



## SNP\_NLMM: A SAS Macro to Implement a Flexible Random Effects Density for Generalized Linear and Nonlinear Mixed Models

David M. Vock  
University of Minnesota

Marie Davidian  
NC State University

Anastasios A. Tsiatis  
NC State University

---

### Abstract

Generalized linear and nonlinear mixed models (GLMMs and NLMMs) are commonly used to represent non-Gaussian or nonlinear longitudinal or clustered data. A common assumption is that the random effects are Gaussian. However, this assumption may be unrealistic in some applications, and misspecification of the random effects density may lead to maximum likelihood parameter estimators that are inconsistent, biased, and inefficient. Because testing if the random effects are Gaussian is difficult, previous research has recommended using a flexible random effects density. However, computational limitations have precluded widespread use of flexible random effects densities for GLMMs and NLMMs. We develop a SAS macro, SNP\_NLMM, that overcomes the computational challenges to fit GLMMs and NLMMs where the random effects are assumed to follow a smooth density that can be represented by the seminonparametric formulation proposed by Gallant and Nychka (1987). The macro is flexible enough to allow for any density of the response conditional on the random effects and any nonlinear mean trajectory. We demonstrate the SNP\_NLMM macro on a GLMM of the disease progression of toenail infection and on a NLMM of intravenous drug concentration over time.

*Keywords:* random effects, nonlinear mixed models, generalized linear mixed models, SAS, SNP.

---

## 1. Introduction

Generalized linear mixed models (GLMMs) are commonly used to represent non-Gaussian longitudinal or repeated measures data. Nonlinear mixed models (NLMMs) have been traditionally employed to model pharmacokinetic/pharmacodynamic, growth curve, disease dynamic, or other longitudinal data where the mean trajectory is an arbitrary non-linear function of a vector of parameters.

Let  $Y_{ij}$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, n_i$ , be the  $j$ th response on the  $i$ th subject or cluster. For both GLMMs and NLMMs we assume that  $Y_{ij}$  conditioned on a  $q$ -dimensional vector of centered subject-specific random effects  $b_i$  has density  $f_{Y_{ij}|b_i}(y_{ij}|b_i; \beta, x_{ij})$ , which depends on a  $p$ -dimensional vector of fixed effects  $\beta$  and covariates  $x_{ij}$ . Common choices for  $f_{Y_{ij}|b_i}$  are the normal density and the Bernoulli and Poisson probability mass functions. We also assume that the random effects  $b_i$ , independent across  $i$ , have density  $g_{b_i}(b_i; \xi)$  which may depend on a vector of parameters  $\xi$ . If we assume that  $Y_{ij}$  are independent across  $i$  and, given  $b_i$ , are independent across  $j$ , then the likelihood of  $\theta = (\beta^\top, \xi^\top)^\top$  is given by

$$\begin{aligned} \mathcal{L}(\theta; \mathbf{Y}) &= \prod_{i=1}^m \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{j=1}^{n_i} \left\{ f_{Y_{ij}|b_i}(y_{ij}|b_i; \beta, x_{ij}) \right\} g_{b_i}(b_i; \xi) db_{1i} \cdots db_{qi} \\ &= \prod_{i=1}^m \int \prod_{j=1}^{n_i} \left\{ f_{Y_{ij}|b_i}(y_{ij}|b_i; \beta, x_{ij}) \right\} g_{b_i}(b_i; \xi) db_i, \end{aligned} \quad (1)$$

where  $\mathbf{Y}$  is the vector of all responses across all  $m$  subjects. For brevity, we write the  $q$ -dimensional integral using a single integral throughout the remainder of the manuscript.

Typically, one assumes that the random effects density  $g_{b_i}(b_i; \xi)$  is a  $q$ -dimensional normal density. The assumption is usually made for convenience because the normal density is the only density for the random effects supported in the SAS/STAT procedures GLIMMIX and NL MIXED used to estimate GLMMs and NLMMs, respectively (SAS Institute Inc. 2008).

However, in many applications the assumption of Gaussian random effects may be too restrictive, and one stands to lose insight into the data generating process if one assumes *a priori* that the random effects are Gaussian. Maximum likelihood estimators of the fixed effects,  $\beta$ , are also, in general, not consistent if the random effects density  $g_{b_i}$  has been misspecified for GLMMs and NLMMs. Furthermore, researchers have noted that there is considerable bias and loss of efficiency in the maximum likelihood parameter estimators assuming Gaussian random effects when either the true random effects density deviates substantially from normality or the variance of the random effects is large and there is moderate misspecification of the random effects (Neuhaus, Hauck, and Kalbfleisch 1992; Hartford and Davidian 2000; Heagerty and Kurland 2001; Agresti, Caffo, and Ohman-Strickland 2004; Litière, Alonso, and Molenberghs 2008).

Several researchers have proposed methods to relax the assumption of Gaussian random effects (or latent traits), and we briefly review the relative merits of the proposed approaches. At one extreme, one can develop consistent and asymptotically normal estimators for the fixed effects in a GLMM or generalized linear latent variable model that do not require the data analyst to correctly posit the density for the random effects or latent traits. For example, Ma and Genton (2010) use semiparametric theory to find the efficient estimating function for parameters in a generalized linear latent variable model, that is, the projection of the score vector onto the complement of the nuisance tangent space (where the density of the latent trait is the infinite dimensional nuisance parameter), and show that the estimating function is unbiased regardless of the true distribution of the latent traits. The general semiparametric approach is explained in Tsiatis (2006). The approach advocated by Ma and Genton (2010) is similar to the conditional likelihood inference proposed by Sartori and Severini (2004) for GLMMs although the derivation is completely different. However, these semiparametric approaches treat the random effect/latent trait density as a nuisance parameter which may be of scientific interest in many applications. Furthermore, deriving the efficient estimating

functions for the fixed effects that do not require estimating the random effects density is not trivial and depends on the conditional density  $f_{Y_{ij}|b_i}(\cdot)$ .

Alternatively, the random effects distribution that does not assume any parametric form may also be estimated by maximum likelihood. Several authors have proposed computational methods to find the nonparametric maximum likelihood estimate of the random effects density (Laird 1978; Bock and Aitkin 1981; Follmann and Lambert 1989; Lesperance and Kalbfleisch 1992; Aitkin 1999). Knott and Tzamourani (2007) propose using the bootstrap to obtain confidence intervals for the estimated nonparametric density. Lindsay (1983) shows that the nonparametric maximum likelihood estimator is discrete and has a limited number of points of support. This is a serious limitation because in many applications we would expect that the distribution of the random effects or latent traits to be continuous and a discrete approximation may not provide sufficient insight into the true data generating mechanism. Furthermore, we may be able to gain substantial efficiency by making the minimal and realistic assumption that the random effects density is continuous.

Magder and Zeger (1996) and Knott and Tzamourani (2007) both propose a smooth nonparametric maximum likelihood estimator of the random effects density. The method developed by the former results in a finite mixture of Gaussian densities while the later suggest smoothing the discrete estimate using kernel methods. Both involve ad hoc approaches to determine the amount of smoothing and the approach advocated by Magder and Zeger (1996) can be computationally intensive. Tao, Palta, Yandell, and Newton (1999) nonparametrically estimate the density of a random effect using a predictive recursive algorithm. The estimate is guaranteed to be continuous if the starting density in the algorithm is chosen to be a continuous density, but unfortunately the authors note that the sample size must be large to obtain promising results and only implement their method on a one-dimensional random effect. Verbeke and Lesaffre (1997) and Litière *et al.* (2008) propose approximating the random effects distribution using a finite mixture of Gaussian distributions and develop an expectation-maximization (EM) algorithm to fit the model. Like many EM algorithms, their algorithm may be computationally intensive and slow. Ghidey, Lesaffre, and Eilers (2004) similarly assume that the random effects are a Gaussian mixture and propose to estimate the mixing proportions using a penalized log-likelihood. However, their method has only been applied to linear mixed models.

Alternatively, many researchers have developed efficient computational methods to fit models with flexible, parametric classes of random effects densities. For example, prior work has developed computational methods to fit models where the random effects distribution is assumed to belong to the class of  $t$ -distributions (Pinheiro, Liu, and Wu 2001; Lin and Lee 2006, 2007; Lee and Thompson 2008), skew extensions of the  $t$ - and normal distributions (Ghosh, Branco, and Chakraborty 2007; Lin and Lee 2008; Ho and Lin 2010; Huang and Dagne 2010; Labra, Garay, Lachos, and Ortega 2012), normal/independent distributions (a class of thick tailed-distributions which contain the multivariate  $t$ , slash, and contaminated normal distributions, see Rosa, Padovani, and Gianola 2003; Lachos, Bandyopadhyay, and Dey 2011), and skew extensions of the normal/independent distributions (Lachos, Dey, and Cancho 2009; Lachos, Ghosh, and Arellano-Valle 2010; Bandyopadhyay, Lachos, Castro, and Dey 2012). Although each of these classes of distributions may be appropriate in certain domain-specific applications, none is likely to be flexible enough for general use. For example, none of the above classes permit multiple modes.

In another approach given extensive attention in the recent literature, one assumes that the

random effects density follows a smooth density that can be represented by the seminonparametric (SNP) formulation proposed by [Gallant and Nychka \(1987\)](#). Although this class of densities does not include all continuous densities, the class is sufficiently flexible to include skewed, multimodal, and thick and thin-tailed densities. The density does not contain densities which have kinks, jumps, or oscillations; densities unlikely to occur in common practice. The SNP representation has been previously used for the random effects density in GLMMs ([Chen, Zhang, and Davidian 2002](#)) and NLMMs ([Davidian and Gallant 1993](#); [Vock, Davidian, Tsiatis, A., and Muir 2012](#)) as well as used for the latent trait density of generalized linear latent variable models ([Irincheeva, Cantoni, and Genton 2012](#)). The SNP representation is flexible enough to cover a broad variety of applications unlike some of the parametric classes discussed above. Furthermore, the SNP representation works for multidimensional random effects and can be implemented as the random effects density for NLMMs unlike some of its competitors. Finally, by assuming that the random effects density is “smooth” without kinks or oscillations, a minimal but realistic assumption, we may gain efficiency over nonparametric methods. Overall, fitting a GLMM or NLMM assuming the random effects follow the SNP representation is a prudent modeling approach when the random effects density is of scientific interest and there is little subject-area knowledge to suggest *a priori* that the random effects are Gaussian. Of course the choice of a method for a flexible random effects density depends on the goal of the analysis. If the random effects density is not of interest to the investigator, then the non-parametric approaches described above may be sufficient.

Previous work has used Fortran programs to estimate NLMMs and GLMMs where the random effects are assumed to follow the SNP representation which has precluded widespread use of the SNP density. There are two main problems in implementing the SNP density using standard software. First, as noted earlier, standard software only permits Gaussian random effects, so the software must be “tricked” to allow non-Gaussian random effects. Secondly, to obtain maximum likelihood parameter estimates of  $\theta$ , we must be able to approximate the integral in Equation 1. This can be difficult for complex densities  $g_{b_i}$ ; as described by [Irincheeva et al. \(2012\)](#) many of the standard approaches for integral approximation work poorly when the random effects follow the SNP representation.

We develop a SAS macro, SNP\_NLMM, that overcomes those computational challenges and allows one to fit easily GLMMs and NLMMs where the random effects density is assumed to follow the SNP density. In Section 2, we discuss the SNP density in detail. Section 3 describes the input parameters for the SNP\_NLMM macro as well as the output datasets. We describe the computational details used in the macro in Section 4. Section 5 illustrates the use of SNP\_NLMM to fit a GLMM of longitudinal measurements on toenail infection. Section 6 describes the results from implementing the macro to estimate a NLMM of intravenous drug concentration over time. We conclude in Section 7.

## 2. Seminonparametric density

To motivate the development of the SNP density, we consider a NLMM of intravenous drug concentration over time. We will return to this example in detail in Section 6. Let  $Y_{ij}$  and  $t_{ij}$  be the concentration and time since infusion began, respectively, for the  $i$ th subject at the

$j$ th measurement. The concentration profile is assumed to follow the model

$$Y_{ij} = \frac{D_i}{Cl_i} \left[ \exp \left\{ -\frac{Cl_i}{V_i} (1 - t_{ij}^*) (t_{ij} - t_{i,\text{inf}}) \right\} - \exp \left( -\frac{Cl_i}{V_i} t_{ij} \right) \right] + \epsilon_{ij} \quad (2)$$

$$= h(x_{ij}; Cl_i, V_i) + \epsilon_{ij},$$

where  $D_i$  is the infusion rate subject  $i$  received;  $Cl_i$  and  $V_i$  are the unknown subject-specific clearance and volume distribution parameters;  $t_{ij}^* = I(t_{ij} < t_{i,\text{inf}})$ ;  $t_{i,\text{inf}}$  is the infusion time;  $x_{ij} = (t_{ij}, t_{i,\text{inf}}, D_i)^\top$ ; and  $\epsilon_{ij} \stackrel{iid}{\sim} N\{0, \sigma^2 h(x_{ij}; Cl_i, V_i)^{2\gamma}\}$ . Because we assume that  $Cl_i$  and  $V_i$  vary across the population, they are subject-specific random effects. Let  $b_i = (\log Cl_i, \log V_i)^\top$ . Throughout the remainder of the manuscript, we write the  $q$ -dimensional vector of centered (non-mean-zero) random effects  $b_i$  as

$$b_i = \mu + RZ_i, \quad (3)$$

where  $\mu$  is a  $q$ -dimensional vector,  $R$  is a  $q \times q$  lower triangular matrix, and  $Z_i$  is a  $q$ -dimensional random effects vector. One would traditionally assume that  $Z_i$  follows a standard  $q$ -dimensional normal density which implies that  $E(b_i) = \mu$  and  $\text{VAR}(b_i) = RR^\top$ . We rewrite the likelihood given in Equation 1 in terms of  $Z_i$  as

$$\begin{aligned} \mathcal{L}(\theta; \mathbf{Y}) &= \prod_{i=1}^m \int \prod_{j=1}^{n_i} \left\{ f_{Y_{ij}|b_i}(y_{ij} | \mu + Rz_i; \beta, x_{ij}) \right\} g_{b_i}(\mu + Rz_i; \xi) |R| dz_i \\ &= \prod_{i=1}^m \int \prod_{j=1}^{n_i} \left\{ f_{Y_{ij}|Z_i}(y_{ij} | z_i; \beta, \mu, r, x_{ij}) \right\} g_{Z_i}(z_i; \xi) dz_i, \end{aligned} \quad (4)$$

where  $r = \text{vech}(R)$  is the vector containing the non-zero components of  $R$ , and, in the example given,  $f_{Y_{ij}|Z_i}$  is the normal density with mean  $h(x_{ij}; Cl_i, V_i)$  and variance  $\sigma^2 h(x_{ij}; Cl_i, V_i)^{2\gamma}$ . Instead of assuming that  $Z_i$  is distributed as standard normal, we assume that  $Z_i$  belongs to the smooth class of densities,  $\mathcal{G}$ , described by [Gallant and Nychka \(1987\)](#). These authors give a mathematical description of densities in this class which are sufficiently differentiable to rule out densities with jumps or oscillations but flexible enough to include thick/thin-tailed, skewed, and multi-modal densities. Mathematically, densities  $g \in \mathcal{G}$  may be expressed as an infinite Hermite series,  $g(z) = P_\infty^2(z)\psi(z)$ , plus a lower bound on the tails, where  $P_\infty(z)$  is an infinite dimensional polynomial and  $\psi(z)$  is the standardized form of a known density with a moment generating function. The density  $\psi(z)$  is usually taken to be the standard normal density as is the case for this macro. For modeling purposes the lower bound is typically disregarded and the polynomial is truncated. Densities in the truncated class which disregard the lower bound have been referred to as seminonparametric or SNP.

The SNP density with degree of truncation  $K$  is given by

$$g_{Z_i}^K(z; \xi) = P_K^2(z) \phi_q(z) = \left\{ \sum_{(j_1 + \dots + j_q) \leq K} a_{j_1, \dots, j_q} z_1^{j_1} \dots z_q^{j_q} \right\}^2 \phi_q(z), \quad (5)$$

where  $\phi_q(\cdot)$  is the  $q$ -dimensional standard normal density,  $j_l \geq 0$  for  $l = 1, \dots, q$ , and  $K$  is the order of the polynomial  $P_K(z)$ . For example, with  $K = 2$  and  $q = 2$ ,  $P_K(z) = a_{00} + a_{10}z_1 + a_{01}z_2 + a_{20}z_1^2 + a_{11}z_1z_2 + a_{02}z_2^2 = a_{00}z_1^0z_2^0 + a_{10}z_1^1z_2^0 + a_{01}z_1^0z_2^1 + a_{20}z_1^2z_2^0 + a_{11}z_1^1z_2^1 + a_{02}z_1^0z_2^2$ . Note

that the subscripts of the coefficients in the polynomial match the powers of the  $q$  random effects in the same term.

Because  $g_{Z_i}^K(z; \xi)$  is a polynomial of order  $K$  squared times a standard normal density,  $g_{Z_i}^K(z; \xi)$  will be non-negative for any choice of  $\xi$ . However, for  $g_{Z_i}^K$  to be a proper density we must constrain the  $d$ -dimensional vector of coefficients  $a = \{a_{j_1, \dots, j_q} : j_1 + \dots + j_q \leq K\}$ , such that  $g_{Z_i}^K$  integrates to 1. This is equivalent to imposing the constraint  $a^\top A a = 1$ , where  $A$  is the matrix with  $(j, k)$  element equal to  $\{E(U_1^{j_1+k_1}) \dots E(U_1^{j_q+k_q})\}$ ,  $j_1, \dots, j_q$  (similarly  $k_1, \dots, k_q$ ) are the subscripts of the  $j$ th ( $k$ th) element of  $a$ , and  $U_1$  is distributed as a standard normal (Zhang and Davidian 2001). Because  $A$  is positive definite, there is a matrix  $B$  so that  $A = B^2$ ; this implies  $c^\top c = 1$  where  $c = Ba$ . Rather than to impose the constraint  $c^\top c = 1$ , we can use a polar coordinate transformation to write  $c$  using only  $d-1$  parameters where  $c_1 = \sin(\xi_1)$ ,  $c_2 = \cos(\xi_1) \sin(\xi_2)$ ,  $\dots$ ,  $c_{d-1} = \cos(\xi_1) \dots \cos(\xi_{d-2}) \sin(\xi_{d-1})$ ,  $c_d = \cos(\xi_1) \dots \cos(\xi_{d-1})$ ,  $-\pi/2 \leq \xi_j \leq \pi/2$  for  $j = 1, \dots, d-1$ , and  $\xi = (\xi_1, \dots, \xi_{d-1})^\top$  (Zhang and Davidian 2001).

We note that when  $K = 0$ , then  $P_K(z)$  is just a constant,  $a_{0, \dots, 0}$ , which must equal 1, and the density  $g_{Z_i}^0$  simplifies to the standard  $q$ -dimensional normal. For  $K > 0$ , the SNP density does not force  $E(Z_i) = 0$ . Therefore,  $E(b_i) = \mu + RE(Z_i)$  and  $\text{VAR}(b_i) = R\text{VAR}(Z_i)R^\top$ . These moments can easily be computed by noting that the moments of  $Z_i$  are just linear combinations of moments of a standard normal density.

The above discussion has assumed a fixed  $K$ , the value of  $K$  controls the departure from the base density, here the standard normal, and, therefore, the flexibility for approximating the true random effects density.  $K$  is not the number of components in a mixture or the number of subpopulations in the underlying population. Instead,  $K$  plays a similar role to the bandwidth in kernel density estimation with smaller bandwidths and larger values of  $K$  capturing more local features of the random effects densities but leading to more variable estimators (Song, Davidian, and Tsiatis 2002). The class of possible densities grows monotonically as  $K$  increases, and Irincheeva *et al.* (2012) relate the number of possible modes to  $K$  and  $q$ . However, none of the extensive methodological work on the SNP representation has related  $K$  to other features of the corresponding class of densities such as the maximum skewness and kurtosis permitted in the class (see, for example, Gallant and Nychka 1987; Davidian and Gallant 1993; Zhang and Davidian 2001; Chen *et al.* 2002; Song *et al.* 2002; Doehler and Davidian 2008; Zhang and Davidian 2008; Vock *et al.* 2012). Although one could *a priori* fix a choice of  $K$ , an objective selection method is often preferred because of the lack of relationship between  $K$  and the features of the distribution permitted. Specifically, most work has advocated treating  $K$  as a tuning parameter and has suggested fitting models for several values of  $K$  and then choosing  $K$  based on information criteria and/or visual inspection of the estimated densities (Davidian and Gallant 1993; Zhang and Davidian 2001; Chen *et al.* 2002; Vock *et al.* 2012). Prior research cited above has indicated that  $K$  need not be larger than two to capture many complicated densities.

We can obtain estimates of  $\theta$  by using the maximum likelihood estimator. For a given  $K$ , the optimization problem is exactly the same as in standard finite dimensional maximum likelihood estimation provided we can approximate the integral in Equation 4 (see Section 4). However, optimization of the likelihood with the SNP density can be highly dependent on starting values for  $\xi$ ; therefore, it is advisable to maximize the likelihood using a few sets of “good” starting values to ensure that one obtains the global maximum. Our macro uses the following routine advocated by Doehler and Davidian (2008) to generate starting values



and parameter estimates. `SNP_NLMM` obtains initial estimates for  $\beta$ ,  $E(b_i)$ , and  $\text{VAR}(b_i)$  by fitting the GLMM or NLMM assuming Gaussian random effects. The likelihood over a grid of values for  $\xi$  is evaluated where  $\beta$  is set to the initial estimate and starting values for  $\mu$  and  $r$  are set to the values that result in the same values of  $E(b_i)$  and  $\text{VAR}(b_i)$  as their initial estimates from assuming Gaussian random effects. The macro then uses a small number of those sets of parameter values with large likelihoods as starting values in the maximization of the likelihood.

### 3. Macro input parameters and output datasets

#### 3.1. Input parameters

The following gives the input parameters needed for `SNP_NLMM` with the default values

```
%SNP_NLMM(dataset = , subject = , response = ,
  prog_stat = , log_cond_dens = , start_value = , q = , Kmax = 2,
  all_K = 1, info_crit = "BIC", quad_pt_K0 = 0, quad_pt_K1 = 0,
  grid_pt_K1 = 9, max_K1 = 1, quad_pt_K2 = 0, grid_pt_K2 = 9,
  max_K2 = 1, quad_pt_K3 = 0, grid_pt_K3 = 9, max_K3 = 1,
  quad_pt_K4 = 0, grid_pt_K4 = 9, max_K4 = 1);
```

where

- **dataset**: Specifies the dataset that contains the response, subject/cluster variable, and any additional covariates needed to specify the conditional likelihood. Because the macro uses the `NLMIXED` procedure in `SAS`, all covariates should be numeric.
- **subject**: Specifies the subject/cluster variable. Observations that have the same value of **subject** are assumed to have the same realization of the random effects.
- **response**: Specifies the response variable of interest.
- **prog\_stat**: Additional programming statements needed to specify the conditional density. The syntax is the same as for the `NLMIXED` procedure. We emphasize that the centered (non-mean zero) random effects must be referred to as **b\_1** and **b\_2**. The programming statements should be enclosed with the `%str()` macro command. Typically one would need to specify any parameters that are part of the conditional density including the mean and variance if the conditional density is Gaussian or the probability of success if the conditional density is Bernoulli.
- **log\_cond\_dens**: The logarithm of the density of the response given the random effects. We emphasize that the centered (non-mean zero) random effects must be referred to as **b\_1** and **b\_2**. The statements should be enclosed with the `%str()` macro command. While having the user specify the conditional density and corresponding programming statements does require greater involvement, this allows for greater flexibility.
- **start\_values**: Starting values for the fixed effect parameters,  $\mu$  and  $\beta$ , as well as variance components,  $R$ , assuming Gaussian random effects. The starting values should

be given using the syntax of the **parms** statement in the **NLMIXED** procedure in SAS, for example **beta1 = 5 beta2 = 1**. The variance of the random effects has been parameterized using  $R$  where  $\text{VAR}(b_i) = RR^\top$  and  $R$  is a lower triangular matrix. The  $(i, j)$  non-zero components of  $R$  should be referred to as **r\_ij**. The  $i$ th component of the vector  $\mu$  should be referred to as **mu\_i**. The statements should be enclosed with the **%str()** macro command.

- **q**: Specifies the dimension of the random effects. Currently only **q = 1** and **q = 2** are supported.
- **Kmax**: Specifies the maximum degree of flexibility considered. As noted above, prior research has indicated that  $K$  need not be larger than two to approximate most densities encountered in practice. Currently **Kmax** must be less than or equal to four for one-dimensional random effects and less than or equal to three for two-dimensional random effects. The default is **Kmax = 2**.
- **all\_K**: If **all\_K = 1**, then models with  $K$ , the degree of flexibility in the random effects density, equal to zero through **Kmax** are fit. If **all\_K = 0**, then models with  $K$  equal to just zero and **Kmax** are fit. The default is **all\_K = 1**.
- **info\_crit**: Specify the information criterion, BIC or AIC, that should be used to select the degree of flexibility in the random effects density. The default is **info\_crit = "BIC"**.
- **quad\_pt\_K0** and similarly for **\_K1**, **\_K2**, **\_K3**, **\_K4**: Specifies the number of quadrature points used for  $K = 0$  and similarly for other values of  $K$ . If **quad\_pt\_K0 = 0** the number of quadrature points is adaptively selected. Specifically, the macro continues to increase the number of quadrature points until the relative change in the log likelihood evaluated at the first set of starting values is below a pre-specified tolerance. We recommend first running the macro with **quad\_pt\_K0 = 0** to select the number of quadrature points adaptively. Then one can re-run the optimization a couple of times with larger numbers of quadrature points to ensure that the maximum likelihood estimates are not sensitive to the number of quadrature points. This is especially helpful if the number of quadrature points selected adaptively is small. The default is to have the number of quadrature points selected adaptively.
- **grid\_pt\_K1** and similarly for **\_K2**, **\_K3**, **\_K4**: Specifies the number of grid points in each dimension of  $\xi$  that should be used in the grid search for starting values. We note that for  $q = 1$  the dimension of  $\xi$  equals  $K$  and for  $q = 2$  the dimension of  $\xi$  equals  $(K + 1)(K + 2)/2 - 1$ . For  $q = 2$  and large values of  $K$ , the number of grid points can be quite large and the time to evaluate the likelihood at each of these grid points can be extensive. However, we have found that a fine grid is the best way to ensure that the global maximum is obtained. We recommend that the number of grid points in each dimension be at least 9 for  $q = 1$  with  $K$  less than 4 and for  $q = 2$  with  $K$  less than or equal to 2. For scenarios with larger values of  $K$ , we would recommend the number of grid points in each dimension to be at least 5–7.
- **max\_K1** and similarly for **\_K2**, **\_K3**, **\_K4**: Specifies the number of optimizations with different sets of starting values that should be performed to ensure that one obtains the global maximum. The default is **max\_K1 = 1**.



### 3.2. Output datasets

The following datasets are generated from the macro, are stored in the `work` directory, and are printed as part of the output:

- `dim_total`: Contains the number of observations used, number of observations not used, number of subjects, number of parameters, and number of quadrature points for each value of  $K$  considered.
- `fitstat_total`: Contains the fit statistics AIC and BIC as well as  $-2 \log$  likelihood for each value  $K$  considered.
- `parm_total`: Contains the parameter estimates and standard errors for each value of  $K$  considered.
- `parm_final_K`: Contains the parameter estimates, standard errors,  $p$  values, and confidence intervals for the value of  $K$  selected by the information criteria in `info_crit`.

In addition to the datasets generated, `SNP_NLMM` plots the estimated density for each value of  $K$  considered.

## 4. Computational details

We describe here some of the computational issues in implementing the SNP density for GLMMs and NLMMs. This section may be skipped by the reader solely interested in applying the macro. However, these issues may be of interest to investigators trying to implement other flexible random effects densities in standard software.

As discussed previously, to obtain the maximum likelihood parameter estimator, we must be able to approximate the contribution to the likelihood from each subject,  $\mathcal{L}_i(\theta, Y_i)$ , where

$$\mathcal{L}_i(\theta, Y_i) = \int \prod_{j=1}^{n_i} \left\{ f_{Y_{ij}|Z_i}(y_{ij}|z_i; \beta, \mu, r, x_{ij}) \right\} g_{Z_i}(z_i; \xi) dz_i. \quad (6)$$

We have found traditional numerical integration techniques available are either too computationally intensive or not appropriate if either  $f_{Y_i|Z_i}$  or  $g_{Z_i}$  are non-Gaussian. For example, the first-order approximation (Beal and Sheiner 1988) requires that the conditional density of the response given the random effects to be Gaussian. Also, the Laplace approximation does not perform well if the random effects distribution has multiple modes. Although Gaussian quadrature can be used to approximate a broad array of integrals, quadrature, as currently implemented, has limitations as well. Non-adaptive Gaussian quadrature centers the quadrature points at the mean of the random effects and scales them based on the variance of the random effects. However, the region over which the integrand in Equation 6 contains most of its mass may not be near the mean of the random effects. Therefore, a large number of quadrature points may be needed to achieve a reasonable approximation. When  $g_{Z_i}$  is a SNP density, we have found the number of quadrature points needed is usually too large to be computationally feasible for  $q \geq 2$ .

To reduce the number of quadrature points needed to obtain a reasonable approximation, Liu and Pierce (1994) suggest centering the quadrature points at the mode of the integrand

and scaling by the inverse of the negative Hessian matrix. Because, in the context of random effects models, both the mode and Hessian depend on the values of the fixed effect parameters, [Pinheiro and Bates \(1995\)](#) suggest the following iterative algorithm to obtain estimates for  $\theta$ . Let  $\hat{Z}_{i,k}$  be the empirical Bayes estimate of  $Z_i$  given the current (i.e.,  $k$ th) estimates of  $\beta$ ,  $\mu$ ,  $r$ , and  $\xi$ ,  $\hat{\beta}^{(k)}$ ,  $\hat{\mu}^{(k)}$ ,  $\hat{r}^{(k)}$ , and  $\hat{\xi}^{(k)}$ , respectively, and  $\hat{G}_{i,k}$  be the estimated covariance matrix of  $\hat{Z}_{i,k}$ . That is,

$$\hat{Z}_{i,k} = \arg \max_{z_i} \prod_{j=1}^{n_i} \left\{ f_{Y_{ij}|Z_i}(y_{ij}|z_i; \hat{\beta}^{(k)}, \hat{\mu}^{(k)}, \hat{r}^{(k)}, x_{ij}) \right\} g_{Z_i}(z_i; \hat{\xi}^{(k)}). \quad (7)$$

We can write  $Z_i = \hat{Z}_{i,k} + \hat{G}_{i,k}^{1/2} U_i$  in terms of the random variable  $U_i$  and then perform a change of variable to have the integral in Equation 6 taken with respect to  $U_i$ :

$$\mathcal{L}_i(\theta, Y_i) = \int |\hat{G}_{i,k}^{1/2}| \prod_{j=1}^{n_i} \left\{ f_{Y_{ij}|Z_i}(y_{ij}|\hat{Z}_{i,k} + \hat{G}_{i,k}^{1/2} u_i; \beta, \mu, r, x_{ij}) \right\} g_{Z_i}(\hat{Z}_{i,k} + \hat{G}_{i,k}^{1/2} u_i; \xi) du_i.$$

This integral can then be approximated by a weighted  $q$ -dimensional sum such that

$$\begin{aligned} & \int |\hat{G}_{i,k}^{1/2}| \prod_{j=1}^{n_i} \left\{ f_{Y_{ij}|Z_i}(y_{ij}|\hat{Z}_{i,k} + \hat{G}_{i,k}^{1/2} u_i; \beta, \mu, r, x_{ij}) \right\} g_{Z_i}(\hat{Z}_{i,k} + \hat{G}_{i,k}^{1/2} u_i; \xi) \frac{\phi_q(u_i)}{\phi_q(u_i^*)} du_i \\ & \approx \sum_{k_1=1}^T \cdots \sum_{k_q=1}^T |\hat{G}_{i,k}^{1/2}| \prod_{j=1}^{n_i} \left\{ f_{Y_{ij}|Z_i}(y_{ij}|\hat{Z}_{i,k} + \hat{G}_{i,k}^{1/2} u_k^*; \beta, \mu, r, x_{ij}) \right\} g_{Z_i}(\hat{Z}_{i,k} + \hat{G}_{i,k}^{1/2} u_k^*; \xi) \times \\ & \quad \left\{ \phi_q(u_k^*) \right\}^{-1} \prod_{l=1}^q w_{k_l}, \end{aligned}$$

where  $u_k^* = (u_{k_1}, \dots, u_{k_q})^\top$  and  $u_{k_l}$  and  $w_{k_l}$ ,  $k_l = 1, \dots, T$ ,  $l = 1, \dots, q$ , are the abscissas and weights, respectively, for the one-dimensional Gaussain quadrature rule based on the standard normal kernel with  $T$  quadrature points.

Using this approximation to Equation 6, we can then maximize Equation 4 with respect to  $\beta$ ,  $\mu$ ,  $r$ , and  $\xi$  to obtain  $\hat{\beta}^{(k+1)}$ ,  $\hat{\mu}^{(k+1)}$ ,  $\hat{r}^{(k+1)}$ , and  $\hat{\xi}^{(k+1)}$ , respectively. Typically one would iterate, until convergence, between finding the empirical Bayes estimates to re-approximate Equation 6, and then updating  $\beta$ ,  $\mu$ ,  $r$ , and  $\xi$ . In many applications, the number of quadrature points needed for an accurate approximation is small; and, therefore, the dimension of the integral can be large and still be computationally feasible (see [Schilling and Bock 2005](#), who considered numerical integration up to 8 dimensions). We have described (adaptive) Gaussian quadrature in the context of a single-level random effect, but the approach may be extended to multi-level models ([Rabe-Hesketh, Skrondal, and Pickles 2005](#)). Additional efficiency may be gained by using spherical quadrature rules although those are not considered as part of SNP\_NLMM.

However, adaptive quadrature requires  $m$  maximizations at each iteration because the integral in Equation 6 is re-centered and re-scaled about the updated empirical Bayes estimates at each iteration. The maximizations can be too time consuming if  $g_{Z_i}$  is sufficiently complex as is the case for the SNP density. Adaptive importance sampling, another numerical integration technique which is a stochastic version of adaptive Gaussian quadrature, also requires  $m$

maximizations at each iteration and, therefore, suffers from the same limitation (Pinheiro and Bates 1995). A novelty of SNP\_NLMM is the computational approach to approximate Equation 6 when  $g_{Z_i}(\cdot)$  is a flexible and complex random effects density, which we describe below.

Heuristically, we would like to center and scale the integral one time at something reasonably close to  $Z_i$ , the true value of the random effect, to avoid  $m$  maximizations at each iteration. As suggested in Vock *et al.* (2012), we propose to center and scale the quadrature points at the empirical Bayes estimates and their estimated variance from assuming Gaussian random effects. The approach advocated here is similar to the computational approach taken by Rizopoulos (2012) in the context of joint longitudinal-survival models. Rizopoulos (2012) suggests using Gaussian quadrature to approximate the integral over the random effects in the joint model with quadrature points centered and scaled at the empirical Bayes estimates and estimated variance obtained from the model of only the longitudinal data. In our application, we need to be careful as  $Z_i$  is assumed to have mean zero when Gaussian random effects are assumed, but this need not be the case for a flexible random effects density, such as the SNP density. Let  $\hat{b}_{i,G}$  be the empirical Bayes estimate of  $b_i$ , the centered random effect, from assuming Gaussian random effects and  $\hat{G}_{i,G}$  the estimated variance of the empirical Bayes estimates.

Then we have

$$\begin{aligned}\mu + RZ_i &= \hat{b}_{i,G} + \hat{G}_{i,G}^{1/2}U_i \\ Z_i &= R^{-1}(\hat{b}_{i,G} - \mu) + R^{-1}\hat{G}_{i,G}^{1/2}U_i,\end{aligned}\tag{8}$$

which we can substitute into Equation 6 and take the integral with respect to  $U_i$ :

$$\begin{aligned}\mathcal{L}_i(\theta, Y_i) &= \int |R^{-1}\hat{G}_{i,G}^{1/2}| \prod_{j=1}^{n_i} \left[ f_{Y_{ij}|Z_i} \left\{ y_{ij} | R^{-1}(\hat{b}_{i,G} - \mu) + R^{-1}\hat{G}_{i,G}^{1/2}u_i; \beta, \mu, r, x_{ij} \right\} \right] \times \\ &\quad g_{Z_i} \left\{ R^{-1}(\hat{b}_{i,G} - \mu) + R^{-1}\hat{G}_{i,G}^{1/2}u_i; \xi \right\} du_i.\end{aligned}\tag{9}$$

As before we can now use Gaussian quadrature to approximate Equation 9; that is,

$$\begin{aligned}&\int |R^{-1}\hat{G}_{i,G}^{1/2}| \prod_{j=1}^{n_i} \left[ f_{Y_{ij}|Z_i} \left\{ y_{ij} | R^{-1}(\hat{b}_{i,G} - \mu) + R^{-1}\hat{G}_{i,G}^{1/2}u_i; \beta, \mu, r, x_{ij} \right\} \right] \times \\ &\quad g_{Z_i} \left\{ R^{-1}(\hat{b}_{i,G} - \mu) + R^{-1}\hat{G}_{i,G}^{1/2}u_i; \xi \right\} \frac{\phi_q(u_i)}{\phi_q(u_i)} du_i \\ &\approx \sum_{k_1=1}^T \cdots \sum_{k_q=1}^T |R^{-1}\hat{G}_{i,G}^{1/2}| \prod_{j=1}^{n_i} \left\{ f_{Y_{ij}|Z_i} \left( y_{ij} | R^{-1}(\hat{b}_{i,G} - \mu) + R^{-1}\hat{G}_{i,G}^{1/2}u_k^*; \beta, \mu, r, x_{ij} \right) \right\} \times \\ &\quad g_{Z_i} \left( R^{-1}(\hat{b}_{i,G} - \mu) + R^{-1}\hat{G}_{i,G}^{1/2}u_k^*; \xi \right) \{ \phi_q(u_k^*) \}^{-1} \prod_{l=1}^q w_{k_l}.\end{aligned}$$

SNP\_NLMM uses the NLMIXED procedure in SAS for numerical integration and optimization of the likelihood. Because NLMIXED does not allow the user to center the quadrature points at arbitrary values and does not permit non-Gaussian random effects, we must creatively use

programming statements to force NL MIXED to fit the GLMM or NLMM with the SNP density using the integration strategy above. More specifically, if we consider

$$q|R^{-1}\hat{G}_{i,G}^{1/2}|\prod_{j=1}^{n_i}\left[f_{Y_{ij}|Z_i}\left\{y_{ij}|R^{-1}(\hat{b}_{i,G}-\mu)+R^{-1}\hat{G}_{i,G}^{1/2}u_i;\beta,\mu,r,x_{ij}\right\}\right]\times$$

$$g_{Z_i}(R^{-1}(\hat{b}_{i,G}-\mu)+R^{-1}\hat{G}_{i,G}^{1/2}u_k^*;\xi)\{\phi_q(u_i)\}^{-1}$$

to be the conditional “density” of the response given the random effect  $u_i$  (which one can specify using the **general** log likelihood function in the **model** statement of the NL MIXED procedure), specify a standard  $q$ -dimensional normal for the random effects density, and use the **noad** option to avoid adaptive quadrature, then the NL MIXED procedure will fit the desired GLMM or NLMM with the SNP random effects density using our proposed method for numerical integration. These steps use the “likelihood reformulation” technique discussed by Liu and Yu (2008) to allow for non-Gaussian random effects densities. Overall, we have found that the above strategy provides an excellent approximation of the integral while being far less computationally demanding than either non-adaptive or adaptive Gaussian quadrature.

## 5. GLMM example: Toenail infection

As a motivating example, we consider data from a multicenter randomized trial on the effect of two oral treatments on dermatophyte onychomycosis, a toenail infection. Investigators measured the degree of onycholysis, a measure of the separation between the nail plate and nail bed, at baseline and 4, 12, 24, 36, and 48 weeks after starting therapy. A total of 294 patients were enrolled in the study and had a total of 1908 measurements taken. These data have been previously analyzed by DeBacker, Vroey, Lesaffre, Scheys, and Keyser (1998) and Lesaffre and Spiessens (2001).

One question of interest to the investigators was whether there was a difference in the proportion of patients with moderate to severe onycholysis over time between the two treatments. Let  $Y_{ij}$  be the indicator if patient  $i$ ,  $i = 1, \dots, m$ , has moderate or severe onycholysis at the  $j$ th measurement,  $j = 1, \dots, n_i$ . Similarly define  $t_{ij}$  to be the time since subject  $i$  started treatment at  $j$ th measurement and  $a_i$  to be the indicator for whether or not subject  $i$  was on treatment  $A$ .

In keeping with previous analyses, we posit that conditioned on a subject-specific intercept  $b_{1i}$ ,  $Y_{ij}$  are independent and follow a logistic regression model with linear predictor  $b_{1i} + \beta_1 t_{ij} + \beta_2 a_i + \beta_3 t_{ij} a_i$ . That is, we assume

$$Y_{ij}|b_{1i} \stackrel{iid}{\sim} \text{bernoulli}(p_{ij}) \tag{10}$$

$$p_{ij} = \exp(b_{1i} + \beta_1 t_{ij} + \beta_2 a_i + \beta_3 t_{ij} a_i) / \{1 + \exp(b_{1i} + \beta_1 t_{ij} + \beta_2 a_i + \beta_3 t_{ij} a_i)\},$$

where  $b_{1i}$  are independent across  $i$  and can be written as  $b_{1i} = \mu + RZ_i$  for constants  $\mu$  and  $R$ . Previous research has assumed that the density of  $Z_i$  is a standard one-dimensional normal density (Lesaffre and Spiessens 2001); however, we assume that the density of  $Z_i$  can be approximated by the SNP representation with  $K = 0, 1$ , and  $2$  (subsequent analysis with larger values of  $K$ , not shown, did not improve the model fit).

The call to SNP\_NLMM is given as

```
%SNP_NLMM(dataset = toenail, subject = id, response = y,
  prog_stat = %str(eta = b_1 + beta1 * time + beta2 * trt +
    beta3 * trt * time; p = exp(eta) / (1 + exp(eta)));),
  log_cond_dens = %str(y * log(p) + (1 - y) * log(1 - p)),
  start_value= %str(mu_1 = -0.5566 beta1 = -0.1703 beta2 = -0.0006
    beta3 = -0.0672 r_11 = 1),
  q = 1, Kmax = 2, all_K = 1, info_crit = "BIC",
  quad_pt_K0 = 31, quad_pt_K1 = 31, grid_pt_K1 = 9, max_K1 = 2,
  quad_pt_K2 = 31, grid_pt_K2 = 9, max_K2 = 4);
```

In the code,  $t_{ij}$  and  $a_i$  are denoted by the variable names `time` and `trt` with the rest of the variable names self-explanatory. We note that the `prog_stat` command defines the linear predictor, `eta`, and the conditional probability, `p`, of having moderate or severe toenail infection given the subject-specific intercept, `b_1`. The starting values were chosen based on the logistic regression model that does not account for within-person correlation. The inputs `Kmax = 2` and `all_K = 1` indicate that we are considering the SNP density with  $K = 0, 1$ , and  $2$ , and the value of  $K$  will be selected by BIC (`info_crit = "BIC"`). For both  $K = 1$  and  $K = 2$  we use a grid of 9 in each dimension of  $\xi$  in the grid search for starting values (`grid_pt_K1 = 9`, `grid_pt_K2 = 9`). The number of quadrature points, `quad_pt_K0 = 31`, `quad_pt_K1 = 31`, `quad_pt_K2 = 31` were selected by first letting NLMIXED select the number of quadrature points adaptively (that is, `quad_pt_K0 = 0`). The macro was re-run with larger number of quadrature points to ensure that the maximum likelihood estimates were not sensitive to the number of quadrature points.

The fit statistics and parameter estimates from fitting the model implied by Equation 10 are given as part of the standard output of `SNP_NLMM` and are re-printed below. We note that `expect_b1` and `var_b1` refer to  $E(b_1)$  and  $\text{VAR}(b_1)$ , respectively; the remaining parameter names of interest are specified by the user in the `prog_stat` statement, above.

Information Criteria  
By Degree of Flexibility

Criteria	Value_K0	Value_K1	Value_K2
-2 Log Likelihood	1250.8	1250.7	1226.0
AIC (smaller is better)	1260.8	1262.7	1240.0
AICC (smaller is better)	1260.8	1262.8	1240.0
BIC (smaller is better)	1279.2	1284.8	1265.8

Parameter Estimates  
By Degree of Flexibility

Parameter	Est_K0	StdErr_K0	Est_K1	StdErr_K1	Est_K2	StdErr_K2
<code>expect_b1</code>	-1.62	0.43	-1.68	0.52	-2.17	0.68
<code>var_b1</code>	16.05	3.03	16.47	3.68	16.17	6.09
<code>beta1</code>	-0.39	0.044	-0.39	0.045	-0.38	0.045
<code>beta2</code>	-0.16	0.58	-0.17	0.59	0.42	0.46

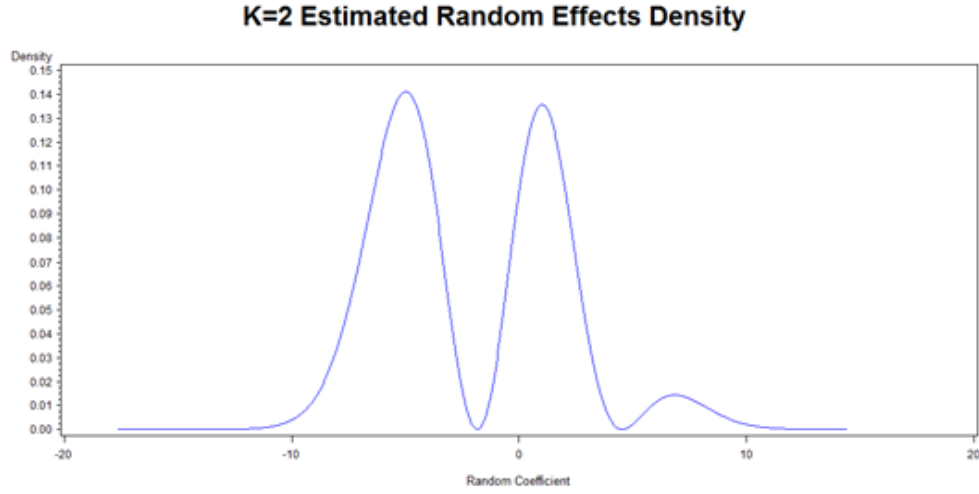


Figure 1: Estimated density of the random intercept from Equation 10 with  $K = 2$  concerning the toenail infection study.

beta3	-0.14	0.068	-0.14	0.068	-0.13	0.070
mu_1	-1.62	0.43	0.70	4.53	0.17	0.46
r_11	4.01	0.38	4.36	1.36	2.46	0.46
xi_1	.	.	-1.28	0.56	-0.58	0.062
xi_2	.	.	.	.	-0.70	0.062

The parameter estimates are similar across all values of  $K$ , although the estimated mean of the random intercept is approximately one standard error smaller for  $K = 2$ . The estimate for  $\beta_2$  is also considerably larger for  $K = 2$  but remains non-significant. Of interest to the investigators was the value of  $\beta_3$ , the interaction between treatment and time, which remained boarder-line significant across each of the three models.

Each of the information criteria prefer  $K = 2$  which, based on the plot of the estimated density given in Figure 1, indicates a strong departure from normality and the presence of two groups. One group of subjects with the mode of  $b_{1i}$  at approximately  $-5$  has a very low probability of moderate to severe onycholysis across all time points. The predicted probability of having moderate or severe onycholysis at each of the scheduled visit times is less than 1.1 percent for both treatment groups. This mode suggests that those who enter the trial without moderate to severe infection have very little probability of transitioning to a more infected state. However, those in the second group with the mode at approximately 1 experience a clear progression through the disease. The estimated probability of having significant onycholysis at weeks 0, 4, 12, 24, 36, and 48 for a patient taking treatment  $A$  with  $b_{1i} = 1$  is 0.81, 0.71, 0.47, 0.16, 0.04, and 0.01, respectively.

Fitting a flexible random effects model not only alters estimates of parameters such as  $E(b_{1i})$ , but perhaps more importantly improves our understanding of the underlying data generating process and disease progression.

## 6. NLMM example: Pharmacokinetics of argatroban

As an example of a NLMM, we consider a pharmacokinetic study of argatroban, an anti-coagulant, reported in [Davidian and Giltinan \(1995, Chapter 9.5\)](#) and introduced in Section 2. In this study, 36 patients received a four-hour intravenous infusion of argatroban at one of nine infusion rates ranging from 1 to 5  $\mu\text{g/kg/min}$  in increments of 0.5  $\mu\text{g/kg/min}$  (four patients at each infusion rate). A 37th patient had an infusion rate of 4.38  $\mu\text{g/kg/min}$ . The intravenous concentration of argatroban was recorded at 30, 60, 90, 115, 160, 200, 240, 245, 250, 255, 260, 275, 295, 320, and 360 minutes after starting infusion. Let  $Y_{ij}$  and  $t_{ij}$  be defined as in Section 2. The concentration profile is assumed to follow the model given by Equation 2. Let  $\log(Cl_i) = b_{1i}$ ,  $\log(V_i) = b_{2i}$ , and  $b_i = (b_{1i}, b_{2i})^\top = \mu + RZ_i$  where  $Z_i = (Z_{1i}, Z_{2i})^\top$ ,  $\mu = (\mu_1, \mu_2)^\top$ , and  $R$  is a  $2 \times 2$  lower triangular matrix. We assume that the density of  $Z_i$  can be represented by the density given in Equation 5 for  $K = 0, 1$ , and 2.

The call to `SNP_NLMM` is given by

```
%SNP_NLMM(dataset = arg, subject = indiv, response = conc,
  prog_stat = %str(c1 = exp(b_1); v = exp(b_2);
    pred = (dose / c1) * (exp(- c1 * (1 - t1) * (time - tinf) / v) -
      exp(- c1 * (time) / v)); var = (sig ** 2) * (pred ** (2 * gamma)));),
  log_cond_dens = %str(log(1 / sqrt(2 * pi * var) *
    exp(- 1 / (2 * var) * (conc - pred) ** 2))),
  start_value = %str(mu_1 = - 5.5 mu_2 = - 2.0 r_11 = 0.374 r_21 = 0.016
    r_22 = 0.076 sig = 23.0 gamma = 0.22),
  q = 2, Kmax = 2, all_K = 1, info_crit = "BIC", quad_pt_K0 = 11,
  quad_pt_K1 = 9, grid_pt_K1 = 9, max_K1 = 5,
  quad_pt_K2 = 9, grid_pt_K2 = 7, max_K2 = 9);
```

In the code above, `dose`, `c1`, `v`, `tinf`, `t1`, `time`, and `conc` refer to  $D_i$ ,  $Cl_i$ ,  $V_i$ ,  $t_{i,\text{inf}}$ ,  $t_{ij}^*$ ,  $t_{ij}$ , and  $Y_{ij}$ , where those quantities were defined in Section 2. The `prog_stat` specifies the mean and variance of the conditional normal density while `log_cond_dens` is just the logarithm of the normal density. The starting values were chosen from fitting a nonlinear model of the concentration profile with no random effects. As in the preceding example, we consider models with  $K = 0, 1$  and 2 (`Kmax = 2`, `all_K = 1`). Following our recommendation given in Section 3, we used a grid of 9 and 7 points in each dimension of  $\xi$  for  $K = 1$  and  $K = 2$ , respectively. The number of quadrature points (`quad_pt_K0 = 11`, `quad_pt_K1 = 9`, and `quad_pt_K2 = 9`) were selected in a similar way as in the example in Section 5.

The fit criteria and parameter estimates from fitting the model implied by Equation 2 assuming  $K = 0, 1$ , and 2 are given as part of the standard output of the macro and are shown below. We should note that with only 37 patients it is difficult to make any definitive statements on the joint density of  $\log(Cl_i)$  and  $\log(V_i)$ . Nonetheless, all the information criteria prefer  $K = 1$  which, based on the bivariate density plot given in Figure 2 and as part of the standard output of the macro, shows strong departure from normality with a group of patients having a much lower log clearance than the others. Fixed effect parameter estimates are nearly identical across all  $K$  considered although the standard errors are larger for  $K = 2$  indicating that we may be overfitting the data.



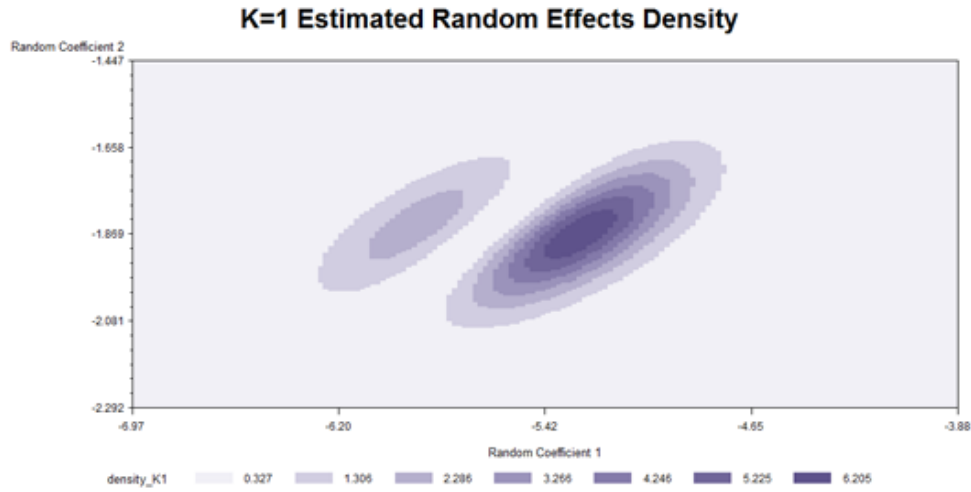


Figure 2: Estimated density of the log clearance and volume parameters from Equation 2 with  $K = 1$  concerning the PK/PD study of argatroban.

Information Criteria  
By Degree of Flexibility

Criteria	Value_K0	Value_K1	Value_K2
-2 Log Likelihood	5712.8	5702.5	5698.8
AIC (smaller is better)	5726.8	5720.5	5722.8
AICC (smaller is better)	5727.0	5720.9	5723.4
BIC (smaller is better)	5738.0	5735.0	5742.1

Parameter Estimates  
By Degree of Flexibility

Parameter	Est_K0	StdErr_K0	Est_K1	StdErr_K1	Est_K2	StdErr_K2
cov_b1b2	0.014	0.012	0.018	0.013	0.017	0.013
expect_b1	-5.42	0.065	-5.42	0.064	-5.43	0.064
expect_b2	-1.87	0.038	-1.86	0.039	-1.86	0.038
var_b1	0.15	0.036	0.15	0.034	0.15	0.035
var_b2	0.011	0.0075	0.013	0.0078	0.017	0.0089
mu_1	-5.42	0.065	-5.56	0.059	-5.53	0.14
mu_2	-1.87	0.038	-1.86	0.035	-1.75	0.064
r_11	0.39	0.046	0.29	0.032	0.29	0.031
r_21	0.035	0.030	0.070	0.028	0.070	0.031
r_22	0.10	0.035	0.087	0.028	0.11	0.027
sig	11.21	3.79	11.72	3.65	11.21	3.43
gamma	0.34	0.056	0.33	0.052	0.34	0.051

xi_1	.	.	-0.34	0.13	0.070	0.22
xi_2	.	.	-0.82	0.20	0.62	0.32
xi_3	.	.	.	.	-0.56	0.19
xi_4	.	.	.	.	-0.12	0.44
xi_5	.	.	.	.	-0.59	0.27

## 7. Conclusion

For many applications of GLMMs and NLMMs, there is little subject area knowledge that would justify assuming *a priori* that the random effects are Gaussian. Given that fixed effect parameter estimators will be inconsistent, biased, and inefficient if the random effects density is misspecified, the one advantage to assuming Gaussian random effects density was computational ease. However, advances in computation and the `SNP_NLMM` macro allow flexible random effect densities to be easily and quickly implemented. The total computation time for all models in the toenail infection example is less than 30 seconds. When  $q = 2$  as in the PK/PD study of argatroban, the grid search for  $K = 2$  can be somewhat time consuming depending on the fineness of the grid and computing speed. But optimization takes less than a couple of minutes for all three models. In addition to improved inference of parameter estimators, one gains an improved understanding of the data generating process which can be crucial for scientific understanding. We hope that the `SNP_NLMM` macro allows flexible random effect densities to be fit routinely for GLMMs and NLMMs.

## References

- Agresti A, Caffo B, Ohman-Strickland P (2004). “Examples in Which Misspecification of a Random Effects Distribution Reduces Efficiency, and Possible Remedies.” *Computational Statistics & Data Analysis*, **47**(3), 639–653.
- Aitkin M (1999). “A General Maximum Likelihood Analysis of Variance Components in Generalized Linear Models.” *Biometrics*, **55**(1), 117–128.
- Bandyopadhyay D, Lachos VH, Castro LM, Dey DK (2012). “Skew-Normal/Independent Linear Mixed Models for Censored Responses with Applications to HIV Viral Loads.” *Biometrical Journal*, **54**(3), 405–425.
- Beal SL, Sheiner LB (1988). “Heteroscedastic Nonlinear-Regression.” *Technometrics*, **30**(3), 327–338.
- Bock RD, Aitkin M (1981). “Marginal Maximum-Likelihood Estimation of Item Parameters: Application of an EM Algorithm.” *Psychometrika*, **46**(4), 443–459.
- Chen J, Zhang D, Davidian M (2002). “A Monte Carlo EM Algorithm for Generalized Linear Mixed Models with Flexible Random Effects Distribution.” *Biostatistics*, **3**(3), 347–360.
- Davidian M, Gallant AR (1993). “The Nonlinear Mixed Effects Model with a Smooth Random Effects Density.” *Biometrika*, **80**(3), 475–488.

- Davidian M, Giltinan DM (1995). *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall, New York.
- DeBacker M, Vroey CD, Lesaffre E, Scheys I, Keyser PD (1998). "Twelve Weeks of Continuous Oral Therapy for Toenail Onychomycosis Caused by Dermatophytes: A Double-Blind Comparative Trial of Terbinafine 250 mg/day Versus Itraconazole 200 mg/day." *Journal of the American Academy of Dermatology*, **38**(5), S57–S63.
- Doehler K, Davidian M (2008). "Smooth Inference for Survival Functions with Arbitrarily Censored Data." *Statistics in Medicine*, **27**(26), 5421–5439.
- Follmann DA, Lambert D (1989). "Generalizing Logistic-Regression by Nonparametric Mixing." *Journal of the American Statistical Association*, **84**(405), 295–300.
- Gallant AR, Nychka DW (1987). "Semi-Nonparametric Maximum Likelihood Estimation." *Econometrica*, **55**(2), 363–390.
- Ghidey W, Lesaffre E, Eilers P (2004). "Smooth Random Effects Distribution in Linear Mixed Model." *Biometrics*, **60**(4), 945–953.
- Ghosh P, Branco MD, Chakraborty H (2007). "Bivariate Random Effect Model Using Skew-Normal Distribution with Application to HIV-RNA." *Statistics in Medicine*, **26**(6), 1255–1267.
- Hartford A, Davidian M (2000). "Consequences of Misspecifying Assumptions in Nonlinear Mixed Effects Models." *Computational Statistics & Data Analysis*, **34**(2), 139–164.
- Heagerty PJ, Kurland BF (2001). "Misspecified Maximum Likelihood Estimates and Generalised Linear Mixed Models." *Biometrika*, **88**(4), 973–985.
- Ho HJ, Lin TI (2010). "Robust Linear Mixed Models Using the Skew t Distribution with Application to Schizophrenia Data." *Biometrical Journal*, **52**(4), 449–469.
- Huang Y, Dagne G (2010). "Skew-Normal Bayesian Nonlinear Mixed-Effects Models with Application to AIDS Studies." *Statistics in Medicine*, **29**(23), 2384–2398.
- Irincheeva I, Cantoni E, Genton MG (2012). "Generalized Linear Latent Variable Models with Flexible Distribution of Latent Variables." *Scandinavian Journal of Statistics*, **39**(4), 663–680.
- Knott M, Tzamourani P (2007). "Bootstrapping the Estimated Latent Distribution of the Two-Parameter Latent Trait Model." *British Journal of Mathematical & Statistical Psychology*, **60**(1), 175–191.
- Labra FV, Garay AM, Lachos VH, Ortega EMM (2012). "Estimation and Diagnostics for Heteroscedastic Nonlinear Regression Models Based on Scale Mixtures of Skew-Normal Distributions." *Journal of Statistical Planning and Inference*, **142**(7), 2149–2165.
- Lachos VH, Bandyopadhyay D, Dey DK (2011). "Linear and Nonlinear Mixed-Effects Models for Censored HIV Viral Loads Using Normal/Independent Distributions." *Biometrics*, **67**(4), 1594–1604.

- Lachos VH, Dey DK, Cancho VG (2009). “Robust Linear Mixed Models with Skew-Normal Independent Distributions from a Bayesian Perspective.” *Journal of Statistical Planning and Inference*, **139**(12), 4098–4110.
- Lachos VH, Ghosh P, Arellano-Valle RB (2010). “Likelihood Based Inference for Skew-Normal Independent Linear Mixed Models.” *Statistica Sinica*, **20**(1), 303–322.
- Laird N (1978). “Nonparametric Maximum Likelihood Estimation of a Mixing Distribution.” *Journal of the American Statistical Association*, **73**(364), 805–811.
- Lee KJ, Thompson SG (2008). “Flexible Parametric Models for Random-Effects Distributions.” *Statistics in Medicine*, **27**(3), 418–434.
- Lesaffre E, Spiessens B (2001). “On the Effect of the Number of Quadrature Points in a Logistic Random-Effects Model: An Example.” *Journal of the Royal Statistical Society C*, **50**(3), 325–335.
- Lesperance ML, Kalbfleisch JD (1992). “An Algorithm for Computing the Nonparametric MLE of a Mixing Distribution.” *Journal of the American Statistical Association*, **87**(417), 120–126.
- Lin TI, Lee JC (2006). “A Robust Approach to  $t$  Linear Mixed Models Applied to Multiple Sclerosis Data.” *Statistics in Medicine*, **25**(8), 1397–1412.
- Lin TI, Lee JC (2007). “Bayesian Analysis of Hierarchical Linear Mixed Modeling Using the Multivariate  $t$  Distribution.” *Journal of Statistical Planning and Inference*, **137**(2), 484–495.
- Lin TI, Lee JC (2008). “Estimation and Prediction in Linear Mixed Models with Skew-Normal Random Effects for Longitudinal Data.” *Statistics in Medicine*, **27**(9), 1490–1507.
- Lindsay BG (1983). “The Geometry of Mixture Likelihoods: A General Theory.” *The Annals of Statistics*, **11**(1), 86–94.
- Litière S, Alonso A, Molenberghs G (2008). “The Impact of a Misspecified Random-Effects Distribution on the Estimation and the Performance of Inferential Procedures in Generalized Linear Mixed Models.” *Statistics in Medicine*, **27**(16), 3125–3144.
- Liu L, Yu Z (2008). “A Likelihood Reformulation Method in Non-Normal Random Effects Models.” *Statistics in Medicine*, **27**(16), 3105–3124.
- Liu Q, Pierce DA (1994). “A Note on Gauss-Hermite Quadrature.” *Biometrika*, **81**(3), 624–629.
- Ma Y, Genton MG (2010). “Explicit Estimating Equations for Semiparametric Generalized Linear Latent Variable Models.” *Journal of the Royal Statistical Society B*, **72**(4), 475–495.
- Magder LS, Zeger SL (1996). “A Smooth Nonparametric Estimate of a Mixing Distribution Using Mixtures of Gaussians.” *Journal of the American Statistical Association*, **91**(435), 1141–1151.
- Neuhaus JM, Hauck WW, Kalbfleisch JD (1992). “The Effects of Mixture Distribution Misspecification when Fitting Mixed-Effects Logistic Models.” *Biometrika*, **79**(4), 755–762.

- Pinheiro JC, Bates DM (1995). “Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model.” *Journal of Computational and Graphical Statistics*, **4**(1), 12–35.
- Pinheiro JC, Liu CH, Wu YN (2001). “Efficient Algorithms for Robust Estimation in Linear Mixed-Effects Models Using the Multivariate  $t$  Distribution.” *Journal of Computational and Graphical Statistics*, **10**(2), 249–276.
- Rabe-Hesketh S, Skrondal A, Pickles A (2005). “Maximum Likelihood Estimation of Limited and Discrete Dependent Variable Models with Nested Random Effects.” *Journal of Econometrics*, **128**(2), 301–323.
- Rizopoulos D (2012). “Fast Fitting of Joint Models for Longitudinal and Event Time Data Using a Pseudo-Adaptive Gaussian Quadrature Rule.” *Computational Statistics & Data Analysis*, **56**(3), 491–501.
- Rosa GJM, Padovani CR, Gianola D (2003). “Robust Linear Mixed Models with Normal/Independent Distributions and Bayesian MCMC Implementation.” *Biometrical Journal*, **45**(5), 573–590.
- Sartori N, Severini TA (2004). “Conditional Likelihood Inference in Generalized Linear Mixed Models.” *Statistica Sinica*, **14**(2), 349–360.
- SAS Institute Inc (2008). *SAS/STAT Software, Version 9.2*. Cary, NC. URL <http://www.sas.com/>.
- Schilling S, Bock R (2005). “High-Dimensional Maximum Marginal Likelihood Item Factor Analysis by Adaptive Quadrature.” *Psychometrika*, **70**(3), 533–555.
- Song X, Davidian M, Tsiatis AA (2002). “A Semiparametric Likelihood Approach to Joint Modeling of Longitudinal and Time-to-Event Data.” *Biometrics*, **58**(4), 742–753.
- Tao H, Palta M, Yandell BS, Newton MA (1999). “An Estimation Method for the Semiparametric Mixed Effects Model.” *Biometrics*, **55**(1), 102–110.
- Tsiatis AA (2006). *Semiparametric Theory and Missing Data*. Springer-Verlag, New York.
- Verbeke G, Lesaffre E (1997). “The Effect of Misspecifying the Random-Effects Distribution in Linear Mixed Models for Longitudinal Data.” *Computational Statistics & Data Analysis*, **23**(4), 541–556.
- Vock DM, Davidian M, Tsiatis, A A, Muir AJ (2012). “Mixed Model Analysis of Censored Longitudinal Data with Flexible Random-Effects Density.” *Biostatistics*, **13**(1), 61–73.
- Zhang D, Davidian M (2001). “Linear Mixed Models with Flexible Distributions of Random Effects for Longitudinal Data.” *Biometrics*, **57**(3), 795–802.
- Zhang M, Davidian M (2008). ““Smooth” Semiparametric Regression Analysis for Arbitrarily Censored Time-to-Event Data.” *Biometrics*, **64**(2), 567–576.

**Affiliation:**

David M. Vock  
University of Minnesota School of Public Health  
Division of Biostatistics  
A466 Mayo Building, MMC 303  
420 Delaware Street S.E.  
Minneapolis MN, 55455, United States of America  
E-mail: [vock@umn.edu](mailto:vock@umn.edu)