



Hierarchical Generalized Linear Models: The R Package HGLMMM

Marek Molas

Emmanuel Lesaffre

Erasmus Medical Center at Rotterdam Erasmus Medical Center at Rotterdam

Abstract

The R package **HGLMMM** has been developed to fit generalized linear models with random effects using the h-likelihood approach. The response variable is allowed to follow a binomial, Poisson, Gaussian or gamma distribution. The distribution of random effects can be specified as Gaussian, gamma, inverse-gamma or beta. Complex structures as multi-membership design or multilevel designs can be handled. Further, dispersion parameters of random components and the residual dispersion (overdispersion) can be modeled as a function of covariates. Overdispersion parameter can be fixed or estimated. Fixed effects in the mean structure can be estimated using extended likelihood or a first order Laplace approximation to the marginal likelihood. Dispersion parameters are estimated using first order adjusted profile likelihood.

Keywords: h-likelihood, complex designs, hierarchical generalized linear models, overdispersion.

1. Introduction

In [Nelder and Wedderburn \(1972\)](#) the class of generalized linear models (GLM) was developed. This class of models allows for the response to follow a distribution from the exponential family, extending modeling capabilities beyond the Gaussian response. In [Henderson, Kempthorne, Searle, and Krosigk \(1959\)](#) the linear mixed model was suggested, which enabled to model correlation in the data. Further, it was extended to the generalized linear mixed model (see e.g., [Molenberghs and Verbeke 2005](#)), where the response from an exponential family is combined with normal random effects. In [Lee and Nelder \(1996\)](#) hierarchical generalized linear models were described, which allows random effects to be not normally distributed. Further [Lee and Nelder \(1996\)](#) proposed the extended likelihood rather than the marginal likelihood to estimate the parameters. Later [Lee and Nelder \(2001\)](#) focused on

the joint modelling of the mean and dispersion structure. This estimation technique relies on the iterative weighted least squares (IWLS) algorithm, where fixed effects and random effects are estimated using the extended likelihood, and dispersion parameters are obtained by maximizing the adjusted profile likelihood. A subsequent adjustment of the algorithm was proposed in [Noh and Lee \(2007\)](#) replacing the extended likelihood by the first order adjusted profile likelihood as a criterion to estimate fixed effects in the mean structure.

The objective of this paper is to present the R ([R Development Core Team 2011](#)) package **HGLMMM** available from the Comprehensive R Archive Network at <http://CRAN.R-project.org/package=HGLMMM>. The package runs under R version 2.9.0 or higher. This package fits the class of generalized linear models with random effects. In the remainder of the paper we first outline the h-likelihood approach to the estimation and statistical inference. Next, we present the capabilities of the **HGLMMM** package on real-life datasets, described in [Lee, Nelder, and Pawitan \(2006\)](#).

2. H-likelihood estimation and inference framework

Standard maximum likelihood estimation for models with random effects is based on the marginal likelihood as objective function. The parameters are estimated by a marginal likelihood procedure (MML) and their standard errors are computed from the inverse of the negative hessian matrix of the marginal likelihood. In the marginal likelihood approach random effects \mathbf{v} are integrated out and only fixed effects in the mean structure $\boldsymbol{\beta}$ and dispersion parameters $\boldsymbol{\lambda}$ are retained in the maximized function. For a mixed effects model the conditional likelihood of the j -th ($j = 1, \dots, n_i$) repeated observation on the i -th subject ($i = 1, \dots, N$), i.e., y_{ij} , is given by $f_{\boldsymbol{\beta}, \boldsymbol{\lambda}}(y_{ij} | \mathbf{v}_i)$. The likelihood of the i -th random effect is denoted as $f_{\boldsymbol{\lambda}}(\mathbf{v}_i)$. Note that $\boldsymbol{\lambda}$ contains dispersion parameters of the random components \mathbf{v}_i as well as the parameters describing the residual dispersion (overdispersion) of the response y_{ij} . The marginal likelihood maximized in the MML procedure is given by

$$L_M(\boldsymbol{\beta}, \boldsymbol{\lambda} | \mathbf{y}) = \prod_{i=1}^N \int \prod_{j=1}^{n_i} f_{\boldsymbol{\beta}, \boldsymbol{\lambda}}(y_{ij} | \mathbf{v}_i) f_{\boldsymbol{\lambda}}(\mathbf{v}_i) d\mathbf{v}_i. \quad (1)$$

Maximizing L_M or equivalently the log-likelihood $\ell_M = \log(L_M)$ yields consistent estimates of the fixed effects parameters. However, the problem lies in computing the integrated likelihood. This may be a time-consuming task especially for complex models since it needs to be done for each subject and each iteration. Further, if the MLE is determined with a Newton-Raphson procedure then integrals need to be computed also for the first and second derivatives. It is also important to fine tune the likelihood calculations, see e.g., [Lesaffre and Spiessens \(2001\)](#).

In [Lee and Nelder \(1996\)](#) another approach to estimating the parameters was proposed. These authors argued to use the joint likelihood L_E for the maximization, which is directly available from the definition of the model. The joint likelihood, called also extended likelihood or h-likelihood, is then maximized jointly with respect to \mathbf{v} and $\boldsymbol{\beta}$ given dispersion parameters $\boldsymbol{\lambda}$. At the maximum, standard errors are obtained in the classical way. In the notation of above, the extended likelihood is given by:

$$L_E(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{v} | \mathbf{y}, \mathbf{v}) = \prod_{i=1}^N \prod_{j=1}^{n_i} f_{\boldsymbol{\beta}, \boldsymbol{\lambda}}(y_{ij} | \mathbf{v}_i) f_{\boldsymbol{\lambda}}(\mathbf{v}_i). \quad (2)$$

The logarithm of (2) is called the extended log-likelihood by Lee *et al.* (2006) and they denoted its logarithm as $h = \log [L_E(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{v} | \mathbf{y}, \mathbf{v})]$. We could say that this extended likelihood reflects the hierarchical character of the data.

In the h-likelihood approach the estimates of dispersion parameters are determined by maximizing the adjusted profile likelihood introduced by Cox and Reid (1987). However, for some models the approach of Lee and Nelder (1996) is not appropriate, therefore Noh and Lee (2007) proposed to replace the joint likelihood as the estimation criterion for $\boldsymbol{\beta}$ with another adjusted profile likelihood. The outline of this procedure is given in the next section.

2.1. Computing marginal MLEs using the h-likelihood approach

In some special cases, i.e., when the random effects are on the canonical scale (see e.g., Lee *et al.* 2006 pp. 112–114), joint maximization of the extended log-likelihood h with respect to all parameters $(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{v}_1, \dots, \mathbf{v}_N)$ is equivalent to maximizing the marginal likelihood with respect to $\boldsymbol{\beta}, \boldsymbol{\lambda}$ and taking the empirical Bayes (EB) estimates for $\mathbf{v}_1, \dots, \mathbf{v}_N$. But, most often the two maximization procedures are not equivalent.

Noh and Lee (2007) suggest in the general case to work with a Laplace approximation to the marginal likelihood (1). Namely, the integral of the function $k(\mathbf{x}, \mathbf{y}) \exp [-ng(\mathbf{x}, \mathbf{y})]$ with respect to \mathbf{x} can be approximated as follows:

$$\int k(\mathbf{x}, \mathbf{y}) \exp [-ng(\mathbf{x}, \mathbf{y})] d\mathbf{x} = \left| \frac{n \frac{\partial^2 g(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}^\top \partial \mathbf{x}}}{2\pi} \right|_{\mathbf{x}=\hat{\mathbf{x}}}^{-\frac{1}{2}} \exp [-ng(\mathbf{x}, \mathbf{y})] k(\mathbf{x}, \mathbf{y})|_{\mathbf{x}=\hat{\mathbf{x}}}, \quad (3)$$

with $\hat{\mathbf{x}}$ the value of \mathbf{x} that maximizes $-g(\mathbf{x}, \mathbf{y})$. This is called the Laplace approximation of the above integral (at $\hat{\mathbf{x}}$). More information on the Laplace approximation can be found in e.g., Severini (2000, Section 2.11).

Taking in expression (3) $k(\mathbf{x}, \mathbf{y}) = 1$; $\exp [ng(\mathbf{x}, \mathbf{y})] = \exp [-h(\boldsymbol{\theta}, \mathbf{v})]$ whereby $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\lambda})$; \mathbf{v} representing the stacked vector of N random effects and finally $n \frac{\partial^2 g(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}^\top \partial \mathbf{x}} = -\frac{\partial^2 h(\boldsymbol{\theta}, \mathbf{v})}{\partial \mathbf{v}^\top \partial \mathbf{v}}$, leads to

$$L_M(\boldsymbol{\beta}, \boldsymbol{\lambda} | \mathbf{y}) = \int \exp [h(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{v})] d\mathbf{v} \approx \left| \frac{-\frac{\partial^2 h(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{v})}{\partial \mathbf{v}^\top \partial \mathbf{v}}}{2\pi} \right|_{\mathbf{v}=\hat{\mathbf{v}}}^{-\frac{1}{2}} \exp [h(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{v})] |_{\mathbf{v}=\hat{\mathbf{v}}}, \quad (4)$$

with $\hat{\mathbf{v}}$ maximizing the extended likelihood for a given (starting) value of $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$, i.e., $\hat{\mathbf{v}}(\boldsymbol{\beta}, \boldsymbol{\lambda})$. Note that the approximation improves when the number of observations per subject, n_i increases, which can be inferred from Severini (2000). Taking the logarithm of the previous expression leads to the adjusted profile (log)-likelihood

$$p_v(h) = h(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{v})|_{\mathbf{v}=\hat{\mathbf{v}}} - 0.5 \log \left| \frac{D(h, \mathbf{v})}{2\pi} \right|_{\mathbf{v}=\hat{\mathbf{v}}}, \quad (5)$$

with $D(h, \mathbf{v}) = -\frac{\partial^2 h(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{v})}{\partial \mathbf{v}^\top \partial \mathbf{v}}$. The term ‘adjusted profile likelihood’ is chosen since $h(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{v})|_{\mathbf{v}=\hat{\mathbf{v}}}$ is a profile (log)-likelihood of $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ and the second term in (5) is a correction term to approximate the marginal log-likelihood.

The next step in the iterative procedure is to maximize the adjusted profile (log-)likelihood (5) with respect to $\boldsymbol{\beta}$. Note that maximizing the profile log-likelihood $h(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{v})|_{\mathbf{v}=\hat{\mathbf{v}}}$ to find

the MLE of β is not appropriate since this is equivalent to joint maximization of h over β and \mathbf{v} which is most often invalid as seen above.

After obtaining $\hat{\beta}_\lambda$ from maximization of (5) for a given dispersion component λ , the estimation algorithm proceeds with estimation of λ . Another adjusted profile likelihood is used as an objective function to find $\hat{\lambda}$.

Let the marginal distribution of the data \mathbf{y} be $f_{\beta,\lambda}(\mathbf{y})$ (marginalized over \mathbf{v}), i.e., the LHS of expression (4) or its approximation, now seen as a probability density function (pdf) of the data. Conditional on the sufficient statistics for β , i.e., $\hat{\beta}_\lambda$, (see Cox and Hinkley 1974, p. 21), the (marginalized) distribution of the data can be derived from:

$$f_\lambda(\mathbf{y}|\hat{\beta}_\lambda) = \frac{f_{\beta,\lambda}(\mathbf{y})}{f_{\beta,\lambda}(\hat{\beta}_\lambda)}, \quad (6)$$

where $f_{\beta,\lambda}(\hat{\beta}_\lambda)$ is the distribution of $\hat{\beta}_\lambda$. One then applies the *p-formula* as derived by Barndorff-Nielsen (1980, 1983), see also Pawitan (2001), to obtain in general the distribution of the ML estimator. Namely,

$$f_{\beta,\lambda}(\hat{\beta}_\lambda) = \left| -\frac{1}{2\pi} \frac{\partial^2 \log f_{\beta,\lambda}(\mathbf{y})}{\partial \beta^\top \partial \beta} \right|_{\beta=\hat{\beta}_\lambda}^{\frac{1}{2}} \frac{f_{\beta,\lambda}(\mathbf{y})}{f_{\hat{\beta}_\lambda,\lambda}(\mathbf{y})}, \quad (7)$$

where $f_{\hat{\beta}_\lambda,\lambda}(\mathbf{y})$ is the marginal profile likelihood of λ , i.e., $\log[f_{\hat{\beta}_\lambda,\lambda}(\mathbf{y})] = \ell_M(\beta_\lambda, \lambda|\mathbf{y})|_{\beta=\hat{\beta}_\lambda}$. After substitution of expression (7) into (6) one obtains:

$$\log [f_\lambda(\mathbf{y}|\hat{\beta}_\lambda)] = \ell_M(\beta_\lambda, \lambda|\mathbf{y})|_{\beta=\hat{\beta}_\lambda} - 0.5 \log \left| \frac{D(\ell_M, \beta)}{2\pi} \right|_{\beta=\hat{\beta}_\lambda}, \quad (8)$$

with $D(\ell_M, \beta) = -\frac{\partial^2 \ell_M}{\partial \beta^\top \partial \beta}$. In the next step, one replaces everywhere the marginal log likelihood ℓ_M by the adjusted profile log-likelihood $p_v(h)$ evaluated in $\beta = \hat{\beta}_\lambda$. This results in:

$$\log [f_\lambda(\mathbf{y}|\hat{\beta}_\lambda)] = h(\beta, \lambda, \mathbf{v})|_{\beta=\hat{\beta}_\lambda, \mathbf{v}=\hat{\mathbf{v}}} - 0.5 \log \left| \frac{D(h, \mathbf{v})}{2\pi} \right|_{\beta=\hat{\beta}_\lambda, \mathbf{v}=\hat{\mathbf{v}}} - 0.5 \log \left| \frac{D[p_v(h), \beta]}{2\pi} \right|_{\beta=\hat{\beta}_\lambda}. \quad (9)$$

Finally, in Appendix 4 of Lee and Nelder (2001) it is shown that the sum of the last two terms in the above expression is equal to

$$-0.5 \log \left| \frac{D[h, (\beta, \mathbf{v})]}{2\pi} \right|_{\beta=\hat{\beta}_\lambda, \mathbf{v}=\hat{\mathbf{v}}},$$

with $D[h, (\beta, \mathbf{v})]$ equal to $\begin{pmatrix} -\frac{\partial^2 h}{\partial \beta^\top \partial \beta} & -\frac{\partial^2 h}{\partial \beta^\top \partial \mathbf{v}} \\ -\frac{\partial^2 h}{\partial \mathbf{v}^\top \partial \beta} & -\frac{\partial^2 h}{\partial \mathbf{v}^\top \partial \mathbf{v}} \end{pmatrix}$ with dimensions equal to the sum of the dimensions of two adjustment terms in (9). As a result one obtains the following adjusted profile likelihood:

$$p_{\beta,v}(h) = h(\beta, \lambda, \mathbf{v})|_{\beta=\hat{\beta}_\lambda, \mathbf{v}=\hat{\mathbf{v}}} - 0.5 \log \left| \frac{D[h, (\beta, \mathbf{v})]}{2\pi} \right|_{\beta=\hat{\beta}_\lambda, \mathbf{v}=\hat{\mathbf{v}}}. \quad (10)$$

The latter adjusted profile likelihood is maximized with respect to $\boldsymbol{\lambda}$ to obtain $\hat{\boldsymbol{\lambda}}$. Note that this objective function is “focussed” solely on the dispersion parameters. This offers an extension of restricted maximum likelihood (REML) estimation and provides inference for the class of generalized linear mixed models, see [Noh and Lee \(2007\)](#). We show below that in the case of linear mixed models this function is exactly the restricted maximum likelihood.

2.2. The linear mixed model case

We illustrate the above estimation framework for the linear mixed model case. While in general the above calculations are approximate, in this case they are exact. Some of the expressions are based on the results shown in [Harville \(1977\)](#). The classical linear mixed effects model assumes:

$$\begin{aligned} f_{\boldsymbol{\beta}, \boldsymbol{\lambda}_e}(\mathbf{y}_i | \mathbf{v}_i) &= \mathcal{N}(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{v}_i; \boldsymbol{\Sigma}_i), \\ f_{\boldsymbol{\lambda}_v}(\mathbf{v}_i) &= \mathcal{N}(0; \boldsymbol{\Lambda}_i). \end{aligned} \quad (11)$$

The design matrix \mathbf{X}_i contains fixed effects for the i -th subject, while \mathbf{Z}_i is a design matrix for the random effects for the i -th subject. Matrices $\boldsymbol{\Sigma}_i$ and $\boldsymbol{\Lambda}_i$ determine the residual variance of \mathbf{y}_i and the variance of random effects \mathbf{v}_i respectively. Note that we have split up $\boldsymbol{\lambda}$ into $\boldsymbol{\lambda}_e$ and $\boldsymbol{\lambda}_v$, the dispersion parameters pertaining to the residual variability $\boldsymbol{\Sigma}_i$ and random effects variability $\boldsymbol{\Lambda}_i$, thereby showing the role of each of the dispersion components. Denote by \mathbf{V} the total (over all subjects) marginal variance-covariance matrix $\mathbf{V} = \mathbf{Z} \boldsymbol{\Lambda} \mathbf{Z}^T + \boldsymbol{\Sigma}$, where \mathbf{X} is the design matrix for fixed effects obtained by stacking \mathbf{X}_1 to \mathbf{X}_N , $\mathbf{Z} = \text{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_N)$ is the design matrix of random effects, $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\Lambda}_1, \dots, \boldsymbol{\Lambda}_N)$ is the variance covariance matrix of the random effects and $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_N)$ is the residual variance covariance matrix.

In this case the adjusted profile likelihood $p_v(h)$ becomes:

$$p_v(h) = \ell_M = -\frac{1}{2} \log |2\pi \mathbf{V}| - \frac{1}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \quad (12)$$

which is the expression for the marginal likelihood of a linear mixed model. Further, the adjusted profile likelihood $p_{v,\boldsymbol{\beta}}(h)$ is equal to:

$$p_{\boldsymbol{\beta},v}(h) = \log[f_{\boldsymbol{\lambda}}(\mathbf{y} | \hat{\boldsymbol{\beta}}_{\boldsymbol{\lambda}})] = \ell_M|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{\boldsymbol{\lambda}}} - 0.5 \log \left| \frac{\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X}}{2\pi} \right|, \quad (13)$$

which is exactly equal to the classical REML likelihood (see e.g., [Verbeke and Molenberghs 2000](#)) for the linear mixed model.

Note that maximization of (5) with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ is actually the maximization of the marginal likelihood computed by a classical Laplace approximation. In the above h-likelihood procedure, estimation of $\boldsymbol{\beta}$ given $\boldsymbol{\lambda}$ is the same as with a classical Laplace approximation, but the estimation of $\boldsymbol{\lambda}$ is essentially different. The h-likelihood procedure gives an elegant set of IWLS equations for the estimation of the dispersion parameters, which possibly depend on covariates.

2.3. Application to the hierarchical generalized linear models

The above theory is applied to generalized linear models with random effects. In this class of models the assumed distribution of the response y_{ij} (conditional on random effects) belongs

to the exponential family:

$$f_{\beta, \lambda_e}(y_{ij}) = \exp \left[\frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{\lambda_e} + c(y_{ij}, \lambda_e) \right]. \quad (14)$$

This distribution is combined with the distribution of the random component, which distribution belongs to the family of conjugate Bayesian distributions for an exponential family, i.e.,

$$f_{\lambda_v}(v_i) = \exp [a_1(\lambda_v)v_i - a_2(\lambda_v)b(v_i) + c_2(\lambda_v)], \quad (15)$$

which can be expressed as the distribution of a pseudo-response ψ_i as follows:

$$f_{\lambda_v}(v_i) = \exp \left[\frac{\psi_i v_i - b(v_i)}{\lambda_v} + c_2(\psi_i, \lambda_v) \right]. \quad (16)$$

In this class of models the response is allowed to follow a Gaussian, binomial, Poisson or gamma distribution. Their corresponding conjugate Bayesian distributions are Gaussian, beta, gamma and inverse-gamma, respectively. Note that both λ_v and λ_e are allowed to depend on covariates.

2.4. Implementation in the HGLMMM package

In this section we document the use of the function `HGLMfit`, together with accompanying functions `HGLMLRTest`, `HGLMLikeDeriv` and `BootstrapEnvelopeHGLM`. First we will describe the use of the fitting function `HGLMfit` and explain the parameters used in the invocation of the function. The following structure of the routine is used:

```
HGLMfit(DistResp = "Normal", DistRand = NULL, Link = NULL, LapFix = FALSE,
        ODEst = NULL, ODEstVal = 0, formulaMain, formulaOD, formulaRand,
        DataMain, DataRand, Offset = NULL, BinomialDen = NULL, StartBeta = NULL,
        StartVs = NULL, StartRGamma = NULL, INFO = TRUE, DEBUG = FALSE,
        na.action, contrasts = NULL, CONV = 1e-04)
```

Below we describe the parameters of the function `HGLMfit` together with the default values in parenthesis.

DistResp ("Normal"): The distribution of the response as defined in (14) is set by the option `DistResp`. The user can set it to: `Binomial`, `Normal`, `Poisson` or `Gamma`. Note that the name of the distribution must start with a capital letter.

DistRand (no default): The next option `DistRand` specifies the distribution of the random components and should be set as a vector of distribution names from the set: `"Beta"`, `"Gamma"`, `"IGamma"` (inverse-gamma) and `"Normal"`. For each random component one entry in the vector must be specified. Therefore for a model with two random effects, whereby the first random effect has a normal distribution and the second random effect has a gamma distribution, you should specify the vector `c("Normal", "Gamma")`.

Link (canonical link): The link option is available for a gamma distribution of the response. The choice is either `"Log"` or `"Inverse"`. For the random variables of a `"Binomial"`, `"Normal"` or `"Poisson"` distribution only the default links are currently available, that is `"Logit"`, `"Identity"` and `"Log"`, respectively.

LapFix (FALSE): Having defined the structure of the model, i.e., (1) the distribution of the response, (2) random effects and (3) the link, one has to specify in the option **LapFix** whether the joint likelihood (2) will be used to estimate fixed effects in the mean structure of the response model, or the effects will be estimated by the adjusted profile likelihood (5), which is an approximation to the marginal likelihood. Set **LapFix** = TRUE for the adjusted profile likelihood and **LapFix** = FALSE for the joint likelihood.

ODEst (likelihood-based analysis), **ODEstVal** (0): Next the option **ODEst** determines whether the dispersion parameter λ_e in (14) will be estimated or held fixed. This is set to NULL by default which implies for the "Poisson" and "Binomial" distribution for the response that it is fixed. For the "Normal" and "Gamma" response, the default option is that it is estimated from the data. Specifying **ODEst** = TRUE implies that λ_e is estimated, while it is fixed for **ODEst** = FALSE. Further, the parameters **ODEstVal** specify either the starting values when λ_e is estimated, or the values used in the estimation when λ_e is set fixed.

formulaMain: The option **formulaMain** requires a two-sided formula to be specified to determine the structure of the linear predictor of the response, e.g., `Outcome ~ Fixed.Effect.1 + Fixed.Effect.2 + (1 | Subject.1)`. This specification sets **Outcome** as the response. Further, in the above example it is specified that there are two fixed effects and a random intercept with subject index **Subject.1**. Correlated random components are currently not allowed, therefore a structure `(1 + time | Subject.1)` is not valid, instead `(1 | Subject.1) + (time | Subject.1)` needs to be entered.

formulaOD, **formulaRand**: Similarly **formulaOD** specifies the covariates in the residual dispersion (overdispersion) structure by a one sided formula e.g., `~ 1 + time`. Further, **formulaRand** requires a list of formulas determining the dependence of the dispersion parameters of the random components on covariates. For a model with two random effects the following code `list(one = ~ 1 + mean.time, two = ~ 1)` means that the dispersion parameter of the first random component depends on the average time, while the dispersion parameter of the second random component is constant (intercept only model).

DataMain, **DataRand**: The option **DataMain** determines which data frame to use for the estimation, i.e., where the data for the mean model and residual dispersion is referred to. Likewise **DataRand** is a list of data frames. They correspond to design matrices for the dispersion parameters of random effects. Therefore for each random effect specified in **formulaMain**, a corresponding dataframe needs to be included.

Offset (1): In Poisson regression, an offset variable t is specified in the form $\log(\mu/t) = \eta$, where η is a linear predictor and μ a mean modeled in the Poisson regression. Therefore it is not necessary to log-transform t before entering in the model.

BinomialDen (1): In binomial regression one has to use option **BinomialDen** to specify the denominator for the binomial distribution. Suppose you toss a coin ten times, then you have 10 independent trials, and the denominator should then be 10. It is allowed that each observation has a different denominator.

StartBeta (NULL), **StartVs** (NULL), **StartGamma** (NULL): The three following options allow to specify the starting values for fixed effects in the mean structure **StartBeta**, random effects in the mean structure **StartVs** (which is a vector of values for all random components together) and dispersion parameters **StartGamma**. Note that starting values for the residual dispersion (overdispersion) are supplied in **ODEstVal**. Recall that the overdispersion parameter may be fixed or estimated by setting the option **ODEst**. When starting values are not supplied, for

the intercept of the mean structure an appropriate sample mean of the response is used, and zeros are used for the other parameters.

CONV (1e-04): Setting option **CONV** determines the criterion for convergence, which is computed as the absolute difference between values of all estimated parameters in the previous iteration and in the current iteration.

The function **HGLMfit** returns an object of class **HGLM**. We refer to the help file of the package for the documentation of all elements of this list. The function **HGLMLikeDeriv** takes an object of the class **HGLM** and returns gradient values for fixed effects in the mean structure and dispersion parameters. The function **HGLMLRTest** compares two **HGLM** objects with respect to h-likelihood values, marginal likelihood $p_v(h)$ and REML likelihood $p_{\beta,v}(h)$. Likelihood ratio tests are produced. Finally, the function **BootstrapEnvelopeHGLM** creates a qq-plot of the standardized deviance residuals for the response together with bootstrap envelopes under the assumption that the given model is correct. If the qq-plot falls within the envelopes we can claim that the assumed distribution of the response is reasonable.

There are many packages/routines/programs for analysis of the hierarchical models. The algorithm based on the h-likelihood approach offers a wider choice for the random effects distributions. Further, dispersion components of random effects as well as an overdispersion parameter can be modeled as a function of covariates. This is combined with relatively modest computational requirements.

3. Analysis of examples

In this section we will present the analysis of the three data sets, also analyzed in [Lee et al. \(2006\)](#), using the package **HGLMMM**.

3.1. Salamander data

In [McCullagh and Nelder \(1989\)](#) a dataset on salamander mating was presented. The dependent variable represents the success of salamanders mating. There were 20 males and 20 females in each of the 3 experiments coming from the two populations called whiteside (denoted by W) and roughbutt (denoted as R). Salamanders were paired for mating six times with individuals of their own kind and from the other population. In total 360 observations were generated. Denote as μ_{ijk} the probability of successful mating. We will consider a model with crossed-random effects (also known as a multi-membership model) to analyze this dataset. The mean structure has the following expression:

$$\log \left(\frac{\mu_{ijk}}{1 - \mu_{ijk}} \right) = \text{Intercept} + \text{TypeF} + \text{TypeM} + \text{TypeF} * \text{TypeM} + v_i + v_j, \quad (17)$$

where TypeF is the type of female (R or W), TypeM is the type of male (R or W), while v_i are the random effects corresponding to males and v_j are the random effects corresponding to females. The following analyses were also performed in [Lee et al. \(2006, pp. 194–195\)](#). The program below is used to fit this model. The design matrices for the dispersion components of male and female random effects contain only an intercept and are created by the command below:

```
R> RSa1 <- data.frame(int = rep(1, 60))
```


The following program is invoked to perform the analysis:

```
R> modBin <- HGLMfit(DistResp = "Binomial", DistRand = c("Normal", "Normal"),
+   Link = "Logit", LapFix = TRUE, ODEst = FALSE, ODEstVal = 0,
+   formulaMain = Mate ~ TypeF + TypeM + TypeF * TypeM + (1 | Female) +
+   (1 | Male), formulaOD = ~ 1, formulaRand = list(one = ~ 1, two = ~ 1),
+   DataMain = salamander, DataRand = list(RSal, RSa1),
+   BinomialDen = rep(1, 360), INFO = TRUE, DEBUG = FALSE)
```

A brief description of the model can be displayed by `modBin`, that uses the `print` method for objects of class `HGLM`. Printing of the object `modBin` gives the following output:

```
===== HGLM Model Information =====
Response Distribution: Binomial
Random Effect 1 : Normal / Female
Random Effect 2 : Normal / Male
Link: Logit
Overdispersion Structure is Fixed
Estimation of fixed effects: Laplace Approximation
Dataset used: salamander
Model Equation:
      Mate ~ TypeF + TypeM + TypeF * TypeM + (1 | Female) + (1 | Male)

Overdispersion Equation:
      ~1

Dispersion Equation(s):
Component 1 :~1

Component 2 :~1
```

The output gives information about the response distribution, the number of random effects, their distribution and the subject index. Further, the link function is reported. Next, information is contained whether overdispersion (residual dispersion) is fixed or estimated. The method of estimating the fixed effects in the regression equation is reported, as well as the model equation, the overdispersion equation and the dispersion equations.

The detailed results of the fit can be obtained by the command:

```
R> summary(modBin, V = TRUE)
```

If option `V = TRUE` is omitted, the results for the random effects are not printed (as in our example). Output for our object `modBin` is as follows:

```
===== Fixed Coefficients - Mean Structure =====

      Estimate Std. Error Z value Pr(>|Z|)
(Intercept)    1.0433     0.4036   2.585  0.00974 **
TypeFW         -3.0055     0.5260  -5.714 1.10e-08 ***
```

```
TypeMW          -0.7290      0.4741  -1.538  0.12413
TypeFW:TypeMW    3.7137      0.5758   6.449 1.12e-10 ***
---
```

```
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
```

The above output gives summary of fixed effects in the mean structure.

```
===== Overdispersion Parameters Fixed =====
              Fixed Value
(Intercept)              0
```

The overdispersion parameter is held fixed at 1, which is $\exp(0)$. Note that the parameters reported for overdispersion and dispersion pertain to the logarithm of the parameters. The output shown below reports values for the parameters in the dispersion structure of random effects. To obtain the value of λ_v one again needs to exponentiate the reported estimate.

```
===== Dispersion Parameters Estimated =====
```

```
Dispersion Component: Female
              Estimate Std. Error Z value Pr(>|Z|)
(Intercept)   0.3183     0.4459   0.714   0.475
```

```
Dispersion Component: Male
              Estimate Std. Error Z value Pr(>|Z|)
(Intercept)   0.1863     0.4602   0.405   0.686
```

Finally the output shown below reports on the values of the extended likelihood (h-likelihood), $p_v(h)$ which is denoted as marginal likelihood, $p_{\beta,v}(h)$ REML likelihood, and the component $f_{\beta,\lambda}(y_{ij}|\mathbf{v}_i)$ C-likelihood.

```
===== Likelihood Functions Value =====
H-likelihood      : -287.8858
Marginal likelihood: -209.3600
REML likelihood    : -209.5131
C-likelihood       : -136.2331
```

The h-likelihood should be used to select the random effects, the marginal likelihood is the choice for the mean structure simplification, while REML likelihood can be used for inference on the variance components. We obtain similar estimates as in [Lee *et al.* \(2006\)](#).

3.2. Cake data

The chocolate cakes preparation experiment was conducted at Iowa State College. Three recipes for preparing the batter were compared. Cakes were prepared at 6 different baking temperatures, which ranged from 175 up to 225 degrees Celsius. For each recipe 6 cakes were prepared baked at different temperatures, therefore 18 cakes were baked all together and they are referred to as replication. There were 15 replications, therefore in total 270 cakes were baked. Further, in each replication there are 3 different recipes. To cope with such a design

we include two random effects, v_i representing the replication, and v_{ij} for the j -th recipe within the replication. Finally e_{ijk} stands for the error term associated with each cake. The linear predictor of the model is defined as follows:

$$\eta_{ijk} = \text{intercept} + \text{recipe}_j + \text{temp}_k + \text{recipe}_j \cdot \text{temp}_k + v_i + v_{ij}, \quad (18)$$

whereby $i = 1, \dots, 15$ refers to the replicate, $j = 1, 2, 3$ is the recipe and $k = 1, \dots, 6$ to the temperature. The analyzes presented can also be found in [Lee et al. \(2006, pp. 163–166\)](#). We will consider two models. In the first model we consider the breaking angle as a normally distributed response, while in the second model we assume that it has a gamma distribution. Further, in both models we assume a normal distribution for both random effects v_i and v_{ij} . The following syntax is used for a gamma response model:

```
R> cake$repbatch <- 100 * cake$Replicate + cake$Batch
R> R1Cake <- data.frame(int = rep(1, 15))
R> R2Cake <- data.frame(int = rep(1, 45))
R> modCake2 <- HGLMfit(DistResp = "Gamma", DistRand = c("Normal", "Normal"),
+   Link = "Log", LapFix = FALSE, ODEst = TRUE, ODEstVal = 0,
+   formulaMain = Angle ~ as.factor(Recipe) + as.factor(Temperature) +
+   as.factor(Recipe) * as.factor(Temperature) + (1 | Replicate) +
+   (1 | repbatch), formulaOD = ~ 1, formulaRand = list(one = ~ 1,
+   two = ~ 1), DataMain = cake, DataRand = list(R1Cake, R2Cake),
+   Offset = NULL, INFO = TRUE, DEBUG = FALSE)
```

The estimated gamma model yields the following (marginal) likelihood value at maximum:

```
===== Likelihood Functions Value =====
H-likelihood          : -676.3907
Marginal likelihood: -808.0586
REML likelihood       : -848.9244
C-likelihood          : -754.2644
```

The normal model yields a (marginal) likelihood value of -819.54 . This value is lower than for a gamma model, and thus preferable according to the AIC criterion. Following [Lee and Nelder \(1998\)](#) we create the model checking plots for both models using the following code. First some graphics manipulation parameters are set to control the layout of the graphs:

```
R> op <- par(mfrow = c(2, 2), oma = c(1, 1, 2, 1), mar = c(3, 3, 4, 1) + 1.2)
```

Next we copy the standardized deviance residuals as suggested by [Lee et al. \(2006, p. 52\)](#) from the model object `modCake2`, and compute the linear predictor of the model, i.e., η together with mean and transformed mean as indicated in [Nelder \(1990\)](#). This is done as follows:

```
R> res <- modCake2$Details$StdDevianceResidualY
R> xmat <- model.matrix(~ as.factor(Recipe) + as.factor(Temperature) +
+   as.factor(Recipe) * as.factor(Temperature), data = cake)
R> eta <- xmat %*% modCake2$Results$Beta
R> mu <- exp(eta)
R> mu <- log(mu)
```

Diagnostics for Cake Model Gamma

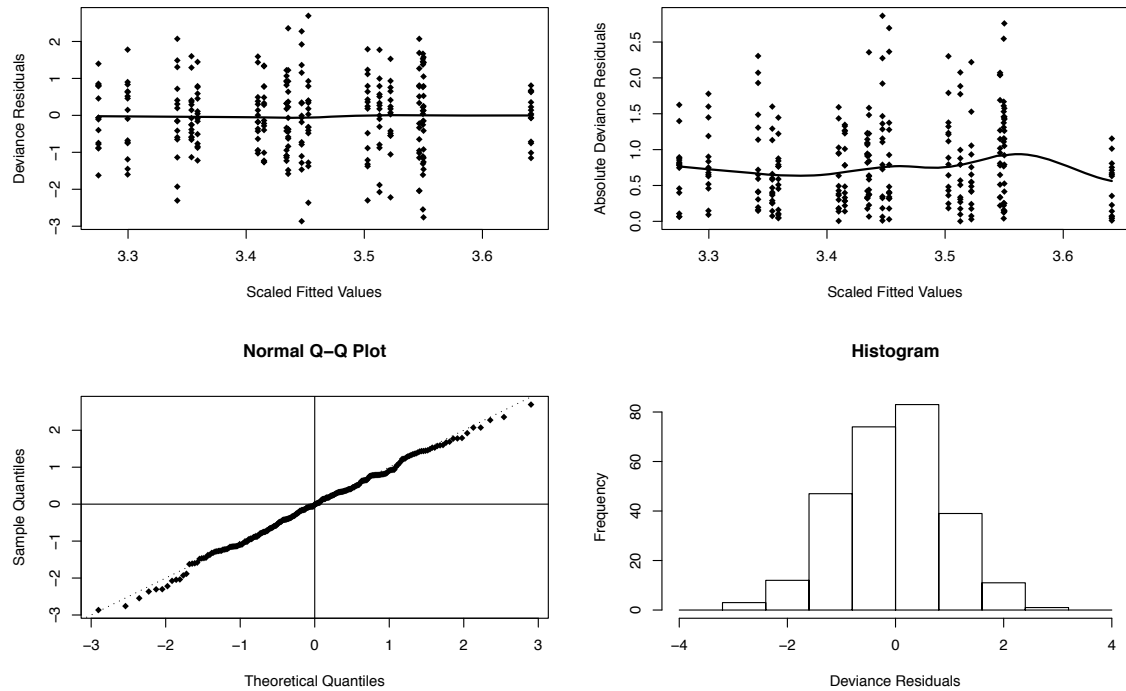


Figure 1: Cake data: Diagnostic plots for the gamma model.

After a few steps, standardized deviance residuals and absolute standardized deviance residuals can be plotted against scaled fitted values, together with a loess curve. In addition a qq-plot of these residuals and histogram are created.

The commands below produce scaled fitted values plot against residuals:

```
R> plot(mu, res, pch = 18, col = "black" xlab = "Scaled Fitted Values",
+       ylab = "Deviance Residuals")
R> los <- loess.smooth(mu, res, span = 1/2, degree = 1,
+       family = c("symmetric", "gaussian"), evaluation = 50)
R> lines(los$x, los$y, col = "black", lwd = 2)
```

Next, we create a plot of scaled fitted values against absolute residuals:

```
R> plot(mu, abs(res), pch = 18, col = "black", xlab = "Scaled Fitted Values",
+       ylab = "Absolute Deviance Residuals")
R> los <- loess.smooth(mu, abs(res), span = 1/2, degree = 1,
+       family = c("symmetric", "gaussian"), evaluation = 50)
R> lines(los$x, los$y, col = "black", lwd = 2)
```

Diagnostics for Cake Model Gamma (Final)

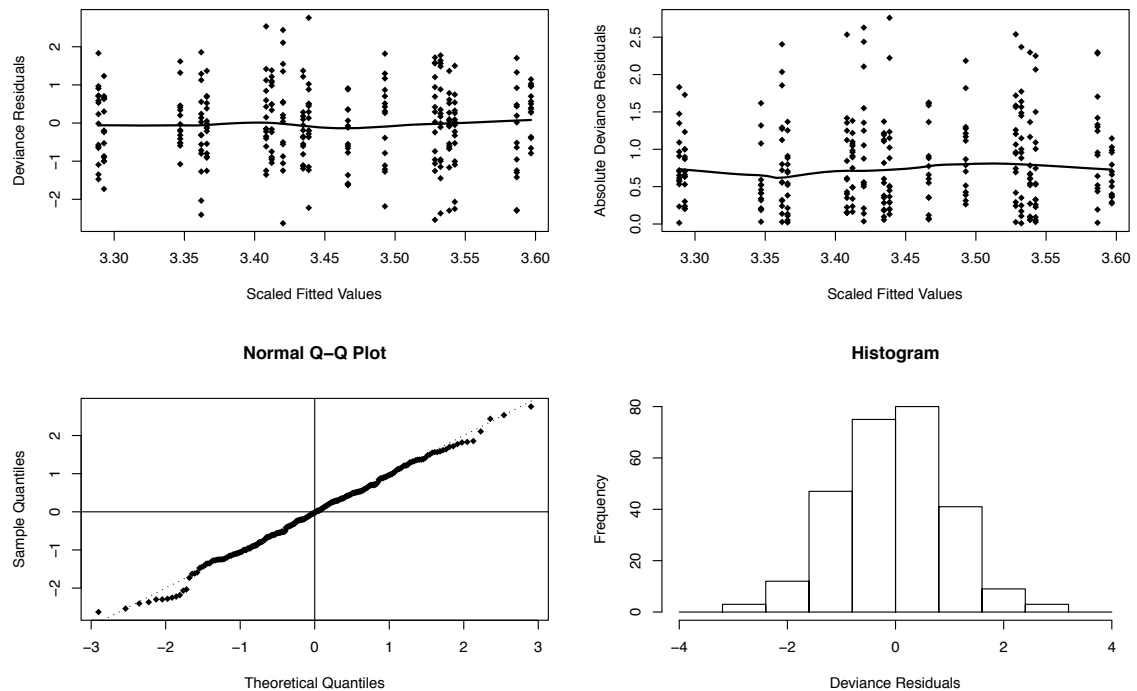


Figure 2: Cake data: Diagnostic plots for the final gamma model.

The qq-plot and histogram of residuals:

```
R> qqnorm(res, pch = 18)
R> abline(h = 0, v = 0)
R> abline(a = 0, b = 1, lty = 3)
R> breaks <- seq(-4, 4, by = 0.8)
R> hist(res, breaks = breaks, xlab = "Deviance Residuals",
+       main = "Histogram")
```

Additional commands provide a title:

```
R> par(op)
R> mtext("Diagnostics for Cake Model Gamma",
+       side = 3, line = 1.5, font = 2, cex = 2, col = "black")
R> par(op)
R> dev.off()
```

The above code produces the diagnostic plots for the gamma model presented in Figure 1.

Further we fitted a gamma model without interaction $\text{recipe}_j * \text{temp}_k$ and compared it with the original model using the likelihood ratio test:

```
R> HGLMLRTest(modCake3, modCake2)
```

The following output is obtained:

```
H-likelihood of model 2 is higher
Marginal likelihood comparison:
LR test p-value: 0.5034955
LR test statistics: 9.304224
LR difference df: 10
REML likelihood comparison:
LR test p-value: NA
LR test statistics: 29.88638
LR difference df: 0
```

The chi-squared test statistics with 10 degrees of freedom indicate that there is not enough evidence for the inclusion of the interaction term into the model (p value: 0.5). Therefore the simplified model is our choice. According to the diagnostics plot on Figure 2 the model seems to be fitting well. Lee *et al.* (2006) analyzed this data using linear mixed model for the original, as well as the transformed response. They also indicated that the gamma model is preferable by the AIC criterion.

3.3. Rat data

Thirty rats were treated with one of three chemotherapy drugs. White blood cell counts (W) and red blood cell counts (R) were taken at each of four different time points. We perform the same analysis as in Lee *et al.* (2006, pp. 224–229). We will focus here on quasi-Poisson model with normal random effects. First we create the design matrix for the dispersion parameter of the random component:

```
R> Rrat <- data.frame(WBC = tapply(rat$WhiteBloodCells, rat$Subject, mean),
+   RBC = tapply(rat$RedBloodCells, rat$Subject, mean))
```

The model can be fitted by:

```
R> modRat2 <- HGLMfit(DistResp = "Poisson", DistRand = c("Normal"),
+   Link = "Log", LapFix = FALSE, ODEst = TRUE, ODEstVal = 0,
+   formulaMain = Y ~ WhiteBloodCells + RedBloodCells + as.factor(Drug) +
+   (1 | Subject), formulaOD = ~ 1, formulaRand = list(
+   one = ~ WBC + I(WBC ^ 2)), DataMain = rat, DataRand = list(Rrat),
+   INFO = TRUE, DEBUG = FALSE)
```

The crucial option here is `ODEst = TRUE`, which requests a quasi-Poisson to be fitted instead of a likelihood-based Poisson model. Note that a summary of the likelihood values is not valid now, as we are not using a likelihood-based technique but extended quasi-likelihood is invoked. By a similar manipulation as in Section 3.2, we ask for the diagnostic plots of the

Diagnostics for Rat Model Quasi-Poisson ylv

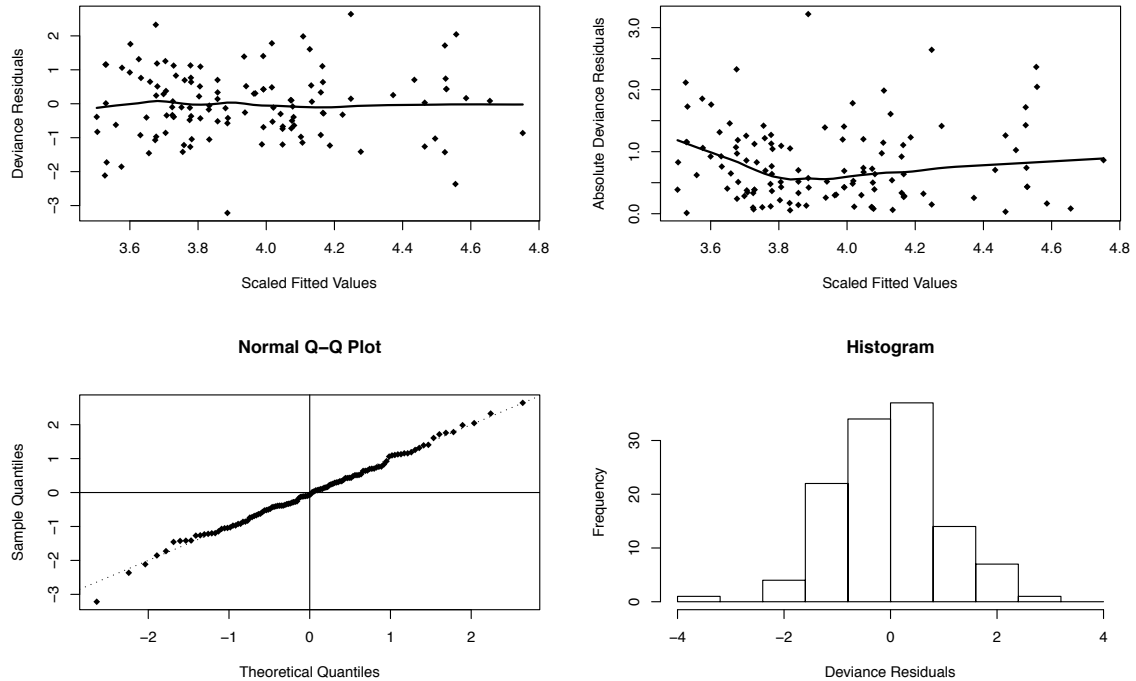


Figure 3: Rat data: Diagnostic plots for the quasi-Poisson model (residual component).

model. Figure 3 presents diagnostic plots for the quasi-Poisson model for the residual error term ($\mathbf{y}|\mathbf{v}$), while Figure 4 gives diagnostics for the choice of the model for the dispersion of the random term (\mathbf{v}). The corresponding diagnostic plots were also presented by Lee *et al.* (2006).

Figure 5 presents the qq-plot for the standardized deviance residuals for the Poisson model (likelihood-based) of the same structure as the above quasi-Poisson model, together with 95% bootstrap envelopes. Clearly the Poisson model does not fit well. The code to generate the plot is the following:

```
R> modRat1 <- HGLMfit(DistResp = "Poisson", DistRand = c("Normal"),
+   Link = "Log", LapFix = FALSE, ODEst = FALSE, ODEstVal = 0,
+   formulaMain = Y ~ WhiteBloodCells + RedBloodCells + as.factor(Drug) +
+   (1 | Subject), formulaOD = ~ 1, formulaRand = ~ list(one = ~ 1),
+   DataMain = rat, DataRand = list(Rrat), INFO = TRUE, DEBUG = FALSE)

R> BootstrapEnvelopeHGLM(modRat1, 19, 9999)
```

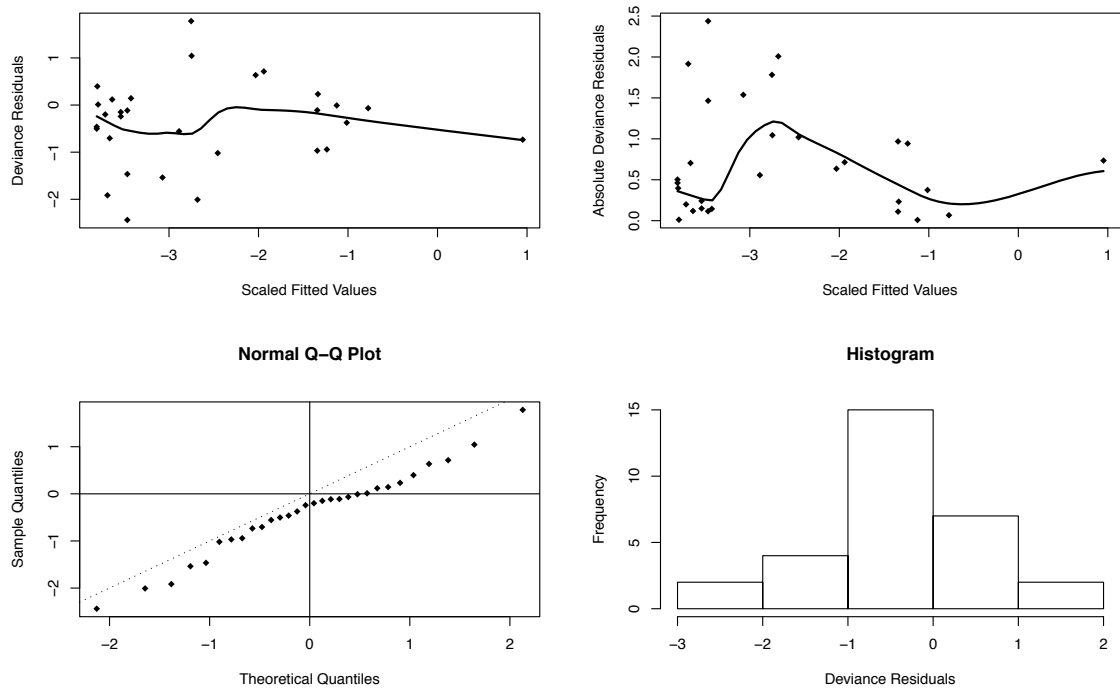

Diagnostics for Rat Model Quasi-Poisson (v)

Figure 4: Rat data: Diagnostic plots for the quasi-Poisson model (dispersion of random component).

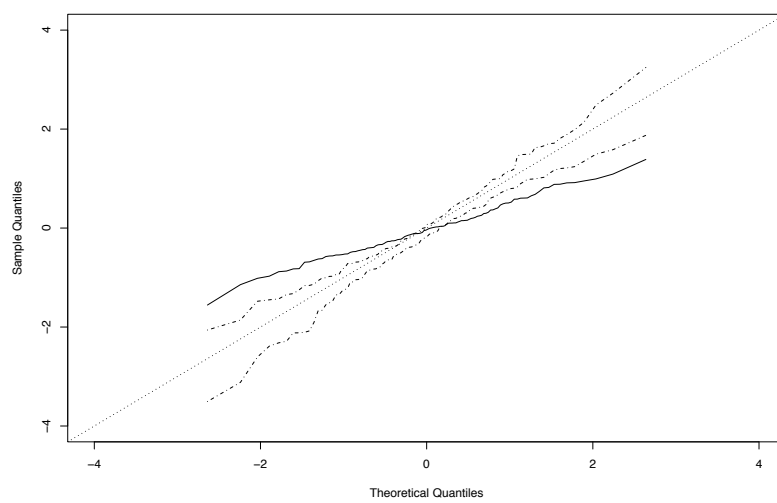


Figure 5: Rat data: Diagnostic plot for the likelihood-based Poisson model.

4. Alternative software for hierarchical data

The methods described in this paper are applied to model hierarchy in the outcomes (resulting in unobserved heterogeneity). There are numerous statistical programs and packages to address this type of modeling. Here we would like to mention a few. Note that that is not our intention to be exhaustive, but rather to focus on most popular programs and which we are acquainted with.

First of all, to our knowledge prior to this R package the h-likelihood algorithms were implemented only in GenStat (Payne, Murray, Harding, Baird, and Soutar 2009) software. The capabilities of **HGLMMM** and GenStat h-likelihood methods are similar. Extra flexibility is provided by **HGLMMM** by the ability to specify several random components, allowing for a different distribution of each component. This is not available in the GenStat program. On the other hand GenStat offers a second order Laplace approximation to estimate variance components, which is not implemented in **HGLMMM**.

To analyze the salamander example the software needs to allow for a crossed random effects design. Gaussian quadrature methods are not applicable in this case as the dimension of integration is often too large. The R package **lme4** (Bates, Mächler, and Bolker 2011) with the function **lmer** allows for the analysis of this example with an ordinary Laplace approximation. The advantage of using **HGLMMM** is the ability to include covariates in the dispersion part of the model. This is not allowed in **lmer**. Packages **lme4** and **HGLMMM** differ in the way the dispersion parameters are estimated. The **HGLMMM** package uses the objective function which is equivalent to the REML approach in linear mixed models. The package **lme4** is however faster and more efficient. Note that **lme4** is a quite popular R package for the analysis of longitudinal data. The salamander data can also be analyzed by the PQL method of Breslow and Clayton (1993) as implemented e.g., in SAS (SAS Institute Inc. 2008) PROC GLIMMIX, which allows the dispersion parameters to depend on categorical covariates.

In the cake example we analyzed the breaking angle of cakes assuming a normal distribution. This analysis can be reproduced in the R package **nlme** (Pinheiro, Bates, DebRoy, Sarkar, and The R Development Core Team 2011) or **lme4** using a random effects model or a linear model with correlated errors. Also the SAS procedures MIXED, NL MIXED or GLIMMIX can be used. When the response follows a gamma distribution SAS procedures GLIMMIX and NL MIXED can be used, but both of them assume the random effects follow a normal distribution. On the other hand, in **HGLMMM** we easily combine non-normal responses with a non-normal random effects distribution. Such a combination can be done in SAS PROC NL MIXED only by using the trick of Liu and Yu (2008) and requires some additional programming. Finally, note that SAS PROC NL MIXED is a likelihood-based procedure, while **HGLMMM** allows the utilization of the extended quasi likelihood, with a overdispersion parameter varying with covariates.

In the rat example an extended quasi likelihood analysis is presented on the number of cancer cell colonies, with the variance of the random effect distribution depending on covariates. Now it is difficult to find standard software which could repeat such an analysis. SAS PROC GLIMMIX comes closest, allowing for overdispersion in Poisson distribution and the variance of random effect to depend on a categorical covariate.

In summary, the package **HGLMMM** blends the distribution of the response from an exponential family with a random effects distribution from the conjugate Bayesian distributions. Further, difficult designs can be handled such as multi-membership designs or more than

2-levels in the hierarchical models. On top of that one can put covariates in mean and (over)dispersion structure. All of these features are not available in one package/software to our knowledge. The limitation of **HGLMMM** is the necessity to assume independent random components, this assumption can be relaxed in R packages **lme4**, **nlme** and SAS PROC MIXED, NLMIXED, GLIMMIX.

5. Future improvements

In future work we would like to adapt the codes to use **Matrix** (Bates and Mächler 2011) package to save memory and to avoid problems with large datasets. Different links for the binomial response such as "Probit" and "CLogLog" need to be implemented as well. In addition, future work will focus on the introduction of the correlation between random effects. In Molas and Lesaffre (2010) the routines were extended to handle hurdle models for count data.

References

- Barndorff-Nielsen O (1980). "Conditionality Resolutions." *Biometrika*, **67**, 293–310.
- Barndorff-Nielsen O (1983). "On a Formula for the Distribution of the Maximum Likelihood Estimator." *Biometrika*, **70**, 343–365.
- Bates D, Mächler M (2011). ***Matrix**: Sparse and Dense Matrix Classes and Methods*. R package version 0.999375-47, URL <http://CRAN.R-project.org/package=Matrix>.
- Bates D, Mächler M, Bolker B (2011). ***lme4**: Linear Mixed-Effects Models Using S4 Classes*. R package version 0.999375-38, URL <http://CRAN.R-project.org/package=lme4>.
- Breslow NE, Clayton DG (1993). "Approximate Inference in Generalized Linear Mixed Models." *Journal of the American Statistical Association*, **88**, 9–25.
- Cox DR, Hinkley DV (1974). *Theoretical Statistics*. Chapman and Hall, London.
- Cox DR, Reid N (1987). "Parameter Orthogonality and Approximate Conditional Inference." *Journal of the Royal Statistical Society B*, **49**, 1–39.
- Harville DA (1977). "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems." *Journal of the American Statistical Association*, **72**, 320–338.
- Henderson CR, Kempthorne O, Searle SR, Krosigk CM (1959). "The Estimation of Environmental and Genetic Trends from Records Subject to Culling." *Biometrics*, **15**, 192–218.
- Lee Y, Nelder JA (1996). "Hierarchical Generalized Linear Models." *Journal of the Royal Statistical Society B*, **58**, 619–678.
- Lee Y, Nelder JA (1998). "Generalized Linear Models for the Analysis of Quality-Improvement Experiments." *Canadian Journal of Statistics*, **26**, 95–105.

- Lee Y, Nelder JA (2001). “Hierarchical Generalised Linear Models: A Synthesis of Generalised Linear Models, Random-Effect Models and Structured Dispersions.” *Biometrika*, **88**, 987–1006.
- Lee Y, Nelder JA, Pawitan Y (2006). *Generalized Linear Models with Random Effects*. Chapman & Hall/CRC, Boca Raton.
- Lesaffre E, Spiessens B (2001). “On the Effect of the Number of Quadrature Points in a Logistic Random-Effects Model: An Example.” *Applied Statistics*, **50**, 325–335.
- Liu L, Yu Z (2008). “A Likelihood Reformulation Method in Non-Normal Random Effects Models.” *Statistics in Medicine*, **27**, 3105–3124.
- McCullagh P, Nelder JA (1989). *Generalized Linear Models*. 2nd edition. Chapman and Hall, London.
- Molas M, Lesaffre E (2010). “Hurdle Models for Multilevel Zero-Inflated Data via H-Likelihood.” *Statistics in Medicine*, **29**, 3294–3310.
- Molenberghs G, Verbeke G (2005). *Models for Discrete Longitudinal Data*. Springer-Verlag, New York.
- Nelder JA (1990). “Nearly Parallel Lines in Residual Plots.” *The American Statistician*, **44**, 221–222.
- Nelder JA, Wedderburn RWM (1972). “Generalized Linear Models.” *Journal of Royal Statistical Society A*, **135**, 370–384.
- Noh M, Lee Y (2007). “REML Estimation for Binary Data in GLMMs.” *Journal of Multivariate Analysis*, **98**, 896–915.
- Pawitan Y (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Clarendon Press, Oxford.
- Payne RW, Murray DA, Harding SA, Baird DB, Soutar DM (2009). *GenStat for Windows (12th Edition) Introduction*. VSN International, Hemel Hempstead, UK.
- Pinheiro J, Bates D, DebRoy S, Sarkar D, The R Development Core Team (2011). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-98, URL <http://CRAN.R-project.org/package=nlme>.
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- SAS Institute Inc (2008). *SAS/STAT Software, Version 9.2*. Cary, NC. URL <http://www.sas.com/>.
- Severini TA (2000). *Likelihood Methods in Statistics*. Oxford University Press, Oxford.
- Verbeke G, Molenberghs G (2000). *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York.

Affiliation:

Marek Molas

Department of Biostatistics

Erasmus Medical Center in Rotterdam

3000 CA Rotterdam, The Netherlands

E-mail: m.molas@erasmusmc.nl