T E C H N I C A L
R E P O R T

0699

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 ‡

January 9, 2007

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. The new approach is compared with the parametric model assuming either the normal distribution or the parametric model with a mixture of two normals for the random effects.

Key words and phrases: generalized linear mixed model, Rasch model, nonparametric method kernel estimator.

---

*Financial support from the IAP research network . P5/24 of the Belgian government (Belgian Science Policy) is gratefully acknowledged

†Associate Researcher to the Institute of Statistics, UCL.

‡Department of Psychology, KULeuven.

1 Introduction

Generalized linear mixed models have become a powerful parametric tool in psychometrics. A common type of psychometric model for binary data, such as the Rasch model, has a random intercept or person parameter θ, and a number of fixed effects, each corresponding to an item difficulty β as indicated in (2.1) and (2.2). 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 are different ways for the estimation and inference one can make depending on how the person specific 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 see 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. An alternative is the conditional maximum likelihood (CML) approach based on sufficient statistics. The CML is applied only for models with sufficient statistics for person parameters. For the Rasch model, the sufficient statistic for the person-specific parameter 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)); see also the biostatistical literature on CML in Chapter 4 of McCulloch and Searle (2001). However, there are some disadvantages associated to CML. First, no inferences are possible on persons since the probability of observing the response patterns depend only on the sufficient statistics and not on the person effect. Second, CML is not the most efficient method, because the conditional likelihood is maximized rather than the full likelihood, although the loss of efficiency may be minor; see Molenaar (1995).

Another way is to consider the person parameters 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:

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

To maximize the MML quantity, we need to specify the density g_p. We can distinguish three cases of the MML approach: nonparametric, semiparametric and parametric (Heinen (1996)). Each method depends on the assumptions one makes about the unobserved population density of the random effects. In the nonparametric maximum likelihood estimation method, no assumptions 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 the semi-parametric estimation method, see Heinen (1996) for the description of the nonparametric and semiparametric approaches), 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 a parametric approach in general. However, since random effects are unmeasurable (latent factors), this is an assumption that cannot be checked in a direct way. In linear mixed models, we assume often that random effects have a normal density. Verbeke and Lesaffre (1996,1997) 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 indeed to the normality assumption. A widely used method to relax the normality assumption is to use a mixture of normals model, see Verbeke and Lesaffre (1996). Recently, the B-spline smoothing approach is proposed for the linear model by Wendimagegn et al, (2004). For the generalized linear mixed model, the normality and mixture of normals of the random effects are also frequently assumed. The effects of misspecification of the random effect distribution was the subject of many studies (Neuhaus et al, (1992); Heagerty et al., (2001); Chen et al (2002); and Agresti et al., (2004)). Recently, for longitudinal data, Litière et al. (2004) have studied through simulations the effect of misspecifying the random effects distribution on the estimation of the fixed effects in the case of a logistic mixed effects model with treatment and time covariates, and with a random intercept. They found that the estimates tend to converge to values different from the true ones, this deviation being the largest for the estimation of the variance components. Asymptotic normality of the estimates holds, but not convergence to the true effects.

It seems then that the semi-parametric and the nonparametric approach Laird (1978) are good alternatives to relax the parametric assumption on the density of the random effects. But the estimators of the random effect density using such maximum likelihoods are discrete, which is not ideal. To overcome this problem and to relax the normality assumption of the random effect, Magder and Zeger (1996) proposed a mixture of Gaussians to smooth the nonparametric maximum likelihood estimate. Zhang and Davidian (2001) suggested the semi-nonparametric, called SNP, method. They maximize the marginal likelihood over a class of smooth densities which may be approximated by a truncated series expansion. The two methods are proposed for linear mixed models. The latter technique is applied by Chen et al (2002) for a more general class of clustered-data GLMMs. In this article, we propose the "smoothed kernel approach" to estimate the density of the random effects in GLMMs based on the smoothed kernel method. This method can be used to decide how many components are needed if a finite mixture is assumed, and it can be used as well for inference on the fixed and random effects. Note that the smoothed kernel approach can be also used for the linear mixed model. This paper will be devoted to the Rasch model, implying that the density of one random effect will be estimated. However, the results of this paper can be extended to other models, such as the 2PL model and multidimensional models.

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 is described and the estimation of the bandwidth parameter is developed in Section 3. In Section 4, we investigate the finite sample properties of the smoothed kernel method. Section 5 concludes.

2 Rasch model and smoothed kernel

2.1 Rasch model

The Rasch models is a random intercept model for binary repeated measurements. The measurements concern the responses of persons to items. The covariates of the measurements are dummy item indicators. The Rasch model is an item response model with one parameter for each person, called the item ability, see van der Linden and Hambleton (1997). The Rasch model is one of the prominent models for binary items (e.g. success/failure on test items) in psychometrics, and it is a special case of the generalized linear mixed model if one assumes that the ability is a random effect. In the Rasch model the log odds of subject p giving a correct response to item i are modeled using a simple additive formula:

$$\eta_{pi} = \theta_p - \beta_i, \tag{2.1}$$

where β_i is a fixed effect and represents the difficulty of item i, and θ_p is a random effect and represents the ability of the person p. In terms of probability, the model is:

$$\pi_{pi} = \exp(\eta_{pi})/(1+\exp(\eta_{pi})) \tag{2.2}$$

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

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

2.2 Smoothed kernel

The most popular and the well documented nonparametric estimator of an unknown probability density function f is the standard kernel estimator; see Silverman (1986). Given a random sample, $X_1, \cdots, X_n$, from a probability distribution F with an unknown density function f, the smoothed kernel estimator is defined as follows:

$$f_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K((x - X_i)/h), \tag{2.3}$$

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". 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.

In this paper we will use the standard kernel to estimate the density of the random effect, but for grouped data, as will be explained next. Suppose that the sample of size n can be subdivided into G distinct subsamples with respective sizes $n_1, n_2, ..., n_G$ and representatives $X_1^*, X_2^*, ..., X_G^*$. For such grouped data, the kernel estimator is adapted:

$$\begin{aligned} \hat{f}_h(x) &= \frac{1}{h}\sum_{g=1}^{G} \frac{n_g}{n} K((x - X_g^*)/h) \\ &= \sum_{i=1}^{G} \pi_g K_h((x - X_g^*)) \end{aligned} \tag{2.4}$$

where $\pi_g = \frac{n_g}{n}$ and $K_h(t) = \frac{1}{h}K(\frac{t}{h})$, $g = 1, \cdots, G$.

This adaptation is needed because the density estimator does not concern an observed continuous variable, but a latent one, which has to be approached in a discrete form. This is explained in the next section.

3 Smoothed kernel for the Rasch model

In linear and generalized linear mixed models, the most common approach to estimate the distribution G of the random effect is to assume that G is a Gaussian distribution and use the maximum likelihood method to estimate the parameters of the distribution. For the linear mixed model, the 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. Although in generalized linear mixed models this property no longer applies, 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, while the assumption may not be valid, in particular if we know that the population comes from several subgroups.

Here, the nonparametric approach is used. No assumptions at all are made about the shape of the distribution of the random effects, i.e., let the data speak for themselves and select the distribution with the highest likelihood. Lindsay (1983) showed that under general conditions, the nonparametric approach is discrete and has a limited number of points of support, which is a disadvantage of this estimator in particular if the shape of the density is of interest. To overcome this problem, Magder and Zeger (1996), in order to control the smoothness of the estimate, proposed to maximize the likelihood not over all distributions but over the class of arbitrary mixtures of Gaussian subject to the constraint that the variances of the Gaussian must be greater than or equal to specific minimum value. As an alternative to the discrete estimator of the nonparametric approach, the smoothed kernel method is presented here. However, a complication arises because the kernel estimation method needs observable data, whereas the random effects are not observed. The solution for this issue 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 model.

3.1 Description of the method

We first describe how the density of the random effect is estimated by the kernel method. 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 the Rasch model. 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 the Rasch model

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

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 (see (3.2)). The number G is estimated but can also be fixed in advance. Let us first consider that G is fixed. To estimate the random-effects density the following adapted kernel estimator will be used (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 the smoothing parameter. For the multidimensional case, we take a multidimensional kernel.

The kernel method to estimate the density of the random function is a two-steps procedure:
Step 1. Estimate $\theta^* = (\theta_1, ..., \theta_G)$ and $\pi^* = (\pi_1, ..., \pi_G)$ using the nonparametric maximum likelihood approach.
Step 1. The modified version of the kernel density estimator g_h in (3.2) of the random effects is:

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

If the kernel is the standard normal density, the smoothed kernel estimator can be viewed as a mixture of G normals with the same variance h^2. The present approach is different from the parametric approach using a mixture of two or more normals. The means and the weights of the components in the kernel approach are determined by the simple nonparametric method. Also the number of components G is selected in the first step by the simple nonparametric maximum likelihood method.

The smoothed kernel can be reduced to be a one-step approach, maximizing the marginal likelihood over the class of densities that can be expressed as (3.3). This is equivalent to maximizing the marginal likelihood over the class of mixtures of normals with the same variance h^2. Then, the smoothed kernel method is similar to the smoothed nonparametric estimate of Magder and Zeger (1996) proposed for linear mixed models.

3.2 Illustration

We will illustrate the discrete nonparametric maximum likelihood estimator and the smoothed kernel estimator, with different examples. Ten random samples of size $n = 500$ are generated from the Rasch model, one per type of density: normal with zero mean and variance equal to 16, a symmetric mixture given by $0.5N(-4, 3) + 0.5N(4, 3)$; a chi-square density with 16 degrees of freedom, and a lognormal distribution with parameters $\mu = 2$ and $\sigma = 0.5$. The distributions were shifted back over their mean to satisfy the mean zero condition of the random effects. Two methods are considered: the simple nonparametric method and the smoothed kernel method. The fixed effects are chosen on the basis of the range of the density of the random effect, for example for β_i the range $[-5, 5]$ was chosen, so that the fixed effects are within the range covered by the body of the random effect, excluding the tails of the distribution. For each sample, eleven item parameters were used. We fixed the number of points for the nonparametric method to G=5, with respective representatives $X_1^*, X_2^*, ..., X_5^*$ and sizes $n_1, n_2, ..., n_5$; but note that implicitly we use $n = n_1 + n_2 + ... + n_5$ observations for the smoothed kernel estimator. Note that in principle the nonparametric method can select this number. We use the standard normal density as a kernel. The value of the smoothing parameter is chosen by the maximum likelihood method, as will be explained in Section 3.3.

Figure 3.1 shows the results of the simulations, for a representative sample. The other samples look very similar. The first column is the histogram of the random effect as generated, the second column is the simple nonparametric estimator. The five bars in the second column are the first step of the procedure. We can see that based on the simple nonparametric estimator it is difficult to detect the underlying model of the random effects. The last column of the figures shows the density estimated by the kernel method, in comparison to the true density. One can see that the smoothed kernel estimator performs as expected and approaches the true density quite well. The full line of the figure represents the true density of the random effect: the normal followed by a mixture of two normals, a chi-square and a lognormal density. With densities such as in these simulations, the problem of the bias at the boundary is not present. Boundary kernels (Jones and Foster, 1993; and Bouezmarni and Scaillet, 2005) are recommended if a boundary bias is present, for example if the random effect would follow an exponential density.

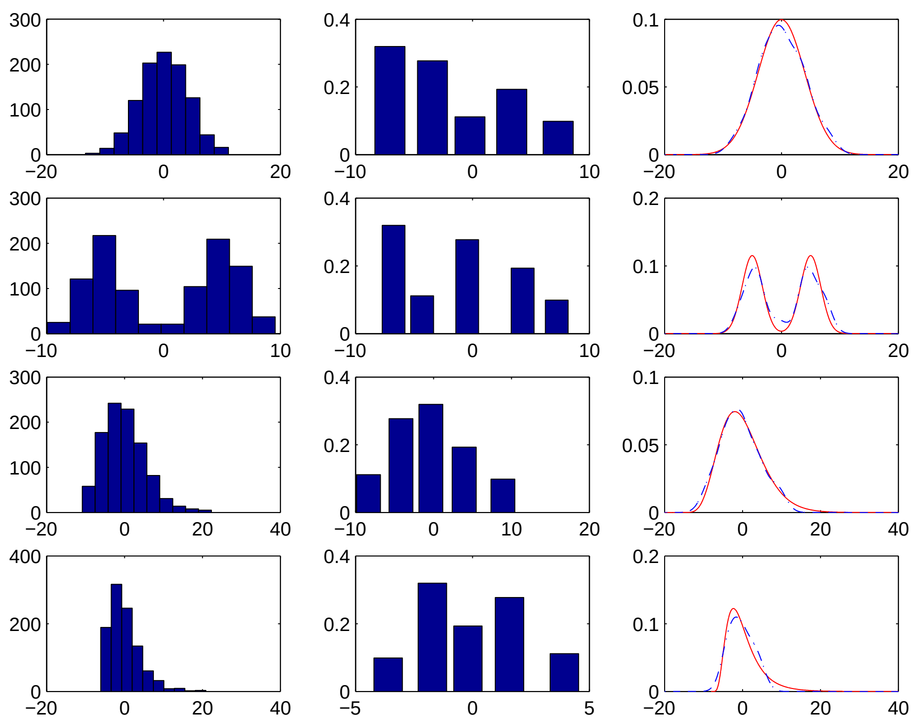

Figure 3.1: The histogram of the random effect generated, the simple nonparametric estimator and the smoothed kernel estimator. The full line represents the true density.

3.3 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, the density is oversmoothed; the variance is reduced, the kernel estimator is stable but the bias is too large, the expected density based on the kernel estimator is too far from the true density. And for a small value of h the density is undersmoothed; the bias is small but the kernel estimator is more variable. An optimal choice is needed for a good estimator of the density. The effect of the bandwidth on the estimation of the random effects density can be illustrated with a symmetric mixture of two normals. The number of points of the simple nonparametric method is fixed to be $G = 4$. In panel a of Figure 3.2, one can see the true density (mixture of two normals, full line), and the smoothed kernel estimator for three different values of the bandwidth parameters: a small one $h = 0.3$, a moderate one $h = 1$ and a large one $h = 3$. For $h = 3$, the smooth estimator has one mode, for $h = 1$ it has two modes and for $h = 0.3$ it has four modes. As the bandwidth parameter goes to zero, the smoothed kernel estimator approaches the simple nonparametric estimator (see panel b of Figure 3.2).

To estimate the optimal bandwidth parameter we have two alternatives. The first one, is to use existing methods, such as cross-validation, bootstrap, plugging, etc. The second, which is better suited to the generalized linear mixed model, is to choose the bandwidth by maximizing the marginal likelihood. Indeed, the bandwidth parameter can be considered as a fixed parameter. As the kernel is supposed to be a normal density, the bandwidth parameter is the standard error of the kernels. The estimation of the bandwidth parameter is equivalent to the estimation of the variance for a mixture of normals. Unfortunately, the maximization of the likelihood is not possible analytically. The EM algorithm can be used to maximize the marginal likelihood and to obtain the estimator of h. The bandwidth parameter $h = 1$ in Figure 3.2 is actually the result of maximizing the marginal likelihood. We can see that the smoothed kernel estimator with this bandwidth approaches the true density and illustrates the two modes.

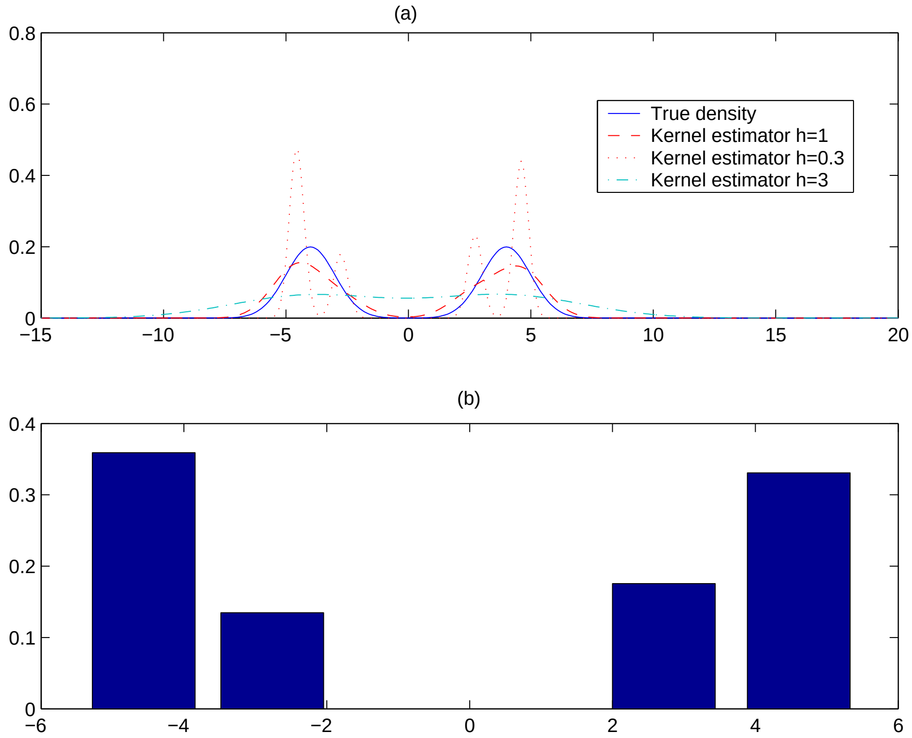

Figure 3.2: (a) The true density and the smoothed kernel estimator for three values of the bandwidth parameter $h = 0.3$, $h = 1$, and $h = 3$; and (b) the first step, the simple nonparametric method

4 Simulation results.

4.1 A comparison of the approaches based on a simulation study.

The properties of the smoothed kernel estimator will be compared with those of the parametric model assuming either the normal distribution or the parametric model with a mixture of two normals for the random effects. We consider four true densities: three densities with a single mode and one with two modes, two symmetric ones and two asymmetric ones.

(a) Normal density with mean zero and variance equal to 16. For this density the fixed effects will be the vector $(-4, -1, 2, 3)$.

(a) Mixture of two normals with equal probability: $0.5N(-4, 3) + 0.5N(4, 3)$. For this density the fixed effects will be the vector $(-5, -3, -2, 1, 3)$.

(a) Chi-square density with 16 degrees of freedom. For this density the fixed effects will be the vector $(-4, -3, 2, 3)$.

(a) Lognormal density with parameters $\mu = 2$ and $\sigma = 0.5$. For this density the fixed effects will be the vector $(-3, -2, 2, 3)$.

The chi-square and the lognormal distributions are shifted to obtain a mean zero condition of the random effects and the fixed effects are chosen to cover the range of random effects.

The study is based on 100 simulations for each density. Throughout the sample is $n = 1000$. The EM algorithm is used to obtain the estimators. For the EM approach, the number G in the first step of the kernel method must be specified. One method to avoid the selection of a number of components is to use a very large number of components. The idea is then that some of the estimated support points may coincide, or some support points will obtain a zero weight. However, this leads to extremely slow convergence of the algorithm. From experience with simulations, it was derived that the smoothed kernel estimator gives a good result with G equal to 3, 4 or 5; these numbers are sufficiently large to obtain a good accuracy and they are sufficiently small for a reasonable computing time. An objective and simple method for the choice of the G remains an open question. Here the number of support points of the simple nonparametric estimator is fixed to $G = 5$.

Global performance is assessed in terms of the mean and deviation error of the 100 simulations of an intuitively appealing global criterion

$$L_\infty(\hat{f}, f) = \sup_I |\hat{f} - f|.$$

where f is the true density, $\hat{f}$ is either the smoothed kernel estimator, the parametric estimator assuming the normal distribution, or the parametric estimator assuming a mixture of two normals, and I is a grid of values of the support of the densities. For each simulation the bandwidth minimizing the L_∞ norm among a grid of values is chosen. Note that the results based on the bandwidth chosen by the maximum likelihood approach are very similar. Table 3.1 lists the results.

The smoothed kernel estimator shows a smaller deviation from the true density than the other methods for the chi-square and the lognormal densities. For the normal density function, the parametric model with normality assumption has a smaller error than the smoothed kernel, as may be expected because the assumption is correct, but the difference

Densities	$\sup \|\hat{f} - f\|$	Smoothed kernel	parametric one normal	parametric two normals
Normal	Mean	0.0159	0.0100	0.0240
	Standard deviation	0.0054	0.0072	0.0250
Mixture	Mean	0.0334	0.0618	0.0301
	Standard deviation	0.0083	0.0022	0.0211
Chi-square	Mean	0.0121	0.0607	0.0195
	Standard deviation	0.0026	0.0112	0.0101
Lognormal	Mean	0.0295	0.1830	0.0305
	Standard deviation	0.0070	0.0232	$1.1.10^{-16}$

Table 4.1: Mean and standard deviation of L_∞ error of the 100 estimations with the smoothed kernel estimator, the parametric model assuming the normality, and the parametric model assuming mixture of two normals.

is very small. For the mixture of two normals, the parametric method with two normals has a smaller error than the smoothed kernel method as may expected because the assumption is correct, but the difference is extremely small.

The standard deviation of the error is rather small for the smoothed kernel estimator in comparison with the other two methods. For two densities (normal, chi-square) the standard deviation of the smoothed kernel is the smallest, and for the two other densities(mixture, the lognormal) it is the second smallest. This means that the method yields stable results.

It may be concluded that the smoothed kernel estimator is the better estimator unless the alternative estimator corresponds with the true model, but also then the smoothed kernel estimator performs quite well. As far as stability is concerned, the smoothed kernel estimator has good qualities as well, and in general performs better than the other estimators. For the mixture density, the parametric model estimator based on one normal seems more stable, but it is much more biased at the same time; and the lognormal is more stable when a mixture of two normals is used.

4.2 Fixed effects estimation

The estimation of the fixed parameters constitutes an important point in the generalized linear mixed model. The smoothed kernel estimator can be used also to estimate the fixed parameters. For this, we replace the true density in (1.1) by the smoothed kernel estimator and the result is:

$$L_{MML}(\phi) = \prod_{p=1}^{P} \int_{-\infty}^{+\infty} \left[\prod_{i=1}^{I} Pr(Y_{pi} = y_{pi} | \theta_p) \right] \hat{g}_p(\theta_p) d\theta_p \tag{4.1}$$

The fixed parameters to be estimated are the bandwidth parameter h and the item parameters β. We maximize the marginal likelihood to obtain the estimators. Unfortunately, we can not obtain an explicit form for the parameter estimates since a direct maximization is not possible analytically. The EM algorithm can be used to maximize the marginal likelihood and to obtain the estimators of the fixed parameters β and h. Here we concentrate on the item parameters β. The smoothed kernel method can be a good alternative for the parametric model using the normality assumption to estimate the item parameters of the model. In this section, we apply this method to estimate the fixed parameters in the Rasch model with the four models of the random effects considered in Section 4.1 and compare the results with those obtained by the parametric model assuming normality of the random effects, which is the common approach. Note that the fixed parameters of the mixture density and the chi-square model are changed. Table 3.2 reports the mean, the standard deviation and the Kolmogorov-Smirnov test (KS-test) of normality of the fixed parameter estimators. For a normal density of the random effects, the two methods provide good results; but as expected, the parametric method gives better results, in terms of mean and standard deviation, than those of using smooth kernel approach. When the random effects are supposed to be mixture of two normals, we remark some bias in the estimation with the parametric method which not the case in the estimation with the kernel method. The latter dominates also in term of variance. For the chi-square and the lognormal model, the parametric approach provides clearly biased estimators of the fixed parameters. These poor results must be attributed to the asymmetric shape of the two densities. The smoothed kernel method does provide quite good results for the two asymmetric densities.

Using the Kolmogorov-Smirnov test for the normality of the fixed effect estimators, the normality assumption is never rejected when the parametric approach is followed, which is in line with the results obtained by Litière et al. (2004). When the smooth kernel approach is followed, the normality is rejected in three cases for $\alpha = .05$, but never for $\alpha = .1$.

A comparison of the kernel method with other methods for the estimation of the item parameters should be a topic of further research, but thus for the kernel method is shown to be a reasonable approach. Also, the bandwidth parameter deserves further investigation. Here we adopt the likelihood method presented in Section 3.3.

5 Discussion and conclusion

A new method was proposed to estimate the random effect density in a GLMM such as the Rasch model. This method is the smoothed kernel estimator. Through simulations it was shown that the smoothed kernel estimator has a good performance, in comparison with the parametric method assuming normality or a mixture of two normals. With the smoothed kernel method one can avoid the normality assumption which is often too restrictive, and at the same time more accurate information on the random effect is provided, including multimodality and skewness. The maximum likelihood estimators based on a normal density or a mixture of normal densities have a poor performance than the smoothed kernel estimator when the density assumption is not correct. The smoothed kernel approach seems to be a reasonable one for the estimation of the fixed effects, and it performs clearly better the normal density approach when the density is asymmetric. A comparison with other approaches such as the conditional maximum likelihood need to be a subject of further research. Although the method seems to work well, further simulations are needed to investigate the consistency of the maximum likelihood estimator based on the smoothed kernel approach. To select the bandwidth parameter we proposed here the maximum likelihood approach. The results were similar to those obtained with minimizing the deviation from a known true density. However, a comparison of the maximum likelihood with the classical methods such as cross-validation, bootstrap, plugging is an interesting topic for further investigation.

Table 4.2: Fixed parameters estimates (means, SEs, p-value of normality) for four different density based models using a kernel approach and a normality approach (based on 100 replications).

		kernel			normal		
Model	Parameter	Mean	SE	p-value[1]	Mean	SE	p-value[1]
Normal	-4	-4.0234	0.2287	0.2254	-4.0069	0.2089	0.5000
	-1	-1.0300	0.2097	0.0173	-1.0077	0.1065	0.3758
	2	2.0292	0.1590	0.1094	1.9970	0.1347	0.1008
	3	3.0242	0.2013	0.1595	3.0176	0.1674	0.2133
Mixture	-3	-2.9762	0.1351	0.1025	-3.0985	0.3128	0.2023
	-1	-0.9446	0.1375	0.4118	-1.0960	0.2333	0.0911
	1	0.9415	0.1127	0.5000	1.0920	0.2218	0.2079
	3	2.9793	0.1726	0.5000	3.1025	0.3253	0.1138
Chi-square	-4	-3.9537	0.1344	0.1130	-4.1508	0.2123	0.1624
	-1	-1.0375	0.2406	0.0442	-1.2272	0.1518	0.0779
	2	2.0165	0.1296	0.5000	2.1870	0.1760	0.2456
	3	2.9747	0.1881	0.3416	3.1911	0.1839	0.5000
Lognormal	-3	-2.9780	0.1222	0.0144	-2.8188	0.1136	0.1548
	-2	-2.0430	0.1424	0.1580	-2.0014	0.1116	0.5000
	2	1.9859	0.1329	0.5000	1.8427	0.1128	0.2091
	3	3.0351	0.1160	0.5000	2.9776	0.1223	0.1806

[1] The p-value is based on the Kolmogorov-Smirnov test of normality for the fixed parameter effects estimates.

References

Agresti, A., Caffo, B., and Ohman-Strickland, P., (2004). Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Computational Statistics and Data Analysis*, 47 (3), 639-653.

Andersen,E.B. (1970). Asymptotic properties of conditional maximum likelihood estimates. *Journal of the Royal Statistical Society, Series B*, **32** , 283-301.

Bouezmarni, T. and Scaillet , O. (2005). Consistency of Asymmetric Kernel Density Estimators and Smoothed Histograms with Application to Income Data. *Econometric Theory*, **21**, 390-412.

Chen, J. Zhang, D. and Davidian, M. (2002) A Monte Carlo EM Algorithm for Generalized Linear Mixed Models with Flexible Random Effects Distribution.. *Biostatistics* , 3, 347-360.

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

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

Heagerty, P.J., and Kurland, B.F. (2001). Misspeciefied maximum likelihood estimates and generalized linear mixed models. *Biometrika*, 88, 973-985.

Heinen, T. (1996). *Latent Class and Discrete Latent Trait Models: Similarities and Differences.* Thousand Oaks, CA: Sage.

Jones, M.C. and Foster, P.J. (1993). Generalized jackknifing and higher order kernels. *Journal of Nonparametric Statistics*, 3, 81-94.

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.

Litière, S., Alonso, A., Molenberghs, G., and Geys, H. (2005) The impact of a misspecified random effects distribution on maximum likelihood estimation in generalized linear mixed models. Submitted to the *Journal of Statistical Computation and Simulations*

McCulloch, C.E., and Searle, S.R. (2001). *Generalized, Linear, and Mixed Models.* New York: Wiley.

Magder, L.S. and Zeger, L. (1996). A smooth nonparametric estimate of a mixing distribution using mixtures of Gaussian. *Journal of the American Statistical Association*, 91, 1141-1151.

Molenaar, I. (1995). Estimation of item parameters. In G.H. Fisher and Molenaar (Eds), *Rasch Models. Foundations, Recent Developments and Application* (pp. 39-57). New York: Spring.

Neyman, J., and Scott, E.L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16, 1-32.

Neuhaus, J.M., Hauck, W.W., and Kalbfleisch, J.D. (1992) The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika*, 79, 755-762.

Silverman, B.W. (1986). *Density estimation for statistics and data analysis.* Chapman and Hall, London.

van der linden, W.J. and Hambleton, R.K.(Eds) (1997) *Handbook of moden item response theory.* New York: Springer.

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

Verbeke, G. and Lesaffre, E (1997) The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics and Data Analysis*, 53, 541-556.

Wendimagegn, G., Lesafre, E., and Eiler, P. (2004). Smooth random effects distribution in a linear mixed model. *Biometrics*, 60, 945-953.

Zhang, D. and Davidian, M. (2001) Linear mixed models with flexible distributions of the random effects for longitudinal data. *Biometrics* , 75, 795-802.