American Journal of Mathematics and Statistics 2016, 6(4): 162-169
DOI: 10.5923/j.ajms.20160604.04

Consistency of Estimators in Mixtures of Stochastic Differential Equations with Additive Random Effects

Alkreemawi Walaa Khazal[1,2,*], Wang Xiang Jun[1]

[1]School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan, China
[2]Department of Mathematics, College of Science, Basra University, Basra, Iraq

Abstract A stochastic differential equation (SDE) defines N independent stochastic processes $(X_i(t), t \in [0, T_i])$, $i = 1, \dots, N$. The drift term depends on the random variable ϕ_i. The distribution of the random effect ϕ_i depends on unknown parameters, which are to be estimated from continuous observation of the processes X_i. When the drift term is defined linearly on the random effect ϕ_i (additive random effect) and ϕ_i has a Gaussian mixture distribution, we obtain an expression of the exact likelihood. When the number of components is known, we prove the consistency of the maximum likelihood estimators $(MLE's)$. The convergence of the EM algorithm described when the algorithm is used to compute (MLE).

Keywords Maximum likelihood estimator, Mixed effects stochastic differential equations, Consistency, EM algorithm, Mixture distribution

1. Introduction

A mixture model (MM) is beneficial for modeling data as output from one of several groups, clusters, or classes; the groups (clusters, classes) might be different from each other, but the observations within the same group are similar to each other. In this paper, we concentrate on the classification problem of longitudinal data modeled by a stochastic differential equation (SDE) with random effects that have a mixture of Gaussian distributions. Some researchers state that the classes are known, whereas other researchers state the opposite. Arribas-Gil et al. [1] and references therein assumed that the classes are known and deal with classification drawbacks of longitudinal data by using random effects models or mixed-effects models. Their aim is to establish a classification rule of longitudinal profiles (curves) into a number that enables dissimilar classes to predict the class of a new individual. Celeux et al. [6] and Delattre et al. [9] assumed that the numbers of classes are unknown. Celeux et al. [6] used maximum likelihood with the EM algorithm to estimate the random effects within a mixture of linear regression models that include random effects (see, Dempster, A. et. al. [10]), and they used Bayesian information criterion (BIC) to select the number of components. Delattreet al. [9] used maximum likelihood to estimate the random effects in SDE with multiplicative random effect in the drift and diffusion terms without random effects with the EM algorithm (Dempster, A. et. al. [10]). They also used BIC to select the number of components. Delattre et al. [9] studied SDEs with the following form:

$$dX_i(t) = \left(\varphi_i' b(X_i(t)) + a(X_i(t))\right) dt + \sigma(X_i(t)) dW_i(t),$$
$$(0) = x, \quad (1)$$

where $(W_1, \dots, W_N)$ are N independent Wiener processes, $(\phi_1, \dots, \phi_N)$ are N independently and identically distributed $(\mathrm{i.i.d})$ random variables. The processes $(W_1, \dots, W_N)$ are also independent on random variables $(\phi_1, \dots, \phi_N)$, and x is a known real value. The drift function $b(x; \varphi)$ is a known function defined on $b(.): \mathbb{R} \to \mathbb{R}^m$ and the functions $\sigma(.), a(.): \mathbb{R} \to \mathbb{R}$. Each process $(X_i(t))$ represents an individual, and the random variable ϕ_i represents the random effect of individual i.

Delattre et al. [2] considered the special case (multiple case) where $\mathrm{b}(\mathrm{x}, \phi_{\mathrm{i}})$ is linear in ϕ_{i}; in other words, $b(x, \phi_i) = \phi_i\, b(x)$, where $b(x)$ is a known real function, and ϕ_i has a Gaussian mixture distribution.

Here, we consider functional data modeled by a SDE with drift term $b(x, \phi_i)$ depending on random effects where $b(x, \phi_i)$ is linear in random effects ϕ_i (addition case), $b(X_i(t), \phi_i) = \phi_i + b(X_i(t))$ and diffusion term without random effects. We consider continuous observations $(X_i(t), t \in [0, T], i = 1, \dots, N)$ with a given T. Here, θ are unknown parameters in the distribution of ϕ_i from the (X_i)'s which will be estimated, but the estimation is not straightforward. Generally, the exact likelihood is not explicit. Maximum likelihood estimation in SDE with

* Corresponding author:
mathematica_walla@yahoo.com (Alkreemawi Walaa Khazal)
Published online at http://journal.sapub.org/ajms

random effects has been studied in a few papers (Ditlevsen and De Gaetano, 2005 [11]; Donnet and Samson, 2008 [12]; Delattre et al. (2013) [8]; Alkreemawi et al. ([2], [3]); Alsukaini et al. ([4], [5]).

In this paper, we assume that the random variables $\phi_1, \dots, \phi_N$ have a common distribution with density $g(\varphi, \theta)$ for all θ, which is given by a mixture of Gaussian distributions; this mixture distribution models the classes. We aim to estimate unknown parameters and study the proportions. $g(\varphi, \theta)$ is a density with respect to a dominant measure on $\mathbb{R}^m$, where $\mathbb{R}$ is the real line, and m is the dimension.

$$g(\varphi, \theta) = \sum_{\ell=1}^{M} \pi_\ell n_\ell(\varphi, \mathcal{T}_\ell),$$

$$n_\ell(\varphi, \mathcal{T}_\ell) d\varphi = N_m(\mu_\ell, \Omega_\ell), \qquad \mathcal{T}_\ell = (\mu_\ell, \Omega_\ell)$$

with π_ℓ as the proportions of the mixture $\sum_{\ell=1}^{M} \pi_\ell = 1$, M as the number of components in the mixture, and $\mu_\ell \in \mathbb{R}^m$ and Ω_ℓ a $m \times m$ as an invertible covariance matrix. Let θ_0 denote the true value of the parameter. M is the number of components and is known. $\theta = ((\pi_\ell, \mathcal{T}_\ell), \ell = 1, \dots, M$ is set for the unknown parameters to be estimated. Our aim is to find estimators of the parameters θ of the density of the random effects from the observations $(X_i(t), t \in [0, T], i = 1, \dots, N)$. We focus on an additional case of linear random effects ϕ_i in the drift term $b(X_i(t), \phi_i) = \phi_i + b(X_i(t))$. In addition, we prove and explain that the observations concerning exact likelihood are explicit. With M as the number of known components, we discuss the convergence of the EM algorithm, and the consistency of the exact maximum likelihood estimator is proven.

The rest of this paper is organized as follows: Section 2 contains the notation and assumptions, and we present the formula of the exact likelihood. In Section 3, we describe the EM algorithm and discuss its convergence. In Section 4, the consistency of the exact maximum likelihood estimator is proved when the number of components is known.

2. Notations and Assumptions

Consider the stochastic processes $(X_i(t), t \geq 0), i = 1, \dots, N,)$, which are defined by (1). The processes $(W_1, \dots, W_N)$ and the random variables $(\phi_1, \dots, \phi_N)$ are defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We use the assumptions (H1, H2, and H3) in Delattre et al. [9]. Consider the filtration $(\mathcal{F}_t, t \geq 0)$ defined by $\mathcal{F}_t = \sigma(\phi_i, W_i(s), s \leq t, i = 1, \dots, N)$.

H1. The functions $x \to a(x)$ and $x \to b(x) = (b_1(x), \dots, b_m(x))'$ are Lipschitz continuous on $\mathbb{R}$ and $x \to \sigma(x)$ is Hölder continuous with exponent $\alpha \in [1/2, 1]$ on $\mathbb{R}$.

By (H1), for $i = 1, \dots, N$, for all $\varphi = (\varphi_1, \dots, \varphi_m)' \in \mathbb{R}^m$, the stochastic differential equation

$$dX_i^\varphi(t) = \left(\varphi + b\left(X_i^\varphi(t)\right) + a\left(X_i^\varphi(t)\right)\right) dt$$
$$+ \sigma\left(X_i^\varphi(t)\right) dW_i(t), \quad X_i^\varphi(0) = x \tag{2}$$

admits a unique solution process $(X_i^\varphi(t), t \geq 0)$ adapted to the filtration $(\mathcal{F}_t, t \geq 0)$, Moreover, the stochastic differential equation (1) admits a unique strong solution adapted to $(\mathcal{F}_t)$ such that the joint process $(\phi_i, X_i(t))$ is strong Markov and the conditional distribution of $(X_i(t))$ given $\phi_i = \varphi$ is identical to the distribution of (2). (1) shows that the Markov property of $(\phi_i, X_i(t))$ is straightforward as the two-dimensional SDE

$$d\phi_i(t) = 0, \qquad \phi_i(0) = \phi_i,$$
$$dX_i(t) = \left(\varphi + b\big(X_i(t)\big) + a\big(X_i(t)\big)\right) dt$$
$$+ \sigma\big(X_i(t)\big) dW_i(t), \quad X_i(0) = x$$

The processes $(\phi_i, X_i(t), t \geq 0), i = 1, \dots, N$ are $i.i.d$ (see Delattre et al. [8]; Genon-Catalot and Larédo [13], Alkreemawi et al. [2], [3]; Alsukaini et al. [4], [5]). To derive the likelihood function of our observations, under (H1), we introduce the distribution $Q_\varphi^{x,T}$ on $(C_T, \mathcal{C}_T)$ of $(X_i^\varphi(t), t \in [0, T])$ given by (2), where C_T denotes the space of real continuous functions $(x(t), t \in [0, T])$ defined on $[0, T]$ endowed with the σ-field $\mathcal{C}_T$ associated with the topology of uniform convergence on $[0, T]$. On $\mathbb{R}^m \times C_T$, let $\mathbb{P}_\theta = g(\varphi, \theta) dv(\varphi) \otimes Q_\varphi^{x,T}$ denote the joint distribution of $(\phi_i, X_i(t), t \in [0, T])$, and let $\mathbb{Q}_\theta$ denote the marginal distribution of $(X_i(t), t \in [0, T])$ on $(C_T, \mathcal{C}_T)$. Also, we denote $\mathcal{T}_\ell = (\mu_\ell, \Omega_\ell)$, $P_{\mathcal{T}_\ell}$ (resp. $Q_{\mathcal{T}_\ell}$) the distribution $n_m(\varphi, \mathcal{T}_\ell) d\varphi \otimes Q_\varphi^{x,T}$ of $\big(\phi_i, X_i(.)\big)$ where ϕ_i has a distribution $N_m(\mu_\ell, \Omega_\ell)$ (resp. of $(X_i(t), t \in [0, T])$), with these notations

$$\mathbb{P}_\theta = \sum_\ell^M \pi_\ell P_{\mathcal{T}_\ell}, \quad \mathbb{Q}_\theta = \sum_\ell^M \pi_\ell Q_{\mathcal{T}_\ell} \tag{3}$$

H2. For all $\varphi \in \mathbb{R}^m$,

$$Q_\varphi^{x,T}\left(\int_0^T \frac{b'(X(t)) b(X(t)) + a^2(X(t))}{\sigma^2(X(t))} dt < +\infty\right) = 1.$$

We denote by (ϕ, X), with $X = (X(t), t \in [0, T])$, which is the canonical process of $\mathbb{R}^m \times C_T$. Under (H1)–(H2), based on Theorem 7.19 p. 294 in [14], the distributions $Q_\varphi^{x,T}$ and $Q_0^{x,T}$ are equivalent. Through an analog approach of [9], the following results are used:

$$\frac{dQ_\varphi^{x,T}}{dQ_0^{x,T}}(X) = L_T(X, \varphi) = exp\left(\varphi' U(X) - \frac{1}{2} \varphi' V(X) \varphi\right),$$

where $U(X)$ is the vector, and $V(X)$ is the $d \times d$ matrix

$$U(X) = \int_0^T \frac{b(X(s))}{\sigma^2(X(s))} (dX(s) - a(X(s)) ds) \tag{4}$$

and $V(X)$ is the $d \times d$ matrix

$$V(X) = \int_0^T \frac{b(X(s)) b'(X(s))}{\sigma^2(X(s))} ds. \tag{5}$$

Thus, the density of $\mathbb{Q}_\theta$ (the distribution of X_i on C_T) with respect to $Q_0^{x,T}$ is obtained as follows:

$$\frac{d\mathbb{Q}_\theta}{dQ_0^{x,T}}(X) = \int_{\mathbb{R}^m} g(\varphi, \theta)\, exp\left(\varphi' U(X) - \frac{1}{2} \varphi' V(X) \varphi\right) d\varphi$$
$$:= \Lambda(X, \theta). \tag{6}$$

The exact likelihood of $(X_i = (X_i(t), t \in [0, T]), i = 1, \dots, N,)$ is

$$L_N(\theta) = \prod_{i=1}^{N} \Lambda(X_i, \theta) \quad (7)$$

Thus, for our situation (addition situation) of drift where $b(x, \phi_i)$ is linear in ϕ_i , $b(X_i(t), \phi_i) = \phi_i + b(X_i(t))$, we have

$$\frac{dQ_{\varphi}^{x,T}}{dQ_0^{x,T}}(X) = L_T(X, \varphi) = exp(\varphi' (Y(X) - D(X)) - \frac{1}{2}\varphi' Z(X)\varphi + (U(X) - \frac{1}{2}V(X))),$$

and,

$$\frac{d\mathbb{Q}_\theta}{dQ_0^{x,T}}(X) = \int_{\mathbb{R}^m} g(\varphi, \theta) \exp(\varphi' (Y(X) - D(X)) - \frac{1}{2}\varphi' Z(X)\varphi + (U(X) - \frac{1}{2}V(X)))d\varphi = \Lambda(X, \theta) \quad (8)$$

where

$$Y(X) = \int_0^T \frac{1}{\sigma^2(X(s))} \left(dX(s) - a(X(s))ds\right), \quad (9)$$

$$Z(X) = \int_0^T \frac{1}{\sigma^2(X(s))} ds, \quad D(X) = \int_0^T \frac{b(X(s))}{\sigma^2(X(s))} ds \quad (10)$$

We have to consider distributions for ϕ_i such that the integral (8) can obtain a tractable formula for the exact likelihood. This is the case when ϕ_i has a Gaussian distribution and the drift term $b(x, \phi_i)$ is linear in ϕ_i, $b(X_i(t), \phi_i) = \phi_i + b(X_i(t))$ (addition situation), as shown in Alkreemawi et al. [2]. This is also the case for the larger class of Gaussian mixtures. The required assumption is defined as follows:

H3. The matrix $Z(X)$ is positive definite $Q_0^{x,T}$-a.s. and $\mathbb{Q}_\theta$-a.s. for all θ.

(H3) is not true when the functions (b_j/σ^2) and $(1/\sigma^2)$ are not linearly independent. Thus, (H3) can ensure a well-defined dimension of the vector ϕ.

Proposition 2.1. Assume that $g(\varphi, \theta)dv(\varphi)$ is a Gaussian mixture distribution, and set $\mathcal{T}_\ell = (\mu_\ell, \Omega_\ell)$, $Y_i = Y(X_i)$, $Z_i = Z(X_i)$, $D_i = D(X_i)$, $U_i = U(X_i)$, $V_i = V(X_i)$. Under (H3), the matrices $Z_i + \Omega_\ell^{-1}$, $I_m + Z_i\Omega_\ell$, $I_m + \Omega_\ell Z_i$ are invertible $Q_0^{x,T} - a.s.$ and $\mathbb{Q}_\theta -$ a.s. for all θ. By setting $R_{i,\ell}^{-1} = (I_m + Z_i\Omega_\ell)^{-1}Z_i$, we obtain,

$$\Lambda_N(\theta) = \sum_{\ell=1}^{M} \pi_\ell \lambda(X_i, \mathcal{T}_\ell) \quad (11)$$

where

$$\lambda(X_i, \mathcal{T}_\ell) = \frac{1}{\sqrt{det(I_m + Z_i\Omega_\ell)}} exp\left[-\frac{1}{2} \times \left(\mu_\ell - Z_i^{-1}(Y_i - D_i)\right)' R_i^{-1}\left(\mu_\ell - Z_i^{-1}(Y_i - D_i)\right)\right] \times \exp\left[\frac{1}{2}(Y_i - D_i)' Z_i^{-1}(Y_i - D_i)\right] \exp\left[U_i - \frac{1}{2}V_i\right]$$
$$= \sqrt{2\pi \det(Z_i)} \exp\left[\frac{1}{2}(Y_i - D_i)' Z_i^{-1}(Y_i - D_i)\right] \times exp\left[U_i - \frac{1}{2}V_i\right] \times n_m\left((Y_i - D_i), (Z_i\mu_\ell, (I_m + \Omega_\ell Z_i)Z_i\right) \quad (12)$$

Here, $n_m\left((Y_i - D_i), (Z_i\mu_\ell, (I_m + \Omega_\ell Z_i)Z_i\right)$ denotes the Gaussian density with mean $Z_i\mu_\ell$ and covariance matrix $(I_m + \Omega_\ell Z_i)Z_i$.

Alkreemawi et al. [2] considered the formula for $\lambda(X_i, \mathcal{T}_\ell)$ (Proposition 3.1.1 and Lemma 4.2). The exact likelihood (7) is explicit. Hence, we can study the asymptotic properties of the exact MLE, which can be computed by using the EM algorithm instead of maximizing the likelihood.

3. Estimation Algorithm

In the situation of mixtures distributions with number of components M, M is known, rather than of solving the likelihood equation, we use the EM algorithm to find a stationary point of the log-likelihood. A Gaussian mixture model (GMM) is helpful for modeling ϕ_i by using a mixture of distributions, which means that the population of individuals is grouped in M clusters. Formally, for the individual i, we (may) introduce a random variable $Y_i \in \{1, \ldots, M\}$, with $P_\theta\ (Y_i = \ell) = \pi_\ell$ and $P_\theta\left(\phi_i \in d_\varphi | Y_i = \ell\right) = \mathcal{N}_\ell(\mu_\ell, \Omega_\ell)$. We assume that (ϕ_i, Y_i) are $i.i.d$ and $(\phi_i, Y_i)_{i=1,\ldots,N}$ independent of $(W_1, \ldots, W_N)$. The concept of the EM algorithm was presented in Dempster et al. [10] which considered the data X_i as incomplete and introduced the unobserved variables $(Y_1, \ldots, Y_N)$. Simply, in the algorithm, we can consider random variables $Z = (Z_i)_{i=1,\ldots,N}$, $Z_i = (Z_{i1}, \ldots, Z_{iM})$ where$Z_{i\ell} = 1_{(Y_i=\ell)}$, for $\ell = 1, \ldots, M$; such values indicate that the density component drives the equation of subject i. For the complete data (X_i, Z_i), the logarithm likelihood function is explicitly given by

$$\mathcal{L}_N\left((X_i, Z_i), \theta\right) = \sum_{i=1}^{N}\sum_{\ell=1}^{M} Z_{il} \log(\pi_\ell \lambda(X_i, \tau_\ell)) \quad (13)$$

The EM algorithm is an iterative method in which the iteration alternates between performing an expectation (E) step, which is the computation of

$$Q\left(\theta \middle| \theta^{(t)}\right) = E(\mathcal{L}_N((X_i, Z_i); \theta) | (X_i); \theta^{(t)})$$

where $E(.|(X_i); \theta^{(t)})$ is the conditional expectation given (X_i) computed with the distribution of the complete data under the value $\theta^{(t)}$ of the parameter, and the maximization (M) step computes parameters that maximize the expected log-likelihood found on the (E) step $Q\left(\theta \middle| \theta^{(t)}\right)$. In the (E) step, we compute

$$Q\left(\theta \middle| \theta^{(t)}\right) = \sum_{i=1}^{N}\sum_{\ell=1}^{M} \tilde{\pi}_\ell(X_i, \theta^{(t)}) \log(\pi_\ell \lambda(X_i, \tau_\ell))$$

where $\tilde{\pi}_\ell(X_i, \theta^{(t)})$ is the posterior probability

$$\tilde{\pi}_\ell\left(X_i, \theta^{(t)}\right) = P\left(Z_{i\ell} = 1 \middle| (X_i), \theta^{(t)}\right) = \frac{\pi_\ell^{(t)} \lambda(X_i, \tau_\ell^{(t)})}{\Lambda(X_i, \theta^{(t)})} \quad (14)$$

In the EM algorithm, at iteration n, we want to maximize $Q\left(\theta \middle| \hat{\theta}^{(n)}\right)$ with respect to θ, where $\hat{\theta}^{(n)}$ is the current value of parameter θ. We can maximize the terms that contain π_ℓ and $\mathcal{T}_\ell = (\mu_\ell, \Omega_\ell)$ separately. We introduce one Lagrange multiplier α to maximize with respect to π_ℓ

with the constraint $\sum_{\ell=1}^{M} \pi_\ell = 1$ and solve the following equation:

$$\frac{\partial}{\partial \pi_\ell}\left[\sum_{i=1}^{N}\sum_{\ell=1}^{M}\tilde{\pi}_\ell\left(X_i,\hat{\theta}^{(n)}\right)log\pi_\ell+\alpha\left(\sum_\ell \pi_\ell-1\right)\right]=0$$

And the classical solution:

$$\hat{\pi}_\ell^{(n+1)}=\frac{1}{N}\sum_{i=1}^{N}\tilde{\pi}_\ell\left(X_i,\hat{\theta}^{(n)}\right)$$

Then, we maximize $\sum_{i=1}^{N}\sum_{\ell=1}^{M}\tilde{\pi}_\ell\left(X_i,\hat{\theta}^{(n)}\right)log\lambda(X_i,\tau_\ell)$, where the derivatives can be computed with respect to the components of μ_ℓ and Ω_ℓ by using some results from matrix algebra. When taking the log of $\lambda(X_i,\tau_\ell)$, substituting it into $\sum_{i=1}^{N}\sum_{\ell=1}^{M}\tilde{\pi}_\ell\left(X_i,\hat{\theta}^{(n)}\right)log\lambda(X_i,\tau_\ell)$, and taking the derivative w.r.t. μ_ℓ, we have

$$\hat{\mu}_\ell^{(n+1)}=\left(\sum_{i=1}^{N}\tilde{\pi}_\ell\left(X_i,\hat{\theta}^{(n)}\right)(I_m+\Omega_\ell Z_i)^{-1}Z_i\right)^{-1}$$
$$\times\sum_{i=1}^{N}\tilde{\pi}_\ell\left(X_i,\hat{\theta}^{(n)}\right)(I_m+\Omega_\ell Z_i)^{-1}(Y_i-D_i) \quad (15)$$

When the Ω_ℓs are known and when the Ω_ℓs are unknown, the maximum likelihood estimators of the parameters are given by the system.

Proposition 3.1 The sequence $\hat{\theta}^{(n)}$ generated by the EM algorithm converges to a stationary point of the likelihood.

Proof. We prove the convergence for $m=1$ to avoid cumbersome details. We employ the results obtained by McLachlan and Krishnan [15]. As the following conditions are given:

1- $\Theta\subset\mathbb{R}^{3M-1}$.
2- $\Theta_{\theta_0}=\{\theta\in\Theta,\ L_N(\theta)\geq L_N(\theta_0)\}$ is a compact set if $L_N(X,\theta_0)>-\infty$.
3- $L_N(\theta)$ is continuous on Θ and differentiable on the interior of Θ.
4- $Q\left(\theta,\theta^{(t)}\right)$ is continuous with respect to both θ and $\theta^{(t)}$.
5- $\frac{\partial Q\left(\theta,\hat{\theta}^{(n)}\right)}{\partial\theta}|_{\theta=\hat{\theta}^{(n+1)}}=0$.
6- $\frac{\partial Q\left(\theta,\theta^{(t)}\right)}{\partial\theta}$ is continuous in both θ and $\theta^{(t)}$.

Conditions 3, 4, 5, and 6 are verified by the regularity of the likelihood (see Proposition 4.2). In a standard Gaussian mixture, condition 2 is usually unverified (see McLachlan and Krishnan [15]. However, here, one has the following result (see (12)):

$$\Lambda(X,\theta)\propto\sum_\ell^M\pi_\ell\, n\Big((Y-D),(\mu_\ell Z,\sigma_\ell^2 Z\Big)$$

Where $\sigma_\ell^2(Z)=Z(1+w_\ell^2)\geq Z>0$. Therefore, the formula of $\Lambda(X,\theta)$ is a mixture of Gaussian distributions that consist of variances all bounded from below. This finding reveals condition 2.

4. Asymptotic Properties of MLE

This section aims to investigate theoretically the consistency and asymptotic normality of the exact maximum likelihood estimator of θ_0 when we assume that the number of components M is known. For simplicity's sake, we consider only the case $m=1$. The parameter set Θ is given by

$$\Theta=\{\ (\pi_\ell,\tau_\ell),\ell=1,\dots,M,$$
$$\forall\ell\in\{1,\dots,M-1\},0<\pi_\ell<1,0<1-\sum_{\ell=1}^{M-1}\pi_\ell<1,$$
$$\tau_\ell=(\mu_\ell,w_\ell^2)\in\mathbb{R}\times\mathbb{R}^+,\ell\neq\ell'\Rightarrow\tau_\ell\neq\tau_\ell'\ \}$$

Now, we set $\pi_M=1-\sum_{\ell=1}^{M-1}\pi_\ell$, but only $3M$-1 parameters need to be estimated. When necessary in notations, we set $\theta=(\theta_1,\dots,\theta_{3M-1})$. The MLE is defined as any solution of

$$\hat{\theta}_N=arg\ max_{\theta\in\Theta}L_N(\theta)$$

where $L_N(\theta)$ is defined by (7)–(11). To prove the identifiability property, the following assumption is required as in Alkreemawi et al. [2]:

(H4) Either the function $b(.)/\sigma(.)$ is constant or not constant, and under $Q_0^{x,T}$, the random variable $(D(X),Y(X),Z(X))$ admits a density $f(d,y,z)$ with respect to the Lebesgue measure on $\mathbb{R}\times\mathbb{R}^+$, which is jointly continuous and positive on an open ball of $\mathbb{R}\times\mathbb{R}^+$.

When $b(.)/\sigma(.)$ is constant, this case is simple. For instance, let $b(.)\equiv\sigma(.)\equiv 1$. Then, $Z(X)=D(X)=V(X)=T$ is deterministic, and under $Q_0^{x,T}$, $U(X)=Y(X)=W_T$. Under Q_θ, $(Y(X)-D(X))$ is a mixture of Gaussian distributions with means $(\mu_\ell T)$, variances $(T(1+w_\ell^2T))$, and proportions (π_ℓ).

The case where $b(.)/\sigma(.)$ is not constant. Under smoothness assumptions on functions b,σ, assumption (H4) will be accomplished by using Malliavin calculus tools (see Alkreemawi et al. [2]). As mixture distributions are utilized, the identifiability of the entire parameter θ can only be obtained in the following concept:

$$\theta\sim\theta_0\Leftrightarrow\ \{(\pi_\ell,\tau_\ell),\ell=1,\dots,M\ \}$$
$$=\left\{\left(\pi_{\ell,0}\right),,\tau_{\ell,0}\right),\ell=1,\dots,M\ \right\}\quad(16)$$

Now, we can prove the following:

Proposition 4.1. Under (H1)-(H2)-(H4), $\mathbb{Q}_\theta=\mathbb{Q}_{\theta_0}$ implies that $\theta\sim\theta_0$.

Proof. First, when $b(.)/\sigma(.)$ is not constant, we consider two parameters θ and θ_0, and aim to prove that $\mathbb{Q}_\theta=\mathbb{Q}_{\theta_0}$ implies $\theta\sim\theta_0$. As $\Lambda(X,\theta)$ and $\lambda(X,\tau_\ell)$ depend on X only through the statistics $Y(X),D(X),Z(X),U(X),V(X)$ with a slight abuse of notation, we set $Y(X)=Y,U(X)=U,Z(X)=Z,D(X)=D,V(X)=V$ and

$$\Lambda(X,\theta)=\Lambda(Y,D,Z,U,V,\theta),$$
$$\lambda(X,\tau_\ell)=\lambda(Y,D,Z,U,V,\tau_\ell)\qquad(17)$$

Under (H4), $\Lambda(y,d,z,u,v,\theta)$ is the density of the distribution of $(Y,D,Z,U,V,)$ under $\mathbb{Q}_\theta$ with respect to the density of $(Y,D,Z,U,V,)$ under $Q_0^{x,T}$ and $\mathbb{Q}_\theta=\mathbb{Q}_{\theta_0}$ implies $\Lambda(y,d,z,u,v,\theta)=\Lambda(y,d,z,u,v,\theta_0)$ a.e., hence, everywhere on $\mathbb{R}\times\mathbb{R}^+$ by the continuity assumption. We deduce that the following equality holds for all $y,d\in\mathbb{R},z>0$:

$$\sum_{\ell=1}^{M} \pi_\ell \frac{1}{\sqrt{1+w_\ell^2 z}} exp\left[-\frac{z\left(\frac{(y-d)}{z}-\mu_\ell\right)^2}{2(1+w_\ell^2 z)}\right]$$

$$= \sum_{\ell=1}^{M} \pi_{\ell,0} \frac{1}{\sqrt{1+w_{\ell,0}^2 z}} exp\left[-\frac{z\left(\frac{(y-d)}{z}-\mu_{\ell,0}\right)^2}{2(1+w_{\ell,0}^2 z)}\right]$$

Let us set

$$p(z) = \prod_{1\le \ell \le M} \sqrt{1+w_\ell^2 z},$$

$$q_\ell(z) = \prod_{1\le \ell' \le M, \ell' \neq \ell} \sqrt{1+w_{\ell'}^2 z},$$

and

$$p_0(z) = \prod_{1\le \ell \le M} \sqrt{1+w_{\ell,0}^2 z},$$

$$q_{\ell,0}(z) = \prod_{1\le \ell' \le M, \ell' \neq \ell} \sqrt{1+w_{\ell',0}^2 z},$$

We note that $q_\ell(z)\sqrt{1+w_\ell^2 z} = p(z)$, $q_{\ell,0}(z)\sqrt{1+w_{\ell,0}^2 z} = p_0(z)$. Thus, such quantities do not depend on ℓ. After reducing to the same denominator, we obtain

$$\frac{p_0(z)}{p(z)} = \frac{\sum_{\ell=1}^{M} \pi_{\ell,0} q_{\ell,0}(z) exp\left[-z\left(\frac{(y-d)}{z}-\mu_{\ell,0}\right)^2/2(1+w_{\ell,0}^2 z)\right]}{\sum_{\ell=1}^{M} \pi_\ell q_\ell(z) exp\left[-z\left(\frac{(y-d)}{z}-\mu_\ell\right)^2/2(1+w_\ell^2 z)\right]}$$

The right-hand side is a function of (y,d,z), whereas the left-hand side is a function of z only. This approach is possible only if $p(z) = p_0(z)$ for all $z > 0$. Therefore,

$$\{w_1^2, \dots, w_M^2\} = \{w_{1,0}^2, \dots, w_{M,0}^2\} \tag{18}$$

and the equality of the variances can be obtained by reordering the terms if required. Then, we have for $\sigma_\ell^2(z) = z(1+w_\ell^2 z)$ and a fixed z,

$$\sum_{\ell=1}^{M} \pi_\ell\, q_\ell(z) exp\left[-\frac{z\left(\frac{(y-d)}{z}-\mu_\ell\right)^2}{2(1+w_\ell^2 z)}\right]$$

$$= p(z)\sqrt{2\pi z}\sum_{\ell=1}^{M} \pi_\ell\, n\big((y-d), (\mu_\ell z, \sigma_\ell^2(z))\big) \tag{19}$$

Here, $n\big((y-d),(m,\sigma^2)\big)$ indicates the Gaussian density with mean m and variance σ^2. Analogously, by using the equality (18),

$$\sum_{\ell=1}^{M} \pi_{\ell,0}\, q_{\ell,0}(z) exp\left[-\frac{z\left(\frac{(y-d)}{z}-\mu_{\ell,0}\right)^2}{2(1+w_{\ell,0}^2 z)}\right] =$$

$$p(z)\sqrt{2\pi z} \times \sum_{\ell=1}^{M} \pi_{\ell,0}\, n\left((y-d), \left(\mu_{\ell,0} z, \sigma_{\ell,0}^2(z)\right)\right)$$

For all fixed $z > 0$, we therefore have for all $y, d \in \mathbb{R}$,

$$\sum_{\ell=1}^{M} \pi_\ell\, n\big((y-d), (\mu_\ell z, \sigma_\ell^2(z))\big) =$$

$$\sum_{\ell=1}^{M} \pi_{\ell,0}\, n\left((y-d), \left(\mu_{\ell,0} z, \sigma_{\ell,0}^2(z)\right)\right).$$

Herein, the equality of two mixtures of Gaussian distributions with proportions (π_ℓ) and $(\pi_{\ell,0})$, expectations $(\mu_\ell z)$ and $(\mu_{\ell,0} z)$, and the same set of known variances $z(1+w_\ell^2 z)$ are given. From the identifiability of Gaussian mixtures, we have the equality

$$\{(\pi_\ell, \mu_\ell), \ell = 1, \dots, M\} = \{(\pi_{\ell,0}, \mu_{\ell,0}), \ell = 1, \dots, M\},$$

and thus, $\theta \sim \theta_0$.

Second, when $b(.)/\sigma(.)$ is constant, for instance, let $b(.) \equiv 1, \sigma(.) \equiv 1$. As noted above, under $Q_0^{x,T}$, $Y = W_T$ and $D = T$. Therefore, $(Y-D)$ is a Gaussian distribution. Under Q_θ, $(Y-D)$ has density $\Lambda((y-d),\theta) \times exp(-((y-d)+T)^2/2T)$ w. r. t. the Lebesgue measure on $\mathbb{R}$. Specifically, the mixture of Gaussian densities can also be deduced based on the identifiability property of Gaussian mixtures.

Proposition 4.2. Let $\mathbb{E}_{\theta_0}$ denote the expectation under $\mathbb{Q}_{\theta_0}$. The function $\theta \to \Lambda(X,\theta)$ is C^∞ on Θ and

1- $\mathbb{E}_{\theta_0}\left(\frac{\partial \log \Lambda(X,\theta)}{\partial \theta_k}|_{\theta=\theta_0}\right)^2 < +\infty$ and
$\mathbb{E}_{\theta_0}\left(\frac{\partial \log \Lambda(X,\theta)}{\partial \theta_k}|_{\theta=\theta_0}\right) = 0$ for all $1 \le k \le 3M-1$.

2- $\mathbb{E}_{\theta_0}\left|\frac{\partial^2 \log \Lambda(X,\theta)}{\partial \theta_k \partial \theta_j}|_{\theta=\theta_0}\right| < +\infty$ for all$1 \le k$, $j \le 3M-1$.

$$\mathbb{E}_{\theta_0}\left(\frac{\partial \log \Lambda(X,\theta)}{\partial \theta_k}|_{\theta=\theta_0} \frac{\partial \log \Lambda(X,\theta)}{\partial \theta_j}|_{\theta=\theta_0}\right)$$

$$= -\mathbb{E}_{\theta_0}\left(\frac{\partial^2 \log \Lambda(X,\theta)}{\partial \theta_k \partial \theta_j}\right)|_{\theta=\theta_0}$$

for all $1 \le k,\ j \le 3M-1$.

The Fisher information matrix can be defined as

$$I(\theta_0) = \left[\mathbb{E}_{\theta_0}\left(\frac{\partial \log \Lambda(X,\theta)}{\partial \theta_k}|_{\theta=\theta_0} \frac{\partial \log \Lambda(X,\theta)}{\partial \theta_j}|_{\theta=\theta_0}\right)\right]$$

for all $1 \le k,\ j \le 3M-1$.

Proof. We use results proved in Alkreemawi el al. [2] **(**Section 3.1, Lemma 3.1.1, Proposition 3.1.2) to prove this proposition. For all $q = (y-d) \in \mathbb{R}$ and all $\tau = (\mu, w^2)$,

$$E_{Q_\tau}\left(exp\left(q\frac{Y-D}{1+w^2 Z}\right)\right) < +\infty$$

Q_τ is the distribution of X_i when ϕ_i has a Gaussian distribution with parameters $\tau = (\mu, w^2)$. This idea implies that $E_{Q_\tau}\left|\frac{Y-D}{1+w^2 Z}\right|^n < +\infty$ for all $n \ge 1$.

Let

$$\eta(\tau_\ell) = \left(\frac{(Y-D)-\mu_\ell Z}{1+w_\ell^2 Z}\right), \quad \xi(w_\ell^2) = \frac{Z}{1+w_\ell^2 Z}$$

The random variable that has moments of any order under Q_τ, $\xi(w_\ell^2)$ is bounded, and the following relations hold:

$$E_{Q_\tau}\eta(\tau_\ell) = 0,\ E_{Q_\tau}\left(\eta^2(\tau_\ell) - \xi(w_\ell^2)\right) = 0 \tag{20}$$

$$E_{Q_\tau}\left[\left(\frac{1}{2}\left(\eta^2(\tau_\ell) - \xi(w_\ell^2)\right)\right)^2 - \eta^2(\tau_\ell)\xi(w_\ell^2) - \frac{1}{2}\xi^2(w_\ell^2)\right] = 0 \tag{21}$$

$$E_{Q_\tau}\left(\frac{1}{2}\eta^3(\tau_\ell) - \frac{3}{2}\eta(\tau_\ell)\xi(w_\ell^2)\right) = 0 \tag{22}$$

All derivatives of $\Lambda(X,\theta)$ are well defined. For

$\ell = 1, \ldots, M-1$, we have

$$\frac{\partial \Lambda(X,\theta)}{\partial \pi_\ell} = \lambda(X,\tau_\ell) - \lambda(X,\tau_M)$$

As for all $\tau = (\mu, w^2)$, $Q_\tau = \lambda(X,\tau) Q_0^{x,T}$, the random variable above is $Q_0^{x,T}$-integrable and

$$\int_{C_T} \frac{\partial \Lambda(X,\theta)}{\partial \pi_\ell} dQ_0^{x,T} = \int_{C_T} \frac{\partial log\, \Lambda(X,\theta)}{\partial \pi_\ell} d\mathbb{Q}_\theta = 0$$

Moreover, as $\lambda(X,\tau_\ell)/\Lambda(X,\theta) \le \pi_\ell^{-1}$, we have

$$\left(\frac{\partial log\, \Lambda(X,\theta)}{\partial \pi_\ell}\right)^2 \Lambda(X,\theta) = \frac{\left(\frac{\partial \Lambda(X,\theta)}{\partial \pi_\ell}\right)^2}{\Lambda(X,\theta)}$$
$$\le \frac{2}{\pi_\ell}\lambda(X,\tau_\ell) + \frac{2}{\pi_M}\lambda(X,\tau_M)$$

Therefore,

$$E_{\mathbb{Q}_\theta}\left(\frac{\partial log\, \Lambda(X,\theta)}{\partial \pi_\ell}\right)^2$$
$$\int_{C_T}\left(\frac{2}{\pi_\ell}\lambda(X,\tau_\ell) + \frac{2}{\pi_M}\lambda(X,\tau_M)\right) dQ_0^{x,T} = \frac{2}{\pi_\ell} + \frac{2}{\pi_M}$$

Higher order derivatives of $\Lambda(X,\theta)$ with respect to the π_ℓ's are nul:

$$\frac{\partial^2 \Lambda(X,\theta)}{\partial \pi_\ell \partial \pi_{\ell'}} = 0$$

Now we find the derivatives with respect to the parameters μ_ℓ, w_ℓ^2. We have:

$$\frac{\partial \Lambda(X,\theta)}{\partial \mu_\ell} = \pi_\ell \frac{\partial \lambda(X,\tau_\ell)}{\partial \mu_\ell} = \pi_\ell \eta(\tau_\ell)\lambda(X,\tau_\ell)$$

We know that:

$$E_{Q_{\tau_\ell}}|\eta(\tau_\ell)| = \int_{C_T} |\eta(\tau_\ell)|\lambda(X,\tau_\ell) dQ_0^{x,T} < +\infty$$

$$E_{Q_{\tau_\ell}}\eta(\tau_\ell) = \int_{C_T} \eta(\tau_\ell)\lambda(X,\tau_\ell) dQ_0^{x,T} = 0.$$

Consequently,

$$\int_{C_T}\left|\frac{\partial \Lambda(X,\theta)}{\partial \mu_\ell}\right| dQ_0^{x,T} < +\infty,$$

$$\int_{C_T}\frac{\partial \Lambda(X,\theta)}{\partial \mu_\ell} dQ_0^{x,T} = E_{\mathbb{Q}_\theta}\frac{\partial log\, \Lambda(X,\theta)}{\partial \mu_\ell} = 0.$$

Now,

$$\left(\frac{\partial log\, \Lambda(X,\theta)}{\partial \mu_\ell}\right)^2 \Lambda(X,\theta) = \frac{\left(\frac{\partial \Lambda(X,\theta)}{\partial \mu_\ell}\right)^2}{\Lambda(X,\theta)} = \frac{\pi_\ell^2 \eta^2(\tau_\ell)\lambda^2(X,\tau_\ell)}{\Lambda(X,\theta)}$$
$$\le \pi_\ell \eta^2(\tau_\ell)\lambda(X,\tau_\ell)$$

Thus,

$$E_{\mathbb{Q}_\theta}\left(\frac{\partial log\, \Lambda(X,\theta)}{\partial \mu_\ell}\right)^2 \le \pi_\ell E_{Q_{\tau_\ell}}\left(\eta^2(\tau_\ell)\right) = \pi_\ell \xi(w_\ell^2)$$

Next, we have

$$\frac{\partial \Lambda(X,\theta)}{\partial w_\ell^2} = \pi_\ell \frac{\partial \lambda(X,\theta)}{\partial w_\ell^2} = \pi_\ell \frac{1}{2}\left(\eta^2(\tau_\ell) - \xi(w_\ell^2)\right)\lambda(X,\tau_\ell)$$

Again, we know that this random variable is $Q_0^{x,T}$-integrable with nul integral, thereby obtaining

$$E_{\mathbb{Q}_\theta}\left|\frac{\partial log\, \Lambda(X,\theta)}{\partial w_\ell^2}\right| < +\infty, \text{ and } E_{\mathbb{Q}_\theta}\frac{\partial log\, \Lambda(X,\theta)}{\partial w_\ell^2} = 0,$$

Moreover,

$$\left(\frac{\partial log\, \Lambda(X,\theta)}{\partial w_\ell^2}\right)^2 \Lambda(X,\theta) = \frac{\left(\frac{\partial \Lambda(X,\theta)}{\partial w_\ell^2}\right)^2}{\Lambda(X,\theta)}$$
$$\le \pi_\ell\left[\frac{1}{2}\left(\eta^2(\tau_\ell) - \xi(w_\ell^2)\right)\right]\lambda(X,\tau_\ell)$$

This finding implies that

$$E_{\mathbb{Q}_\theta}\left(\frac{\partial log\, \Lambda(X,\theta)}{\partial w_\ell^2}\right)^2 < +\infty.$$

Now we look at second-order derivatives. The successive derivatives with respect to μ_ℓ, $\mu_{\ell'}$, w_ℓ^2, $w_{\ell'}^2$ with $\ell \neq \ell'$ are nul. We obtain

$$\frac{\partial^2 \Lambda(X,\theta)}{\partial \mu_\ell^2} = \pi_\ell \frac{\partial^2 \lambda(X,\theta)}{\partial \mu_\ell^2} = \pi_\ell\left(\eta^2(\tau_\ell) - \xi(w_\ell^2)\right)\lambda(X,\tau_\ell).$$

This random variable is integrable with respect to $Q_0^{x,T}$ with the nul integral. Thus,

$$\frac{\partial^2 log\, \Lambda(X,\theta)}{\partial \mu_\ell^2}\Lambda(X,\theta) = \frac{\partial^2 \Lambda(X,\theta)}{\partial \mu_\ell^2} - \frac{\left(\frac{\partial \Lambda(X,\theta)}{\partial \mu_\ell}\right)^2}{\Lambda(X,\theta)},$$

We find that this random variable is integrable with respect to $Q_0^{x,T}$, and computing the integral obtains

$$E_{\mathbb{Q}_\theta}\frac{\partial^2 log\, \Lambda(X,\theta)}{\partial \mu_\ell^2} = -E_{\mathbb{Q}_\theta}\left(\frac{\partial log\, \Lambda(X,\theta)}{\partial \mu_\ell}\right)^2.$$

Next,

$$\frac{\partial^2 \Lambda(X,\theta)}{\partial \mu_\ell \partial w_\ell^2} = \pi_\ell \frac{\partial^2 \lambda(X,\theta)}{\partial \mu_\ell \partial w_\ell^2}$$
$$= \pi_\ell\left(\frac{1}{2}\eta^3(\tau_\ell) - \frac{3}{2}\eta(\tau_\ell)\xi(w_\ell^2)\right)\lambda(X,\tau_\ell)$$

$$\frac{\partial^2 \Lambda(X,\theta)}{\partial (w_\ell^2)^2} = \pi_\ell\left[\left(\frac{1}{2}\left(\eta^2(\tau_\ell) - \xi(w_\ell^2)\right)\right)^2 - \right.$$
$$\left. \eta^2(\tau_\ell)\xi(w_\ell^2) - \frac{1}{2}\xi^2(w_\ell^2)\right]\lambda(X,\tau_\ell).$$

Thus, we conclude the proof analogously using (21) and (22).

Proposition 4.3. Assume that $I(\theta_0)$ is invertible and (H1)–(H2). Then, an estimator $\widehat{\theta}_N$ solves the likelihood estimating equation $\partial L_N(\theta)/\partial\theta = 0$ with a probability tending to 1 and $\widehat{\theta}_N \to \theta_0$ in probability.

Proof. For weak consistency following the standard steps, the uniformity condition needs to be proven. We prove that

An open convex subset S of Θ exists, which contains θ_0 and functions $G_{k,j,\ell}(X)$ such that, on S,

$\left|\frac{\partial^3 log\Lambda(X,\theta)}{\partial\theta_k \partial\theta_j \partial\theta_r}\right| \le G_{k,j,\ell}(X)$ and $E_{\theta_0}\left|G_{k,j,\ell}(X)\right| < +\infty$ for all $1 \le k,j,r \le 3M-1$.

$K, \alpha, \beta, c_0, c_1$ are set as positive numbers such that $0 < \alpha < \beta < 1, 0 < c_0 < c_1$, and θ_0 is assumed to belong to

$$S = \{(\pi_\ell, \mu_\ell, w_\ell^2)_{1<\ell<M}, \alpha < \pi_\ell < \beta,$$
$$|\mu_\ell| < K, c_0 < w_\ell^2 < c_1, 1 < \ell < M\}$$

where $\pi_M = 1 - \sum_{\ell=1}^{M-1} \pi_\ell$. We have to study

$$\frac{\partial^3 log\,\Lambda(X,\theta)}{\partial\theta_k \partial\theta_j \partial\theta_r} =$$
$$\frac{1}{\Lambda(X,\theta)} \frac{\partial^3 \Lambda(X,\theta)}{\partial\theta_k \partial\theta_j \partial\theta_r} - \frac{1}{\Lambda^2(X,\theta)} \frac{\Lambda(X,\theta)}{\partial\theta_r} \frac{\partial^2 \Lambda(X,\theta)}{\partial\theta_k \partial\theta_j}$$
$$- \frac{1}{\Lambda^2(X,\theta)} \left(\frac{\Lambda(X,\theta)}{\partial\theta_k} \frac{\partial^2 \Lambda(X,\theta)}{\partial\theta_j \partial\theta_r} - \frac{\Lambda(X,\theta)}{\partial\theta_j} \frac{\partial^2 \Lambda(X,\theta)}{\partial\theta_k \partial\theta_r} \right)$$
$$+ \frac{2}{\Lambda^3(X,\theta)} \frac{\partial\Lambda(X,\theta)}{\partial\theta_k} \frac{\partial\Lambda(X,\theta)}{\partial\theta_j} \frac{\partial\Lambda(X,\theta)}{\partial\theta_r}$$

Therefore, we have, for j, k, r distinct indexes

$$\frac{\partial^3 log\,\Lambda(X,\theta)}{\partial\pi_j \partial\pi_k \partial\pi_r} = \frac{2}{\Lambda^3(X,\theta)} \left(\lambda(X, \tau_j) - \lambda(X, \tau_M) \right)$$
$$\times \left(\lambda(X, \tau_k) - \lambda(X, \tau_M) \right) \left(\lambda(X, \tau_r) - \lambda(X, \tau_M) \right)$$

As $\lambda(X, \tau_j)/\Lambda(X, \theta) \le \pi_j^{-1} < \alpha^{-1}$,

$$\frac{\partial^3 log\,\Lambda(X,\theta)}{\partial\pi_j \partial\pi_k \partial\pi_r} \le 2^4 \alpha^{-3}$$

We use $\lambda(X, \tau_j)/\Lambda(X, \theta) \le \pi_j^{-1} < \alpha^{-1}$ again to bound the other third-order derivatives. Then, in the derivatives, random variables appear

$$\eta^n(\tau) = \left(\frac{(Y-D) - \mu Z}{1 + w^2 Z} \right)^n$$

for different values of n. We now bound $\eta(\tau)$ by an r. v. independent of τ and have moments of any order under $\mathbb{Q}_{\theta_0}$. We have

$$\frac{(Y-D)}{1+w^2 Z} = \frac{(Y-D)}{1+c_1 Z} \left(1 + \frac{(c_1 - w^2) Z}{1 + w^2 Z} \right)$$

Thus,

$$|\eta(\tau)| \le \frac{c_1}{c_0} \left| \frac{(Y-D)}{1+c_1 Z} \right| + \frac{K}{c_0}$$

Now, in the same method

$$\left| \frac{(Y-D)}{1+c_1 Z} \right| = \left| \frac{(Y-D)}{1+w^2 Z} \left(1 + \frac{(w^2 - c_1) Z}{1 + c_1 Z} \right) \right| \le 3 \left| \frac{(Y-D)}{1+w^2 Z} \right|$$

For all τ, this finding implies that

$$E_{Q_\tau} \left| \frac{(Y-D)}{1+c_1 Z} \right|^n < +\infty.$$

Consequently,

$$E_{\mathbb{Q}_{\theta_0}} \left| \frac{(Y-D)}{1+c_1 Z} \right|^n = \sum_{\ell=1}^M \pi_{\ell,0} E_{\mathbb{Q}_{\tau_0}} \left| \frac{(Y-D)}{1+c_1 Z} \right|^n < +\infty$$

The proof of Proposition 4.3 is complete.

5. Conclusions

In stochastic differential equations based random effects model framework. We considered the addition case in the drift where $b(x, \phi_i)$ is linear in ϕ_i($b(x, \phi_i) = \phi_i + b(x)$), where ϕ_i has a mixture of Gaussian. We obtain an expression of the exact likelihood. When the number of components is known, we prove the consistency of the maximum likelihood estimators (MLEs). Properties of the EM algorithm are described when the algorithm is used to compute MLE.

ACKNOWLEDGEMENTS

The Author is grateful to Dr. Wang Xiang jun of School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan, Hubei, 430074 P.R. China. Also, to Mr. Alsukaini M. S. of the department of Mathematics, College of Science, Basra University, Basra, Iraq for their constructive comments.

REFERENCES

[1] Arribas-Gil, A., De la Cruz, R., Lebarbier, E. and Meza, C. (2015). Classification of longitudinal data through a semiparametric mixed-effects model based on lasso-type estimators. Biometrics 71, 333–343.

[2] Alkreemawi W. K., Alsukaini M. S. and Wang X. J. “On Parameters Estimation in Stochastic Differential Equations with Additive Random Effects,” Journal of Advances in Mathematics, Vol. 11, no.3, 5018 – 5028, 2015.

[3] Alkreemawi W. K., Alsukaini M. S. and Wang X. J. “Asymptotic Properties of MLE in Stochastic Differential Equations with Random Effects in the Drift Coefficient,” International Journal of Engineering, Science and Mathematics (IJESM), Vol. 5, Issue. 1, 1 – 14, 2016.

[4] Alsukaini M. S., Alkreemawi W. K. and Wang X. J., “Asymptotic Properties of MLE in Stochastic Differential Equations with Random Effects in the Diffusion Coefficient,” International Journal of Contemporary Mathematical Sciences, Vol. 10, no. 6, 275 – 286, 2015.

[5] Alsukaini M. S., Alkreemawi W. K. and Wang X. J., “ Maximum likelihood Estimation for Stochastic Differential Equations with two Random Effects in the Diffusion Coefficient,” Journal of Advances in Mathematics, Vol. 11, no.10, 5697 – 5704, 2016.

[6] Celeux, G., Martin, O. and Lavergne, C. (2005). Mixture of linear mixed models application to repeated data clustering. Statistical Modelling 5, 243–267.

[7] Comte, F., Genon-Catalot, V. and Samson, A. (2013). Nonparametric estimation for stochastic differential equations with random effects. Stoch. Proc. Appl. 123, 2522–2551.

[8] Delattre M., Genon Catalot V. and Samson A., “Maximum Likelihood Estimation for Stochastic Differential Equations with Random Effects,” Scandinavian Journal of Statistics, 40, 322-343, 2013.

[9] Delattre M., Genon Catalot V. and Samson A., “Mixtures of stochastic differential equations with random effects: application to data clustering,” Journal of statistical planning and inference, Publication MAP5-2015-36.

[10] Dempster, A., Laird, N. and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. Jr. R. Stat. Soc. B 39, 1–38.

[11] Ditlevsen, S. and De Gaetano, A. (2005). Stochastic vs. deterministic uptake of dodecanedioic acid by isolated rat livers. Bull. Math. Biol. 67, 547–561.

[12] Donnet, S. and Samson, A. (2008). Parametric infer ence for mixed models defined by stochastic differential equations. ESAIM P&S 12, 196–218.

[13] Genon-Catalot, V. and Larédo, C. (2015). Estimation for stochastic differential equations with mixed effects. Hal-00807258 V2.

[14] Lipster, R. and Shiryaev, A. (2001). Statistics of random processes I: general theory. Springer.

[15] McLachlan, G. and Krishnan, T. (2008). The EM Algorithm and Extensions.