Supplementary material for ‘A family of linear mixed-effects models using the generalized Laplace distribution’

Marco Geraci1*    Alessio Farcomeni2

1Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina
2Department of Economics and Finance, University of Rome “Tor Vergata”

Abstract

This report contains supplemental material for the paper entitled ‘A family of linear mixed-effects models using the generalized Laplace distribution’, hereinafter referred to as the ‘Manuscript’. Section 1 reports supplementary tables and figures from the simulation study, and from the real data analyses. Section 2 contains a brief tutorial using R code.

1 Supplementary tables and figures

Table 1: Mean bias of the estimator for the generalized Laplace mixed-effects model with data generated under the Normal-Normal and Laplace-Laplace scenarios.

<table>
<thead>
<tr>
<th>Sample size (n, M)</th>
<th>β0</th>
<th>β1</th>
<th>β2</th>
<th>θ11</th>
<th>θ12</th>
<th>θ22</th>
<th>ψ2</th>
<th>α1</th>
<th>α2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal-Normal scenario</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5, 20)</td>
<td>-0.010</td>
<td>-0.003</td>
<td>0.030</td>
<td>-0.100</td>
<td>-0.004</td>
<td>-0.033</td>
<td>-0.002</td>
<td>0.144</td>
<td>0.076</td>
</tr>
<tr>
<td>(10, 20)</td>
<td>0.007</td>
<td>-0.007</td>
<td>0.003</td>
<td>-0.067</td>
<td>-0.009</td>
<td>-0.019</td>
<td>-0.007</td>
<td>0.154</td>
<td>0.026</td>
</tr>
<tr>
<td>(5, 100)</td>
<td>0.012</td>
<td>0.006</td>
<td>0.003</td>
<td>-0.003</td>
<td>0.013</td>
<td>0.012</td>
<td>0.007</td>
<td>0.055</td>
<td>0.021</td>
</tr>
<tr>
<td>(10, 100)</td>
<td>0.008</td>
<td>0.004</td>
<td>-0.001</td>
<td>-0.019</td>
<td>0.003</td>
<td>0.019</td>
<td>0.000</td>
<td>0.051</td>
<td>0.016</td>
</tr>
<tr>
<td>Laplace-Laplace scenario</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5, 20)</td>
<td>-0.001</td>
<td>-0.000</td>
<td>-0.012</td>
<td>-0.232</td>
<td>-0.105</td>
<td>-0.105</td>
<td>-0.032</td>
<td>-0.430</td>
<td>-0.190</td>
</tr>
<tr>
<td>(10, 20)</td>
<td>0.004</td>
<td>-0.003</td>
<td>0.001</td>
<td>-0.213</td>
<td>-0.091</td>
<td>-0.077</td>
<td>-0.121</td>
<td>-0.339</td>
<td>-0.150</td>
</tr>
<tr>
<td>(5, 100)</td>
<td>-0.001</td>
<td>0.004</td>
<td>0.003</td>
<td>-0.029</td>
<td>-0.021</td>
<td>-0.016</td>
<td>-0.054</td>
<td>-0.145</td>
<td>-0.092</td>
</tr>
<tr>
<td>(10, 100)</td>
<td>-0.008</td>
<td>0.004</td>
<td>-0.003</td>
<td>-0.038</td>
<td>-0.023</td>
<td>-0.026</td>
<td>-0.160</td>
<td>-0.128</td>
<td>-0.127</td>
</tr>
</tbody>
</table>

*Address for correspondence: Marco Geraci, Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, 915 Greene Street, Columbia, SC 29208, USA. E-mail: geraci@mailbox.sc.edu
Table 2: Mean bias of the estimator for the generalized Laplace mixed-effects (GLME) model, linear mixed-effects (LME) model, and heavy-tailed LME (tLME) model with data generated under the Laplace-Laplace and t-t scenarios.

<table>
<thead>
<tr>
<th>Sample size $(n, M)$</th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\varsigma_{11}$</th>
<th>$\varsigma_{12}$</th>
<th>$\varsigma_{22}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Laplace-Laplace scenario</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLME</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5, 20)</td>
<td>$-0.001$</td>
<td>$-0.000$</td>
<td>$-0.012$</td>
<td>$-0.232$</td>
<td>$-0.105$</td>
<td>$-0.105$</td>
</tr>
<tr>
<td>(10, 20)</td>
<td>$0.004$</td>
<td>$-0.003$</td>
<td>$0.001$</td>
<td>$-0.213$</td>
<td>$-0.091$</td>
<td>$-0.077$</td>
</tr>
<tr>
<td>(5, 100)</td>
<td>$-0.001$</td>
<td>$0.004$</td>
<td>$0.003$</td>
<td>$-0.029$</td>
<td>$-0.021$</td>
<td>$-0.016$</td>
</tr>
<tr>
<td>(10, 100)</td>
<td>$-0.008$</td>
<td>$0.004$</td>
<td>$-0.003$</td>
<td>$-0.038$</td>
<td>$-0.023$</td>
<td>$-0.026$</td>
</tr>
<tr>
<td>LME</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5, 20)</td>
<td>$-0.017$</td>
<td>$-0.007$</td>
<td>$0.009$</td>
<td>$3.621$</td>
<td>$-5.743$</td>
<td>$7.204$</td>
</tr>
<tr>
<td>(10, 20)</td>
<td>$0.001$</td>
<td>$-0.007$</td>
<td>$0.002$</td>
<td>$1.475$</td>
<td>$-3.406$</td>
<td>$4.606$</td>
</tr>
<tr>
<td>(5, 100)</td>
<td>$-0.003$</td>
<td>$-0.001$</td>
<td>$0.002$</td>
<td>$0.473$</td>
<td>$-2.134$</td>
<td>$2.802$</td>
</tr>
<tr>
<td>(10, 100)</td>
<td>$-0.004$</td>
<td>$0.003$</td>
<td>$-0.006$</td>
<td>$0.353$</td>
<td>$-1.998$</td>
<td>$2.623$</td>
</tr>
<tr>
<td>tLME</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5, 20)</td>
<td>$1.444$</td>
<td>$-0.992$</td>
<td>$-0.007$</td>
<td>$8.19 \times 10^4$</td>
<td>$7.98 \times 10^3$</td>
<td>$2.75 \times 10^3$</td>
</tr>
<tr>
<td>(10, 20)</td>
<td>$-0.399$</td>
<td>$0.125$</td>
<td>$0.003$</td>
<td>$8.05 \times 10^3$</td>
<td>$-376.15$</td>
<td>$-26.50$</td>
</tr>
<tr>
<td>(5, 100)</td>
<td>$-0.055$</td>
<td>$-0.010$</td>
<td>$0.001$</td>
<td>$2.26 \times 10^3$</td>
<td>$72.67$</td>
<td>$4.11$</td>
</tr>
<tr>
<td>(10, 100)</td>
<td>$1.119$</td>
<td>$0.096$</td>
<td>$-0.002$</td>
<td>$1.42 \times 10^3$</td>
<td>$70.00$</td>
<td>$6.34$</td>
</tr>
<tr>
<td><strong>t-t scenario</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLME</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5, 20)</td>
<td>$-0.000$</td>
<td>$-0.009$</td>
<td>$0.002$</td>
<td>$1.621$</td>
<td>$0.646$</td>
<td>$0.715$</td>
</tr>
<tr>
<td>(10, 20)</td>
<td>$0.005$</td>
<td>$-0.003$</td>
<td>$0.007$</td>
<td>$1.524$</td>
<td>$0.628$</td>
<td>$0.775$</td>
</tr>
<tr>
<td>(5, 100)</td>
<td>$0.002$</td>
<td>$-0.007$</td>
<td>$-0.005$</td>
<td>$1.750$</td>
<td>$0.717$</td>
<td>$0.844$</td>
</tr>
<tr>
<td>(10, 100)</td>
<td>$0.004$</td>
<td>$-0.000$</td>
<td>$-0.002$</td>
<td>$1.771$</td>
<td>$0.714$</td>
<td>$0.851$</td>
</tr>
<tr>
<td>LME</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5, 20)</td>
<td>$0.004$</td>
<td>$0.004$</td>
<td>$0.015$</td>
<td>$5.07 \times 10^6$</td>
<td>$-5.13 \times 10^6$</td>
<td>$5.19 \times 10^6$</td>
</tr>
<tr>
<td>(10, 20)</td>
<td>$0.010$</td>
<td>$0.003$</td>
<td>$0.011$</td>
<td>$4.66$</td>
<td>$-5.64$</td>
<td>$12.41$</td>
</tr>
<tr>
<td>(5, 100)</td>
<td>$-0.003$</td>
<td>$-0.013$</td>
<td>$-0.011$</td>
<td>$3.51$</td>
<td>$-4.23$</td>
<td>$7.63$</td>
</tr>
<tr>
<td>(10, 100)</td>
<td>$-0.006$</td>
<td>$-0.008$</td>
<td>$0.001$</td>
<td>$2.88$</td>
<td>$-3.42$</td>
<td>$6.55$</td>
</tr>
<tr>
<td>tLME</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5, 20)</td>
<td>$-7.057$</td>
<td>$1.896$</td>
<td>$0.001$</td>
<td>$1.23 \times 10^3$</td>
<td>$2.65 \times 10^3$</td>
<td>$1.20 \times 10^4$</td>
</tr>
<tr>
<td>(10, 20)</td>
<td>$-8.804$</td>
<td>$-0.565$</td>
<td>$0.008$</td>
<td>$1.98 \times 10^4$</td>
<td>$663.64$</td>
<td>$114.19$</td>
</tr>
<tr>
<td>(5, 100)</td>
<td>$-2.531$</td>
<td>$-0.130$</td>
<td>$-0.003$</td>
<td>$8.24 \times 10^3$</td>
<td>$176.51$</td>
<td>$12.98$</td>
</tr>
<tr>
<td>(10, 100)</td>
<td>$-0.857$</td>
<td>$-0.016$</td>
<td>$-0.000$</td>
<td>$3.94 \times 10^3$</td>
<td>$152.02$</td>
<td>$9.81$</td>
</tr>
</tbody>
</table>
Figure 1: Absolute bias of the estimator for the generalized Laplace mixed-effects model with data generated under the Normal-Normal (NN) and Laplace-Laplace (LL) scenarios.
Figure 2: Absolute bias of the estimator for the generalized Laplace mixed-effects (GLME) model, linear mixed-effects (LME) model, and heavy-tailed LME (tLME) model with data generated under the Laplace-Laplace (LL) and $t$-$t$ (TT) scenarios. The y-axis is on the log$_{10}$-scale.
Figure 3: Ordinary least squares estimates of intercepts and slopes for individual growth curves in the rats weight gain data. The scatterplots on the top show the pairwise estimates with LOESS smoothing superimposed (dashed grey lines mark mean values). The plots on the bottom depict the estimated densities of intercepts (solid line) and slopes (dashed line) centred and scaled using their respective means and standard deviations.
2 R code

2.1 Brief tutorial

In this section, we provide an example on how to do inference on generalized Laplace mixed-effects (GLME) models using the R package `nlmm`. The latter is available on GitHub and can be installed as follows:

```r
library(devtools)
install_github("marco-geraci/nlmm")
```

We consider the dataset `rats`, which contains data analyzed in Section 6.2 of the Manuscript and is available in the package. After loading the package, the following code shows how to attach the dataset and access the R documentation describing the variables:

```r
library(nlmm)
data(rats)
?rats
```

The following code fits a GLME model with treatment and time effects, their interaction, and random intercepts:

```r
fit <- nlmm(y ~ trt*time, random = ~ 1, group = id, data = rats,
            control = nlmmControl(multistart = TRUE))
```

```r
> fit
Call: nlmm(fixed = y ~ trt * time, random = ~1, group = id, data = rats,
          control = nlmmControl(multistart = TRUE))

Generalized Laplace mixed-effects model

Alpha:

Random effects          Error
 0.01073                0.96213

Fixed effects:

(Intercept)  trt2  trt3  time  trt2:time  trt3:time
  50.715     7.737 -1.574  27.635   -11.221     1.834

Covariance matrix of the random effects:

(Intercept)
(Intercept)  61.96

Residual variance:  53.98

Log-likelihood: -479.7

Number of observations: 135
Number of groups: 27
The output provides basic information about the fit. Starting from the top, we find the estimate of
the shape parameter $\alpha$ (‘Alpha’), then $\beta$ (‘Fixed effects’), $\Psi_1$ (‘Covariance matrix of the random
effects’), and $\Psi_2$ (‘Residual variance’). The log-likelihood of the model at convergence and information
on the sample size are provided at the bottom. Note that the object fit contains more
information and this is described in the R documentation of the function nlmm.

The estimates of $\alpha_1$ and $\alpha_2$, are, respectively, 0.011 and 0.962, suggesting that the data come from
a Normal-Laplace (NL) model, that is, a normal distribution for the random effects and a Laplace
distribution for the error term. We can test the NL hypothesis using the function lrt_nlmm. We
already have obtained the unconstrained fit, but we also need fitting a constrained model:

```r
fit0 <- nlmm(y ~ trt*time, random = ~ 1, group = id, data = rats,
             control = nlmmControl(alpha = c(0,1), alpha.index = 0))

> fit0
Call: nlmm(fixed = y ~ trt * time, random = ~1, group = id, data = rats,
          control = nlmmControl(alpha = c(0, 1), alpha.index = 0))

Generalized Laplace mixed-effects model
Special case: Normal-Laplace

Alpha:
Random effects          Error
 0                    1

Fixed effects:
(Intercept)    trt2    trt3     time  trt2:time  trt3:time

Covariance matrix of the random effects:
   (Intercept)
(Intercept) 68.68

Residual variance: 50.52

Log-likelihood: -481

Number of observations: 135
Number of groups: 27
```

By setting the argument alpha.index = 0 in the function nlmmControl, the estimation algorithm
constrains the values of $\alpha$ to those provided in the argument alpha (see ?nlmmControl for additional
details). Note that in the printed output for fit0, the NL model has been detected and reported as
a special case. The constrained model has a smaller log-likelihood than that of the unconstrained
model, while the parameter estimates are not (much) dissimilar. A formal comparison between the
two models is done as follows:

```r
> lrt_nlmm(fit0, fit)
Likelihood ratio test for constrained
   generalized Laplace mixed-effects models
**H0:** alpha random effects is equal to 0 and alpha error is equal to 1

LRT statistic = 2.5287
df or weights = 0.48 0.5 0.02
p-value (Chi-bar-squared) = 0.0609

At the 5% level, we would not reject the null hypothesis that the model is NL. However, the evidence in favor of the null is not overwhelming.

The package provides several S3 methods for fitted *nlmm* objects including: **summary**, which gives standard errors, *p*-values, and confidence intervals; **fixef** and **ranef** to extract, respectively, fixed and random effects; **VarCorr** to extract the variance-covariance matrix of the random effects; **vcov** to extract the variance-covariance matrix of the maximum likelihood estimator $\hat{\theta}$ as defined in Section 4.1 of the Manuscript; and **predict**, **residuals**, and **logLik**, whose names are self-explanatory. The *nlmm* package makes use of some functions available from *nlme* (as, for example, the variance function structures). However the estimation algorithm, as described in Section 4.1 of the Manuscript, is completely independent from the fitting routines used in *nlme*. On the other hand, the *nlmm* function does make a call to *lme* in two cases: when the Normal-Normal model is requested (using *nlmmControl(alpha = c(0,0), alpha.index = 0)*), and to get the starting values for GLME models (with *nlmmControl(lme = TRUE)*), although the latter is optional. Finally, we draw attention on the availability in the *nlmm* package of density, probability, quantile, and random number generation functions for the generalized Laplace distribution.

### 2.2 Simulation study

In this section, we provide the minimum R code to run the simulation study in Section 5 of the Manuscript.

```r
library(nlmm)
library(nlme)
library(heavy)
library(lqmm)
library(Qtools)
library(xtable)
library(Hmisc)
library(mvtnorm)
library(statmod)
library(numDeriv)
library(MASS)
```

```
# R code to run Part (a) of the simulation study

# Generate data

R <- 500
b <- c(1, 2, 0)
P <- length(b)
```
M <- c(20, 100)
n <- c(5, 10)
sizes <- expand.grid(n = n, M = M)
samp.lab <- apply(sizes, 1, paste, collapse = "-")
nsamp <- nrow(sizes)
distr <- c("NN", "LL")

S1 <- matrix(c(2, 0.8, 0.8, 1), 2, 2)
cov_name <- "pdSymm"
npar <- P + theta.z.dim(type = cov_name, n = 2) + 2 + 1

DATA_NN <- DATA_LL <- list()
for(j in 1:nsamp){
  S2 <- diag(1, sizes[j, 'n'], sizes[j, 'n'])
  # NN
  DATA_NN[[j]] <- generate.data(R = R, n = sizes[j, 'n'], M = sizes[j, 'M'], sigma_1 = S1,
                              sigma_2 = S2, dist.u = "norm", dist.e = "norm", beta = b, gamma = c(1,0),
                              shape_1 = NULL, shape_2 = NULL, seed = 123)
  # LL
  DATA_LL[[j]] <- generate.data(R = R, n = sizes[j, 'n'], M = sizes[j, 'M'], sigma_1 = S1,
                              sigma_2 = S2, dist.u = "genlaplace", dist.e = "genlaplace", beta = b, gamma = c(1,0),
                              shape_1 = 1, shape_2 = 1, seed = 123)
}

# control parameters
ctrl <- nlmmControl(multistart = TRUE, grid = c(0.001, 0.5, 0.999), alpha.index = 9,
                     lme = TRUE, lmeMethod = "REML", nK = 8)

# fit nlmm
RES <- array(NA, dim = c(R, npar, nsamp, length(distr)))
V <- array(NA, dim = c(R, npar*npar, nsamp, length(distr)))
LRT <- array(NA, dim = c(R, 2, nsamp, length(distr)))

for(k in 1:length(distr)){
  tmp <- get(paste("DATA", distr[k], sep = "_"))
  for(j in 1:nsamp){
    for(i in 1:R){
      id <- tmp[[j]]$group
      x <- tmp[[j]]$X[,,i]
      y <- tmp[[j]]$Y[,,i]
      dd <- data.frame(y = y, x = x[, 'x'], z = x[, 'z'], id = id)

      fit <- try(nlmm(y ~ x + z, random = ~ x, group = id, covariance = "pdSymm",
                      data = dd, weights = NULL, control = ctrl), silent = TRUE)

      if(!inherits(fit, "try-error")){
        RES[i, , j, k] <- fit$par
      }
    }
  }
}

9
V[i, j, k] <- as.numeric(vcov(fit))

fit0 <- try(update(fit, control = nlmmControl(alpha.index = 0,
          alpha = c(0,0))), silent = TRUE)

val <- try(lrt_nlmm(fit0, fit), silent = TRUE)

try(LRT[i, j, k] <- unlist(val[[c('statistic', 'p.value')]]),
  silent = TRUE)

}

}

}

}

#################################################
# R code to run Part (b) of the simulation study
#################################################

# Generate data

R <- 500
b <- c(1, 2, 0)
P <- length(b)

M <- c(20, 100)
n <- c(5, 10)
sizes <- expand.grid(n = n, M = M)
samp.lab <- apply(sizes, 1, paste, collapse = "-")
nsamp <- nrow(sizes)

distr <- c("TT", "LL")

S1 <- matrix(c(2, 0.8, 0.8, 1),2,2)
cov2cor(S1)
cov_name <- "pdSymm"

npar <- P + theta.z.dim(type = cov_name, n = 2) + 2 + 1

DATA_TT <- DATA_LL <- list()
for(j in 1:nsamp){

S2 <- diag(1, sizes[j,'n'], sizes[j,'n'])

# TT
DATA_TT[[j]] <- generate.data(R = R, n = sizes[j,'n'], M = sizes[j,'M'], sigma_1 = S1,
    sigma_2 = S2, dist.u = "t", dist.e = "t", beta = b, gamma = c(1,0), shape_1 = 4,
    shape_2 = 4, seed = 123)

# LL
DATA_LL[[j]] <- generate.data(R = R, n = sizes[j,'n'], M = sizes[j,'M'], sigma_1 = S1,
    sigma_2 = S2, dist.u = "genlaplace", dist.e = "genlaplace", beta = b, gamma = c(1,0),
    shape_1 = 4, shape_2 = 4, seed = 123)
shape_1 = 1, shape_2 = 1, seed = 123)

# control parameters
ctrl <- nlmmControl(multistart = TRUE, grid = c(0.001, 0.5, 0.999), alpha.index = 9, lme = TRUE, lmeMethod = "REML", nK = 8)

# fit nlmm and heavyLme
RES <- array(NA, dim = c(R, npar, nsamp, length(distr), 2))
V <- array(NA, dim = c(R, npar*npar, nsamp, length(distr), 2))

for(k in 1:length(distr)){
tmp <- get(paste("DATA", distr[k], sep = "_"))
for(j in 1:nsamp){
for(i in 1:R){
id <- tmp[[j]]$group
x <- tmp[[j]]$X[,,i]
y <- tmp[[j]]$Y[,i]
dd <- data.frame(y = y, x = x[,'x'], z = x[,'z'], id = id)

fit1 <- try(nlmm(y ~ x + z, random = ~ x, group = id, covariance = "pdSymm", data = dd, weights = NULL, control = ctrl), silent = TRUE)
fit2 <- try(heavyLme(y ~ x + z, random = ~ x, groups = ~ id, data = dd, family = Student(df = 2)), silent = TRUE)

if(!inherits(fit1, "try-error") & !inherits(fit2, "try-error")){
RES[i, , j, k, 1] <- fit1$par
V[i, , j, k, 1] <- as.numeric(vcov(fit1))
RES[i, 1:(npar - 1), j, k, 2] <- c(fit2$coefficients,
   fit2$theta[lower.tri(fit2$theta, diag = TRUE)], fit2$family$call[[2]],
   fit2$scale)
V[i, 1:P, j, k, 2] <- summary(fit2$coefficients[,2])
}
}
}

References