Journal of Statistical Planning and Inference 150 (2014) 30–48

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

New adaptive strategies for nonparametric estimation in linear mixed models

Charlotte Dion [a,b,*]

[a] LJK, UMR CNRS 5224, Université Joseph Fourier, 51 rue des Mathématiques, 38041 Grenoble, France
[b] MAP5, UMR CNRS 8145, Université Paris Descartes, Sorbonne Paris Cité, 45 rue des Saints Pères, 75006 Paris, France

ARTICLE INFO

Article history:
Received 19 November 2013
Received in revised form
3 March 2014
Accepted 17 March 2014
Available online 26 March 2014

Keywords:
Deconvolution
Linear mixed models
Model selection
Nonparametric estimation

ABSTRACT

This paper surveys new estimators of the density of a random effect in linear mixed-effects models. Data are contaminated by random noise, and we do not observe directly the random effect of interest. The density of the noise is supposed to be known, without assumption on its regularity. However it can also be estimated. We first propose an adaptive nonparametric deconvolution estimation based on a selection method set up in Goldenshluger and Lepski (2011). Then we propose an estimator based on a simpler model selection deviced by contrast penalization. For both of them, non-asymptotic $\mathbb{L}^2$−risk bounds are established implying estimation rates, much better than the expected deconvolution ones. Finally the two data-driven strategies are evaluated on simulations and compared with previous proposals.

1. Introduction

In order to analyze repeated measures in time for individuals with the same behavior, we focus on a linear mixed-effects model. The observation value at time t_j with $j \in \{0, \dots, J\}$ for individual k, $k \in \{1, \dots, N\}$, denoted $Y_{k,j}$, follows the model

$$Y_{k,j} = \alpha_k + \beta_k t_j + \epsilon_{k,j}, \tag{1}$$

where the coefficients α_k and β_k are the random effects of regression and depend on the subject k. The times of observation t_j are known and equidistant with a time step Δ: $t_j = j\Delta$. The random variables $(\epsilon_{k,j})_{1 \le k \le N,\ 0 \le j \le J}$ are the measurement errors of zero mean. They are supposed independent and identically distributed (i.i.d.) with common density f_ϵ. We suppose α_k i.i.d. with density f_α and β_k i.i.d. with density f_β. Moreover we assume that $(\epsilon_{k,j})_{1 \le k \le N,\ 0 \le j \le J}$ and $(\alpha_k, \beta_k)_{1 \le k \le N}$ are independent sequences.

Mixed models with random effects are often used, for example, in pharmacokinetics. They describe both individual behavior and variability between individuals. The distribution of random effects is of special interest. It allows for example to describe the heterogeneity of the drug kinetics in the population of individuals. Mixed models have been widely studied, often with parametric strategies and Gaussian random effects and noise (see Pinheiro and Bates, 2000). However it is not clear that this normality assumption of the random effects is truly satisfied in practice. The aim of this paper is to produce nonparametric estimation of the density of the random effects from the observations $Y_{k,j}$.

* Correspondence address: MAP5, UMR CNRS 8145, Université Paris Descartes, Sorbonne Paris Cité, 45 rue des Saints Pères, 75006 Paris, France.
E-mail address: charlotte.dion1@gmail.com

http://dx.doi.org/10.1016/j.jspi.2014.03.006
0378-3758/

Several papers consider this problem. Wu and Zhu (2010) relax the Gaussian assumption proposing an orthogonality-based estimation of moments estimating the third and fourth moments of the random effects and errors. Komárek and Lesaffre (2008) suppose that the random effects are a mixture of Gaussian distributions, and estimate the weights of the mixture components using a penalized approach. Papageorgiou and Hinde (2012) propose semi-parametric density models of random effects. But it is the estimation of the complete random effects density which should be appropriate. This is the framework of nonparametric estimation that is adopted in this paper.

Complete nonparametric estimation of f_β and f_α in model (1) has been little studied in the literature. The main reference is Comte and Samson (2012) who set up a nonparametric estimation with deconvolution ideas. Indeed, model (1) can be seen as a measurement error model. Deconvolution methods are various and have been initiated by the kernel study of Fan (1991), followed by adaptive strategies (Pensky and Vidakovic, 1999; Comte et al., 2006; Delaigle and Gijbels, 2004) and optimality of rates (Butucea and Tsybakov, 2007). Except the penalized contrast estimator of Comte and Samson (2012), none of these previous estimators has been used in the mixed model framework. However, for all these methods, the noise density is usually assumed to be known with a certain regularity. The resulting rates are thus very specific and depend on the regularity of the noise and of the function under estimation. For example the noise is supposed to be ordinary smooth in Comte and Samson (2012).

This paper focuses on improving existing nonparametric strategies for the estimation of f_β. We focus on f_β and we refer to Comte et al. (2006) for f_α, noticing that for $t_0=0$ model (1) writes $Y_{k,0}=\alpha_k+\epsilon_{k,0}$ which is a standard convolution equation. We first propose an estimator which uses all the available observations. A Mean Integrated Squared Error (MISE) bound is computed and it shows that a bias-variance compromise must be performed. The particularity here is the selection model procedure: we adapt a method set up in Goldenshluger and Lepski (2011) for kernel estimators. Then we establish a non-asymptotic oracle risk bound using Talagrand's inequality. In a second time, another estimator is proposed, built using the last and the first time of observations. If two times of observation are separated by a quite long time, enough information is available to obtain a good estimation. As in the first case, risk bound leads us to propose a model selection strategy. In that case a more classical and easy to implant strategy can be adopted, namely a penalized criterion (Massart, 2007; Birgé and Massart, 1998).

What is new is that our two strategies take advantage of both nonparametric and deconvolution ideas. With very mild assumption on the noise regularity (we only require that the characteristic function of the noise is nonzero), the two estimators recover standard rates in density estimation for f_β (Stone, 1980; Donoho et al., 1996) while deconvolution rates were expected. Precisely logarithmic speed of convergence implied by deconvolution methods is avoided. Therefore our two estimators have better rates. Furthermore we do not assume that the noise is ordinary smooth, which was assumed in Comte and Samson (2012).

This paper is organized as follows. We proceed with the construction of the estimators in Sections 2.1 and 13. Bias-variance decompositions of the $\mathbb{L}^2$−risk are proved (Propositions 1, 5) and lead us to propose adaptive strategies. The main results (Theorems 3, 7) prove that the resulting estimators are adaptive. A short Section 3 provides a comparison of our estimators with Comte and Samson (2012). Lastly, we illustrate the different methods by simulation experiments presented in Section 4. We briefly discuss therein the case of unknown noise density (see Sections 4.3 and A.8) which is also implemented for comparison. Proofs are gathered in Appendix A.

2. Construction of two estimators, risk bounds and adaptive results

Let us introduce some notations. For two functions f and g in $\mathbb{L}^1(\mathbb{R})\cap\mathbb{L}^2(\mathbb{R})$, the scalar product is defined by $\langle f,g\rangle=\int_{\mathbb{R}} f(x)\overline{g(x)}\,dx$ and the associated norm is $\|f\|^2=\int_{\mathbb{R}}|f(x)|^2\,dx$. The Fourier transform of f is $f^*(x)=\int_{\mathbb{R}} e^{ixu}f(u)\,du$ for all $x\in\mathbb{R}$. When f is the density of a random variable X, $f^*(u)=\mathbb{E}[e^{iuX}]$ is called the characteristic function of X. Then the convolution product of f and g for all $x\in\mathbb{R}$, is $f\star g(x)=\int_{\mathbb{R}} f(x-y)g(y)\,dy$. Finally we remind Plancherel–Parseval's formula: $\forall f\in\mathbb{L}^1(\mathbb{R})\cap\mathbb{L}^2(\mathbb{R})$, $2\pi\|f\|^2=\|f^*\|^2$.

2.1. Estimator using all the observations: $\widehat{f}$

In this section we build an estimator of the density f_β using all the available observations. We consider the normalized variables $Z_{k,m}$ defined by the difference between two observations, for all $k\in\{1,\dots,N\}$, and for all $m\in\{1,\dots,J\}$

$$Z_{k,m}:=\frac{Y_{k,m}-Y_{k,0}}{m\Delta}=\beta_k+\frac{\epsilon_{k,m}-\epsilon_{k,0}}{m\Delta}=:\beta_k+W_{k,m}. \tag{2}$$

For a given m, the variables $W_{k,m}$, $k\in\{1,\dots,N\}$, are i.i.d. with density f_{W_m}. Due to the independence between the ϵ_k's and the β_k's, the $Z_{k,m}$ are i.i.d. with density denoted by f_{Z_m}, and definition (2) implies

$$f_{Z_m}=f_\beta\star f_{W_m}. \tag{3}$$

The Fourier transform of (3) yields

$$f^*_{Z_m}=f^*_\beta f^*_{W_m}.$$