorithms are powerful tools in data analysis processes, as they group elements and points in data. Despite their incredible benefits, they are not without some disadvantages, including (Brian et al., 2011):

1. Clustering algorithms are affected by noise and outliers, reflected in their performance.
2. Some algorithms are not suitable for extensive data. They may require a lot of time and computational resources, making them impractical for some applications.
3. Some clustering algorithms, especially those that rely on distance measurements, work well with spherical or convex clusters, but they have difficulty with clusters that have irregular or complex shapes.
4. Similarity criteria vary depending on the type of data and the algorithm used, making the selection of the appropriate criterion challenging in some cases.
5. Clustering results are sometimes difficult to interpret, and there are no clear indications of why the data falls into specific groups.
6. Only binary vectors are capable of being classified. However, they are included due to some advantages of the algorithm (Allyn and Robert, 2001:48).

Types of clustering algorithms:

There are many clustering algorithms that researchers rely on to analyse data, but we will highlight the most important and most commonly used of these algorithms, which are as follows:

1. **K-Means algorithm:** The (K-Means) algorithm is one of the most widely used algorithms in data clustering. This algorithm works by identifying and clustering data in K different groups, where the number of groups K is determined in advance. The algorithm determines the centre of each group and tries to reduce the distance between the elements in the group and the specified centre. The algorithm continues to iterate and improve the clustering until an ideal clustering is reached. The primary step in clustering using K-Means is to determine the number of groups K and assume the center of these groups. We can take any random observations as the initial center, or the first K observations in the sequence can also be the initial center (Puzicha et al., 2000: 605).

   K-Means algorithm steps (Christopher, 2006: 423-428):

   Step 1. We start by deciding on the value of k = the number of clusters, i.e., impose a value of (k).

   Step 2. Set any initial part that divides the data into k clusters, and you can assign samples randomly or systematically as follows:

   - Take the first k training sample as single-element clusters.
   - Assign each remaining sample (N-k) to the cluster with the closest center of gravity. After each assignment, recalculate the center of gravity of the acquired cluster.

   Step 3. Take each sample in sequence and calculate its distance from the center of gravity of each of the clusters. If the sample is not currently in the cluster with the closest center of gravity, swap this sample to that cluster and update the center of gravity of the cluster that gains the new sample and the cluster that loses the sample.

   Step 4. Repeat step 3 until convergence is achieved, i.e., until passing through the sample does not cause new assignments.

2. **MeanShift algorithm:**

The MeanShift algorithm aims at the clustering process, which discovers clusters in the density of a series of samples. It is an algorithm based on the center of gravity, which updates the candidates for the center of gravity to be the average of the points within a particular region. The candidates are filtered in the post-processing stage to remove duplicate repetitions to form the final set of data center of gravity for the central algorithm; in other words, the goal is to determine the location of the central points of each group by updating the sliding window, updating the points within it, and then taking their averages as the center of the group in the case of Mixture distributions, the data can be derived from several elemental distributions. The MeanShift algorithm is used to identify clusters within these Mixture distributions without knowing the number of distributions or their shapes (Chunxia and Meng, 2010).

MeanShift algorithm steps (Abdulmunen, 2012: 46):

Step 1: Choose a set of existing data points, as all or some of the points can be used.

Step 2: Choose the search range and specify the size of the window, whether circular or spherical, surrounding each point.

Step 3: Calculate each point's moving mean (mean-shift) and specify the neighbours within the search range.

Step 4: Calculate the center of mass for all points within the search range.

Step 5: Move the points towards the new center.

Step 6: Repeat steps (2) and (3) until the points stop moving.

Step 7: The points will gather, generating a cluster shape, and each group that approaches each other can be considered a single cluster, and the classification will be based on the centers of the clusters gathered.

Finally, these clusters with their centers will be gathered as points close to each other within a specific range.

3. DBSCAN algorithm:

This algorithm clusters asymmetric data that does not follow the usual pattern. This algorithm works by identifying dense points in the data and grouping them. The algorithm depends on finding the direct neighbours of each end and expanding the clusters based on specific rules.

The DBSCAN algorithm is one of the practical density-based clustering algorithms. It is simple and creates a cluster in a random form. It also collects information such as noise and outliers. The DBSCAN algorithm is also characterised by speed and efficiency with large databases. It has two inputs: the first is the radius, and the second is the minimum number of points in the radius. The general idea of DBSCAN is based on finding the minimum number of points in the radius space to form clusters. Otherwise, if the minimum number of points is not in the radius, it will be marked as noise (Raval and Jani, 2015: 72 – 76). DBSCAN algorithm steps (Schubert, 2017):

Step 1: A point is randomly selected within the data.

Step 2: Determine the points that fall within the selected point.

Step 3: If the surrounding and neighbouring points of the selected point are greater than or equal to (MinPts), it is a fundamental point, and a cluster is formed that includes all points adjacent to the previously selected essential point.

Step 4: If the surrounding and neighbouring points of the selected point are more minor than (MinPts), it is a noise point and does not form any cluster.

Step 5: Repeat the process to ensure all points are added to the cluster.

Step 6: Move to a new unclassified point and repeat the process until all points are classified.

Step 7: End the process after classifying all points as part of a specific cluster within the data or as noise points.

4. Agglomerative Hierarchical Clustering Algorithm:

This algorithm works by merging similar elements to form larger clusters. The algorithm starts by assigning each element to a single cluster and then combines the clusters based on similarity. This process is repeated until large clusters containing all similar elements are formed. Agglomerative hierarchical clustering is a hierarchical clustering algorithm and one of the cluster analysis algorithms used to collect data into groups or clusters. It starts by merging individual points, and each point is a cluster in itself in larger groups until a single cluster that includes all points is formed.

The advantages of this algorithm are that it does not need to specify the number of clusters in advance; in addition to that, it gives a comprehensive view of the data in the form of a hierarchical tree. Still, its disadvantage is that it is expensive in terms of calculations, especially with extensive data. It also depends on the researcher's distance, which may significantly affect the final results if inaccurate (LaPlante, 2015).

Steps of the Agglomerative Hierarchical Clustering Algorithm: (Mok et al., 2012)

Step 1: Each data point is considered a separate cluster from the rest of the data.

Step 2: The distance between the clusters is calculated.

Step 3: The closest clusters whose data is most similar to each other are merged to form a new cluster.

Step 4: The process of calculating distances and merging clusters is repeated until a single cluster, which includes all points, is formed.

Step 5: Representing the merging process as a hierarchical tree.

5. EM Algorithm:

According to the study (Nilashi et al. 2015: 542), EM Algorithm includes applications for calculating maximum likelihood parameters in a statistical model. This application is applicable in cases where it is difficult to solve equations directly, as the models use latent variables and unknown parameters in addition to known data observations. This means that the data either contain missing values, or it is possible to reformat the model more only by assuming the presence of additional unobserved data points and a latent variable that determines a component of the mixture that belongs to each data point. The maximum likelihood solution requires using derivatives of the likelihood function, considering every unknown value, i.e. both the parameters and the latent variables, which involves solving the resulting equations simultaneously. However, this is not usually possible in the case of statistical models containing latent variables. Instead, it results in a set of nested equations where the solution to the parameters must contain the exact values of the latent variables and vice versa. However, while replacing one set of equations, when they are replaced in the other set. The advantages of this algorithm are that it can handle incomplete or missing data effectively and is suitable for a wide range of problems in statistics and machine learning. The disadvantages are that it depends on the choice of initial parameters, and performance may be poor if these parameters are far from the optimal solution. They may sometimes reach a minimum of the objective function and cannot necessarily guarantee reaching the maximum. Steps of the EM (Trevor et al., 2001):

Step 1: Initialization, where the algorithm starts by choosing initial estimates for the parameters we want to estimate. These estimates can be chosen randomly or using any prior knowledge if available.

Step 2: The expectation step where we calculate the expected values for the unobserved data based on the current parameters.

Step 3: The maximisation step, where the expected values calculated in the expectation step are used, updates the parameter estimates by maximising the probability function.

Step 4: The iteration step where the expectation and maximisation steps are repeated until the parameter change stops and reaches a specific stopping value (convergence).

Step 5: Upon reaching the convergence limit, the calculated parameters are the parameter estimates that maximise the probability function based on the observed and expected data.

Applications and uses of assembly algorithms:

Since ancient times, scientists in mathematics have used algorithms to solve complex mathematical equations and problems, and their use continues to this day. Algorithms are integral to mathematical operations, but algorithms have developed significantly to include their use more widely in computer science, statistics, and others. Today, all programs, applications, and operating systems are algorithms that depend directly on data processing, performing calculations and logical analyses, and solving mathematical equations. On the one hand, algorithms are used in medical, engineering, physical, and economic fields from scientific aspects. We will mention some uses, including (Al-Dhafri, 2024):

- Using algorithms in social media.
- Using them in search engines (Google and others).
- They are used in statistical fields to predict the future, such as weather, temperature, rain, etc.
- Detecting criminals through facial recognition.
- Used in data encryption.
- Used in geographic and spatial analysis (GPS).
- Smart robots.
- Artificial intelligence applications.
- Detecting data patterns.
- Genetic data analysis and discovering new types of genes.

Mixed Distributions:

The concept of mixed distributions Mixture Distribution:

In most statistical applications, the probability behaviour of observations is heterogeneous, as these heterogeneous observations are clustered into homogeneous subsets. The observations of subsets within the community may have a probability behaviour with a similar probability density function but with different parameters, i.e. they have the same distribution but with other parameters, and the probability function of these subsets may be different, i.e. they have a different distribution. In this case, the distribution of the total community is a linear combination of the probability density functions of the subsets, which is called the Mixture distribution (Abdul Hussein and Hussein, 2017: 124).

Mixture probability distributions result from mixing multiple probability distributions and can be binary, triple, or more distributions depending on the nature of the data. These distributions are characterised by the fact that they result from more than one different probability distribution or similar distributions with other parameters, which makes them more explanatory in describing the phenomena that result from multiple probability distributions. These distributions have many applications in medicine, engineering, agriculture, industry, and others (Garavaglia et al., 2011). (Al-Bayati, 2012: 11-12) defined Mixture distributions as "a collection of heterogeneous components of statistical data, which occurs when a sample is drawn from a heterogeneous community whose probability functions are different or similar but with different parameters for each partial community. In this case, statistical tests are required to determine whether the Mixture distribution is from the same family instead of the single distribution.

Some mixed distributions:

There are many Mixture distributions, but our study will be limited to three Mixture distributions, which are:

Mixed exponential distribution: Mixture Exponential Distribution:

It is a Mixture probability model, called the Mixture exponential distribution, and its probability and cumulative density function are as follows (Al-Duri and Bahiya, 2018: 2-3):

$$f(z,\varphi) = \sum_{j=1}^{k} \frac{1}{\lambda_i} \exp\left(-z/\lambda_i\right) p_i \qquad i = 1,2,\dots,k \qquad \dots (3)$$

Whereas:

$\varphi = (\lambda_1, \dots, \lambda_k, p_1, \dots, p_k)$

λ_i : represents the scaling parameter of the exponential distribution.

p_i: represents the mixing ratio parameter for compound i.

$0 < p_i < 1 \qquad \sum_{i=1}^{k} p_i = 1 \qquad \int f(x)\, d(x) = 1$

Mixture Weibull Distribution :

The Mixture Weibull distribution is a statistical model that combines two or more Weibull distributions. This model accurately describes data that exhibits complex or multi-model behaviour. Generally, it is a probability distribution commonly used in reliability data analysis. The Mixture Weibull distribution is characterised by a probability density function in the following form (Al-Wakeel, 2010: 150-152):

$$f(x) = w\left\{\frac{\beta_1 x^{\beta_1 - 1}}{\alpha_1^{\beta_1}} e^{-\left[\left(\frac{x}{\alpha_1}\right)^{\beta_1}\right]}\right\} + (1-w)\left\{\frac{\beta_2 (x-y)^{\beta_2 - 1}}{\alpha_2^{\beta_2}} e^{-\left[\left(\frac{x-y}{\alpha_2}\right)^{\beta_2}\right]}\right\}$$

Mixture Pareto Distribution

It is a Mixture probability model called the general Mixture Pareto distribution. It is a probability distribution often used to model phenomena characterised by a heavy-tailed distribution, meaning that the probability of large values is more significant than what might be expected from other distributions. The probability density function and the aggregate are given in the following form (Garavaglia et al., 2011: 521-524):

$$f(z,\varphi) = \sum_{i=1}^{k} \frac{1}{\lambda_i} \left(1 + \varepsilon_i \frac{y}{\lambda_i}\right)^{-\left(\frac{1+\varepsilon_i}{\varepsilon_i}\right)} p_i \quad , \quad i = 1,2,\dots,k$$

Estimating mixed distributions:

Estimating the mixed exponential distribution:

Estimating the mixed exponential distribution using the Maximum Likelihood Estimation method:

Assuming $(x_1, x_2, \dots, x_n)$Observations follow a Mixture exponential distribution, so its estimate will be as follows: (Bhat et al.,2018:44):

$$1 = \sum_{i=1}^{n} log\left[p\lambda e^{-\lambda x_i} + \frac{(1-p)\lambda^{\omega+1} X_i^{\omega} e^{-\lambda X_i}}{\Gamma(\omega+1)}\right]$$

Estimates of the unknown parameters are obtained by partially differentiating the parameters of interest and setting them equal to zero; we get the following equation:

$$\frac{\partial l}{\partial \lambda} = \sum_{i=1}^{n}\left[\frac{\frac{\partial}{\partial \lambda}\left\{p\lambda e^{-\lambda x_i} + \frac{(1-p)\lambda^{(\omega+1)} X_i^{\omega} e^{-\lambda X_i}}{\Gamma(\omega+1)}\right\}}{p\lambda e^{-\lambda x_i} + \frac{(1-p)\lambda^{(\omega+1)} X_i^{\omega} e^{-\lambda X_i}}{\Gamma(\omega+1)}}\right] = 0$$

$$\frac{\partial l}{\partial \omega} = \sum_{i=1}^{n}\left[\frac{\frac{\partial}{\partial \omega}\left\{p\lambda e^{-\lambda x_i} + \frac{(1-p)\lambda^{(\omega+1)} X_i^{\omega} e^{-\lambda X_i}}{\Gamma(\omega+1)}\right\}}{p\lambda e^{-\lambda x_i} + \frac{(1-p)\lambda^{(\omega+1)} X_i^{\omega} e^{-\lambda X_i}}{\Gamma(\omega+1)}}\right] = 0$$

$$\frac{\partial l}{\partial P} = \sum_{i=1}^{n}\left[\frac{\frac{\partial}{\partial P}\left\{p\lambda e^{-\lambda x_i} + \frac{(1-p)\lambda^{(\omega+1)} X_i^{\omega} e^{-\lambda X_i}}{\Gamma(\omega+1)}\right\}}{p\lambda e^{-\lambda x_i} + \frac{(1-p)\lambda^{(\omega+1)} X_i^{\omega} e^{-\lambda X_i}}{\Gamma(\omega+1)}}\right] = 0$$

The above system of equations is non-linear and cannot be solved analytically. The Newton-Raphson procedure is implemented to obtain the parameter estimates to overcome this drawback.

Estimating the mixed exponential distribution using the least squares method:

This method is based on the existence of a regression relationship between the experimental F and the F distribution, taking into account the ordered observations. ($x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$) versus experimental distribution, $F(x_{(i)}) \equiv \frac{i}{n+1}$ The equation can be given as follows (Yilmaz and Buse,2015:56):

$$Q(\emptyset) = \sum_{i=1}^{n} (F(x_{(i);}\emptyset) - \hat{F}(x_{(i)}))^2$$

$$\frac{dQ}{d\alpha} = \sum_{i=1}^{n} \left(\frac{i}{n+1} - \alpha \left(1 - e^{-x_{(i)}\frac{1}{\theta_1}}\right) - (1-\alpha)\left(1 - e^{-x_{(i)}\frac{1}{\theta_2}}\right)\right)\left(e^{-x_{(i)}\frac{1}{\theta_1}} - e^{-x_{(i)}\frac{1}{\theta_2}}\right) = 0$$

$$\frac{dQ}{d\theta_1} = \sum_{i=1}^{n} \left(\frac{i}{n+1} - \alpha \left(1 - e^{-x_{(i)}\frac{1}{\theta_1}}\right) - (1-\alpha)\left(1 - e^{-x_{(i)}\frac{1}{\theta_2}}\right)\right)\left(\frac{\alpha x_{(i)}}{\theta_1^2} e^{-x_{(i)}\frac{1}{\theta_1}}\right) = 0$$

$$\frac{dQ}{d\theta_2} = \sum_{i=1}^{n} \left(\frac{i}{n+1} - \alpha \left(1 - e^{-x_{(i)}\frac{1}{\theta_1}}\right) - (1-\alpha)\left(1 - e^{-x_{(i)}\frac{1}{\theta_2}}\right)\right)\left(\frac{(1-\alpha) x_{(i)}}{\theta_2^2} e^{-x_{(i)}\frac{1}{\theta_2}}\right) = 0$$

Since the equations obtained after differentiation are related to θ, it isn't easy to get solutions. Therefore, it is necessary to use numerical methods.

Estimating the mixed Weibull distribution:

The probability function of the Mixture Weibull distribution can be shown as follows (Al-Wakeel et al., 2009: 6-7):.

$$L(\alpha, \beta) = \prod_{i=1}^{n} f(X_1, \alpha, \beta)$$

Estimating the mixed Weibull distribution using the Maximum Likelihood Method:

The maximum likelihood estimator (MLE) for a two-parameter Weibull distribution can be obtained by solving the equations resulting from setting the partial derivatives of L(α,β) equal to zero.

$$L(\alpha, \beta) = \prod_{i=1}^{n} f(\frac{\beta X_i^{\beta-1}}{\alpha^{\beta}}) e^{-(\frac{X_i}{\alpha})^{\beta}}$$

$$\frac{\partial lnL}{\partial \beta} = \frac{n}{\beta} + \sum_{i=1}^{n} lnX_i - \frac{1}{\alpha}\sum_{i=1}^{n} X_i^{\beta} \ln X_i = 0$$

$$\frac{\partial lnL}{\partial \alpha} = \frac{n}{\alpha} + \frac{1}{\alpha^2}\sum_{i=1}^{n} X_i^{\beta} = 0$$

Then $\hat{\beta}$ Is the solution of :

$$\frac{\sum_i^n (X_i^{\hat{\beta}} lnX_i)}{\sum_{i=1}^n X_i^{\hat{\beta}}} - \frac{1}{\hat{\beta}} - \frac{1}{n}\sum_{i=1}^{n} lnX_i = 0$$

When the shape parameter is estimated, the scale parameter can also be calculated by:

$$\hat{\alpha} = (\frac{1}{n}\sum_{i=1}^{n} X_i^{\beta})^{\frac{1}{\hat{\beta}}}$$

As for the three-parameter Mixture Weibull distribution, its estimate will be as follows (Al-Wakeel et al., 2009: 7):

$$L(\alpha, \beta, \gamma) = \frac{\beta^n}{\alpha^{\beta\alpha}} \left[\prod_{i=1}^{n} (X_i - \gamma)^{\beta-1} \right] e^{\frac{1}{\alpha^{\beta}} \sum_{i=1}^{n} (X_i - \gamma)^{\beta}}$$

By adjusting the partial derivative of the previous equation concerning α ,β and γ On zero and by solving the following set of equations simultaneously, we obtain the following estimate:

$$\hat{\alpha}^{\hat{\beta}} - \frac{1}{n}\sum_{i=1}^{n} (X_i - \hat{\gamma})^{\hat{\beta}} = 0$$

$$\frac{\sum_{i=1}^n (X_i - \hat{\gamma})^{\hat{\beta}} ln(X_i - \hat{\gamma})}{\sum_{i=1}^n (X_i - \hat{\gamma})^{\hat{\beta}}} - \frac{1}{\beta} - \frac{1}{n}\sum_{i=1}^{n} ln(X_i - \hat{\gamma}) = 0$$

$$\left(\hat{\beta}-1\right)\sum_{i=1}^{n}(X_i-\ddot{y})^{-1}-\hat{\beta}\hat{\alpha}^{-\hat{\beta}}\sum_{i=1}^{n}(X_i-\hat{y})^{\hat{\beta}-1}=0$$

Estimating the mixed Weibull distribution using the least squares method (Al-Wakeel et al., 2009: 8-9):

This linear equation is as follows:

$$\ln\ln\left[\frac{1}{1-F(x)}\right]=\beta\ln x-\beta\ln\alpha$$

Then:

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}ln\left\{ln\left[\frac{1}{(1-\frac{i}{n+i})}\right]\right\}$$

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}\ln x_i$$

$$\beta=\frac{\left\{n\sum_{i=1}^{n}(\ln x_i).\left(ln\left[ln\left[\frac{1}{(1-\frac{i}{n+1})}\right]\right]\right)\right\}-\left\{\sum_{i=1}^{n}ln\left[\ln\left(\frac{1}{1-\frac{i}{n+1}}\right]\sum_{i=1}^{n}\ln x_i\right\}\right.}{\{n\sum_{i=1}^{n}(\ln x_i)^2\}-\{\sum_{i=1}^{n}(\ln x_i)\}^2}$$

Then:

$$\hat{\alpha}=e^{(\frac{\bar{y}-\bar{x}}{\hat{\beta}})}$$

Least squares estimates of the Weibull distribution can be obtained. α, β By choosing the estimates, which are as follows:

$$\hat{\alpha}(\gamma)=\left\{\prod_{i=1}^{n}[X_i-\gamma]\right\}^{\frac{1}{n}}\left\{\prod_{i=1}^{n}\left[-\log\left[E(\mu_i)\right]\right]\right\}^{\frac{1}{n}\beta(\gamma)}$$

And

$$\hat{\beta}(\gamma)=\frac{\sum_{i=1}^{n}(R_i-\bar{R}_i)[\log(x_i-\gamma)-(logx)^t]}{[\sum_{i=1}^{n}(\log(x_i-\gamma)-(logX)^t]^2}\qquad where\ i=1,2,\dots.,n$$

$$R_i=\log[-\log E(\mu_i)]\quad and\quad \bar{R}=\frac{1}{n}\left(\sum_{i=1}^{n}R_i\right)\ and\ \ (logx)^t=(\frac{1}{n})\left\{\sum_{i=1}^{n}\log(x_i-\gamma)\right\}$$

Estimating the mixed Pareto distribution:

Estimating the mixed Pareto distribution using the maximum likelihood method:

In this topic, we will discuss the maximum likelihood method for estimating the Pareto distribution as follows:

Let's assume$\theta=(\alpha,\beta,\omega)$ It is a vector of model parameters$(X_1,X_2,\dots,X_n)$ (Random variables and (θ) minimise the logarithmic probability function as follows (Nareerat et al., 2015: 203-204):

$$L(\theta)=\prod_{i=1}^{n}\left\{\frac{1}{\beta}\left(\frac{X_i}{\beta}\right)^{-(\alpha+1)}\left[(1-\omega)\alpha+\frac{\omega(\alpha-1)X_i}{\beta}\right]\right\}$$

$$logL(\theta)=\sum_{i=1}^{n}log\left[1-\omega)\alpha+\frac{\omega(\alpha-1)X_i}{\beta}\right]^{-1}-nlog\beta-(\alpha+1)\sum_{i=1}^{n}log\left(\frac{x_i}{\beta}\right)$$

The components corresponding to the model parameters are calculated, and the results are normalised to zero. We will obtain the following equation:

$$\frac{\partial logL(\theta)}{\partial\alpha}=\sum_{i=1}^{n}\left[\alpha-\frac{\omega X_i}{\beta-\omega\beta+\omega X_i}\right]^{-1}-\sum_{i=1}^{n}\log\left(\frac{X_i}{\beta}\right)$$

$$\frac{\partial logL(\theta)}{\partial\omega}=\sum_{i=1}^{n}\left[\omega+\frac{\alpha\beta}{\alpha X_i-X_i-\alpha\beta}\right]^{-1}$$

since it's $X\geq\beta$ Maximum probability estimate β It is a first-class statistic $X_{(1)}$ The maximum probability estimates for the parameters α and ω .We use the command nlm in the program R statistical package.

Estimating the mixed Pareto distribution using the Least Squares (LS) and Weighted Least Squares (WLS):

The LS method minimises the squared difference between the experimental and theoretical CDFs. This method is easy to implement and interpret, making it accessible for various applications. However, LS assumes homogeneity, which is difficult to achieve, and here we need the weighting scheme used in the WLS method to heteroscedasticity by assigning larger weights to observations closer to the center of the sample and smaller weights to observations closer to the edges of the sample. In addition, both LS and WLS methods are computationally intensive, as they rely on the CDF, which needs to be calculated numerically.

Least Squares (LS) Estimator of $\boldsymbol{\theta}$: (Frederico and Mina, 2024)

$\hat{\theta}^{LS} = \left(\hat{c}^{LS}, \hat{\lambda}_1^{LS}, \hat{\lambda}_2^{LS}\right)$ *can be obtained as* :

$$\hat{\theta}^{LS} = argmin\left\{\sum_{i=1}^{n}\left(F\left(x_{(i)}\backslash\theta\right) - \frac{i}{n+1}\right)^2\right\}$$

Furthermore, the parameters were estimated using the weighted least squares (WLS) method, which is denoted by

$\hat{\theta}^{WLS} = \left(\hat{c}^{WLS}, \hat{\lambda}_1^{WLS}, \hat{\lambda}_2^{WLS}\right)$ *can be determined by*:

$$\hat{\theta}^{WLS} = argmin\left\{\sum_{i=1}^{n}\frac{(n+1)^2(n+2)}{i(n-i+1)}\left(F\left(x_{(i)}\backslash\theta\right) - \frac{i}{n+1}\right)^2\right\}$$

Simulation Study:

The clustering algorithms were applied through Mixture distributions on a set of data, which is unrealistic data generated for the study. Its sample size was (40, 80, 120, 160, 200) n = single, and it was applied with parameters (α=1) and the value of (β) was changed to be (0.5, 1, 2) respectively. Note Table (1), (2), and (3), respectively, noting that the same data was applied to all types of Mixture distributions to measure the effect of clustering algorithms on types of Mixture distributions, and the results were as follows in order:

Simulation application on the Mixture exponential distribution according to the parameters and sample size: Data were taken to simulate the Mixture exponential distribution with parameters (β=0,5, α=1), (β=1, α=1)and (β=2, α=1) (and with sample sizes of (n= 40, 80, 120, 160, 200)) and the results were as shown in Table (1) below.

Table (1) Simulation of clustering algorithms and samples taken from the mixed exponential distribution

Estimation of Mixture exponential distribution with parameters β = 0.5, α=1 Through clustering algorithms					
n	K-Means	MeanShift	DBSCAN	A. H. C.	EM
40	0.5010	0.5015	0.5345	0.6022	1.0637
80	0.4935	0.4988	0.5189	0.5714	1.0451
120	0.4912	0.4989	0.5168	0.5611	1.0391
160	0.4846	0.4986	0.5081	0.5358	1.0224
200	0.4848	0.4997	0.5008	0.5202	1.0128
Estimation of Mixture exponential distribution with parameters β = 1, α=1 Through clustering algorithms					
n	K-Means	MeanShift	DBSCAN	A. H. C.	EM
40	1.0174	1.0127	1.0891	1.2502	1.4356
80	1.0116	1.0042	1.0689	1.1989	1.3941
120	0.9894	0.9986	1.0536	1.1636	1.3616
160	0.9867	0.9987	1.0403	1.1402	1.3428
200	0.9861	0.9937	1.0314	1.1234	1.3281

Estimation of Mixture exponential distribution with parameters β = 2, α=1 Through clustering algorithms					
n	K-Means	MeanShift	DBSCAN	A. H. C.	EM
40	4.0580	4.0971	4.3751	4.9782	4.6165
80	3.9909	4.0506	4.2492	4.8226	4.5363
120	3.9613	4.0104	4.2008	4.6629	4.4321
160	3.9497	3.9628	4.1440	4.5641	4.3710
200	3.9285	3.9787	4.1361	4.5300	4.3645

Source: Prepared by the researcher based on the results of the clustering algorithms and samples taken from the Mixture exponential distribution.

Applying the simulation to the mixed Weibull distribution according to the parameters and sample size:

Data were taken simulating a Mixture Weibull distribution with parameters (β=0,5, α=1), (β=1, α=1)and (β=2, α=1) (With sample sizes (n= 40, 80, 120, 160, 200) the results were as shown in Table (2) below.

Table (2) Simulation of clustering algorithms and samples taken from the mixed Weibull distribution

Estimation of the Mixture Weibull distribution with parameters β = 0.5, α=1 Through clustering algorithms					
n	K-Means	MeanShift	DBSCAN	A. H. C.	EM
40	0.5109	0.5078	0.5436	0.6305	1.0816
80	0.5010	0.5015	0.5345	0.6022	1.0637
120	0.4979	0.5029	0.5289	0.5871	1.0534
160	0.4935	0.4988	0.5189	0.5714	1.0451
200	0.4912	0.4989	0.5168	0.5611	1.0391
Estimation of the Mixture Weibull distribution with parameters β = 1, α=1 Through clustering algorithms					
n	K-Means	MeanShift	DBSCAN	A. H. C.	EM
40	1.0174	1.0127	1.0891	1.2502	1.4356
80	0.9867	0.9987	1.0403	1.1402	1.3428
120	0.9709	0.9938	1.0107	1.0715	1.2709
160	0.9745	0.9951	1.0025	1.0416	1.2313
200	0.9791	0.9965	1.0012	1.0232	1.2022
Estimation of the Mixture Weibull distribution with parameters β = 2, α=1 Through clustering algorithms					
n	K-Means	MeanShift	DBSCAN	A. H. C.	EM
40	3.9909	4.0506	4.2492	4.8226	4.5363
80	3.9613	4.0104	4.2008	4.6629	4.4321
120	3.8721	3.9817	4.0418	4.2893	4.2038
160	3.8739	3.9894	4.0209	4.1789	4.1349
200	3.8939	3.9896	3.9949	4.1013	4.0772

Source: Prepared by the researcher based on the results of the clustering algorithms and samples taken from the Mixture exponential distribution.

Simulation application on the mixed Pareto distribution according to the parameters and sample size:

Data were taken simulating a Mixture Pareto distribution with parameters (β=0,5, α=1), (β=1, α=1)and (β=2, α=1) (With sample sizes (n= 40, 80, 120, 160, 200) the results were as shown in Table (3) below.

Table (3) Simulation of clustering algorithms and samples taken from the mixed Pareto distribution

Estimation of Mixture Pareto distribution with parameters $\beta = 0.5, \alpha=1$ Through clustering algorithms					
n	K-Means	MeanShift	DBSCAN	A. H. C.	EM
40	0.4896	0.4986	0.5156	0.5571	1.0356
80	0.4874	0.4978	0.5040	0.5272	1.0169
120	0.4880	0.4993	0.5000	0.5183	1.0111
160	0.4872	0.4983	0.5003	0.5126	1.0082
200	0.4883	0.4994	0.4988	0.5099	1.0064
Estimation of Mixture Pareto distribution with parameters $\beta = 1, \alpha=1$ Through clustering algorithms					
n	K-Means	MeanShift	DBSCAN	A. H. C.	EM
40	0.9793	0.9943	1.0265	1.1145	1.3186
80	0.9713	0.9959	1.0039	1.0580	1.2512
120	0.9712	0.9994	0.9999	1.0320	1.2185
160	0.9791	0.9965	1.0012	1.0232	1.2022
200	0.9759	0.9969	1.0005	1.021	1.1938
Estimation of Mixture Pareto distribution with parameters $\beta = 2, \alpha=1$ Through clustering algorithms					
n	K-Means	MeanShift	DBSCAN	A. H. C.	EM
40	3.9143	3.9774	4.1161	4.4426	4.3002
80	3.8895	3.9792	4.0276	4.2194	4.1599
120	3.9039	3.9820	4.0067	4.1358	4.0977
160	3.8939	3.9896	3.9949	4.1013	4.0772
200	3.9071	4.0021	3.9994	4.0788	4.0602

Source: Prepared by the researcher based on the results of the clustering algorithms and samples taken from the Mixture Pareto distribution.

Study of actual data:

The clustering algorithms were applied through Mixture distributions on a set of data, which is actual data taken from Al-Sadr Teaching Hospital in Najaf Governorate for people with diabetes for the study. Its sample size was the same as the sample in the applied example (simulation) and was n= (40, 80, 120, 160, 200) single, and it was applied with parameters (α=1) and the value of (β) was changed to be (0.5, 1, 2) respectively. See Table (4), (5), (6) respectively. The same data sample size and parameters α, β were applied to all types of Mixture distributions to measure the effect of clustering algorithms on types of Mixture distributions. The results were as follows:

Applying actual data to the mixed exponential distribution according to the parameters and sample size:

Actual data were taken that follow a Mixture exponential distribution with parameters (β=0,5, α=1), (β=1, α=1)and (β=2, α=1) (With sample sizes (n= 40, 80, 120, 160, 200) the results were as shown in Table (3) below.

Table (4) Clustering algorithms and samples taken from the mixed exponential distribution

Estimation of Mixture exponential distribution with parameters β = 0.5, α=1 Through clustering algorithms					
n	K-Means	MeanShift	DBSCAN	A. H. C.	EM
40	0.9866	0.9637	0.9765	1.2520	1.1350
80	0.9782	0.9658	0.9622	1.1986	1.1041
120	0.9623	0.9684	0.9633	1.1659	1.0864
160	0.9583	0.9740	0.9614	1.1422	1.0736
200	0.9540	0.9765	0.9600	1.1261	1.0657
Estimation of Mixture exponential distribution with parameters β = 1, α=1 Through clustering algorithms					
n	K-Means	MeanShift	DBSCAN	A. H. C.	EM
40	0.9581	0.9743	0.9514	1.1112	1.0297
80	0.9633	0.9771	0.9534	1.0921	1.0241
120	0.9534	0.9801	0.9574	1.0763	1.0179
160	0.9571	0.9836	0.9586	1.0673	1.0161
200	0.9562	0.9847	0.9567	1.0593	1.0138
Estimation of Mixture exponential distribution with parameters β = 2, α=1 Through clustering algorithms					
n	K-Means	MeanShift	DBSCAN	A. H. C.	EM
40	0.9835	0.9923	0.9809	1.0257	1.0009
80	0.9827	0.9939	0.9809	1.0212	1.0006
120	0.9825	0.9945	0.9832	1.0181	1.0003
160	0.9842	0.9951	0.9833	1.0160	1.0004
200	0.9846	0.9959	0.9864	1.0140	1.0003

Source: Prepared by the researcher based on the results of the clustering algorithms and samples taken from the Mixture exponential distribution.

Applying actual data to the mixed Weibull distribution according to the parameters and sample size:

Actual data were taken that follow a Mixture Weibull distribution with parameters (β=0,5, α=1), (β=1, α=1)and (β=2, α=1) (With sample sizes (n= 40, 80, 120, 160, 200) the results were as shown in Table (3) below.

Table (5) Clustering algorithms and samples taken from the mixed Weibull distribution

Estimation of the Mixture Weibull distribution with parameters β = 0.5, α=1 Through clustering algorithms					
n	K-Means	MeanShift	DBSCAN	A. H. C.	EM
40	0.9782	0.9658	0.9622	1.1986	1.1041
80	0.9583	0.9740	0.9614	1.1422	1.0736
120	0.9540	0.9765	0.9611	1.1261	1.0657
160	0.9405	0.987	0.9662	1.0706	1.0357
200	0.9437	0.9919	0.9672	1.0418	1.0213
Estimation of the Mixture Weibull distribution with parameters β = 1, α=1 Through clustering algorithms					
n	K-Means	MeanShift	DBSCAN	A. H. C.	EM
40	0.9633	0.9781	0.9554	1.1921	1.1246
80	0.9575	0.9856	0.9596	1.1673	1.1161
120	0.9562	0.9847	0.9637	1.1593	1.1138

160	0.9579	0.9913	0.9645	1.1352	1.1074
200	0.9671	0.9954	0.9751	1.1206	1.1038
Estimation of the Mixture Weibull distribution with parameters $\beta = 2, \alpha=1$ Through clustering algorithms					
n	K-Means	MeanShift	DBSCAN	A. H. C.	EM
40	0.9842	0.9939	0.9809	1.016	1.0006
80	0.9846	0.9959	0.9864	1.0212	1.0003
120	0.9825	0.9945	0.9832	1.0181	1.0003
160	0.9858	0.9978	0.9884	1.0084	1.0001
200	0.9889	0.9990	0.9922	1.0000	1.0005

Source: Prepared by the researcher based on the results of the clustering algorithms and samples taken from the Mixture Weibull distribution.

Applying actual data to the mixed Pareto distribution according to the parameters and sample size:

Actual data were taken that follow the Mixture Pareto distribution with parameters (β=0,5, α=1), (β=1, α=1)and (β=2, α=1) (With sample sizes (n= 40, 80, 120, 160, 200) the results were as shown in Table (3) below.

Table (6) Clustering algorithms and samples taken from the mixed Pareto distribution

Estimation of Mixture Pareto distribution with parameters $\beta = 0.5, \alpha=1$ Through clustering algorithms					
n	K-Means	MeanShift	DBSCAN	A. H. C.	EM
40	0.9497	0.9778	0.9599	1.1101	1.0564
80	0.948	0.9913	0.9705	1.053	1.0271
120	0.9522	0.9943	0.9684	1.0349	1.0178
160	0.9529	0.9949	0.9749	1.0255	1.0128
200	0.9562	0.9965	0.977	1.0208	1.0107
Estimation of Mixture Pareto distribution with parameters $\beta = 1, \alpha=1$ Through clustering algorithms					
n	K-Means	MeanShift	DBSCAN	A. H. C.	EM
40	0.9568	0.9872	0.9618	1.0522	1.0113
80	0.9626	0.9942	0.9705	1.0259	1.0051
120	0.9676	0.9965	0.9758	1.0172	1.0032
160	0.9723	0.9976	0.9813	1.0125	1.0009
200	0.9732	0.9979	0.9842	1.0101	1.0016
Estimation of Mixture Pareto distribution with parameters $\beta = 2, \alpha=1$ Through clustering algorithms					
n	K-Means	MeanShift	DBSCAN	A. H. C.	EM
40	0.9854	0.9963	0.9853	1.0128	1.0004
80	0.9879	0.9984	0.9907	1.0063	1.0001
120	0.9906	0.9991	0.9927	1.0042	1.0001
160	0.9918	0.9994	0.9943	1.0031	1.0000
200	0.9927	0.9995	0.9955	1.0025	1.0000

Source: Prepared by the researcher based on the results of the clustering algorithms and samples taken from the Mixture Pareto distribution.

4. Conclusion

1. Each clustering algorithm has characteristics and features suitable for different data types.
2. In the case of the Mixture exponential distribution, we notice a clear division of data in the algorithms and the discovery of some heterogeneous groups and groups with different densities.
3. When applying different clustering algorithms to data generated using the Mixture Weibull distribution. Each algorithm gives different results based on how the data is processed and the groups are discovered, as follows:

   K-Means: The algorithm depends on the pre-specified number of groups and may be sensitive to the initial configuration.

   MeanShift: It does not require specifying the number of groups in advance but depends on the width of the kernel.

   DBSCAN: It determines the groups based on the density and identifies the outliers.

   Agglomerative Hierarchical Clustering: It builds a tree for clustering, and the number of groups can be specified.

   Gaussian Mixture Model: Uses distributed models to find clusters and relies on a diverse data distribution.

4. Applying different clustering algorithms to data generated using the Mixture pytro distribution. Each algorithm gives different results depending on how the data is processed and discovered for clusters, such as:

   K-Means: This algorithm groups data into clusters based on the distance from the cluster centers. It can be sensitive to the shape of the distribution.

   Mean Shift: Relies on the density of points and works well on irregular data shapes.

   DBSCAN: Relies on the density of points and can handle noise. It can discover clusters with irregular shapes.

   Agglomerative Hierarchical Clustering: It is useful when clusters with hierarchical structures are needed.

   EM (Gaussian Mixture Model): It assumes that the data comes from a Mixture distribution of Gaussian distributions. It can be helpful if the data follows multiple distributions.

REFERENCES

[1] Abdul-Hussein, Ali and Hussein, Ali Nasser, "Comparison between the (EM) algorithm "NR Algorithm for Finding the Estimators of the Hybrid Weibull-Kamma Distribution Using Simulation" Journal of Economic Sciences, 2017.

[2] Abdulmunen, Ashwan Anwer, "Converting Color Real Image to Cartoon Image Using NonParametric Mean-Shift Technique", Jornal of Kerbala University, vol. 10 no.3 scientific, 2012.

[3] Al-Bayati, Khader Naseef, "Comparison of Methods for Estimating the Reliability Function of the Mixture Exponential Distribution Using the Simulation Method" PhD Thesis, University of Baghdad, College of Administration and Economics, 2012.

[4] Al-Dhafri, Abdul-Jabbar Hussein, "Comprehensive Introduction to Algorithms and Flowcharts", Ibb University, College of Education, Department of Educational and Information Technology, Republic of Yemen, 2024.

[5] Al-Douri, Intisar Arabi Faddam and Bahiya, Ward Basem, "Comparison of Some Robust Methods for the Mixture Exponential and Mixture Pareto Distributions", University of Baghdad, College of Administration and Economics, 2018.

[6] Allyn Treshansky and Robert McGraw, "AN OVERVIEW OF CLUSTERING ALGORITHMS", Enabling Technology for Simulation Science V, Vol. 4367, .2001

[7] Al-Wakeel, Ali Abdul-Hussein, "Finding the Mixture Weibull Distribution", Journal of Administrative and Economic Sciences, Volume 16, Issue 59, 2010.

[8] Al-Wakeel, Ali A. Salih, Ahamd Mahir Razali, Asaad Mahdi, "Estimation Accuracy of Weibull Distribution Parameters", Journal of Applied Sciences Research, 2009.
[9] Backlund, H., Hedblom, A., and Neijman, N.. DBSCAN A Density-Based Spatial Clustering of Application with Noise. Linköpings Universitet, 2011.
[10] Bhat, A. A., Sofi Mudasir, and S.P. Ahmad, "Mixture of Exponential and Weighted Exponential Distribution: Properties and Applications", International Journal of Scientific Research inMathematical and Statistical Sciences, Volume-5, Issue-6, pp.38-46, 2018.
[11] Brian S. Everitt, Sabine Landau, Morven Leese, Daniel Stahl, "Cluster Analysis", King's College London, UK, 5th Edition, 2011.
[12] Christopher M. Bishop, "Pattern Recognition and Machine Learning", Microsoft Research Ltd Cambridge CB3 0FB, U.K., 2006.
[13] Chunxia Xiao, Meng Liu, "Efficient Mean-shift Clustering Using Gaussian KD-Tree", computer graphics forum, Volume29, Issue7, September 2010.
[14] Frederico, Caeiro and Mina Norouzirad, "Comparing Estimation Methods for the Power–Pareto Distribution", Journals Econometrics, Volume 12, issue 3, 2024.
[15] Garavaglia, F., Lang, M., Paquet, E., Gaillard, J., Garcon, R. and Renard, B. "Reliability and robustness of rainfall compound distribution model based on weatherpattern sub-sampling", 2011.
[16] http://research.microsoft.com/cmbishop
[17] http://research.microsoft.com/cmbishop
[18] https://doi.org/10.1111/j.1467-8659.2010.01793.x
[19] https://doi.org/10.3390/econometrics12030020
[20] LaPlante, F., Belacel, N., Kardouchi, M, (2015). A Heuristic Automatic Clustering Method Based on Hierarchical Clustering. In: Duval, B., van den Herik, J., Loiseau, S., Filipe, J. (eds) Agents and Artificial Intelligence, 2015.
[21] Mok, P., Huang, H., Kwok, Y., Au, J, "A robust adaptive clustering analysis method for automatic identification of clusters", Pattern Recogn, 45, 2012.
[22] Nareerat Nanuwong, Winai Bodhisuwan, and Chookait Pudprommarat, "A New Mixture Pareto Distribution and Its Application", Thailand Statistician, 2015.
[23] Nilashi, M, O. bin Ibrahim, N. Ithnin, & N. H. Sarmin," A multi-criteria collaborative filtering recommender system for the tourism domain using Expectation Maximization (EM) and PCA–ANFIS", Electronic Commerce Research and Applications, 2015.
[24] Puzicha, J., Hofmann, T., and Buhmann, J., Histogram Clustering for Unsupervised Image Segmentation. IEEE Proceedings of the Computer Vision and Pattern Recognition, Vol. 2, 2000.
[25] Raval, R. U., and Jani C, "Implementing and Improvisation of K-means Clustering", International Journal of Computer Science and Mobile Computing, 4(11): 72 – 76, 2015.
[26] Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X, "DBSCAN revisited, revisited: why and how you should (still) use DBSCAN", 2017.
[27] Trevor Hastie, Robert Tibshirani, Jerome Friedman, "The Elements of Statistical Learning Data Mining, Inference, and Prediction", Second Edition, Stanford, California, May 2001.
[28] Yilmaz, Mehmet and Buse Buyum, "Parameter Estimation Methods For Two-Component Mixture Exponential Distributions", Journal Of The Turkish Statistical Association, Vol. 8, No. 3, 2015.