

Introduction

Linear mixed models (LMMs) are widely used for analyzing repeated measures data, as they flexibly handle within-subject correlations. Typically, both random effects and errors are assumed to have a normal distribution for simplicity. However, these assumptions can reduce robustness, especially with heavy tails or skewness. Additionally, errors are often assumed to be conditionally independent, which may not hold in longitudinal data, where serial correlation is common. To address this, Schumacher et al. [1] proposed an EM-based approach for fitting scale mixture of skew-normal LMMs (SMSN-LMMs) with serially correlated errors, incorporating structures like AR(p) and DEC. Their methods are implemented in the R package *skewlmm*, available on CRAN [2]. This work introduces the model and highlights the features of the *skewlmm* package, which provides tools for robust analysis of longitudinal data.

Motivation

To illustrate the need for a more flexible modeling approach, we consider a real dataset on reaction times from a sleep deprivation study, available in the R package *lme4*. In this study, the average daily reaction times of subjects were recorded. On day 0, subjects had their normal amount of sleep, after which they were restricted to just 3 hours of sleep per night for the next nine days. Reaction times were measured each day based on a series of tests.

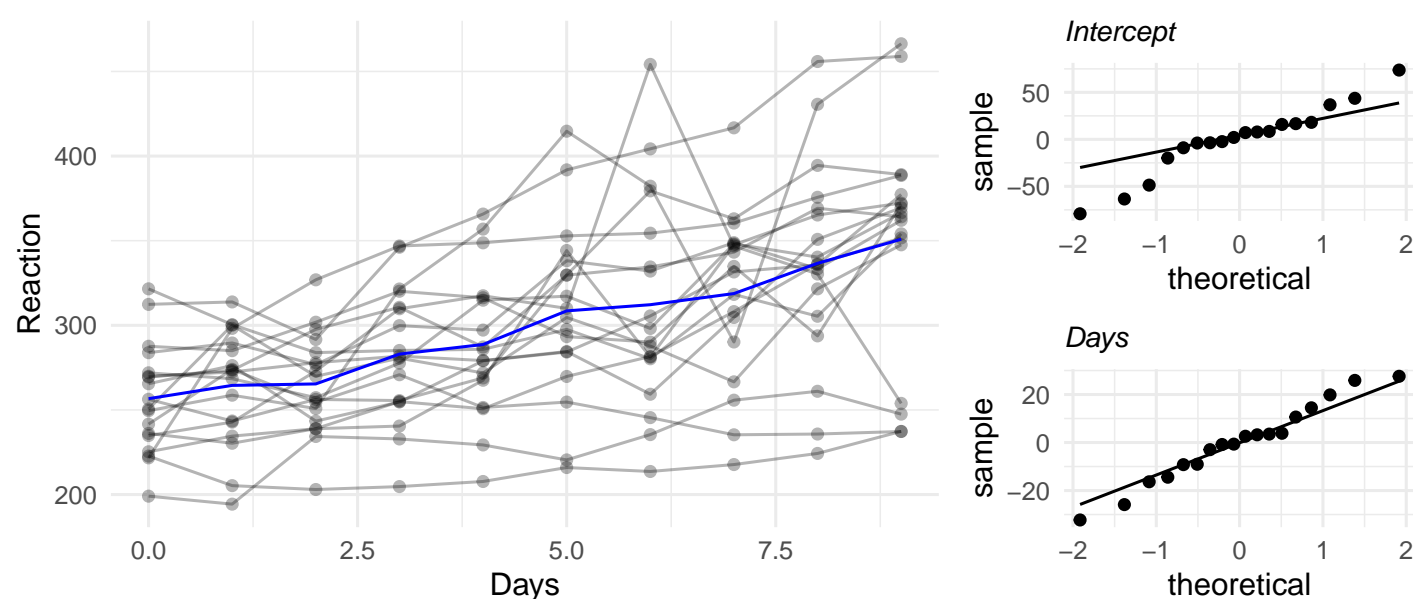


Figure 1: Sleep Deprivation Study.

Figure 1 displays the individual reaction time trajectories over time along with the overall mean profile. It also shows the empirical Bayes estimates of random effects obtained by fitting a normal LMM. The quantile plots reveal that the normality assumption is questionable, indicating the presence of heavy tails and suggesting that a more robust or flexible model may be necessary.

Model Formulation

For the i th subject, $i = 1, \dots, n$, let \mathbf{Y}_i be a $n_i \times 1$ vector of observed continuous responses. The SMSN-LMM can be defined by considering

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad (1)$$

$$\text{and } \begin{pmatrix} \mathbf{b}_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \stackrel{\text{ind.}}{\sim} \text{SMSN}_{q+n_i} \left(\begin{pmatrix} c\boldsymbol{\Delta} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_i \end{pmatrix}, \begin{pmatrix} \boldsymbol{\lambda} \\ \mathbf{0} \end{pmatrix}; H \right), \quad (2)$$

where \mathbf{X}_i of dimension $n_i \times l$ is the design matrix corresponding to the fixed effects, $\boldsymbol{\beta}$ of dimension $l \times 1$ is a vector of fixed effects, \mathbf{Z}_i of dimension $n_i \times q$ is the design matrix corresponding to the $q \times 1$ random effects vector \mathbf{b}_i , $\boldsymbol{\epsilon}_i$ of dimension $n_i \times 1$ is the vector of random errors, $c = c(\boldsymbol{\nu}) = -\sqrt{2/\pi}k_1$, with $k_1 = E\{U^{-1/2}\}$, $\boldsymbol{\Delta} = \mathbf{D}^{1/2}\boldsymbol{\delta}$, and $\boldsymbol{\delta} = \boldsymbol{\lambda}/\sqrt{1 + \boldsymbol{\lambda}^\top \boldsymbol{\lambda}}$.

The $q \times q$ random effects scale matrix \mathbf{D} can be unstructured or diagonal, and we consider the $n_i \times n_i$ error scale matrix as $\boldsymbol{\Sigma}_i = \sigma_e^2 \mathbf{R}_i$, with $\mathbf{R}_i = \mathbf{R}_i(\boldsymbol{\phi})$, $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)^\top$, being one of the following: a structured matrix, such as the covariance matrix of an AR(p) processes, a DEC structure, or an identity matrix for conditionally uncorrelated (UNC) processes.

The SMSN-LMM has a convenient hierarchical representation, which is useful for the implementation of an EM-type algorithm, and therefore the package *skewlmm* uses the ECME algorithm for parameter estimation through the functions `smsn.lmm` and `smn.lmm`, where the latter refers to the special case of $\boldsymbol{\lambda} = \mathbf{0}$. An introduction to the package and its use in fitting an SMSN-LMM to the sleep study data is provided next.

The R package *skewlmm*

The package *skewlmm* provides tools for fitting and evaluating the SMSN-LMM given in (1)-(2) in R using S3 class, with a user-friendly interface. The basic syntax of the main functions is as follows:

```
smsn.lmm(data, formFixed, groupVar, formRandom, depStruct, distr, covRandom, ...)
smn.lmm(data, formFixed, groupVar, formRandom, depStruct, distr, covRandom, ...)
```

The functions return an object of the class SMSN and SMN, respectively, containing a list of elements. Additionally, some estimation options can be controlled using the argument `control` with the function `lmmControl`.

For example, an SL-LMM and an SSL-LMM, respectively, can be fitted as shown below. Additionally, a likelihood ratio test for testing $H_0 : \boldsymbol{\lambda} = \mathbf{0}$ can be performed using the `lr.test` function.

```
fit1 <- smn.lmm(data = sleepstudy,
  formFixed = Reaction~Dayst,
  distr = 'sl',
  formRandom = ~Dayst,
  groupVar = "Subject",
  control = lmmControl(quiet = TRUE))

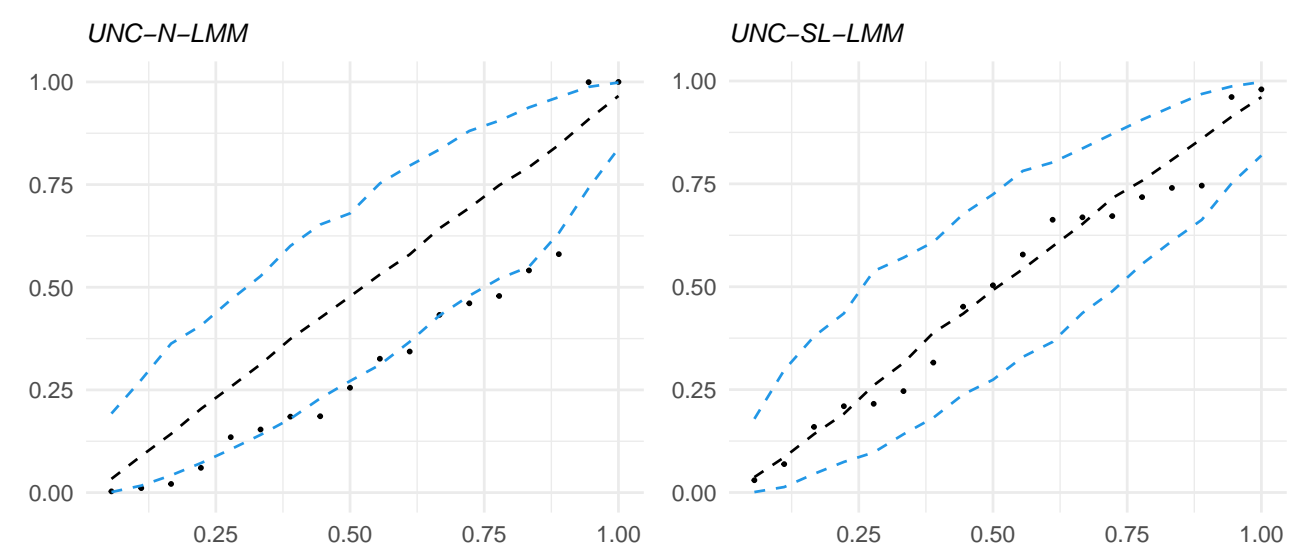
fitskew1 <- smsn.lmm(data = sleepstudy,
  formFixed = Reaction~Dayst,
  distr = 'ssl',
  formRandom = ~Dayst,
  groupVar = "Subject",
  control = lmmControl(quiet = TRUE))

lr.test(fit1, fitskew1)
```

```
## Model selection criteria:
##          logLik    AIC    BIC
## fit1      -861.088 1736.175 1758.526
## fitskew1 -860.745 1739.490 1768.226
##
## Likelihood-ratio Test
## chi-square statistics = 0.6854122
## df = 2
## p-value = 0.7098468
## The null hypothesis that both models represent the
## data equally well is not rejected at level 0.05
```

To evaluate the adequacy of the distributional assumption, a Healy-type plot can be used as illustrated next, where the gain in considering a heavy-tailed distribution for modeling this dataset is evidenced.

```
grid.arrange(healy.plot(fit0, calcCI = TRUE), healy.plot(fit1, calcCI = TRUE), nrow=1)
```



Furthermore, to verify if the uncorrelation assumption, used by default, is appropriate, a possible approach is to use the update function to refit the model considering different correlation structures and then compare AIC and BIC to select the most appropriate model. Since the data are equally spaced and sorted by time, the use of `timeVar` in here is optional (the function will use the position if `timeVar` is not provided).

```
fitar1 <- update(fit1, depStruct = "ARp", pAR = 1)
fitar2 <- update(fit1, depStruct = "ARp", pAR = 2)
fitDEC <- update(fit1, depStruct = "DEC",
  timeVar = "Days")
```

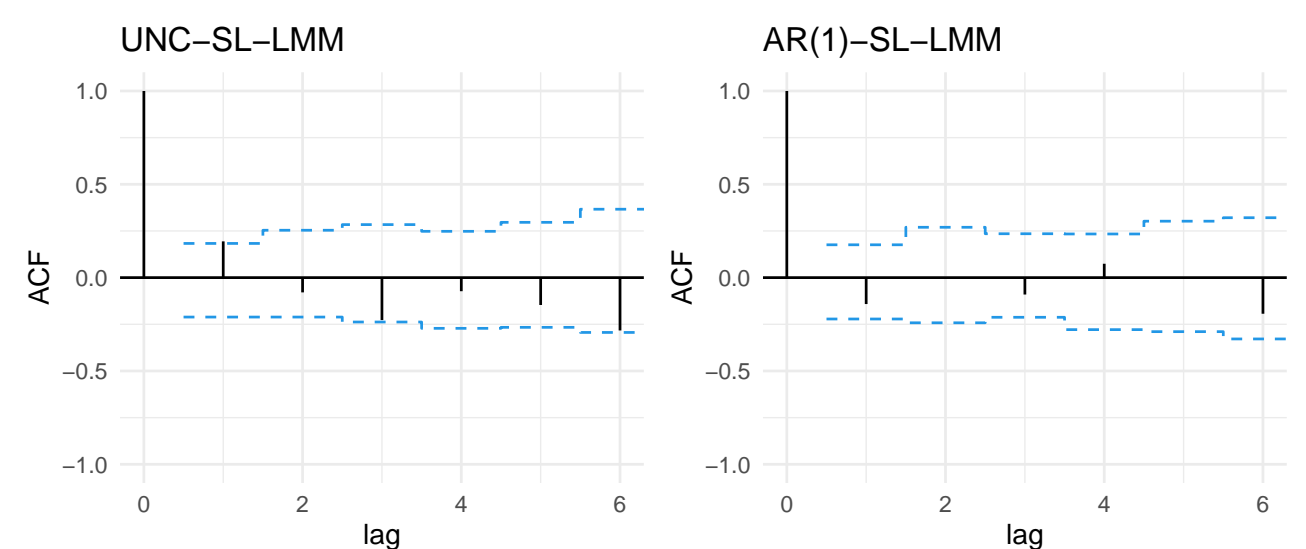
depStruct	AIC	BIC
UNC	1736.2	1758.5
AR(1)	1716.8	1742.3
AR(2)	1717.3	1746.0
DEC	1718.6	1747.3

Additionally, we can compute the empirical autocorrelation function (ACF) for standardized marginal residuals, which at lag l can be defined as

$$\hat{\rho}(l) = \frac{\sum_{i=1}^n \sum_{j=1}^{n_i} \{ (j,k) | t_k - t_j = l \} r_{it_j} r_{it_k} / N(l)}{\sum_{i=1}^n \sum_{j=1}^{n_i} r_{it_j}^2 / N(0)},$$

where $\mathbf{r}_i = \hat{\boldsymbol{\Gamma}}_i^{-1/2} (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})$ is the standardized marginal residual vector for subject i , with $\boldsymbol{\Gamma}_i = \text{Var}(\mathbf{Y}_i)$, and $N(\cdot)$ is the number of pairs used in the respective numerator summation.

```
grid.arrange(plot(acfresid(fit1, calcCI = TRUE, maxLag = 6)) + ggtitle("UNC-SL-LMM"),
  plot(acfresid(fitar1, calcCI = TRUE, maxLag = 6)) + ggtitle("AR(1)-SL-LMM"),
  nrow = 1)
```



Finally, methods such as `print`, `summary`, `plot`, `fitted`, `residuals` and `predict` are implemented.

Referências

- [1] F. L. Schumacher, V. H. Lachos, and L. A. Matos. Scale mixture of skew-normal linear mixed models with within-subject serial dependence. *Statistics in Medicine*, 40(7):1790–1810, 2021.
- [2] F. L. Schumacher, L. A. Matos, and V. H. Lachos. *skewlmm: Scale Mixture of Skew-Normal Linear Mixed Models*, 2021.



Agradecimentos/Financiamento

This study was partially financed by the São Paulo Research Foundation (FAPESP), Brazil. Project Number: 2020/16713-0.