Filomat 32:19 (2018), 6575–6598
https://doi.org/10.2298/FIL1819575Z

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

Asymptotic Properties of the Estimator for a Finite Mixture of Exponential Dispersion Models

Mouna Zitouni[a], Mourad Zribi[b], Afif Masmoudi[c]

[a]*Laboratory of Probability and Statistics. Sfax University. B.P. 1171, Tunisia.*
[b]*Laboratoire d'Informatique Signal et Image de la Côte d'Opale (LISIC-EA4491), ULCO, 50 Rue Ferdinand Buisson BP 719, 62228 Calais Cedex France.*
[c]*Laboratory of Probability and Statistics. Sfax University. B.P. 1171, Tunisia.*

Abstract. This paper is concerned with a class of exponential dispersion distributions. We particularly focused on the mixture models, which represent an extension of the Gaussian distribution. It should be noted that the parameters estimation of mixture distributions is an important task in statistical processing. In order to estimate the parameters vector, we proposed a formulation of the Expectation-Maximization algorithm (EM) under exponential dispersion mixture distributions. Also, we developed a hybrid algorithm called Expectation-Maximization and Method of moments algorithm (EMM). Under mild regularity, several convergence results of the EMM algorithm were obtained. Through simulation studies, the robustness of the EMM was proved and the strong consistency of the EMM sequence appeared when the data set size and the number of iterations tend to infinity.

1. Introduction

Analyzing and interpreting statistical data and preparing them for application is an important step in order to have the best approximation of the parameters in different fields, such as finance, medicine, and insurance [14, 15].
In 1977, Dempster, Laird, and Rubin derived several convergence properties of the Maximum Likelihood from incomplete data via the EM algorithm in a general context [1]. Dempster's results were applicable to many practical problems, such as the exponential dispersion models mixture which has been extensively developed in the field of statistics and classification literature in particular in image segmentation [9, 8, 15, 2, 11, 22]. This mixture was a convex combination of two or more probability density functions. So, by combining the properties of the individual probability density functions, the mixture models were able to approximate any arbitrary distribution. Consequently, finite mixture models constitute a powerful and flexible tool for modeling complex data [11]. In our work, we introduced a finite mixture of exponential dispersion distributions [3, 4] which represent a generalization of the natural exponential family which mixture has been developed in [22]. The exponential dispersion distributions include several well-known

2010 *Mathematics Subject Classification*. Primary 62F10; Secondary 62H30, 62F12, 62M40, 62J12
Keywords. Asymptotic properties, EM algorithm, Exponential dispersion distribution, Finite mixture, Maximum likelihood, Method of moments.
Received: 14 May 2016; Revised: 11 September 2016; Accepted: 11 December 2016
Communicated by Marko Petković
Email addresses: zitounimounaa@yahoo.fr (Mouna Zitouni), Mourad.Zribi@lisic.univ-littoral.fr (Mourad Zribi), afif.masmoudi@fss.rnu.tn (Afif Masmoudi)

families of distributions as special cases, giving a convenient general framework like the exponential family and the Tweedie distributions. This finite mixture represents a natural extension of the finite Gaussian mixture of distributions, with a special emphasis on the reproductive and additive cases and their use of both continuous and discrete data. In fact, we considerd the basic examples of exponential dispersion models such as the closed Normal, Gamma, Poisson, Inverse Gaussian, Laplace and the Tweedie exponential dispersion models closed under a scale transformation. Indeed, the exponential dispersion models have proved to be very successful in terms of increasing the modeling flexibility while remaining within a well understood inferential framework. A good treatment of this theory and application can be found in [3, 4]. It is well known that the exponential dispersion models constitute the major tools of the parametric modeling theory. Hence, in order to estimate the mixture of exponential dispersion distributions, we proposed a new algorithm called the Expectation and Method of moments-Maximization algorithm (EMM). This algorithm can be obtained by combining the EM algorithm [1, 13, 21] and the method of moments [16]. The exponential dispersion model is parameterized by the mean parameter μ and the dispersion parameter λ. For the parameters estimation, it employs the EM traditional mixture, but equipped with a method-of-moments procedure to estimate the unknown dispersion parameter. Furthermore, this combination was achieved because some of the exponential dispersion models (like Laplace exponential dispersion model) have a closed form where the maximum likelihood estimation of the dispersion parameters do not exist [3, 4, 17]. In fact, the Laplace exponential dispersion distribution is a convolution family given by

$$f(x;\mu,\lambda) = c(x,\lambda)e^{\lambda\left[\frac{\left[\sqrt{\left(1+\left(\frac{\mu}{\lambda}\right)^2\right)}-1\right]}{\left(\frac{\mu}{\lambda}\right)}x-\log\left(1-\left(\frac{\left[\sqrt{\left(1+\left(\frac{\mu}{\lambda}\right)^2\right)}-1\right]}{\left(\frac{\mu}{\lambda}\right)}\right)^2\right)\right]}.$$

For a suitable function, c is given in Table 1.
These models are very useful in several applications. Indeed, they can be used in modeling and analyzing real world data.
The proposed model has the following advantages: the simplicity and efficiency of its computation using the EMM algorithm. The latter constitutes an iterative approach which components are the standard tool for maximum likelihood estimation and the moment estimation in a probability distribution. Therefore, it can easily be extended to the finite mixture of exponential dispersion models. So, it can become a very popular computational method in statistics. The implementation of the E-step and the MM-step is easy for many statistical problems, thanks to the complete likelihood function form. However, the main drawbacks of the EMM algorithm are its slow convergence and the dependence of the solution on both the stopping criterion and the initial values used. Hence, studying the asymptotic properties of the EMM sequence represents the major issue of this work. We proved that the propose estimators of a finite mixture of exponential dispersion models are asymptotically consistent. The finite sample property of these estimators is proved. By using a set of simulation studies of Tweedie exponential dispersion model, we evaluated the finite sample performance by varying the sample size. From the simulated data, we estimated the mixture models parameters. As a result, the EMM sequence estimators are asymptotically convergent and this has been proved by computing the mean squared error MSE. So, we are able to support our theoretical results regarding consistency of the EMM algorithm. The remaining of this paper was organized as follows. Section 2 reviewed some properties of the Exponential Dispersion Models (EDMs). Section 3 introduced a finite mixture of Exponential Dispersion Models and described the proposed EMM algorithm. The asymptotic properties of the different EMM sequences were studied in section 4. Section 5 was devoted to the conclusion.

2. Properties of the Exponential Dispersion Models

The EDMs constitute some statistical models in which the probability distributions are characterized by a special form [3, 4]. In fact, this class of models represents a generalization of the Natural Exponential Families (NEFs) which play an important role in the statistical theory as a result of its special structure that allows making deductions dealing with appropriate statistical inference.

The EDMs represent a large class of models, characterized by a high number of important mathematical properties [5, 3, 4].
Let ν be a σ- finite positive measure on $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$. We denote by

$$L_\nu(\theta) = \int_{\mathbb{R}} e^{\theta y} \nu(dy) \tag{1}$$

the Laplace transform of ν, whose effective domain and cumulant function are, respectively, given by $D(\nu) = int\{\theta \in \mathbb{R};\ L_\nu(\theta) < +\infty\}$, and $\mathbb{C}_\nu(\theta) = \log(L_\nu(\theta))$.
We denote by $\mathcal{M}(\mathbb{R})$ the set of σ - finite positive measures, such that $D(\nu)$ is not empty and ν is not a Dirac measure.
For all measures $\nu \in \mathcal{M}(\mathbb{R})$, the NEF [4, 12, 18] generated by ν is defined by

$$F = F(\nu) = \{\mathbb{P}(\theta, \nu)(dy) = e^{\theta y - \mathbb{C}_\nu(\theta)} \nu(dy); \theta \in D(\nu)\}. \tag{2}$$

Let us notice that $\theta \longrightarrow \mathbb{C}_\nu(\theta)$ is strictly convex, infinitely differentiable, and its differential is given by $\mathbb{C}_\nu^{'}(\theta) = \int_{\mathbb{R}} y \mathbb{P}(\theta, \nu)(dy)$. The open interval $M_F = \mathbb{C}_\nu^{'}(D(\nu))$ is called the mean domain of F. We state that

$$\begin{array}{rcl} \mathbb{C}_\nu^{'} : D(\nu) & \longrightarrow & M_F \\ \theta & \longmapsto & \mathbb{C}_\nu^{'}(\theta) = \mu \end{array}$$

defines a bijection. So, we denote its inverse function by ψ_ν, and $\mu \longrightarrow \mathbb{V}_F(\mu) = [\psi^{'}{}_\nu(\mu)]^{-1} = \mathbb{C}_\nu^{''}(\psi_\nu(\mu))$ represents the variance function of the NEF F [4, 12, 18].
We obtain a new parametrization called the mean parametrization [18]. So, the NEF is given by $F = \{\mathbb{P}(\mu, F) = \mathbb{P}(\psi_\nu(\mu), \nu), \mu \in M_F\}$. Note that the variance function $\mathbb{V}_F$ characterizes the NEF F [10].
We define the Jørgensen set of ν by

$$\Lambda(\nu) = \{\lambda > 0, \exists \nu_\lambda \in \mathcal{M}(\mathbb{R})\ ;\ L_{\nu_\lambda}(\theta) = (L_\nu(\theta))^\lambda \quad \forall \theta \in D(\nu) = D(\nu_\lambda)\}. \tag{3}$$

Hence, the measure ν is infinitely divisible, if $\Lambda(\nu) =]0, +\infty[$.
For a fixed $\lambda \in \Lambda(\nu)$, we equivalently say that

$$F_\lambda = F(\nu_\lambda) = \{\mathbb{P}(\mu, F_\lambda) = e^{\psi_\nu(\frac{\mu}{\lambda}) y - \lambda \mathbb{C}_\nu(\psi(\frac{\mu}{\lambda}))} \nu_\lambda(dy); \mu \in \lambda M_F\} \tag{4}$$

is the NEF generated by ν_λ which verifies

$$\mathbb{V}_{F_\lambda}(\mu) = \lambda \mathbb{V}_F(\frac{\mu}{\lambda})\ \forall\ \mu \in M_{F_\lambda} = \lambda M_F. \tag{5}$$

In what follows, we consider $F = F(\nu)$ as a natural exponential family, and we denote by $\mathbb{C} = \mathbb{C}_\nu$, $\psi = \psi_\nu$, $\mathbb{V} = \mathbb{V}_F$.
Let Y be a random variable with distribution $\mathbb{P}(\theta, \nu_\lambda)$ and $\theta = \psi_\nu(\frac{\mu}{\lambda})$.
Then, the expectation and the variance of Y are, respectively, given by

$$\mathbb{E}(Y) = \lambda \mathbb{C}'(\theta),\ \textit{and}\ Var(Y) = \lambda \mathbb{C}''(\theta) = \lambda \mathbb{V}\left(\frac{\mu}{\lambda}\right). \tag{6}$$

These models draw their richness from a dispersion parameter $\sigma^2 = \frac{1}{\lambda}$ which is equal to 1 in the case of NEF.
Consequently, the EDM constitutes a useful generalization of the NEF, and it is a generated form given by $ED(\nu) = \{\mathbb{P}(\mu, F_\lambda), \mu \in \lambda M_F, \lambda \in \wedge(\nu)\}$.
For both theoretical and practical reasons, the Gaussian distribution is probably the most important distribution in statistics for example in modeling applications, such as linear and non-linear regressions. So, Jørgensen [4] proposes the reproductive EDMs as an extension of the Gaussian distribution given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2\sigma^2}\left(x-\mu\right)^2}, \quad x \in \mathbb{R} \tag{7}$$

indeed Jørgensen [4] defines a dispersion model by extending the Euclidean distance $(x-\mu)^2$ to a general discrepancy function $d(x;\mu)^2$ between the observation x and the mean μ. So, many commonly used parametric distributions, such as those in Table 1, Table 2 and Table 3 below, are included as special cases of this extension. From more on, we used the reproductive EDMs.
According to Jørgensen (1997), we considered the following transformation [4]. If $Y \sim \mathbb{P}(\mu, F_\lambda)$, where $\mu = \mathbb{E}(Y) = \lambda\mathbb{C}'(\theta)$ is the expectation and λ is the dispersion parameter. Then, $X = \frac{Y}{\lambda}$ follows the reproductive distribution defined by

$$\mathbb{P}^*(\mu, \nu_\lambda)(dx) = e^{\lambda[\psi(\frac{\mu}{\lambda})x - \mathbb{C}(\psi(\frac{\mu}{\lambda}))]}\nu^*_\lambda(dx) \tag{8}$$

and is characterized by its expectation and its variance function given by

$$\mathbb{E}(X) = \mathbb{C}'(\theta) = \frac{\mu}{\lambda}, \textit{ and } Var(X) = \frac{\mathbb{V}\left(\frac{\mu}{\lambda}\right)}{\lambda}, \tag{9}$$

where $\mu \in \lambda M_F, \lambda \in \wedge(\nu)$, and ν^*_λ denotes the image measure of ν_λ by the map $y \longrightarrow \frac{y}{\lambda} = x$.
Let us assume that there exist a σ finite measure ζ and a positive function $c(x, \lambda)$, such that $\nu^*_\lambda(dx) = c(x, \lambda)\zeta(dx)$. Then,

$$\mathbb{P}^*(\mu, \nu_\lambda)(dx) = e^{\lambda\left[\psi(\frac{\mu}{\lambda})x - \mathbb{C}(\psi(\frac{\mu}{\lambda}))\right]}c(x, \lambda)\zeta(dx), \tag{10}$$

where ζ denotes the Lebesgue measure, if $\mathbb{P}^*$ is a continuous probability, and it is the counting measure, if $\mathbb{P}^*$ is a discrete probability. Moreover, if $\mathbb{P}^*$ is the sum of discrete and continuous measures, then ζ is the sum of a Lebesgue measure and a counting measure.
Some examples of absolutely continuous EDMs [4, 5] are summarized in the following Table.

	Gaussian	*Gamma*	*Inverse Gaussian*	*Laplace*
$c(x, \lambda)$	$\frac{\sqrt{\lambda}}{\sqrt{2\pi}}e^{\frac{-\lambda x^2}{2}}$	$\frac{\lambda^\lambda x^{\lambda-1}}{\Gamma(\lambda)}$	$\frac{\sqrt{\lambda}}{\sqrt{2\pi}}x^{\frac{-3}{2}}e^{\frac{-\lambda}{2x}}$	$\frac{\lambda e^{\lambda x}}{\Gamma(\lambda)^2}\int_{\lambda x}^{+\infty} e^{-2t}t^{\lambda-1}(t-\lambda x)^{\lambda-1}dt$
$D(\nu)$	$\mathbb{R}$	$(-\infty, 0)$	$(-\infty, 0)$	$]-1, 1[$
$\mathbb{C}(\theta)$	$\frac{\theta^2}{2}$	$-\log(-\theta)$	$-\sqrt{-2\theta}$	$-\log(1-\theta^2)$
$\psi(\mu)$	μ	$\frac{-1}{\mu}$	$-\frac{1}{2\mu^2}$	$\frac{[\sqrt{(1+\mu^2)}-1]}{\mu}$
$\mathbb{V}(\mu)$	1	μ^2	μ^3	$\frac{[\sqrt{(1+\mu^2)}-1]}{\mu^2\sqrt{(1+\mu^2)}}$

Table 1: Examples of absolutely continuous EDMs.

The Stable and the Tweedie Compound Poisson EDMs are presented in Table 2 below.

	Stable, $\alpha \in]0, 1[$	*Tweedie Compound Poisson*, $p \in]1, 2[$
$c(x, \lambda)$	$\frac{1}{\pi x}\sum_{k=1}^{\infty}\frac{\Gamma(1+\alpha k)\lambda^k k^k(\frac{-1}{\lambda x})}{k!}\sin(-\alpha k\pi)$	$\frac{1}{x}\sum_{k=1}^{\infty}\frac{\lambda^k k^k(\frac{-1}{\lambda x})}{\Gamma(-\alpha k)k!}$
$D(\nu)$	$(-\infty, 0)$	$(-\infty, 0)$
$\mathbb{C}(\theta)$	$\frac{\alpha-1}{\alpha}(\frac{\theta}{\alpha-1})^\alpha$	$\frac{1}{2-p}[\theta(1-p)]^{\frac{2-p}{1-p}}$
$\psi(\mu)$	$(\alpha-1)\mu^{\frac{1}{\alpha-1}}$	$\mu^{1-p}\frac{1}{1-p}$
$\mathbb{V}(\mu)$	$\mu^{\frac{\alpha-2}{\alpha-1}}$	μ^p

Table 2: The Stable (α) and the Tweedie Compound Poisson (p) EDMs.

Finally, in the last following Table, we presented the Cosinus Hyperbolic EDM and some discrete EDMs.

	Binomial	*NegativeBinomial*	*Poisson*	*CosinusHyperbolic*
$c(x,\lambda)$	$\lambda\begin{pmatrix}\lambda\\ \lambda x\end{pmatrix}$	$\lambda\begin{pmatrix}\lambda+\lambda x-1\\ \lambda x\end{pmatrix}$	$\frac{\lambda^{\lambda x}}{x!}$	$\lambda 2^{\lambda-2}\frac{\lvert\Gamma(\frac{(\lambda+i\lambda y)}{2})\rvert^2}{\Gamma(\lambda)\Gamma(\frac{\lambda}{2})^2}$
$D(\nu)$	$\mathbb{R}$	$\mathbb{R}^-$	$\mathbb{R}$	$(-\frac{\pi}{2},\frac{\pi}{2})$
$\mathbb{C}(\theta)$	$\log(1+e^\theta)$	$-\log(1-e^\theta)$	e^θ	$-\log(\cos(\theta))$
$\psi(\mu)$	$\log(\frac{\mu}{1-\mu})$	$\log(\frac{\mu}{1+\mu})$	$\log(\mu)$	$\arctan(\mu)$
$\mathbb{V}(\mu)$	$\mu(1-\mu)$	$\mu(1+\mu)$	μ	$1+\mu^2$

Table 3: The Cosinus Hyperbolic EDM and some discrete EDMs.

We now consider the estimation of the vector parameters (μ,λ) of the exponential dispersion distribution. So, if $(x_1, x_2, ..., x_N)$ are independent and identically distributed observations from the distribution considered here, then the maximum likelihood estimate for μ is given by

$$\widehat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i. \tag{11}$$

In general case, the maximum likelihood estimate of $\lambda = \frac{1}{\sigma^2}$ doesn't exist and it is estimated by the Pearson estimator (see [3, 4]). In the last part of this work, we focused particulary on the mixture of exponential dispersion distributions.

3. A Finite mixture of Exponential Dispersion Models

In the following, we showed the generalization of the EDMs for a finite mixture [9, 8, 15, 2, 11, 20, 21, 23]. In this section, we introduced the finite mixture of EDMs, especially the finite mixture of distributions.
Let $\mu \in M_F$, $\lambda \in \Lambda(\nu)$, and let X be a random variable with a probability distribution $\mathbb{P}^*(\mu, \nu_\lambda)$. This random variable has a probability density function, with respect to ζ, which is given by

$$f(x;\mu,\lambda) = c(x,\lambda)e^{\lambda[\psi(\frac{\mu}{\lambda})x - \mathbb{C}(\psi(\frac{\mu}{\lambda}))]}. \tag{12}$$

We say that a density f is a mixture of K components of exponential dispersion distributions $f_1, f_2, ..., f_K$, if

$$f(x) = \sum_{k=1}^{K} \pi_k f_k\left(x;\mu_k,\lambda_k\right), \tag{13}$$

where π_k represents the mixing weights, $0 < \pi_k < 1$, $\sum_{k=1}^{K}\pi_k = 1$ and $f_k(.;\mu_k,\lambda_k)$ denotes the exponential dispersions density function with parameters μ_k and λ_k.
Let $\Theta = (\mu_1,\ldots,\mu_K,\pi_1,\ldots,\pi_K,\lambda_1,...,\lambda_K)$ be the parameter vector of the mixture models. So, the mixture density of the exponential dispersion distributions is given by

$$f(x) = \sum_{k=1}^{K} \pi_k c(x,\lambda_k)e^{\lambda_k\left[\psi(\frac{\mu_k}{\lambda_k})x - \mathbb{C}(\psi(\frac{\mu_k}{\lambda_k}))\right]}. \tag{14}$$

Now, we are interested in estimating the vector parameter Θ of the mixture models. To resolve this problem, we proposed an iterative algorithm. This new algorithm was obtained by combining the EM algorithm [1, 13] and the method of moments [16].
The proposed, so called EMM algorithm, consists of the following two steps: Expectation and Method of moments-Maximization.

The issue of using a combined maximum-likelihood and moment estimator could first be considered for a single distribution and then can easily be extended to the finite mixture case.
So, we applied the EM algorithm in order to estimate the parameter vector

$$\beta = (\pi_1, \ldots, \pi_K, \mu_1, \ldots, \mu_K), \tag{15}$$

and we proposed to estimate the unknown dispersion parameter λ_k by using the method of moments.

In the statistical world, there exist some statistical dispersion models for which their density function has a special form, and the maximum likelihood of the dispersion parameter λ_k cannot be found as a mixture of Laplace EDMs (See Figure 1). So, to resolve this problem, we resorted to use the method of moments.

From the Laplace exponential dispersion density(see Figure 1) the maximum is not achieved.
So, a modification of the M step of the EM algorithm was proposed to estimate the dispersion parameter λ. This modification can lead to a new algorithm: Expectation and Method of moments-Maximization (EMM) which constitutes the solution.

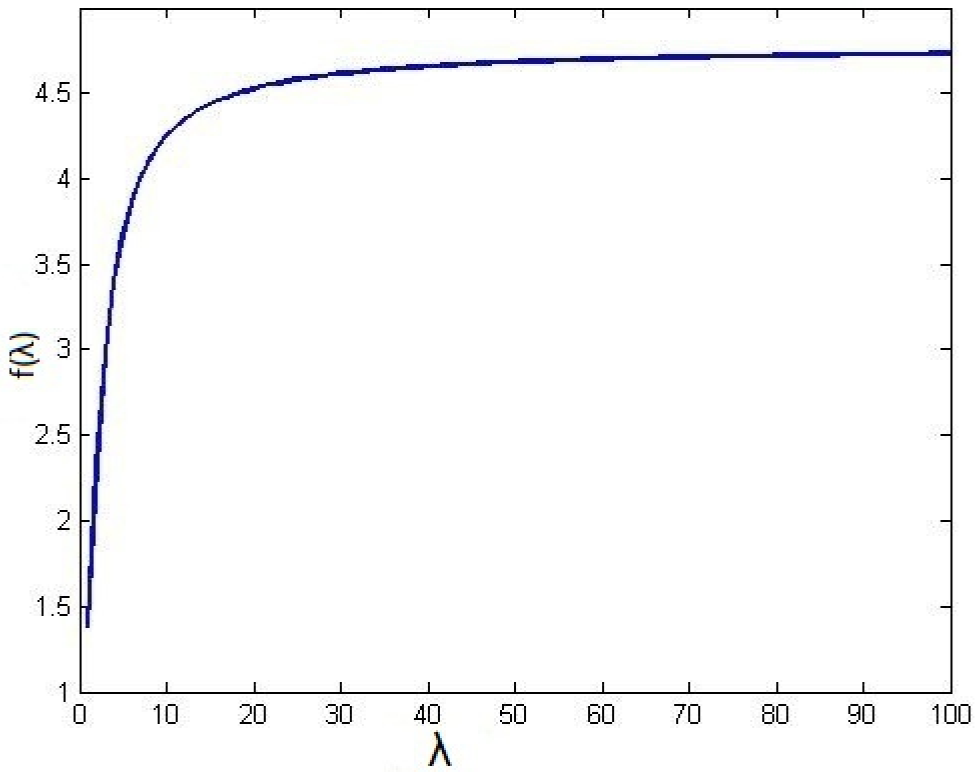

Figure 1: Curve of the Laplace exponential dispersion density distribution with respect to λ.

It is impossible to determine the maximum likelihood of the dispersion parameter associated with a mixture of Laplace EDMs.
In fact, the conditional expectation of the complete-data log-likelihood $\mathbb{Q}$ given the observed data X and a parameterizations $\Theta^{(l)}$ (the parameter vector in the l^{th} iteration) associated with the mixture of Laplace exponential dispersion distributions found in the Expectation step of the EM algorithm [1]. It is a multimodal

function with respect to the dispersion parameter λ, it has more than one global optima (see Figure 2). Consequently, the estimator of the dispersion parameter does not exist. So, the maximum likelihood estimator is undefined.

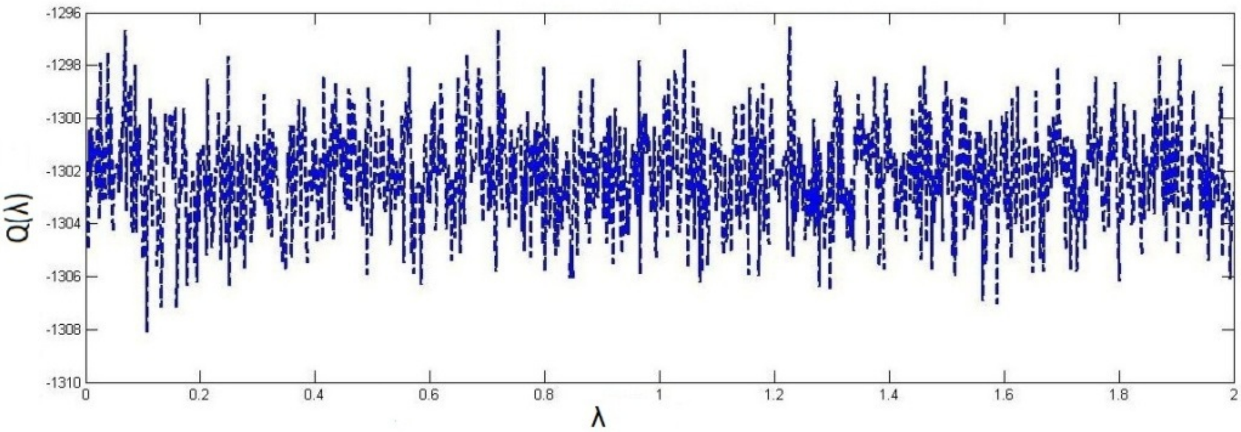

Figure 2: Illustration of the $\mathbb{Q}$ conditional expectation of Laplace exponential dispersion distribution with respect to λ.

3.1. Parameters estimation

Let $(X_1, ..., X_N)$ be N independent random variables with the same density function f (i.i.d) given by (14), and let $(x_1, ..., x_N)$ be N associated observations.
The incomplete likelihood function l of $(x_1, ..., x_N)$ is given by

$$l(x_1, x_2, \dots, x_N; \Theta) = \prod_{i=1}^{N} f(x_i) = \prod_{i=1}^{N} \left(\sum_{k=1}^{K} \pi_k c(x_i, \lambda_k) e^{\lambda_k \left[\psi(\frac{\mu_k}{\lambda_k}) x_i - \mathbb{C}(\psi(\frac{\mu_k}{\lambda_k})) \right]} \right). \tag{16}$$

Since, we need to use the logarithm in order to turn multiplication into addition, then the log-likelihood L is

$$L(x_1, x_2, \dots, x_N; \Theta) = \log l(x_1, x_2, \dots, x_N; \Theta) = \sum_{i=1}^{N} \log \left(\sum_{k=1}^{K} \pi_k f_k(x_i; \lambda_k, \mu_k) \right). \tag{17}$$

A mixture model is a convex combination of K distributions. It has a complex distribution of an observed variable X_i given by (13). Therefore, maximizing the incomplete likelihood function over Θ is a difficult task. So, in order to estimate the parameter vector Θ, for each observed data point X_i we associate a discrete random vector $Z_i = (Z_{i1}, ..., Z_{iK})$ following a multivariate Bernoulli distribution with vector parameters $(\pi_1, ..., \pi_K)$. That is

$$\mathbb{P}(Z_i = z_i) = \prod_{k=1}^{K} \pi_k^{z_{ik}}, \tag{18}$$

where $z_i = (z_{i1}, ..., z_{iK}) \in \{0, 1\}^K$, and $\sum_{k=1}^{K} z_{ik} = 1$. The random variables $(Z_i)_{1 \leq i \leq N}$ are called latent variables or missing values which indicate which component X_i is drawn from. Furthermore, if $Z_{ik} = 1$ this means that

the observation X_i exists in the class k.
Note that

$$\mathbb{E}(X_i|Z_{ik}=1)=\frac{\mu_k}{\lambda_k} \text{ and } Var(X_i|Z_{ik}=1)=\frac{\mathbb{V}(\frac{\mu_k}{\lambda_k})}{\lambda_k}. \tag{19}$$

The maximum likelihood function from complete data l_c is given by

$$\begin{aligned} l_c(x_1,\dots,x_N,z_1,\dots,z_N;\Theta) &= \prod_{i=1}^{N}\mathbb{P}(Z_i=z_i)\prod_{k=1}^{K}f_k^{z_{ik}}(x_i;\lambda_k,\mu_k) \\ &= \prod_{i=1}^{N}\left(\prod_{k=1}^{K}\left(\pi_k f_k(x_i;\lambda_k,\mu_k)\right)^{z_{ik}}\right). \end{aligned}$$

In addition, the log-likelihood function L_c from complete data is written as

$$\begin{aligned} L_c(x_1,x_2,\dots,x_N,z_1,z_2,\dots,z_N;\Theta) &= \sum_{i=1}^{N}\sum_{k=1}^{K}z_{ik}\log(\pi_k)+\sum_{i=1}^{N}\log\left(\sum_{k=1}^{K}\pi_k f_k(x_i;\lambda_k,\mu_k)\right) \\ &= \sum_{i=1}^{N}\sum_{k=1}^{K}z_{ik}\log\left(\pi_k f_k(x_i;\lambda_k,\mu_k)\right), \end{aligned}$$

where $\sum_{k=1}^{K}z_{ik}=1$. The EMM algorithm is an iterative approach for finding estimators for models with latent variables. It consists of two steps: the expectation step (E), which computes the expectation of the log-likelihood evaluated using the current estimate for the parameters, and the Method of moments-Maximization step (MM), which computes the parameters maximizing the expected log-likelihood found at the E step and estimate the dispersion parameter by using the Method of moments.
The Expectation step computes the conditional expectation of the complete data log likelihood [14, 1] for $Z_1,\dots,Z_N$ given $X_1=x_1,\dots,X_N=x_N$ in the l^{th} iteration.

$$\begin{aligned} \mathbb{Q}(\Theta||\Theta^{(l)}) &= \mathbb{E}\left(L_c(X_1=x_1,\dots,X_N=x_N,Z_1,\dots,Z_N;\Theta)\,|\Theta^{(l)},X_1=x_1,\dots,X_N=x_N\right) \\ &= \sum_{i=1}^{N}\sum_{k=1}^{K}\tau_{ik}{}^{(l)}\log\left(\pi_k f_k(x_i;\lambda_k,\mu_k)\right), \end{aligned}$$

where

$$\begin{aligned} \tau_{ik}^{(l)} &= \mathbb{E}\left(Z_{ik}|\Theta^{(l)},X_1=x_1,\dots,X_N=x_N\right)=\mathbb{E}\left(Z_{ik}|\Theta^{(l)},X_i=x_i\right) \\ &= \frac{\pi_k^{(l)}f_k(x_i;\lambda_k^{(l)},\mu_k^{(l)})}{\sum_{k=1}^{K}\pi_k^{(l)}f_k(x_i;\lambda_k^{(l)},\mu_k^{(l)})}, \end{aligned}$$

with $i=1,\dots,N,\ k=1,\dots,K$ and $\Theta^{(l)}=\Theta^{(l)}(X_1,\dots,X_N)$.
So, the conditional expectation of the complete-data log likelihood $\mathbb{Q}$ associated with the mixture of exponential dispersion distributions [14, 1] is given by

$$\mathbb{Q}(\Theta||\Theta^{(l)})=\sum_{i=1}^{N}\sum_{k=1}^{K}\tau_{ik}^{(l)}\left[\log(\pi_k)+\log c(x_i,\lambda_k)+\lambda_k x_i\psi(\frac{\mu_k}{\lambda_k})-\lambda_k\mathbb{C}(\psi(\frac{\mu_k}{\lambda_k}))\right]. \tag{20}$$

The MM step (Method of moments-Maximization) consists in estimating both the dispersion parameter $\lambda=(\lambda_1,\dots,\lambda_K)$ using the method of moments, and the vector parameter $\beta=(\mu_1,\dots,\mu_K,\pi_1,\dots,\pi_K)$ using

the maximum likelihood method.
Hence, by maximizing the concave function $\beta \longrightarrow \mathbb{Q}(\beta||\beta^{(l)})$, we obtain the estimated vector parameter $\beta^{(l+1)} = \underset{\beta}{\text{Argmax}}\mathbb{Q}(\beta||\beta^{(l)})$ in the $(l+1)^{th}$ iteration, such as

$$\mathbb{Q}(\beta^{(l+1)}||\beta^{(l)}) \geq \mathbb{Q}(\beta^{(l)}||\beta^{(l)}). \tag{21}$$

In the $(l+1)^{th}$ iteration of the EM algorithm, it follows that the updated estimate $\beta^{(l+1)}$ for β is obtained by solving

$$\frac{\partial \mathbb{Q}(\beta||\beta^{(l)})}{\partial \beta} = 0. \tag{22}$$

According to Dempster, Laird, and Rubin [1] and by applying Jensen's inequality, the Incomplete-likelihood function was not decreased after each EM iteration.
In the last part of this section, we consider two cases of the EMM algorithm where the dispersion parameter λ is known and unknown.

3.1.1. Known dispersion parameter

Theorem 3.1. *Suppose that $\lambda_1, ..., \lambda_K$ are known. Then, in the $(l+1)^{th}$ iteration, the EM results are*

$$\pi_k^{(l+1)} = \frac{1}{N}\sum_{i=1}^{N} \tau_{ik}^{(l)}, \tag{23}$$

and

$$\mu_k{}^{(l+1)} = \lambda_k \frac{\sum_{i=1}^{N} \tau_{ik}^{(l)} x_i}{\sum_{i=1}^{N} \tau_{ik}^{(l)}}. \tag{24}$$

Proof:
Let us recall that the conditional expectation of the complete data log-likelihood is defined by

$$\begin{aligned} \mathbb{Q}(\beta||\beta^{(l)}) &= \sum_{i=1}^{N}\sum_{k=1}^{K} \tau_{ik}^{(l)} \log\left(\pi_k f_k\left(x_i;\lambda_k,\mu_k\right)\right) \\ &= \sum_{i=1}^{N}\left[\sum_{k=2}^{K} \tau_{ik}^{(l)} \log\left(\pi_k f_k\left(x_i;\lambda_k,\mu_k\right)\right) + \tau_{i1}^{(l)} \log\left(\left(1-\sum_{k=2}^{K}\pi_k\right) f_1\left(x_i;\lambda_1,\mu_1\right)\right)\right]. \end{aligned}$$

We calculate the first derivative function of $\mathbb{Q}$, with respect to π_k, which is given by

$$\begin{aligned} \frac{\partial \mathbb{Q}(\beta||\beta^{(l)})}{\partial \pi_k} &= \sum_{i=1}^{N} \tau_{ik}^{(l)} \frac{1}{\pi_k} - \sum_{i=1}^{N} \tau_{i1}^{(l)} \frac{f_1\left(x_i;\lambda_1,\mu_1\right)}{\left[1-\sum_{k=2}^{K}\pi_k\right] f_1\left(x_i;\lambda_1,\mu_1\right)} \\ &= \sum_{i=1}^{N} \tau_{ik}^{(l)} \frac{1}{\pi_k} - \sum_{i=1}^{N}\left(\frac{\tau_{i1}^{(l)}}{1-\sum_{k=2}^{K}\pi_k}\right). \end{aligned}$$

By making it equal to zero, we obtain

$$\frac{\pi_k}{\pi_1} = \frac{\sum_{i=1}^{N} \tau_{ik}^{(l)}}{\sum_{i=1}^{N} \tau_{i1}^{(l)}}. \tag{25}$$

Furthermore, we have

$$\sum_{k=1}^{K} \left(\frac{\pi_k}{\pi_1}\right) = \sum_{k=1}^{K} \left(\frac{\sum_{i=1}^{N} \tau_{ik}^{(l)}}{\sum_{i=1}^{N} \tau_{i1}^{(l)}} \right).$$

As, $\sum_{k=1}^{K} \pi_k = 1$ and $\sum_{k=1}^{K} \tau_{ik}^{(l)} = 1$ we obtain

$$\frac{1}{\pi_1} = \frac{\sum_{i=1}^{N} \sum_{k=1}^{K} \tau_{ik}^{(l)}}{\sum_{i=1}^{N} \tau_{i1}^{(l)}}.$$

Consequently, we have

$$\pi_1 = \frac{1}{N} \sum_{i=1}^{N} \tau_{i1}^{(l)}.$$

By replacing it in equation (25), and in the $(l+1)^{th}$ iteration, we get

$$\pi_k^{(l+1)} = \frac{1}{N} \sum_{i=1}^{N} \tau_{ik}^{(l)}. \tag{26}$$

$\pi_k^{(l+1)}$ represents the maximum likelihood estimator of the mixing proportion π_k.
Since ψ is a bijective function, then $(\mu_1, ..., \mu_K) \longrightarrow \left(\psi\left(\frac{\mu_1}{\lambda_1}\right), ..., \psi\left(\frac{\mu_K}{\lambda_K}\right)\right) = (t_1, ..., t_K)$ is bijective. So, a new parametrization by (t_k, λ_k, π_k) of the conditional expectation of the complete-data log-likelihood $\mathbb{Q}$ is obtained. Hence,

$$\begin{aligned} \underset{\mu=(\mu_1,...,\mu_K)\in M_{F_\lambda}^K}{\text{Argmax}} \mathbb{Q}(\beta \| \beta^{(l)}) &= \underset{\mu=(\mu_1,...,\mu_K)\in M_{F_\lambda}^K}{\text{Argmax}} \left(\sum_{i=1}^{N} \sum_{k=1}^{K} \tau_{ik}^{(l)} \lambda_k \{ x_i \psi(\frac{\mu_k}{\lambda_k}) - \mathbb{C}(\psi(\frac{\mu_k}{\lambda_k})) \} \right) \\ &= \underset{t=(t_1,...,t_K)\in D(\nu)^K}{\text{Argmax}} \left(\sum_{i=1}^{N} \sum_{k=1}^{K} \tau_{ik}^{(l)} \lambda_k \{ x_i t_k - \mathbb{C}(t_k) \} \right). \end{aligned}$$

Therefore, we calculate the Hessian matrix of

$$(t_1, ..., t_K) \longrightarrow \sum_{i=1}^{N} \sum_{k=1}^{K} \tau_{ik}^{(l)} \lambda_k \{ x_i t_k - \mathbb{C}(t_k) \} = H(t_1, ..., t_k)$$

which is a diagonal matrix. So, $\forall k = 1, ..., K$ we get

$$\frac{\partial H(t_1, ..., t_K)}{\partial t_k} = \sum_{i=1}^{N} \tau_{ik}^{(l)} \lambda_k \{ x_i - \mathbb{C}'(t_k) \}.$$

Therefore,

$$\frac{\partial^2 H(t_1,...,t_K)}{\partial^2 t_k} = -\sum_{i=1}^{N} \tau_{ik}^{(l)} \lambda_k \mathbb{C}^{''}(t_k) \ \ and \ \ \frac{\partial^2 H(t_1,...,t_K)}{\partial t_k \partial t_j} = 0.$$

So, we deduce that the Hessian matrix is given by

$$H^{''}(t_1,...,t_K) = diag\left(-\sum_{i=1}^{N} \lambda_1 \tau_{i1}^{(l)} \mathbb{C}^{''}(t_1),...,-\sum_{i=1}^{N} \lambda_K \tau_{iK}^{(l)} \mathbb{C}^{''}(t_K)\right),$$

where *diag* denotes the diagonal matrix.
Since $\mathbb{C}$ is strictly convex on $D(\nu)$ then, H is strictly concave on $(D(\nu))^K$. So, the maximum of $\mathbb{Q}$ is achieved when $\frac{\partial H(t_1,...,t_K)}{\partial t_k} = 0, \ \forall k = 1,...,K.$

$$\begin{aligned} \frac{\partial Q(\Theta \| \Theta^{(l)})}{\partial t_k} &= \sum_{i=1}^{N} \tau_{ik}^{(l)} \left[\lambda_k x_i - \lambda_k \mathbb{C}^{'}(t_k)\right] \\ &= \sum_{i=1}^{N} \tau_{ik}^{(l)} \lambda_k \left[x_i - \mathbb{C}^{'}(t_k)\right] = 0. \end{aligned}$$

This means that

$$\sum_{i=1}^{N} \tau_{ik}^{(l)} x_i = \sum_{i=1}^{N} \tau_{ik}^{(l)} \mathbb{C}^{'}(t_k).$$

By using equation (9) we obtain $\forall k = 1, ..., K, \ \mathbb{C}^{'}(t_k) = \frac{\mu_k}{\lambda_k}$.
Hence,

$$\sum_{i=1}^{N} \tau_{ik}^{(l)} x_i = \sum_{i=1}^{N} \tau_{ik}^{(l)} \frac{\mu_k}{\lambda_k}.$$

Therefore, in the $(l+1)^{th}$ iteration, the estimator of the mean parameter μ_k is given by

$$\mu_k^{(l+1)} = \lambda_k \frac{\sum_{i=1}^{N} \tau_{ik}^{(l)} x_i}{\sum_{i=1}^{N} \tau_{ik}^{(l)}}. \tag{27}$$

3.1.2. *Unknown dispersion parameter*

The method of moments [16] is an estimation method of population parameters, using the law of large numbers. We introduced this technique in order to approximate the unknown parameter λ_k. Since the conditional variance of X_1, given $Z_{1k} = 1$, is defined by

$$Var(X_1|Z_{1k}) = \frac{\mathbb{V}\left(\frac{\mu_k}{\lambda_k}\right)}{\lambda_k} \tag{28}$$

then, by applying the method of moments, we get

$$\lambda_k = \frac{\mathbb{V}\left(\frac{\mu_k}{\lambda_k}\right)}{S_k^{(l+1)}}, \tag{29}$$

where

$$S_k{}^{(l+1)} = \frac{\sum_{i=1}^{N} \tau_{ik}^{(l)} \left(x_i - \frac{\mu_k^{(l)}}{\lambda_k^{(l)}} \right)^2}{\sum_{i=1}^{N} \tau_{ik}^{(l)}} \tag{30}$$

denotes the empirical variance data in the $(l+1)^{th}$ iteration.

We will prove that the empirical variance $S_k{}^{(l)}$ converges in probability to $Var(X_1/Z_{1k}) = \frac{\mathbb{V}\left(\frac{\mu_k}{\lambda_k}\right)}{\lambda_k}$ as data size N tends to infinity.
So, the estimator of λ_k in the $(l+1)^{th}$ iteration is given by

$$\lambda_k{}^{(l+1)} = \frac{\mathbb{V}\left(\frac{\sum_{i=1}^{N} \tau_{ik}^{(l)} x_i}{\sum_{i=1}^{N} \tau_{ik}^{(l)}} \right)}{S_k^{(l+1)}}, \tag{31}$$

and will converge, as N tends to infinity, to λ_k.
Maximization of $\mathbb{Q}$ with unknown dispersion parameter was proved in the following Theorem.

Theorem 3.2. *Suppose that $\lambda_1, ..., \lambda_K$ are unknown. Let $\lambda_k^{(l+1)}$ the estimator of λ_k in the $(l+1)^{th}$ iteration is given by equation* (3.29). *Then, in the $(l+1)^{th}$ iteration, we have*

$$\pi_k^{(l+1)} = \frac{1}{N} \sum_{i=1}^{N} \tau_{ik}^{(l)}, \tag{32}$$

and

$$\mu_k{}^{(l+1)} = \lambda_k{}^{(l+1)} \frac{\sum_{i=1}^{N} \tau_{ik}^{(l)} x_i}{\sum_{i=1}^{N} \tau_{ik}^{(l)}}. \tag{33}$$

We now considered a description of the EMM algorithm with unknown dispersion parameter which combines the EM algorithm and the method of moments.

Algorithm 3.3. *The EMM algorithm*

1: ***begin***

2: ***Initialisation:*** $\Theta^{(0)} = \left(\lambda_k{}^{(0)}, \mu_k{}^{(0)}, \pi_k{}^{(0)} \right)$

3: $S_k{}^{(0)} \longleftarrow \frac{\sum_{i=1}^{N} \tau_{ik}^{(0)} \left(x_i - \frac{\mu_k^{(0)}}{\lambda_k^{(0)}} \right)^2}{\sum_{i=1}^{N} \tau_{ik}^{(0)}}$

$\lambda_k^{(1)} \longleftarrow \dfrac{\mathbb{V}\left(\dfrac{\sum_{i=1}^{N}\tau_{ik}^{(0)}x_i}{\sum_{i=1}^{N}\tau_{ik}^{(0)}}\right)}{S_k^{(0)}}$

4: $\mu_k^{(1)} \longleftarrow \lambda_k^{(1)} \dfrac{\sum_{i=1}^{N}\tau_{ik}^{(0)}x_i}{\sum_{i=1}^{N}\tau_{ik}^{(0)}}$

5: $\pi_k^{(1)} \longleftarrow \frac{1}{N}\sum_{i=1}^{N}\tau_{ik}^{(0)}$

6: $\Theta^{(1)} \longleftarrow \left(\lambda_k^{(1)}, \mu_k^{(1)}, \pi_k^{(1)}\right)$

7: ***Iteration*** $(\boldsymbol{l}+1)$:

8: ***Expectation:***

9: $\tau_{ik}^{(l)} \longleftarrow \dfrac{\pi_k^{(l)} f_k(x_i;\lambda_k^{(l)},\mu_k^{(l)})}{\sum_{k=1}^{K}\pi_k^{(l)} f_k(x_i;\lambda_k^{(l)},\mu_k^{(l)})}$

10: ***Method of moments-Maximization:***

11: $S_k^{(l+1)} \longleftarrow \dfrac{\sum_{i=1}^{N}\tau_{ik}^{(l)}\left(x_i - \dfrac{\mu_k^{(l)}}{\lambda_k^{(l)}}\right)^2}{\sum_{i=1}^{N}\tau_{ik}^{(l)}}$

12: $\lambda_k^{(l+1)} \longleftarrow \dfrac{\mathbb{V}\left(\dfrac{\sum_{i=1}^{N}\tau_{ik}^{(l)}x_i}{\sum_{i=1}^{N}\tau_{ik}^{(l)}}\right)}{S_k^{(l+1)}}$

13: $\mu_k^{(l+1)} \longleftarrow \lambda_k^{(l+1)} \dfrac{\sum_{i=1}^{N}\tau_{ik}^{(l)}x_i}{\sum_{i=1}^{N}\tau_{ik}^{(l)}}$

14: $\pi_k^{(l+1)} \longleftarrow \frac{1}{N}\sum_{i=1}^{N}\tau_{ik}^{(l)}$

15: $\Theta^{(l+1)} \longleftarrow \{\lambda_k^{(l+1)}, \mu_k^{(l+1)}, \pi_k^{(l+1)}\}$

16: *If* $\|\Theta^{(l+1)} - \Theta^{(l)}\| < \epsilon$ *is not satisfied return to step* 2.

17: ***end***

3.2. Example

Now, let us consider an estimation example of the dispersion parameters Tweedie models [3, 4] with variance function $\mathbb{V}(\mu) = \mu^p$ and $1 < p < 2$. The estimators of the parameters (μ, λ, π) in the $(l+1)^{th}$ iteration are, respectively, given by

the mean estimator

$$\mu_k{}^{(l+1)} = \lambda_k{}^{(l+1)} \frac{\sum_{i=1}^{N} \tau_{ik}^{(l)} x_i}{\sum_{i=1}^{N} \tau_{ik}^{(l)}}, \tag{34}$$

the dispersion parameter estimator

$$\lambda_k{}^{(l+1)} = \left(\frac{\left(\mu_k{}^{(l+1)}\right)^p}{S_k{}^{(l+1)}} \right)^{\frac{1}{p+1}}, \text{ and} \tag{35}$$

the mixing weight estimator

$$\pi_k{}^{(l+1)} = \frac{1}{N} \sum_{i=1}^{N} \tau_{ik}^{(l)}. \tag{36}$$

In order to illustrate the EMM algorithm performances, we considered an example of mixture with 3 components of Tweedie distributions (see Figure 3) with true parameters $\mu = [\mu_1, \mu_2, \mu_3]$, $\lambda = [\lambda_1, \lambda_2, \lambda_3]$ and $\pi = [\pi_1, \pi_2, \pi_3]$ (see Table 4). We simulated a sample with different sizes $N = 10000, 15000, 20000, 30000$ from a mixture of 3 components of Tweedie distributions.

Note that the histograms have the form of the mixture of 3 components Tweedie distributions. We generated $n = 10000$ samples of size $N = 10000$ observations $(x_1{}^{(i)}, x_2{}^{(i)}, ..., x_N{}^{(i)}), 1 \leq i \leq n$ from a mixture of 3 components Tweedie distributions and we, then, computed the estimated parameters using the EMM algorithm.
In order, to prove the consistency of the EMM sequence and to further test the accuracy of the proposed

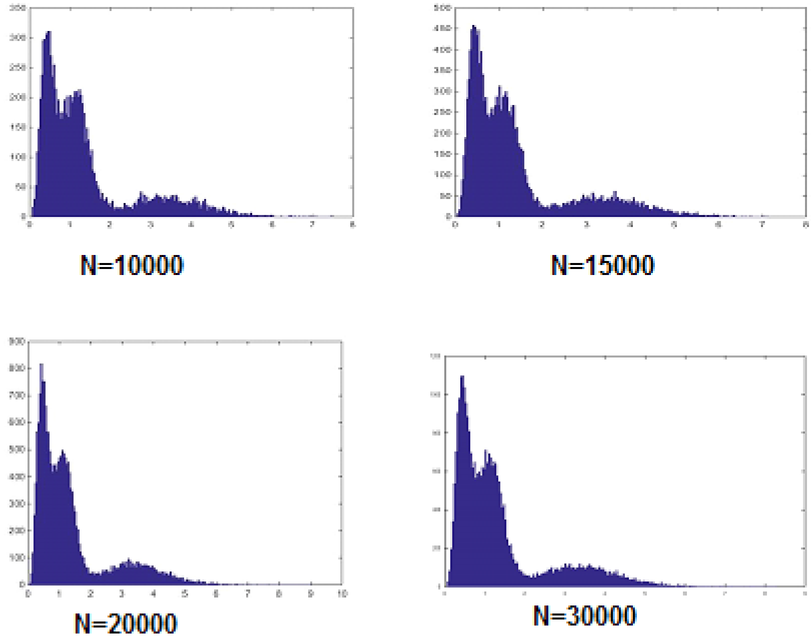

Figure 3: Histograms of mixture of 3 components of Tweedie distributions with $p = 1.7$

algorithm, we calculated the mean squared error between the estimated parameter $\widehat{\theta}$ and the true value of the parameter
$\theta\left(i.e.\ MSE(\theta, \widehat{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\theta_i} - \theta\right)^2\right)$ we then computed the mean squared error between the estimated density $\widehat{f}$ and the true density f
$\left(MSE(f, \widehat{f}) = \frac{1}{nN}\sum_{i=1}^{n}\sum_{j=1}^{N}\left(\widehat{f}\left(x_j^{(i)}\right) - f\left(x_j^{(i)}\right)\right)^2\right)$ which are reported in Table 4 below.

True parameters	Estimated parameters	$MSE(\theta, \widehat{\theta})$	$MSE(f, \widehat{f})$
$\mu = [0.1\ 0.39\ 0.5]$	[0.1002 0.3931 0.5011]	[0.00231 0.00323 0.003432]	0.0063
$\lambda = [1\ 4\ 6]$	[1.0041 4.00207 6.1032]	[0.00202 0.00133 0.00426]	
$\pi = [0.29\ 0.31\ 0.4]$	[0.2871 0.3011 0.410]	[0.01101 0.00150 0.01210]	

Table 4: Estimated parameters by EMM algorithm and the MSE values.

For different true values of parameters vector Θ and for a small sample size $N = 10000$, Table 4 shows that the MSE values for the EMM approach are clearly low. We notice that the latter gives good practical results.

In the second experiment, we proved the performance of the EMM algorithm by the curve of the true and the estimated mixture density distributions of Tweedie EDMs with $p = 1.7$. These are very close, the estimated density $\widehat{f}$ converges to the true density f (see Figure 4).

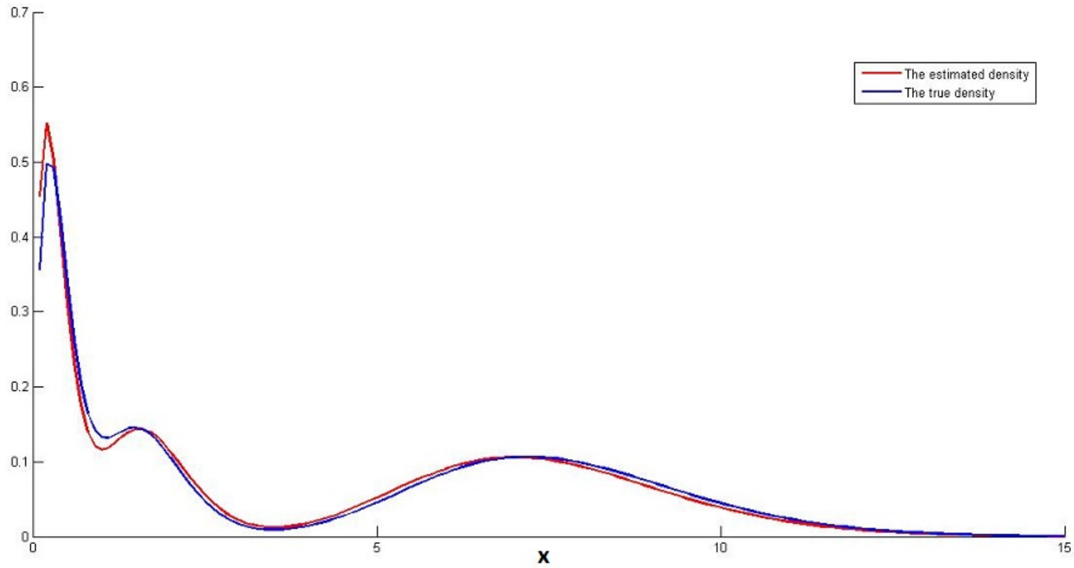

Figure 4: Curve of the true and the estimated mixture density of 3 Tweedie EDMs.

When comparing the true and the estimated parameters of a Tweedie mixture with $p = 1.7$, we concluded from the MSE that there exists a tiny variation between the true and the estimated parameters. From these results and the curve of the true and the estimated mixture density distributions of Tweedie EDMs with $p = 1.7$, we suggested performing the EMM algorithm for the exponential dispersion mixture densities.
Another experiment is to evaluate the algorithm convergence with respect to the sample size N. So, several simulations are created using samples with different sizes N varying from 100 to 10000.

N	100	200	500	1000	2000	5000	10000
MSE	0.12369	0.02208	0.01002	0.009514	0.0078	0.0071	0.00631

Table 5: The MSE values with respect to the sample size N.

From the Table 5, it is clearly seen that the proposed algorithm EMM converges. Therefore, we observe a rapid decrease of the MSE values when the sample size N is greater than 500.

4. Asymptotic properties

In this section, we studied the asymptotic properties [9, 14, 19, 21, 7, 23] of the maximum likelihood estimators of π_k and μ_k determined by the EM algorithm [1] and the estimator of the dispersion parameter λ_k using the method of moments [16] for the mixture model of exponential distributions, with fixed number K of components.
The EMM algorithm is the combination of the EM algorithm (based on the method of maximum likelihood) and the method of moments. However, it does not have the same properties of the EM algorithm [1, 6] . Using the moment estimates here makes little difference. Indeed, in our work there is no guarantee that the incomplete likelihood increases after each EMM iteration with the unknown dispersion parameter. In this section, we provided a proof of the EMM sequence consistency [17, 1, 16].

This consistency is a condition which ensures that, for large data set size N tends to infinity, the estimators will converge to the true parameters. In what follows, we denote by $\pi = (\pi_1, \pi_2, ..., \pi_K)$, $\lambda = (\lambda_1, ..., \lambda_K)$ and $\mu = (\mu_1, \mu_2, ..., \mu_K)$.

Theorem 4.1. *Suppose that $\Theta^{(l+1)} = \Theta^{(l+1)}(X_1, ..., X_N) = (\pi^{(l+1)}, \mu^{(l+1)}, \lambda^{(l+1)})$ is the EMM sequence. Then, in probability we have*

$$\lim_{N\to+\infty} \mu^{(l+1)} = \mu. \tag{37}$$

$$\lim_{N\to+\infty} \pi^{(l+1)} = \pi. \tag{38}$$

$$\lim_{N\to+\infty} \lambda^{(l+1)} = \lambda. \tag{39}$$

The proof of Theorem 4.1 necessitates the following technical lemmas.

Lemma 4.2. *Let $(W_N)_{N\geq 1}$ be a sequence of i-i-d random variables in the space of square integrable functions L^2 with a common mean μ, and let $(V_N)_{N\geq 1}$ be a sequence of random variables. Then,*

$$\mathbb{E}(\overline{W}_N|V_N) \underset{N\to+\infty}{\longrightarrow} \mu, \textit{ in } L^2 \tag{40}$$

where $\overline{W}_N = \frac{1}{N}\sum_{i=1}^{N} W_i$.

Proof of Lemma 4.2**:**
Let us denote $\sigma^2 = Var(W_i) < +\infty$.
Since $(W_N)_{N\geq 1}$ are i-i-d, then

$$Var(\overline{W}_N) = \frac{\sigma^2}{N} \underset{N\to+\infty}{\longrightarrow} 0.$$

It is well known that

$$\frac{\sigma^2}{N} = Var(\overline{W}_N) = \mathbb{E}\left(Var(\overline{W}_N|V_N)\right) + Var\left(\mathbb{E}(\overline{W}_N|V_N)\right).$$

Then, we have

$$0 \leq Var\left(\mathbb{E}(\overline{W}_N|V_N)\right) \leq \frac{\sigma^2}{N}.$$

This implies that

$$Var\left(\mathbb{E}(\overline{W}_N|V_N)\right) \underset{N\to+\infty}{\longrightarrow} 0.$$

Hence,

$$\mathbb{E}(\overline{W}_N|V_N) \underset{N\to+\infty}{\longrightarrow} \mathbb{E}(W_1) = \mu, \textit{ in } L^2$$

which completes this proof.

Proof of Theorem 4.1**:**
Let $\mathbb{F}_N{}^{(l)} = \sigma(\Theta^{(l)}(X_1, ..., X_N), (X_1, ..., X_N))$ be a σ-algebra .
1) We will prove that $\lim\limits_{N\to+\infty} \pi_k^{(l+1)} = \pi_k$, *in probability*.
Recall that

$$\pi_k^{(l+1)} = \frac{1}{N}\sum_{i=1}^{N} \tau_{ik}^{(l)}.$$

Mention that

$$\tau_{ik}^{(l)} = \mathbb{E}\left(Z_{ik}|\mathbb{F}_N{}^{(l)}\right).$$

By applying Lemma 4.2, we obtain

$$\pi_k^{(l+1)} = \frac{1}{N}\sum_{i=1}^{N}\tau_{ik}^{(l)} = \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N} Z_{ik}|\mathbb{F}_N{}^{(l)}\right) \underset{N\to+\infty}{\longrightarrow} \mathbb{E}(Z_{1k}) = \pi_k, \textit{in } L^2.$$

Therefore,

$$\lim_{N\to+\infty}\pi_k^{(l+1)} = \pi_k, \textit{in probability}.$$

2) We will prove that $\lim\limits_{N\to+\infty}\mu_k^{(l+1)} = \mu_k$, *in probability*.
Recall that, the mean estimator in the $(l+1)^{th}$ iteration is given by

$$\mu_k{}^{(l+1)} = \lambda_k{}^{(l+1)}\frac{\sum_{i=1}^{N}\tau_{ik}^{(l)}X_i}{\sum_{i=1}^{N}\tau_{ik}^{(l)}} = \lambda_k^{(l+1)}\frac{\frac{1}{N}\sum_{i=1}^{N}\tau_{ik}^{(l)}X_i}{\frac{1}{N}\sum_{i=1}^{N}\tau_{ik}^{(l)}},$$

and we have

$$\begin{aligned}\mathbb{E}(Z_{1k}X_1) &= \mathbb{E}(Z_{1k}X_1|Z_{1k}=1)\mathbb{P}(Z_{1k}=1) + \mathbb{E}(Z_{1k}X_1|Z_{1k}=0)\mathbb{P}(Z_{1k}=0)\\ &= \mathbb{E}(Z_{1k}X_1|Z_{1k}=1)\mathbb{P}(Z_{1k}=1) + 0 = \frac{\mu_k}{\lambda_k}\pi_k. \end{aligned} \tag{41}$$

Let us note that

$$\frac{1}{N}\sum_{i=1}^{N}\tau_{ik}^{(l)}X_i = \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N} Z_{ik}X_i|\mathbb{F}_N{}^{(l)}\right).$$

Then, according to Lemma 4.2, we have

$$\lim_{N\to+\infty}\mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N} Z_{ik}X_i|\mathbb{F}_N{}^{(l)}\right) = \frac{\mu_k}{\lambda_k}\pi_k, \textit{ in probability}$$

which implies that

$$\lim_{N\to+\infty}\frac{\mu_k^{(l+1)}}{\lambda_k^{(l+1)}} = \lim_{N\to+\infty}\frac{\frac{1}{N}\sum_{i=1}^{N}\tau_{ik}^{(l)}X_i}{\frac{1}{N}\sum_{i=1}^{N}\tau_{ik}^{(l)}} = \frac{\mu_k\pi_k}{\lambda_k\pi_k} = \frac{\mu_k}{\lambda_k}, \textit{ in probability}.$$

3) We will now prove that $\lim\limits_{N\to+\infty}\lambda_k{}^{(l+1)} = \lambda_k$, *in probability*.

We see that, $\lambda_k{}^{(l+1)}$ is the solution of the equation $\lambda_k{}^{(l+1)} = \dfrac{\mathbb{V}\left(\frac{\mu_k{}^{(l+1)}}{\lambda_k{}^{(l+1)}}\right)}{S_k{}^{(l+1)}}$.
Observe that

$$\begin{aligned}\mathbb{E}(Z_{1k}X_1^2) &= \mathbb{E}(Z_{1k}X_1^2|Z_{1k}=1)\mathbb{P}(Z_{1k}=1) + \mathbb{E}(Z_{1k}X_1^2|Z_{1k}=0)\mathbb{P}(Z_{1k}=0)\\ &= (Var(X_1|Z_{1k}=1) + (\mathbb{E}(X_1|Z_{1k}=1))^2)\pi_k = (\frac{1}{\lambda_k}\mathbb{V}(\frac{\mu_k}{\lambda_k}) + \frac{\mu_k{}^2}{\lambda_k{}^2})\pi_k.\end{aligned} \tag{42}$$

After some calculations, we have

$$\begin{aligned}
\lim_{N\to+\infty}\frac{1}{N}\sum_{i=1}^{N}\tau_{ik}^{(l)}\left(X_i-\frac{\mu_k{}^{(l)}}{\lambda_k{}^{(l)}}\right)^2 &= \lim_{N\to+\infty}\mathbb{E}(\frac{1}{N}\sum_{i=1}^{N}Z_{ik}(X_i-\frac{\mu_k^{(l)}}{\lambda_k^{(l)}})^2|\mathbb{F}_N{}^{(l)})\\
&= \lim_{N\to+\infty}\mathbb{E}(\frac{1}{N}\sum_{i=1}^{N}Z_{ik}X_i^2|\mathbb{F}_N{}^{(l)})\\
&- \lim_{N\to+\infty}2(\frac{\mu_k^{(l)}}{\lambda_k^{(l)}})\mathbb{E}(\frac{1}{N}\sum_{i=1}^{N}Z_{ik}X_i|\mathbb{F}_N{}^{(l)})\\
&+ \lim_{N\to+\infty}(\frac{\mu_k{}^{(l)}}{\lambda_k{}^{(l)}})^2\mathbb{E}(\frac{1}{N}\sum_{i=1}^{N}Z_{ik}|\mathbb{F}_N{}^{(l)})\\
&= \mathbb{E}(X_1{}^2Z_{1k})-2\frac{\mu_k{}^2}{\lambda_k{}^2}\pi_k+\frac{\mu_k{}^2}{\lambda_k{}^2}\pi_k\\
&= \frac{1}{\lambda_k}\mathbb{V}(\frac{\mu_k}{\lambda_k})\pi_k.
\end{aligned}$$

Consequently, we obtain

$$\lim_{N\to+\infty}S_k^{(l+1)}=\frac{1}{\lambda_k}\mathbb{V}(\frac{\mu_k}{\lambda_k}).$$

As

$$\lim_{N\to+\infty}\frac{\mu_k{}^{(l+1)}}{\lambda_k{}^{(l+1)}}=\frac{\mu_k}{\lambda_k},\ \textit{in probability},$$

and since $\mathbb{V}$ is continuous on M_F, then we have

$$\lim_{N\to+\infty}\mathbb{V}\left(\frac{\mu_k{}^{(l+1)}}{\lambda_k{}^{(l+1)}}\right)=\mathbb{V}\left(\frac{\mu_k}{\lambda_k}\right),\ \textit{in probability.}$$

Hence,

$$\lim_{N\to+\infty}\lambda_k^{(l+1)}=\lim_{N\to+\infty}\frac{\mathbb{V}\left(\frac{\mu_k{}^{(l+1)}}{\lambda_k{}^{(l+1)}}\right)}{S_k^{(l+1)}}=\frac{\mathbb{V}(\frac{\mu_k}{\lambda_k})}{\frac{1}{\lambda_k}\mathbb{V}(\frac{\mu_k}{\lambda_k})}=\lambda_k.$$

We conclude that, for large data, the estimator of the dispersion parameter converges in probability to the true parameter λ_k. Which completes this proof.

Theorem 4.3. *Suppose that, almost surely, we have*

$$\lim_{l\to+\infty}\lambda^{(l)}=\widehat{\lambda},\tag{43}$$

$$\lim_{l\to+\infty}\mu^{(l)}=\widehat{\mu}\tag{44}$$

and

$$\lim_{l\to+\infty}\pi^{(l)}=\widehat{\pi}.\tag{45}$$

Then, $(\widehat{\pi_k},\frac{\widehat{\mu_k}}{\widehat{\lambda_k}})$ *is the maximum likelihood estimator of* $(\pi_k,\frac{\mu_k}{\lambda_k}),\ k=1,...,K$ *and* $\widehat{\lambda_k}$ *is the estimator of* λ_k *by the method of moments.*

Proof:
Note that, the mapping

$$(\mu_1, ..., \mu_K) \longrightarrow (\frac{\mu_1}{\lambda_1}, ..., \frac{\mu_K}{\lambda_K}) = (t_1, ..., t_K) = t$$

is bijective. Hence, we obtain a new parametrization by (t, λ, π) of the mixture model. Then, the density mixture is given by

$$f(x) = \sum_{k=1}^{K} \pi_k c(x, \lambda_k) e^{\lambda_k [\psi(t_k)x - \mathbb{C}(\psi(t_k))]}. \tag{46}$$

Let $X_1, X_2, ..., X_N$ be a sample with mixture density $f(x)$ given by (46). So, the log-likelihood with $\Theta = (t, \lambda, \pi)$ is given by

$$L(X_1, X_2, \ldots, X_N; \Theta) = \sum_{i=1}^{N} \log\Big(\sum_{k=1}^{K} \pi_k^{(l)} f_k(X_i; t_k, \lambda_k)\Big).$$

Note that, $t_k \longrightarrow \psi(t_k)x - \mathbb{C}(\psi(t_k))$ is a strictly concave function. Then, the mixture density f is also strictly concave with respect to $t = (t_1, ...t_k)$.
Observe that, a-s

$$\lim_{l\to+\infty} \lambda^{(l)} = \widehat{\lambda}$$

and

$$\lim_{l\to+\infty} \mu^{(l)} = \widehat{\mu}.$$

Therefore, $t_k^{(l)} = \frac{\mu_k^{(l)}}{\lambda_k^{(l)}}$ converges to $\widehat{t_k} = \frac{\widehat{\mu_k}}{\widehat{\lambda_k}}$ as the iteration l tends to $+\infty$.
As ψ is a continuous function, we obtain

$$\lim_{l\to+\infty} \psi(t_k^{(l)}) = \psi(\widehat{t_k}).$$

Now we show that,

$$\frac{\partial L(X_1, X_2, \ldots, X_N; \widehat{t_k})}{\partial t_k} = 0,$$

gives that the maximum likelihood estimator for t_k is $\widehat{t_k}$.
Indeed, we have for $k \in \{1, ..., K\}$,

$$\begin{aligned} \frac{\partial L(X_1, X_2, \ldots, X_N; t^{(l)})}{\partial t_k} &= \sum_{i=1}^{N} \pi_k^{(l)} c(X_i, \lambda_k^{(l)}) \frac{[\mathbb{V}(t_k^{(l)})]^{-1} \big[\lambda_k^{(l)} X_i - \lambda_k^{(l)} t_k^{(l)}\big] e^{\lambda_k^{(l)}[\psi(t_k^{(l)})X_i - \mathbb{C}(\psi(t_k^{(l)}))]}}{\sum_{k=1}^{K} \pi_k^{(l)} f_k(X_i; t_k^{(l)}, \lambda_k^{(l)})} \\ &= \sum_{i=1}^{N} \tau_{ik}^{(l)} [\mathbb{V}(t_k^{(l)})]^{-1} \lambda_k^{(l)} [X_i - t_k^{(l)}], \end{aligned}$$

by letting $l \longrightarrow +\infty$, we obtain

$$\lim_{l\to+\infty} \tau_{ik}^{(l)} = \lim_{l\to+\infty} \frac{\pi_k^{(l)} f_k(X_i; \lambda_k^{(l)}, t_k^{(l)})}{\sum_{k=1}^{K} \pi_k^{(l)} f_j(X_i; \lambda_k^{(l)}, t_k^{(l)})} = \frac{\widehat{\pi_k} f_k(X_i; \widehat{\lambda_k}, \widehat{t_k})}{\sum_{k=1}^{K} \widehat{\pi_k} f_k(X_i; \widehat{\lambda_k}, \widehat{t_k})}$$

because f_k is a continuous function.

By using, $\widehat{\tau_{ik}} = \dfrac{\widehat{\pi_k} f_k(X_i; \widehat{\lambda_k}, \widehat{t_k})}{\sum_{k=1}^{K} \widehat{\pi_k} f_k(X_i; \widehat{\lambda_k}, \widehat{t_k})}$, and computing the first derivative of L with respect to t_k, we obtain,

$$\begin{aligned}
\frac{\partial L(X_1, X_2, \ldots, X_N; \widehat{t})}{\partial t_k} &= \sum_{i=1}^{N} \widehat{\tau_{ik}} \widehat{\lambda_k} [\mathbb{V}(\widehat{t_k})]^{-1} [X_i - \widehat{t_k}] \\
&= \sum_{i=1}^{N} \widehat{\tau_{ik}} \widehat{\lambda_k} [\mathbb{V}(\widehat{t_k})]^{-1} X_i - \sum_{i=1}^{N} \widehat{\tau_{ik}} \widehat{\lambda_k} [\mathbb{V}(\widehat{t_k})]^{-1} \widehat{t_k} \\
&= 0.
\end{aligned}$$

This result implies that, $\widehat{t_k} = \frac{\widehat{\mu_k}}{\widehat{\lambda_k}}$ is the maximum likelihood estimator of $t_k = \frac{\mu_k}{\lambda_k}$, $\forall k \in \{1, ..., K\}$.

2) Now, we consider the log-likelihood function with $\Theta = (\mu, \lambda, \pi)$ which is given by

$$L(X_1, X_2, \ldots, X_N; \Theta) = \sum_{i=1}^{N} \log(f(X_i)) = \sum_{i=1}^{N} \log\left(\sum_{k=1}^{K} \pi_k f_k(X_i; \lambda_k, \mu_k) \right),$$

where the component density f_k is given by

$$f_k(x; \lambda_k, \mu_k) = c(x, \lambda_k) e^{\lambda_k [\psi(\frac{\mu_k}{\lambda_k}) x - \mathbb{C}(\psi(\frac{\mu_k}{\lambda_k}))]}.$$

Clearly, the mixture density f is a concave function with respect to the mixing weight. On the other hand, observe that

$$\lim_{l \to +\infty} \pi_k{}^{(l)} = \lim_{l \to +\infty} \frac{\sum_{i=1}^{N} \tau_{ik}^{(l)}}{N} = \widehat{\pi_k}.$$

Now, we show for $k \in \{1, ..., K\}$,

$$\frac{\partial L(X_1, X_2, \ldots, X_N; \widehat{\pi})}{\partial \pi_k} = 0$$

gives that the maximum likelihood estimator for π_k is $\widehat{\pi_k}$.

Recall that, the posterior probabilities are given by

$$\tau_{ik}^{(l)} = \frac{\pi_k^{(l)} f_k(X_i; \lambda_k^{(l)}, \mu_k^{(l)})}{\sum_{k=1}^{K} \pi_k^{(l)} f_k(X_i; \lambda_k^{(l)}, \mu_k^{(l)})}$$

by letting $l \longrightarrow +\infty$, we obtain

$$\frac{\partial L(X_1, X_2, \ldots, X_N; \widehat{\pi})}{\partial \pi_k} = \sum_{i=1}^{N} \widehat{\tau_{ik}} \frac{1}{\widehat{\pi_k}} - N = 0.$$

From this result, it follows that $\widehat{\pi_k}$ is the maximum likelihood estimator of π_k, $\forall k \in \{1, ..., K\}$.

3) By hypothesis, $\lim_{l \to +\infty} \lambda_k{}^{(l)} = \widehat{\lambda_k}$, a-s

where

$$\lambda_k{}^{(l+1)} = \frac{\mathbb{V}\left(\frac{\sum_{i=1}^{N} \tau_{ik}^{(l)} X_i}{\sum_{i=1}^{N} \tau_{ik}^{(l)}} \right)}{S_k^{(l+1)}},$$

and by letting $l \longrightarrow +\infty$, we obtain

$$S_k{}^{(l+1)} = \frac{\sum_{i=1}^{N} \tau_{ik}^{(l)} \left(X_i - \frac{\mu_k^{(l)}}{\lambda_k^{(l)}} \right)^2}{\sum_{i=1}^{N} \tau_{ik}^{(l)}} \underset{l \to +\infty}{\longrightarrow} \widehat{S_k} = \frac{\sum_{i=1}^{N} \widehat{\tau_{ik}} \left(X_i - \frac{\widehat{\mu_k}}{\widehat{\lambda_k}} \right)^2}{\sum_{i=1}^{N} \widehat{\tau_{ik}}},$$

and

$$\mathbb{V}\left(\frac{\sum_{i=1}^{N} \tau_{ik}^{(l)} X_i}{\sum_{i=1}^{N} \tau_{ik}^{(l)}} \right) \underset{l \to +\infty}{\longrightarrow} \mathbb{V}\left(\frac{\sum_{i=1}^{N} \widehat{\tau_{ik}} X_i}{\sum_{i=1}^{N} \widehat{\tau_{ik}}} \right) = \mathbb{V}(\frac{\widehat{\mu_k}}{\widehat{\lambda_k}}).$$

Then, $\lambda_k{}^{(l+1)}$ converges to $\widehat{\lambda_k} = \frac{\mathbb{V}(\frac{\widehat{\mu_k}}{\widehat{\lambda_k}})}{\widehat{S_k}}$ as $l \longrightarrow +\infty$, which is the estimator of λ_k by the method of moments [16], which completes this proof.

Theorem 4.4. *Suppose that*

$$\Theta^{(l)}(X_1, ..., X_N) = (\mu^{(l)}, \pi^{(l)}, \lambda^{(l)}) \underset{l \to +\infty}{\overset{a-s}{\longrightarrow}} \widehat{\Theta}(X_1, ..., X_N) = (\widehat{\mu}, \widehat{\pi}, \widehat{\lambda}). \tag{47}$$

Then,

$$\widehat{\Theta}(X_1, ..., X_N) = (\widehat{\mu}, \widehat{\pi}, \widehat{\lambda}) \underset{N \to +\infty}{\longrightarrow} \Theta(X_1, ..., X_N) = (\mu, \pi, \lambda), \textit{in probability}. \tag{48}$$

Proof:
Note that

$$\lim_{l \to +\infty} \tau_{ik}^{(l)} = \lim_{l \to +\infty} \frac{\pi_k^{(l)} f_k(X_i; \lambda_k^{(l)}, \mu_k^{(l)})}{\sum_{k=1}^{K} \pi_k^{(l)} f_k(X_i; \lambda_k^{(l)}, \mu_k^{(l)})} = \frac{\widehat{\pi_k} f_k(X_i; \widehat{\lambda_k}, \widehat{\mu_k})}{\sum_{k=1}^{K} \widehat{\pi_k} f_k(X_i; \widehat{\lambda_k}, \widehat{\mu_k})} = \widehat{\tau_{ik}}$$

because f_k is a continuous distribution.

1) First we will prove that, $\lim_{N \to +\infty} \widehat{\pi_k} = \frac{1}{N} \sum_{i=1}^{N} \widehat{\tau_{ik}} = \pi_k$, *in probability.*
Mention that

$$\widehat{\tau_{ik}} = \mathbb{E}\left(Z_{ik} | \widehat{F_N} \right)$$

where, $\widehat{F_N} = \sigma(\widehat{\Theta}(X_1, ..., X_N), (X_1, ..., X_N))$ be a σ- algebra.
By applying Lemma 4.2, we obtain

$$\mathbb{E}\left(\frac{1}{N} \sum_{i=1}^{N} Z_{ik} | \widehat{F_N} \right) \underset{N \to +\infty}{\longrightarrow} \mathbb{E}(Z_{1k}) = \pi_k, \textit{in } L^2.$$

Therefore,

$$\lim_{N \to +\infty} \widehat{\pi_k} = \pi_k, \textit{in probability}.$$

2) In the second part of this proof, we will show that $\lim_{N\to+\infty} \widehat{\mu_k} = \mu_k$, *in probability* .
On the other hand, we have

$$\frac{1}{N}\sum_{i=1}^{N} \widehat{\tau_{ik}} X_i = \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N} Z_{ik} X_i | \widehat{F_N}\right).$$

Then, according to Lemma 4.2 and equation (41) we have

$$\lim_{N\to+\infty} \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N} Z_{ik} X_i | \widehat{F_N}\right) = \mathbb{E}(Z_{1k} X_1) = \frac{\mu_k}{\lambda_k}\pi_k, \textit{ in probability.}$$

This implies that,

$$\lim_{N\to+\infty} \frac{\widehat{\mu_k}}{\widehat{\lambda_k}} = \lim_{N\to+\infty} \frac{\frac{1}{N}\sum_{i=1}^{N} \widehat{\tau_{ik}} X_i}{\frac{1}{N}\sum_{i=1}^{N} \widehat{\tau_{ik}}} = \frac{\mu_k \pi_k}{\lambda_k \pi_k} = \frac{\mu_k}{\lambda_k}, \textit{ in probability.}$$

3) We will now prove that $\lim_{N\to+\infty} \widehat{\lambda_k} = \lambda_k$, *in probability*.

We see that, $\widehat{\lambda_k}$ is the solution of the equation $\widehat{\lambda_k} = \dfrac{\mathbb{V}\left(\frac{\widehat{\mu_k}}{\widehat{\lambda_k}}\right)}{\widehat{S_k}}$.

And

$$\lim_{N\to+\infty} \widehat{\lambda_k} = \lim_{N\to+\infty} \frac{\mathbb{V}(\frac{\widehat{\mu_k}}{\widehat{\lambda_k}})}{\widehat{S_k}},$$

according to equation (42) and by using the fact that

$$\begin{aligned}
\lim_{N\to+\infty} \frac{1}{N}\sum_{i=1}^{N} \widehat{\tau_{ik}}\left(X_i - \frac{\widehat{\mu_k}}{\widehat{\lambda_k}}\right)^2 &= \lim_{N\to+\infty} \mathbb{E}\Big(\frac{1}{N}\sum_{i=1}^{N} Z_{ik}(X_i - \frac{\widehat{\mu_k}}{\widehat{\lambda_k}})^2 | \widehat{F_N}\Big) \\
&= \lim_{N\to+\infty} \mathbb{E}\Big(\frac{1}{N}\sum_{i=1}^{N} Z_{ik} X_i^2 | \widehat{F_N}\Big) - \lim_{N\to+\infty} 2\Big(\frac{\widehat{\mu_k}}{\widehat{\lambda_k}}\Big)\mathbb{E}\Big(\frac{1}{N}\sum_{i=1}^{N} Z_{ik} X_i | \widehat{F_N}\Big) \\
&\quad + \lim_{N\to+\infty} \Big(\frac{\widehat{\mu_k}}{\widehat{\lambda_k}}\Big)^2 \mathbb{E}\Big(\frac{1}{N}\sum_{i=1}^{N} Z_{ik} | \widehat{F_N}\Big) \\
&= \mathbb{E}({X_1}^2 Z_{1k}) - 2\frac{{\mu_k}^2}{{\lambda_k}^2}\pi_k + \frac{{\mu_k}^2}{{\lambda_k}^2}\pi_k \\
&= \frac{1}{\lambda_k}\mathbb{V}\Big(\frac{\mu_k}{\lambda_k}\Big)\pi_k,
\end{aligned}$$

we obtain

$$\lim_{N\to+\infty} \widehat{S_k} = \frac{1}{\lambda_k}\mathbb{V}\Big(\frac{\mu_k}{\lambda_k}\Big).$$

Since

$$\lim_{N\to+\infty} \frac{\widehat{\mu_k}}{\widehat{\lambda_k}} = \frac{\mu_k}{\lambda_k}, \textit{ in probability}$$

and since $\mathbb{V}$ is continuous on M_F, then we have

$$\lim_{N\to+\infty} \mathbb{V}\left(\frac{\widehat{\mu_k}}{\widehat{\lambda_k}}\right) = \mathbb{V}\left(\frac{\mu_k}{\lambda_k}\right), \textit{ in probability.}$$

Consequently,

$$\lim_{N\to+\infty}\widehat{\lambda_k}=\lim_{N\to+\infty}\frac{\mathbb{V}\left(\frac{\widehat{\mu_k}}{\widehat{\lambda_k}}\right)}{\widehat{S_k}}=\frac{\mathbb{V}\left(\frac{\mu_k}{\lambda_k}\right)}{\frac{1}{\lambda_k}\mathbb{V}\left(\frac{\mu_k}{\lambda_k}\right)}=\lambda_k,$$

which completes this proof.

5. Conclusion

In this paper, we tried to answer two major issues raised in the introduction. First, we proposed an iterative algorithm called the EMM algorithm in order to estimate the parameters of the finite mixture of exponential dispersion distributions. The EMM algorithm is the combination of the EM algorithm (based on the method of maximum likelihood) and the method of moments. The proposed algorithm determines the estimators of the mixing weight π, the mean μ, and the dispersion parameter λ. Second, we presented the asymptotic properties of the EMM sequence when the data set of size N tends to infinity, and the number of iterations l tends to infinity. These results suggest that the mixture of exponential dispersion distributions is proved to be an efficient solution for an estimation framework. In a future work, we propose to use the EDM in an unsupervised image segmentation.

References

[1] A. Dempster, N. Laird and D. Rubin, Maximum likelihood from Incomplete Data via the EM Algorithme, Journal of the Royal Statistical Society B (Methodological)39 (1977) 1-38.

[2] A.T. Mendoza-Rosas and S. De la Cruz-Reyna, A mixture of exponential distributions for a simple and precise assessment of the volcanic hazard, Natural Hazards Earth System Sciences 9 (2009) 425-431.

[3] B.Jorgensen, Exponential dispersion models (with discussion), Journal of the Royal Statistical Society Series B 49 (1987) 127-162.

[4] B. Jorgensen, The Theory of Dispersion Models, London, Chapman & Hall/CRC Monographs on statistics & Applied Probability, 1997.

[5] C.C. Kokonendji, S.Dossou-Gbété and C.G.B. Demétrio, Some discrete exponential dispersion models: Poisson-Tweedie and Hinde-Demétrio classes 28 (2004) 201-214.

[6] C.F.Jeff Wu, On the Convergence Properties of the EM Algorithm, The Annals of Statistics 11 (1983) 95-103.

[7] C.Yeojin and B.G.Lindsay, Convergence of the EM algorithm for continuous mixing distributions, Statistics & Probability Letters 96(2015) 190 – 195.

[8] D. M. Titterington, A. F. M. Smith, U. E. Makov, Statistical analysis of finite mixture distributions, Wiley series in probability and mathematical statistics, 1985.

[9] D. Nityasuddhi, D.Böhning, Asymptotic properties of the EM algorithm estimate for normal mixture models with component specific variances, Computational Statistics & Data Analysis 41 (2003) 591-601.

[10] G. Abdelaziz, M.Afif, Characterization of multinomial exponential families by generalized variance, Statistics and Probability Letters 80 (2010) 939-944.

[11] G.J.McLachlan, D. Peel, Finite Mixture Models, Wiley series in probability and statistics, New York, 2000.

[12] G. Letac, Lectures on Natural Exponential Families and their Variance Functions, Numéro 50 de Monografias de Matemática Pura e Aplicada, 1992.

[13] G. Maclachlan, T. Krishnan, The EM Algorithm and Extensions, (2nd edition), Wiley Series in Probability and Statistics, 2007.

[14] J-L. Foulley, Algorithme EM: théorie et application au modèle mixte, Journal de la société Française de statistique 143 (2002) 57-109.

[15] J-M. Marin, K.Mengersen, C P.Robert, Bayesian modelling and inference on mixtures of distributions, In Dey, D. & Rao,C.R.(Eds.) Handbook of Statistics, 2005.

[16] K. O. Bowman and L. R. Shenton, Estimator: Method of Moments, Encyclopedia of statistical sciences, 1998.

[17] K. Tanaka and A. Takemura, Strong Consistency of the Maximum Likelihood Estimator for Finite Mixtures of Location - Scale Distributions When the Scale Parameters are exponentially Small, International Statistical Institute (ISI) and the Bernoulli Society for Mathematical Statistics and Probability 12 (2006) 1003-1017.

[18] O. Barndorff-Nielsen, Information and Exponential Families: In Statistical Theory, John Wiley and Sons, New York, 1978.

[19] R. A. Boyles, On the convergence of the EM algorithm, Journal of the Royal Statistical Society, Series B (Methodological) 45 (1983) 47-50.

[20] R.A. Redner and H F. Walker, Mixture Densities, Maximum Likelihood and the EM Algorithm, SIAM Review 26 (1984) 195-239.

[21] R. Sundberg, Maximum Likelihood Theory for Incomplete Data from an Exponential Family, Scandinavian Journal of Statistics 1(1974) 49-58.

[22] V. Seshadri, Finite Mixtures of Natural Exponential Families, The Canadian Journal of Statistics 19 (1991) 437-445.

[23] Y. Miyata, Maximum likelihood estimators in finite mixture models with censored data, Journal of statistical Planning and inference 141 (2011) 56-64.