T E C H N I C A L

R E P O R T

0533

SMOOTHED KERNEL FOR GENERALIZED LINEAR MIXED MODEL

BOUEZMARNI, T., RIJMEN, F., and P. DE BOECK

I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Smoothed Kernel for Generalized Linear Mixed Model*

T. Bouezmarni ‡† F. Rijmen ‡ and P. De Boeck ‡

April 25, 2005

Abstract

We propose a new method to estimate the distribution function of the random effects in a generalized linear mixed model of the logistic type. The method is the smoothed kernel method. The distribution estimate based on a nonparametric maximum likelihood approach is discrete, and the kernel method is used to smooth this density estimate. Simulation results are shown for Rasch model. Also a comparison of the new approach with the nonparametric maximum likelihood approach, methods based on the normality assumption and a mixture of normal densities is presented.

Key words and phrases: ...
JEL Classification : ...

*The research support from "Interuniversity Attraction Pole", Phase V from Belgian Government ..... is gratefully acknowledged.

†Associate Researcher to the Institute of Statistics, UCL.
‡Department of Psychology, KULeuven.

1 Introduction

Generalized linear mixed model have become a powerful parametric tool in psychometric. Estimation is based on maximum likelihood theory which assumes that the underlying probability model determining the behavior of the event investigated is correctly specified. There different ways for estimation and inference one can make depending on how the random effect is seen. The first approach is to suppose that the person parameters θ are fixed and have the same status as the item parameter β. To estimate the item and person parameters, the likelihood is maximized jointly. This method is called joint maximum likelihood estimation. However, the JML estimators for item parameters are not consistent (Neyman and Scott, 1948). This is due to the fact that the number of parameters increases at the same rate as the sample size increases because each new person implies a new parameter. Another alternative is the conditional maximum likelihood (CML) approach based on sufficient statistics. The CML is applied only for models with sufficient statistics for person parameter. For Rasch model, the sufficient statistics for a person-specific is the sum score. The item parameters are estimated by maximizing the so-called conditional likelihood. The CML estimators are consistent and asymptotically normally distributed (Andersen, 1970). However, there are some disadvantages associated to CML. First, no inference are possible on persons since the probability of observing the response depends only on the sufficient statistics and not on the person effect. Second, CML may not be the most efficient method, because the conditional likelihood is maximized rather than the full likelihood. Another way is to consider the person parameter as independent random draws from a density defined over the population of persons denoted by g_p. This approach is widely used. The item parameters are estimated by maximizing the so-called marginal likelihood:

$$L_{MML}(\phi) = \prod_{p=1}^{P} \int_{-\infty}^{+\infty} \prod_{=1}^{I} Pr(Y_{pi} = y_{pi}|\theta_p) g_p(\theta_p) \tag{1.1}$$

To maximize the MML quantity, we need to specifies the model g_p. We can distinguish three cases of the MML approach: nonparametric, semiparametric and parametric. Each method depends on the assumptions one makes about the unobserved population density of the random effects. The nonparametric maximum likelihood estimation, no assumption are made about the density g_p. Laird (1978) has shown that in this case the estimate of the cumulative distribution function G_p is a step function with a finite number of steps. In semi parametric estimation method, Heinen (1996), the location of steps are assumed to be known but the probability masses at these fixed nodes have to be estimated. In the parametric estimation method, the population density is taken to be a parametric density for which the parameters have to be estimated. The normal density and a mixture of normal densities are the two models used in parametric approach in general. However, since random effects are unmeasurable (latent factors), this is an assumption that can not be checked. In linear mixed models, we assume often that random effects have a normal density. Verbeke and Lesafre (1996) studied the impact of the normality assumption for random effects. They showed the robustness of the inference on the fixed effects to the nonnormality of the random effects. However, inference on the random effects is sensitive to the normality assumption. One way, widely used, to relax the normality assumption is to use a mixture model, see

Verbeek and Lesaffre (1996). Recently, the B-spline smoothing approach is proposed by Wendimagegn et al, (2004).

For the generalized linear mixed models, the normality and mixture of normals of the random effects are frequently assumed. For Longitudinal data, Saskia et al. (2004) have studied through simulations the effect of mispecifying the random effects distribution on the estimation of the fixed effects in the case of logistic mixed effects model with a treatment and time covariant and a random intercept. They found that the estimates tends to converge to value different from the true effects, this deviation being the largest for the estimation of the variance components. Asymptotic normality of the estimates holds, but not respect to the true effects. It seems then that the nonparametric approach (Laird 1978) is a good alternative to relax the parametric assumption on the density of the random effects. But the estimator using the nonparametic maximum livelihood is discrete, that is not appreciated. To overcome this problem Magder and Zeger (1996) proposed a mixtures of Gaussians to smooth nonparametric maximum likelihood estimate and Zhang and Davidian (2001) suggested the semi-nonparametric(SNP) method. In this article, we propose a new approach to estimate the density of the random effects in GLMM by the smoothed kernel approach. This method can be either used to decide how many component if a finite mixture is assumed or to be used for inference on the fixed and random effects. Note that smoothed kernel approach her can be used even for linear mixed model. This paper will be devoted to the Rach model, i.e., density of one random effects will be estimated. Al thought, the results of this paper can be extended to the multivariate case.

The paper is organized as follows. In Section 2, we outline the framework and present the Rasch model and the smoothed kernel estimators. The new method to estimate the density of the random effects and the estimation of the parameters are described in Section 3. Section 4 provides Monte Carlo results concerning the finite sample properties of the estimator. A comparison of the new approach with the nonparametric approach and the parametric model based on normality and mixture of normal assumptions.

2 Rasch model and smoothed Kernel

2.1 Rasch Model

Item response models may be used to model the responses of subjects to a number of questions or test items. An item response model with one parameter for item difficulty is known as a Rasch model. Rasch Models are one of the dominant models for binary items (e.g. success/failure on test items) in Psychometrics and is a special case of the generalized linear mixed model. In Rasch model the log odds of subjects p giving a correct response to item i ma be modeled using a one-parameter logistic response model:

$$\eta_{pi} = \theta_p - \beta_i$$

where the β_i represent the difficulty of the item and θ_p represent the ability of the subjects. In terms of probability the model is:

$$\pi_{pi} = \exp(\theta_p - \beta_i)/(1 + \exp(\theta_p - \beta_i))$$

where π_{pi} is the probability of a 1-response. i.e.

$$f_p(y_{pi}|\theta_p) = \Pr(Y_{pi} = y_{pi}) = \frac{exp[(\theta_p - \beta_i)y_{pi}]}{1 + exp(\theta_p - \beta_i)} \tag{2.1}$$

The Rasch model will be the central example for which the various estimation methods will be explained. The generalization of the method presented to other models is straightforward.

2.2 Smoothed Kernel

The most popular nonparametric estimator of an unknown probability density function f is the standard kernel estimator. Given a random sample, $X_1, \cdots, X_n$, from a probability distribution F with an unknown density function f. The smoothed kernel estimator:

$$f_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K((x - X_i)/h),$$

where the kernel K is a symmetric density function and h is a smoothing parameter, called the bandwidth. Without loss of generality, assume that K is the standard normal density. The standard kernel estimate is a sum of "bumps" placed at the observations. The shape of the bumps is determined by the kernel K and it is the same for all the points where the density is estimated. The bandwidth parameter controls the width of the bumps. Various measures have been studied to state the closeness of the estimator to the true density. In this paper we will use he standard kernel to estimate the density of the random effect. But, we need a kernel smoothing for grouped data. Suppose that the sample of size n can be subdivided into G distinct sub samples with respective sizes $n_1, n_2, ..., n_G$ and representatives $X_1^*, X_2^*, ..., X_G^*$. For such grouped data, Kernel estimator is adapted:

$$\begin{aligned} \hat{f}_h(x) &= \frac{1}{h}\sum_{i=1}^{G} \frac{n_i}{n} K((x - X_i^*)/h) \\ &= \sum_{i=1}^{G} \pi_i K_h((x - X_i^*)) \end{aligned}$$

where $\pi_i = \frac{n_i}{n}$ and $K_h(t) = \frac{1}{h}K(\frac{t}{h})$.

3 Smoothed Kernel for Rasch Model

In linear and generalized linear mixed models, the most used approach to estimate the distribution G of the random effect is to assume that G is a Gaussian distribution and use the maximum likelihood to estimate the parameters of the distribution. For linear mixed model, normality assumption is mathematically convenient, because under this assumption, both the marginal distribution and the posterior distribution of the random effects are Gaussian. This simplifies the maximization of the likelihood and the estimation of the random effects. In Generalized linear mixed models, even this properties disappear, the normality assumption is still used. The disadvantage of the Gaussian approach is that it depends on a strong assumption about the shape of this distribution hat may be not valid, in particular if we know that the population comes from different groups. Here, we use the nonparametric approach. no assumptions at all about the shape of the distribution of the random effects, i.e., let the data speak from and for themselves and select the distribution with highest likelihood of all distribution. Lindsay(1983) showed that under general conditions, this estimator is discrete and has a limited number of points of support, which is an disadvantage of this estimator in particular if the shape of the density is the intersect object. To overcomes this problem, Magder and Zeger (1996), in order to control the degree of smoothness of the estimate, proposed to maximize the likelihood not over all distribution but over the class of arbitrary mixtures of Gaussians subject to the constraint that the variances of the Gaussians must be greater than or equal to specific minimum value. In this work, we propose another alternative to perform the discrete estimator of the nonparametric approach. the smooth kernel method. However, the kernel estimate need the observable data and the random effects are not. The idea behind our method is to use the points of the discrete nonparametric estimator as the data, in other words we smooth the discrete nonparametric estimator. In this section we present the method for the generalized linear mixed models. The method is still valid for the linear mixed models.

3.1 Description of the Method

We first describe how the density of the random effect is estimated by kernel. Let y_{pi} denote the ith observation for person p and $Y_p = (Y_{p1}, ..., Y_{pI})$. Conditionally on the random effects θ_p ($J \times 1$ vector). The elements Y_{pi} of Y_p are assumed to be independent, with density function of an exponential family form $f_p(y_{pi}|\theta_p)$. The density is replaced by probabilities in the case of discrete outcomes $\Pr(y_{pi} = y_{pi})$, for example for Rasch models. The marginal joint distribution of Y_p is then

$$f_p(y_p) = \int f_p(y_p|\theta_p)g(\theta_p)d\theta_p, \tag{3.1}$$

where for Rasch models

$$f_p(y_p|\theta_p) = \Pr(y_{pi} = y_{pi}) = \frac{exp[(\theta_p - \beta_i)y_{pi}]}{1 + exp(\theta_p - \beta_i)}$$

We denote by $\theta_1, ..., \theta_G$ the limited number of points of the simple nonparametric maximum likelihood estimate and $\pi_1, ..., \pi_G$ the corresponding probabilities. The number G is estimated but can be fixed in advance. In our case, we consider that G is fixed. In order to get a smooth estimator, and then a shape of the density of the random effects, A naive idea is to smooth this estimator by kernel. In this paper, we propose, to estimate random effects density,the adapted kernel estimator (see the previous section):

$$g_h(\theta) = \sum_{g=1}^{G} \pi_g \, K_h(\theta - \theta_g) \tag{3.2}$$

where $K_h(\theta - \theta_g) = h^{-1/2}K(h^{-1/2}(\theta - \theta_g))$, $K(.)$ is a kernel function, and h is a positive value called he smoothing parameter. For multivariate case, we take a multivariate kernel.

The kernel the density of the random function we propose is a two steps procedure:

Step 1. Estimate $\theta^* = (\theta_1, ..., \theta_G)$ and $\pi^* = (\pi_1, ..., \pi_G)$ using the nonparametric maximum likelihood

Step 2. The modified version of the kernel density estimator g_h in (3.2) of the random effects is:

$$\hat{g}_h(\theta) = \sum_{g=1}^{G} \pi_g \, K_h(\theta - \theta_g) \tag{3.3}$$

We show in two simple examples, the discrete nonparametric maximum likelihood estimator, and the adaptive kernel estimator compared with the true density function of the random effects. All the simulations in this work are for Rasch models, the true density are of mean zero and the programs are implemented in Matlab.

Example 1: Random samples of size $n = 500$ are generated from the Rasch models:

$$\Pr(y_{pi} = y_{pi}) = \frac{exp[(\theta_p - \beta_i)y_{pi}]}{1 + exp(\theta_p - \beta_i)}$$

where the fixed effects β_i, of size 21, are ranged in $[-1, 1]$. Random effects are generated from normal with zero mean and variance equal to four. Two methods are considered: the simple nonparametric method and the smoothed kernel method. The smoothed kernel requires a value of the bandwidth parameter, we choose $h = 1.53$, we come back how the optimal bandwidth parameter is selected for these models. We fixed the number of points for the nonparametric method, G=3. Note that you can let the nonparametric method select this number. Only ten different data sets are simulated and the mean is considered for the smoothed kernel estimator:

$$g_h(\theta) = \sum_{g=1}^{BG} \pi_g \, K_h(\theta - \theta_g).$$

Example 2: Random samples of size $n = 500$ are generated from the Rasch models. The same fixed as in example 1. But the random effects are generated from a mixture of two normals with the same weigh(0.5), one normal with mean -4 and variance 2 and another normal with mean four and variance 2. Note that the mean of the random effects density is zero. For the smoothed kernel method the bandwidth parameter is $h = 1.4$. We fixed the number of points for the nonparametric method, G=5. As in example 1, ten different data sets are simulated.

The upper parts of Figure 3.1 and Figure 3.2 show the simple nonparametric method. The upper parts of the figures, we can see the shape of the smoothed kernel estimator and its closeness to true density of the random effects.

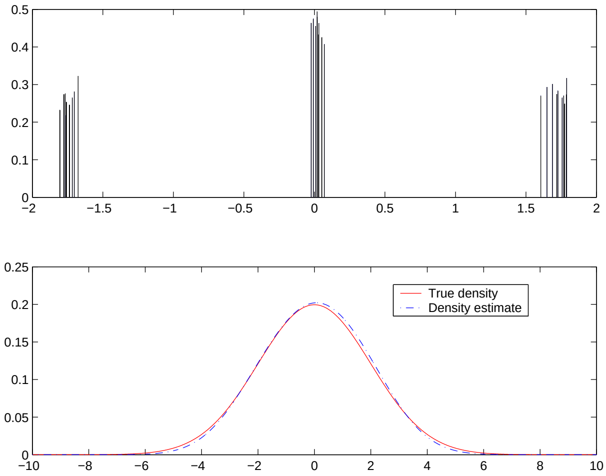

Figure 3.1: The simple nonparametric and the smoothed kernel estimation for the density of the random effect drawn from normal density $N(0, 2)$

3.2 Choice of h

Nonparametric density estimation requires the specification of the smoothing parameter. Many theoretical and practical studies have been carried out for choosing the optimal smoothing parameter for the standard kernel density estimators. Indeed, the kernel estimator is very sensitive to the choice of the bandwidth parameter because with a large value of the bandwidth parameter h, we over smooth the density, i.e., the variance is reduced but the bias is too large. And for a small value of h we under smooth the density, i.e. the bias is small but the variance becomes large. Then, the optimal choice is needed to get a good estimator of the density. In our case, by an example, we can show the effect of the bandwidth on the estimation of the random effects density. The same data in example 2are considered. the number of points of the simple nonparametric method is fixed to be $G = 4$ just in order to vary the value of G. In Figure 3.2, we plot the true density (mixture of two normals), smoothed kernel estimator with two different value of the bandwidth parameters, a small one $h = 0.3$ and a large one $h = 3$. For $h = 3$, we get a smooth estimator with one mode and for $h = 0.3$ we have a more variable estimator, with four modes which are just the values of the simple nonparametric estimators.

To estimate the optimal bandwidth parameter we have two alternatives to select the optimal bandwidth. The first one, is to use one methods already used, i.e., cross validation, bootstrap, plugging, etc. Another method, more adapted for the generalized linear mixed model is to choose the bandwidth by maximizing the marginal likelihood. In the next section we describes this methods with details.

3.3 Estimation of the Model Parameter

The smoothed kernel estimator for the random effects density can be used for inference on the random effect or to known the number of components of the mixture and the shape

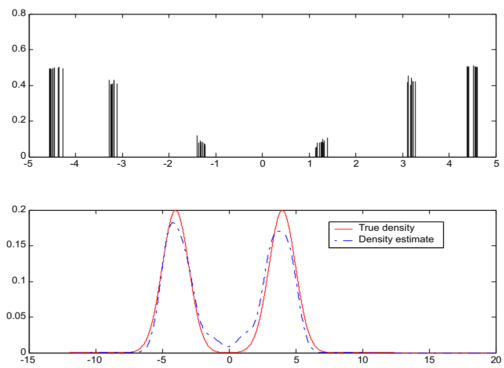

Figure 3.2: The simple nonparametric and the smoothed kernel estimation for the density of the random effect drawn from a mixture of normal densities $0.5N(-4, \sqrt{2}) + 0.5N(4, \sqrt{2})$

of the densities in the mixture . But we can also use this estimator to estimate the fixed parameters. In this section we describe the estimation of the parameters. The idea, like other methods, we replace g_h in (1.1) by $\hat{g}_h$ and then estimate the fixed parameters, i.e. β and h by maximizing the marginal likelihood:

$$L(\zeta) = \prod_{p=1}^{P} \int f_p(y_p|\theta_p)\hat{g}_h(\theta_p)d\theta_p,$$

where

$$f_p(y_p|\theta_p) = \prod_{i=1}^{I} f_p(y_{pi}|\theta_p)$$

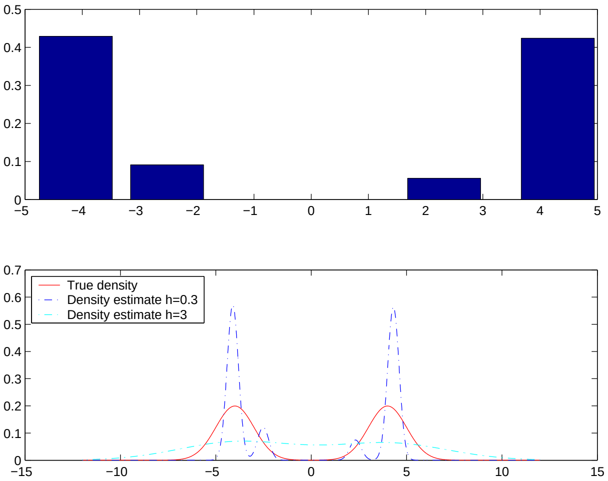

Figure 3.3: The true density and the smoothed kernel estimator with two value of the bandwidth parameters $h = 0.3$ and $h = 3$

References

De Boeck, P., & Wilson, M. (2004) *Explanatory Item Response Models: A generalized Linear and Nonlinear Approach*. New york, Springer.

Ghidey, W., Lesafre, E. & Eiler, P. (2004) Smoothed Random Effects Distribution in a Linear Mixed Model. *Biometrics* 60, 945-653.

Laird, N. (1978) Nonparametric Maximum Likelihood Estimation of a Mixing Distribution. *Journal of the American Statistical Association* 73, 805-811.

Lindsay, B.G. (1983) Semiparametric Estimation in the Rasch Model and Related Exponential Response Models, Including a Simple Latent Class Model for Item Analysis. *Journal of the American Statistical Association* 86, 96-107.

Lindsay, B.The Geometry of Mixture Likelihoods: A General Theory. *The Annals of Statistics* 11, 86-95.

Magder, L.S. & Zeger, L. (1996) A smoothe Nonparametric Estimate of a Mixing Distribution Using Mixtures of Gaussians. *Journal of the American Statistical Association* 91, 1141-1151.

Verbeke, G. & Lesaffre, E (1996) A linear Mixed-Effects Model With Heterogeneity in the Random-Effects Population. *Journal of the American Statistical Association* 91, 217-221.

⋮

	β	Kernel	Nonpar.	Mixed(C=1)	Mixed(C=2)	Mixed(C=3)
$n = 100$	-1	-1.1016	-1.0284	-0.8648	-1.0384	-1.0322
	-0.5	-0.5587	-0.5098	-0.4128	-0.5080	-0.5089
	0	-0.1757	-0.1666	-0.1329	-0.1580	-0.1650
	0.5	0.6411	0.5886	0.4805	0.5919	0.5897
	1	1.1949	1.1162	0.9299	1.1124	1.1165
$\sum$errors	-	0.6719($h^* = 1.18$)	0.4096	0.4449	0.4087	0.4123
$n = 500$	-1	-0.9981	-1.0413	-0.8547	-1.0357	-1.0390
	-0.5	-0.4324	-0.4433	-0.3693	-0.4473	-0.4450
	0	0.0244	0.0303	0.0212	0.0262	0.0306
	0.5	0.3541	0.3713	0.3027	0.3673	0.3709
	1	1.0521	1.0830	0.9001	1.0895	1.0824
$\sum$errors	-	0.2919($h^* = 1.369$)	0.3400	0.5944	0.3369	0.3361
$n = 1000$	-1	-1.0004	-0.9853	-0.9604	-0.9856	-0.9852
	-0.5	-0.4296	-0.4227	-0.4132	-0.4239	-0.4243
	0	-0.0631	-0.0632	-0.0609	-0.0624	-0.0629
	0.5	0.4525	0.4441	0.4346	0.4459	0.4457
	1	1.0405	1.0272	0.9999	1.0261	1.0267
$\sum$errors	-	0.2219($h^* = 1.435$)	0.2383	0.2528	0.2331	0.2344

Table 3.1: estimation of the fixed parameters by the simple nonparametic approach, smoothed kernel, and parametric approach assuming the normality (C=1) and a mixture of two (C=2) and three (C=3) normals, the random effects are $N(0, 2)$

	β	Kernel	Nonpar.	Mixed(C=1)	Mixed(C=2)	Mixed(C=3)
$n = 100$	-1	-1.0156	-1.0801	-0.7594	-1.0500	-1.0675
	-0.5	-0.4091	-0.4159	-0.3086	-0.4293	-0.4357
	0	-0.2610	-0.2342	-0.1904	-0.2529	-0.2488
	0.5	0.7614	0.7890	0.5757	0.8004	0.8023
	1	0.9244	0.9412	0.6826	0.9318	0.9498
$\sum$errors	-	$0.7045(h^* = 1.63)$	0.7462	1.0155	0.7422	0.7330
$n = 500$	-1	-1.1192	-1.0461	-0.8437	-1.0388	-1.0495
	-0.5	-0.5748	-0.5416	-0.4360	-0.5427	-0.5413
	0	0.1351	0.1233	0.0962	0.1228	0.1256
	0.5	0.4593	0.4307	0.3477	0.4294	0.4314
	1	1.0995	1.0337	0.8359	1.0292	1.0338
$\sum$errors	-	$0.4693(h^* = 1.58)$	0.3141	0.6327	0.3042	0.3189
$n = 1000$	-1	-1.0069	-0.9740	-0.7028	-0.9609	-0.9731
	-0.5	-0.6183	-0.6006	-0.4209	-0.5994	-0.6034
	0	0.0203	0.0250	0.0084	0.0239	0.0265
	0.5	0.5322	0.5138	0.3701	0.5116	0.5176
	1	1.0726	1.0358	0.7451	1.0247	1.0325
$\sum$errors	-	$0.2505(h^* = 1.52)$	0.2012	0.7694	0.1987	0.20695

Table 3.2: estimation of the fixed parameters by the simple nonparametric approach, the smoothed kernel, and parametric approach assuming the normality (C=1) and a mixture of two (C=2) and three (C=3) normals, the random effects are $0.5N(-1, 2) + 0.5N(1, 2)$