ARTICLE IN PRESS

Journal of Statistical Planning and Inference ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Mixtures of stochastic differential equations with random effects: Application to data clustering

Maud Delattre [a,b], Valentine Genon-Catalot [c], Adeline Samson [d,*]

[a] AgroParisTech, UMR518 MIA-Paris, F-75231 Paris Cedex 05, France

[b] INRA, UMR518 MIA-Paris, F-75231 Paris Cedex 05, France

[c] UMR CNRS 8145, Laboratoire MAP5, Université Paris Descartes, Sorbonne Paris Cité, France

[d] UMR CNRS 5224, Laboratoire Jean Kuntzmann, Université Grenoble Alpes, France

ARTICLE INFO

Article history:
Received 7 December 2015
Accepted 9 December 2015
Available online xxxx

Keywords:
Mixed-effects models
Stochastic differential equations
BIC
Classification
EM algorithm
Mixture distribution
Maximum likelihood estimator

ABSTRACT

We consider N independent stochastic processes $(X_i(t), t \in [0, T_i]), i = 1, \ldots, N$, defined by a stochastic differential equation with drift term depending on a random variable ϕ_i. The distribution of the random effect ϕ_i is a Gaussian mixture distribution, depending on unknown parameters which are to be estimated from the continuous observation of the processes X_i. The likelihood of the observation is explicit. When the number of components is known, we prove the consistency of the exact maximum likelihood estimators and use the EM algorithm to compute it. When the number of components is unknown, BIC (Bayesian Information Criterion) is applied to select it. To assign each individual to a class, we define a classification rule based on estimated posterior probabilities. A simulation study illustrates our estimation and classification method on various models. A real data analysis is performed on growth curves with convincing results.

1. Introduction

The goal of clustering methods is to discover structures among individuals: data are grouped into a few clusters such that the observations in the same cluster are more similar to each other than those from the other clusters. In this paper we focus on individuals described by longitudinal data or functional data: data is represented by curves and the random variable underlying data is a stochastic process. Some papers deal with the problem of classification of longitudinal data through mixed-effects models or models with random effects, assuming that the classes are known (see Arribas-Gil et al., 2015, and references therein). Their purpose is to build a classification rule of longitudinal curves/profiles into a given number of different classes to be able to predict the class of a new individual. This is very different from the problem of classification when the classes and the number of classes are unknown. Here, we adopt the latter point of view. We consider functional data modeled by a stochastic differential equation (SDE) with random effects. This is a new approach which is very different from usual functional data analysis methods (see e.g. Jacques and Preda, 2014, for a recent review). The clustering of the trajectories is then obtained by modeling the distribution of the random effects as a mixture of distributions (with unknown number of components).

Mixture of linear regression models with random effects is considered in Celeux et al. (2005). Unknown parameters are estimated by maximum likelihood, with the EM algorithm and BIC (Bayesian Information Criterion) for selecting the number

* Corresponding author.
E-mail addresses: maud.delattre@agroparistech.fr (M. Delattre), valentine.genon-catalot@parisdescartes.fr (V. Genon-Catalot), adeline.leclercq-samson@imag.fr (A. Samson).

http://dx.doi.org/10.1016/j.jspi.2015.12.003

Please cite this article in press as: Delattre, M., et al., Mixtures of stochastic differential equations with random effects: Application to data clustering. J. Statist. Plann. Inference (2016), http://dx.doi.org/10.1016/j.jspi.2015.12.003

of components. Here, we consider functional data modeled by a stochastic differential equation with drift term depending on random effects and diffusion term without random effects. More precisely, we consider N real valued stochastic processes $(X_i(t), t \geq 0), i = 1, \ldots, N$, with dynamics ruled by the following SDEs:

$$dX_i(t) = (\phi_i' b(X_i(t)) + a(X_i(t)))dt + \sigma(X_i(t))\, dW_i(t), \quad X_i(0) = x, \tag{1}$$

where $(W_1, \ldots, W_N)$ are N independent Wiener processes, $\phi_1, \ldots, \phi_N$ are N *i.i.d.* $\mathbb{R}^d$-valued random variables, $(\phi_1, \ldots, \phi_N)$ and $(W_1, \ldots, W_N)$ are independent and x is a known real value. The functions $\sigma(.), a(.) : \mathbb{R} \to \mathbb{R}$ and $b(.) : \mathbb{R} \to \mathbb{R}^d$ are known. Each process $(X_i(t))$ represents an individual and the random variable ϕ_i represents the random effect of individual i.

We consider continuous observations $(X_i(t), t \in [0, T], i = 1, \ldots, N)$ with a given T. The estimation of unknown parameters in the distribution of ϕ_i from the (X_i)'s is not straightforward, as the exact likelihood is generally not explicit. Maximum likelihood estimation in SDEs with random effects has been studied in a few papers (Ditlevsen and De Gaetano, 2005; Donnet and Samson, 2008; Picchini et al., 2010). In Delattre et al. (2013), model (1) is considered with ϕ_i having a Gaussian distribution. This has the advantage of leading to an explicit formula for the exact likelihood.

In this paper, we assume that the random effects ϕ_i have distribution given by a mixture of Gaussian distributions, this mixture distribution modeling the classes. We want to estimate the number of components of the mixture, as well as the parameters and the proportions. More precisely, we assume that the random variables $\phi_1, \ldots, \phi_N$ have a common distribution with density $g(\varphi, \theta)$ on $\mathbb{R}^d$, which is given by a mixture of Gaussian distributions:

$$g(\varphi, \theta) = \sum_{\ell=1}^{M} \pi_\ell\, n_d(\varphi, \tau_\ell), \qquad n_d(\varphi, \tau_\ell)d\varphi = \mathcal{N}_d(\boldsymbol{\mu}_\ell, \Omega_\ell), \qquad \tau_\ell = (\boldsymbol{\mu}_\ell, \Omega_\ell)$$

with M the number of components in the mixture and π_ℓ the proportions of the mixture ($\sum_{\ell=1}^{M} \pi_\ell = 1$), $\boldsymbol{\mu}_\ell \in \mathbb{R}^d$ and Ω_ℓ a $d \times d$ invertible covariance matrix. Set $\theta = ((\pi_\ell, \tau_\ell), \ell = 1, \ldots, M)$ for the unknown parameters to be estimated when M is known. Below, we denote by θ_0 the true value of the parameter.

Our aim is to estimate the parameters θ of the density of the random effects from the observations $\{X_i(t), 0 \leq t \leq T, i = 1, \ldots, N\}$. We prove that the exact likelihood of observations is explicit. This allows to use the EM-algorithm to compute the maximum likelihood estimator when the number of components is known. We discuss the convergence of the algorithm. Then BIC is applied for selecting the number of mixture components. The EM algorithm also enables to define a classification rule of individuals. As a theoretical result, we prove the consistency of the exact maximum likelihood estimator when the number M of components is known. Our methods show good results on simulated data, both for the parameter estimation and the classification rule. An implementation on real data coming from growth chicken curves (Jaffrézic et al., 2006) is performed.

In Section 2, we introduce notations, assumptions and give the formula of the exact likelihood. In Section 3, the EM algorithm and its properties are described. We present BIC to select the number of components and the classification rule. In Section 4, we prove the consistency of the exact maximum likelihood estimator when the number of components is known. Section 5 is devoted to a simulation study on various models. Section 6 concerns the implementation on real data. Some concluding remarks are given in Section 7. Theoretical proofs are gathered in the Appendix.

2. Model, assumptions and notations

Consider N real valued stochastic processes $(X_i(t), t \geq 0), i = 1, \ldots, N$, with dynamics ruled by (1). The processes $(W_1, \ldots, W_N)$ and the r.v.'s $\phi_1, \ldots, \phi_N$ are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Consider the filtration $(\mathcal{F}_t = \sigma(\phi_i, W_i(s), s \leq t, i = 1, \ldots, N), t \geq 0)$. We introduce the following assumptions:

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

Under (H1), for $i = 1, \ldots, N$, for all $\varphi = (\varphi_1, \ldots, \varphi_d)' \in \mathbb{R}^d$, the stochastic differential equation (SDE)

$$dX_i^\varphi(t) = (\varphi' b(X_i^\varphi(t)) + a(X_i^\varphi(t)))dt + \sigma(X_i^\varphi(t))\, dW_i(t), \quad X_i^\varphi(0) = x \tag{2}$$

admits a unique strong solution process $(X_i^\varphi(t), t \geq 0)$ adapted to the filtration $(\mathcal{F}_t)$. Moreover, the SDE (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). The Markov property of $(\phi_i, X_i(t))$ is straightforward by looking at (1) as the two-dimensional SDE:

$$d\phi_i(t) = 0, \qquad \phi_i(0) = \phi_i,$$
$$dX_i(t) = (\phi_i(t)' b(X_i(t)) + a(X_i(t)))dt + \sigma(X_i(t))\, dW(t), \quad X_i(0) = x.$$

The processes $(\phi_i, X_i(t), t \geq 0), i = 1, \ldots, N$ are *i.i.d.* (see *e.g.* Delattre et al., 2013; Genon-Catalot and Larédo, 2015; Comte et al., 2013).

To define the likelihood of the observations, let us introduce the associated canonical model. Let C_T denote 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