Sensitivity analysis in Bayesian generalized linear mixed models for binary data

Małgorzata Roos, Leonhard Held

Division of Biostatistics, Institute for Social and Preventive Medicine, University of Zurich, 8001 Zurich, Switzerland

Abstract

Generalized linear mixed models (GLMMs) enjoy increasing popularity because of their ability to model correlated observations. Integrated nested Laplace approximations (INLAs) provide a fast implementation of the Bayesian approach to GLMMs. However, sensitivity to prior assumptions on the random effects precision parameters is a potential problem. To quantify the sensitivity to prior assumptions, we develop a general sensitivity measure based on the Hellinger distance to assess sensitivity of the posterior distributions with respect to changes in the prior distributions for the precision parameters. In addition, for model selection we suggest several cross-validatory techniques for Bayesian GLMMs with a dichotomous outcome. Although the proposed methodology holds in greater generality, we make use of the developed methods in the particular context of the well-known salamander mating data. We arrive at various new findings with respect to the best fitting model and the sensitivity of the estimates of the model components.

Keywords:

Bayesian Analysis, Binary data, Generalized Linear Mixed Models, Hellinger distance, Integrated nested Laplace approximations, Model choice, Sensitivity analysis.

Email address: `mroos@ifspm.uzh.ch` (Małgorzata Roos)

1. Introduction

Generalized linear mixed models (GLMMs) allow for correlated responses through the inclusion of random effects in the linear predictor. They greatly extend the range of possible applications beyond ordinary generalized linear models but involve more difficult and challenging computational issues. Several inference procedures have therefore been proposed. A Bayesian analysis is often performed using Markov chain Monte Carlo (MCMC). However, recently Rue et al. (2009) introduced a novel numerical inference approach, so-called integrated nested Laplace approximations (INLAs). With this approach, MCMC sampling becomes redundant as the posterior marginal distributions are accurately approximated in a fully automated way. In addition, INLA allows the computation of many Bayesian GLMMs in a reasonable amount of time, enabling an extensive comparison of different models and prior distributions.

In this paper we make use of a new method for sensitivity investigations exclusively for GLMMs with binary response and describe our approach in the particular context of the salamander mating data, although it applies more widely. The well-known salamander mating data set was first described and analyzed by McCullagh and Nelder (1989). It has drawn the attention of statisticians because of the demanding crossed study design and has been used by numerous researchers (a list with 24 references can be obtained on request from the authors) to illustrate GLMM estimation procedures using different model specifications.

Zeger and Karim (1991) and Karim and Zeger (1992) adopted the Gibbs sampling method for GLMMs to analyze the salamander mating data. Recently, Fong et al. (2010) proved the applicability of INLA to analyze Bayesian GLMMs. Karim and Zeger (1992) computed the marginal likelihood of the models and Lee and Nelder (1996) referred to the analysis of deviance when discussing model choice techniques in the context of the salamander mating data. Based on the most recent developments on predictive measures and their availability in INLA we suggest several alternative model selection techniques for Bayesian GLMMs with a binary outcome.

A crucial problem in the formulation of Bayesian GLMMs is the specification of the prior distribution for the random effects precision parameters. Lunn et al. (2009a) argue that the choice of the gamma prior G(ϵ,ϵ) with small ϵ is generally inappropriate. They regret that BUGS users frequently rely on these priors as a default choice. However, Wakefield (2009) recommended a probabilistic derivation of the gamma priors by considering the residual odds for binary data. Fong et al. (2010) proposed a particular choice of the gamma prior distributions for the random effects precisions for the salamander mating data set.

In the rejoinder Lunn et al. (2009b) emphasized that their preference is to avoid gamma priors for precision parameters altogether. They recommend to follow Gelman (2006) in using half-normal or half-Cauchy priors for random-effect standard deviations.

The choice of the prior for random-effect precision can be critical when the number of groups is large compared to the number of observations in each group (Box and Tiao, 1973, Chapter 7). We face an important practical question: "How sensitive is the posterior distribution of the random effect to changes in the prior distribution?" To address this problem Geisser (1993) suggested perturbation of the model in potentially conceivable directions to determine the effect of such alterations on the analysis. He discussed both the Kullback-Leibler divergence and Hellinger distance as possible perturbation diagnostics.

Sensitivity to prior assumptions for the random effect precision was discussed in Browne and Draper (2006) and discussion articles following it (Gelman, 2006; Kass and Natarajan, 2006; Lambert, 2006). Geisser (1993) admitted that a Bayesian analysis may depend critically on the modeling assumptions. Sensitivity analysis has been also suggested using MCMC. Besag et al. (1995) discuss that a sensitivity analysis using MCMC can be carried out using importance sampling, which speeds up the computations considerably. Narasimhan (2005) describes software facilitating dynamic exploration of posteriors by means of importance sampling. The changes in the posterior distributions given the changes in the priors can be assessed visually.

A general measure of sensitivity based on the Kullback-Leibler divergence together with its calibration was developed in McCulloch (1989). Here we adopt this method but resort to the Hellinger distance for reasons which will be discussed. We propose a general sensitivity measure based on the Hellinger distance together with its calibration. This calibration enables us to assess the relevance of the observed discrepancies between the posterior distributions. Moreover, we study micro- and macro-sensitivity of the posterior mean estimates of the fixed effects and the random effects precisions on a grid of prior values for the hyperparameters. We arrive at interesting new findings regarding the sensitivity of the estimates for the salamander mating data set.

The remainder of this article is organized as follows: Section 2 reviews the salamander mating data and discusses several random-effects specifications together with the hyperprior distributional assumptions. Section 3 briefly discusses the INLA approach in the context of the salamander data set and presents several model selection measures. In Section 4 a sensitivity measure based on the Hellinger distance together with its calibration is proposed, the selection of gamma and half-normal priors is discussed and a short description of the sensitivity analysis on the grid of the hyperprior values is provided. In Section 5 results are presented. Some concluding remarks are given in Section 6.

2. Data and models

The detailed description of 11 crosses between salamanders from nine geographically isolated populations can be found in Verrell and Arnold (1989). The salamander mating data set, as introduced by McCullagh and Nelder (1989), deals with one particular cross of two populations. The salamanders were named after the locality of the population from which the animals in the study were collected from: the Rough Butt Bald in Great Balsam Mountains and the Whiteside Mountain on Highlands Plateau.

Data consist of three separate experiments, each performed according to the design described by McCullagh and Nelder (1989) in Table 14.3. The first one was conducted in summer of 1986 and the other two in fall 1986. The animals used for the first and the second experiment were identical. A new set of salamanders was used in the third experiment.

In one particular experiment there were two groups of 20 salamanders each. Every group comprised five male Rough Butt, five male Whiteside, five female Rough Butt and five female Whiteside salamanders. Within each group, all animals were paired to three animals of the opposite sex from the same population and to three animals from the other population. Therefore, $10 \times 6 = 60$ male-female pairs were formed within one group leading to $2 \times 60 = 120$ binary observations for each experiment. A successful mating was indicated by 1 and an unsuccessful one by 0. McCullagh and Nelder (1989) report in Tables 14.4 through 14.6 on the 360 observed binary outcomes on mating success in these three experiments.

2.1. Review of the models

The main scientific question addressed in the study was whether the mating of both geographically isolated species of salamanders was as successful as the one between the animals from the same population. Moreover, there was some interest if a seasonal effect could be identified. Therefore, two factors `fW` (Whiteside female "yes": 1, "no": 0) and `mW` (Whiteside male "yes": 1, "no": 0) together with their interaction `WW` and a seasonal effect `fall` (experiment conducted in fall "yes": 1, "no": 0) were defined. The interaction coefficient, representing the cross effect, was of primary interest.

Let Y_{ijk} denote the binary outcome of the mating for female i and male j in experiment k where $i = 1, \ldots, 20$, $j = 1, \ldots, 20$ and $k = 1, 2, 3$. Let Y_{ijk} denote a Bernoulli random variable with success probability π_{ijk} and logit link function

$$\log \frac{\pi_{ijk}}{1 - \pi_{ijk}} = \boldsymbol{x}_{ijk}^T \boldsymbol{\beta} + b_{ik}^F + b_{jk}^M,$$

where $\boldsymbol{x}_{ijk}$ is a vector comprising the intercept, $\texttt{fW}_{ik}$, $\texttt{mW}_{jk}$, $\texttt{WW}_{ijk}$ and $\texttt{fall}_{ijk}$ variables, $\boldsymbol{\beta}$ is the corresponding vector of the fixed effects parameters and b_{ik}^F and b_{jk}^M are normally distributed random effects with mean zero of females and males respectively. Their precise specifications are discussed below and can be found in Table 1.

We have fitted several models to these data. In model A it is ignored that the same salamanders were used in the first two experiments. The data are modeled as if different

Table 1: Definition of female and male random effects for models A, B1, B2, B3, B4 and C. Here the precision is denoted by τ. Notation (k) in model C indicates that the data from each experiment are analyzed separately.

Model	Female	Male
A	$b^F_{ik} \overset{iid}{\sim} \mathrm{N}(0, \tau^{-1}_{F(1,2,3)})$	$b^M_{jk} \overset{iid}{\sim} \mathrm{N}(0, \tau^{-1}_{M(1,2,3)})$
B1	$(b^F_{i1}, b^F_{i2})^T \sim \mathrm{N}_2(\mathbf{0}, \boldsymbol{W}_F^{-1})$, $\boldsymbol{W}_F \sim \mathrm{Wishart}_2(r, \boldsymbol{R}_F^{-1})$, $b^F_{i3} \overset{iid}{\sim} \mathrm{N}(0, \tau^{-1}_{F3})$	$(b^M_{j1}, b^M_{j2})^T \sim \mathrm{N}_2(\mathbf{0}, \boldsymbol{W}_M^{-1})$, $\boldsymbol{W}_M \sim \mathrm{Wishart}_2(r, \boldsymbol{R}_M^{-1})$, $b^M_{j3} \overset{iid}{\sim} \mathrm{N}(0, \tau^{-1}_{M3})$
B2	$b^F_{ik} \overset{iid}{\sim} \mathrm{N}(0, \tau^{-1}_{Fk})$, for $k = 1, 2, 3$	$(b^M_{j1}, b^M_{j2})^T \sim \mathrm{N}_2(\mathbf{0}, \boldsymbol{W}_M^{-1})$, $\boldsymbol{W}_M \sim \mathrm{Wishart}_2(r, \boldsymbol{R}_M^{-1})$, $b^M_{j3} \overset{iid}{\sim} \mathrm{N}(0, \tau^{-1}_{M3})$
B3	$b^F_{ik} \overset{iid}{\sim} \mathrm{N}(0, \tau^{-1}_{Fk})$, for $k = 1, 2, 3$	$b^M_{j12} \overset{iid}{\sim} \mathrm{N}(0, \tau^{-1}_{M12})$, $b^M_{j3} \overset{iid}{\sim} \mathrm{N}(0, \tau^{-1}_{M3})$
B4	$b^F_{ik} \overset{iid}{\sim} \mathrm{N}(0, \tau^{-1}_{Fk})$, for $k = 1, 2, 3$	$b^M_{jk} \overset{iid}{\sim} \mathrm{N}(0, \tau^{-1}_{Mk})$, for $k = 1, 2, 3$
C	$b^F_{ik} \overset{iid}{\sim} \mathrm{N}(0, \tau^{-1}_{F(k)})$, for $k = 1, 2, 3$	$b^M_{jk} \overset{iid}{\sim} \mathrm{N}(0, \tau^{-1}_{M(k)})$, for $k = 1, 2, 3$

sets of 40 salamanders were used in each experiment. The random effects b_{ik}^F and b_{jk}^M are allowed to have different precisions, which are assumed to be the same from one experiment to the next.

Model B1 accounts for the fact that the same animals were utilized in two experiments by assuming that the corresponding random effects for each animal are bivariate normal. Different but correlated effects for a single animal used in the first and in the second experiment are allowed. The precision matrix is gender-specific.

Karim and Zeger (1992), Breslow and Clayton (1993) and Chan and Kuk (1997) found evidence for correlation of the random effects for males but not for females. Therefore, we consider three additional variations of model B1. In model B2 we assume that the random effects of females between the two experiments are independent. Model B3 assumes that the male random effects for the first two experiments are identical and the female random effects are independent. In model B4 independence of random effects in each experiment is assumed.

Finally, in model C it is assumed that the salamanders from different experiments are independent and the data from each experiment are analyzed separately. Consequently, separate fixed and random effects are estimated for each experiment. The estimates of the fixed effects ($\boldsymbol{\beta}_k$) in model C hence depend on the experiment (k) in contrast to the models A and B discussed above.

2.2. Hyperpriors

Fong et al. (2010) suggest in "Supplementary material" available at *Biostatistics online* the use of the gamma distribution $\mathrm{G}(a_1, a_2)$ with mean a_1/a_2 and variance a_1/a_2^2 and density $f(\tau) = a_2^{a_1}\Gamma^{-1}(a_1)\tau^{a_1-1}\exp(-a_2\tau)$, where $a_1 = 1$ and $a_2 = 0.622$ as a prior distribution for the precision τ of the random effects in the salamander mating data. We therefore consider $\boldsymbol{\theta}_0 = (1, 0.622)$ as a default choice. Alternatively, we assume that $\kappa_0 = 0.01$ is the default value of the parameter κ_0 of the half-normal hyperprior $\mathrm{HN}(\kappa_0)$ for the random-effect standard deviation $\sigma = \tau^{-1/2}$ with density $f(\sigma) = 2\kappa_0^{1/2}(2\pi)^{-1/2}\exp(-\kappa_0\sigma^2/2)$. Note that assuming a $\mathrm{HN}(\kappa_0)$ distribution for the standard deviation of the hyperparameter corresponds to assuming a gamma distribution with parameters $1/2$ and $\kappa_0/2$ for the variance (rather than for the precision as in Fong et al., 2010).

3. Estimation and model selection

3.1. INLA and inference based on deterministic approximations

The `inla` program, available at `http://www.r-inla.org`, allows the user to conveniently perform approximate Bayesian inference in latent Gaussian models. An R package called INLA serves as an interface to the `inla` program and its usage is similar to the `glm` function in R (see `http://www.r-inla.org/examples/volume-ii/code-for-model-b-of-salamander-data`). The R interface is easy to use. Its standard output encompasses marginal posterior densities of all parameters in the model together with summary characteristics. Furthermore, several model choice criteria (see Section 3.2) are available in `inla`.

Latent Gaussian Markov random fields (GMRF) models, which underlie INLA, are described in detail in Rue and Held (2005). A hierarchical GMRF model is characterized through three stages of observables and parameters. First, the distributional assumption for the observables dependent on latent parameters is formulated. Second, an a priori model for the unknown parameters is assigned and the corresponding GMRF $\boldsymbol{x}$ is specified. For the models described in Section 2.1, $\boldsymbol{x} = (\boldsymbol{\beta}^T, (\boldsymbol{b}^F)^T, (\boldsymbol{b}^M)^T)$.

By default a flat improper prior for the intercept β_0 is assumed in `inla`. All other components of $\boldsymbol{\beta}$ are assumed to be independent zero-mean Gaussian $\mathrm{N}(0, \sigma^2)$ with fixed precision $\sigma^{-2} = 0.0001$ *a priori*.

The definition of the latent model is completed by assigning prior distributions to the hyperparameters. A gamma prior for τ with values $(a_1,\ a_2)$ in the independent random effects model is the default choice. The prior distribution of the precision matrix $\boldsymbol{W}_F$ ($\boldsymbol{W}_M$) for correlated random effects is assumed to be $\mathrm{Wishart}_2(r, \boldsymbol{R}^{-1})$ (see `http://www.math.ntnu.no/~hrue/r-inla.org/doc/latent/iid123d.pdf`). By setting $r = 2a_1 + 1$, $R_{ii} = 2a_2$ for $i = 1, 2$ and $R_{12} = 0$ (Fong et al., 2010) we ensure that the marginal distribution of the precisions in the bivariate specification is equal to the gamma prior in the independent random effects model. Alternatively, a half-normal prior with parameter κ can be assumed for $\tau^{-1/2}$ in the independent random effects model.

We used the default INLA settings: the simplified Laplace approximation strategy and the Central Composite Design (CCD). These options give rise to both quick and reliable estimates of the models. Throughout we have used the function `inla.hyperpar()` to obtain improved estimates of the marginal posterior densities of the hyperparameters. This manuscript has been generated with the INLA-R-Interface-Version generated on the 15th of December 2010. For a detailed description of the INLA methodology we refer to Rue et al. (2009) and Fong et al. (2010).

3.2. Model selection

The Deviance Information Criterion (DIC) is contained in the standard `inla` output. It is a well-known Bayesian model choice criterion for comparing complex hierarchical models (Spiegelhalter et al., 2002). Lower DIC values correspond to better models. However, DIC is problematic in models with many random effects (Plummer, 2008).

Alternatively, `inla` also provides the log marginal likelihood (LML). The marginal likelihood for a certain model M is defined as $\pi(\mathbf{y}|M) = \int \pi(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta}|M) d\mathbf{x} d\boldsymbol{\theta}$ and can be used as a basis for model comparison. Larger values of the LML correspond to a better model. However, the LML reacts severely when nearly improper priors are used. Nevertheless, the LML can be used to compare models with identical improper priors on the same parameters.

The conditional predictive ordinate (CPO) as discussed in Geisser (1993) can be obtained from `inla` as a predictive measure. The CPO value given by $\text{CPO}_{ijk} = \pi(y_{ijk,obs}|\mathbf{y}_{-(ijk),obs})$ is defined as the cross-validated predictive density $\pi(y_{ijk}|\mathbf{y}_{-(ijk)})$ at the observation $y_{ijk,obs}$, where $\mathbf{y}_{-(ijk),obs}$ denotes the data without the ijk'th observation. For the salamander data we use the mean logarithmic CPO defined as

$$\overline{\text{LCPO}} = -\frac{1}{360} \sum_{i,j,k} \log(\text{CPO}_{ijk}).$$

This can be identified as the cross-validated logarithmic score (Gneiting and Raftery, 2007; Held et al., 2010), which measures the predictive quality of a model. Stone (1977) showed its asymptotic equivalence to AIC. Lower values of $\overline{\mathrm{LCPO}}$ indicate a better model.

For binary response the goodness of fit of a model is the degree to which the fitted probabilities of a successful mating for the ijk'th observation $\hat{P}[y_{ijk} = 1]$ coincide with the observed binary outcomes $y_{ijk,obs}$. The mean Brier score (e.g. Schmid and Griffith, 2005)

$$\frac{1}{360} \sum_{i,j,k} (y_{ijk,obs} - \hat{P}[y_{ijk} = 1])^2 \tag{1}$$

is often used in this context. The cross-validated Brier Score $\overline{\mathrm{BS}}$ is defined as in (1) but using

$$\hat{P}[y_{ijk} = 1 | y_{ijk,obs}] = \begin{cases} \pi(y_{ijk} = y_{ijk,obs} | \mathbf{y}_{-(ijk),obs}), & \text{if } y_{ijk,obs} = 1 \\ 1 - \pi(y_{ijk} = y_{ijk,obs} | \mathbf{y}_{-(ijk),obs}), & \text{if } y_{ijk,obs} = 0 \end{cases} \tag{2}$$

instead of $\hat{P}[y_{ijk} = 1]$. Note that (2) is a function of CPO_{ijk} and y_{ijk} alone, so directly available in INLA. Lower values of $\overline{\mathrm{BS}}$ indicate a better model.

To evaluate discrimination of a particular model the area under the curve (AUC) is often used in a ROC analysis (e.g. Pepe, 2003). Here we suggest the use of the cross-validated probability $\hat{P}[y_{ijk} = 1 | y_{ijk,obs}]$ defined in (2) for comparison with the observed values of success to obtain a cross-validated AUC. Higher values of AUC indicate a better discrimination of the model.

4. Sensitivity analysis and choice of prior distributions

4.1. Sensitivity measure and its calibration

There are several ways of investigating the sensitivity of the estimates and their posterior distributions to the choice of the hyperprior distributions. One of the questions is if the posterior distributions of the fixed effects change and how the change can be quantified. Another question is to what extent the posterior distribution of the random effects precision is sensitive to the changes in their priors.

McCulloch (1989) developed a simple but fairly general method for comparing the influence of the prior distribution on the posterior distribution based on the Kullback-Leibler divergence. He suggested a measure of sensitivity which warns if the results are overly sensitive to the choices of the priors. Moreover, he provided a useful calibration of both the Kullback-Leibler divergence and the sensitivity measure. Unfortunately, the Kullback-Leibler divergence turns infinity as soon as one of the densities attains the value 0 and the other one not. This happens quite frequently in our application as the estimated marginal posterior distributions provided numerically by `inla` attain nonzero values only on a finite discrete set of points. Consequently, the Kullback-Leibler divergence is not applicable in our setting. In contrast the Hellinger distance does not have this undesirable feature.

For a default $\boldsymbol{\theta}_0$ and a shifted $\boldsymbol{\theta}$ prior value let

$$S(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = \frac{H(post(\boldsymbol{\theta}_0), post(\boldsymbol{\theta}))}{H(pri(\boldsymbol{\theta}_0), pri(\boldsymbol{\theta}))} \tag{3}$$

denote the relative change in the posterior distribution with respect to the change in the prior distribution as measured by the Hellinger distance H, where $pri(\boldsymbol{\theta})$ is the prior (no-data) distribution and $post(\boldsymbol{\theta})$ is the corresponding posterior distribution. The Hellinger distance (Le Cam, 1986) is a symmetric measure of discrepancy between two densities f and g:

$$H(f, g) = \sqrt{\frac{1}{2} \int_{-\infty}^{\infty} \left(\sqrt{f(u)} - \sqrt{g(u)}\right)^2 du} = \sqrt{1 - BC(f, g)}.$$

Here $BC(f, g) = \int_{-\infty}^{\infty} \sqrt{f(u)g(u)} du$ denotes the Bhattacharayya coefficient (Bhattacharyya, 1943). The Hellinger distance is equal to 0 if and only if both densities are equal. It takes maximal value equal to 1 if BC is equal to 0. This happens whenever the density f assigns probability 0 to every set to which the density g assigns a positive probability and vice versa.

Note that the Hellinger distance is invariant to any one-to-one transformation (for example logarithmic, inverse or square-root) of both densities (Jeffreys, 1961, p. 180). To see this, let X_1 and X_2 be random variables with densities f_{X_1} and f_{X_2} respectively.

Let $Y_1 = g(X_1)$ and $Y_2 = g(X_2)$, where $g(x)$ is a strictly monotonic (increasing) function with inverse $g^{-1}(y)$ having a continuous derivative $dg^{-1}(y)/dy$. Then

$$\begin{aligned} BC(f_{Y_1}, f_{Y_2}) &= \int_{g(-\infty)}^{g(\infty)} \sqrt{f_{Y_1}(y) f_{Y_2}(y)} dy \\ &= \int_{g(-\infty)}^{g(\infty)} \sqrt{f_{X_1}(g^{-1}(y)) \Big| \frac{dg^{-1}(y)}{dy} \Big| f_{X_2}(g^{-1}(y)) \Big| \frac{dg^{-1}(y)}{dy} \Big|} dy \\ &= \int_{g(-\infty)}^{g(\infty)} \Big| \frac{dg^{-1}(y)}{dy} \Big| \sqrt{f_{X_1}(g^{-1}(y)) f_{X_2}(g^{-1}(y))} dy \\ &= \int_{g^{-1}(g(-\infty))}^{g^{-1}(g(\infty))} \sqrt{f_{X_1}(x) f_{X_2}(x)} dx \\ &= \int_{-\infty}^{\infty} \sqrt{f_{X_1}(x) f_{X_2}(x)} dx = BC(f_{X_1}, f_{X_2}). \end{aligned}$$

The invariance of the Hellinger distance is important in our application since we specify prior distributions both for precisions and corresponding standard deviations.

Similar to McCulloch (1989) for the Kullback-Leibler divergence we provide a calibration of the Hellinger distance H. Let $\text{Ber}(q)$ denote a Bernoulli distribution with success probability q. Then it can be shown that

$$H(\text{Ber}(1/2), \text{Ber}(q)) = \sqrt{1 - \sqrt{(1-q)/2} - \sqrt{q/2}}. \tag{4}$$

Note that (4) attains its maximum $\sqrt{1 - \sqrt{1/2}} = 0.541$ at $q = 0$ and $q = 1$. Conversely, for a given Hellinger distance h smaller than 0.541 we can derive $q(h)$ such that $H(\text{Ber}(1/2), \text{Ber}(q(h))) = h$. Let $h(q)$ be the Hellinger distance corresponding to the probability $q = q(h)$. Using (4) we obtain

$$q(h) = \left(1 + \sqrt{1 - 4((1-h^2)^2 - 1/2)^2}\right)/2.$$

Assuming that $q(h)$ is between 0.5 and 1, this expression is invertible:

$$h(q) = \sqrt{1 - \sqrt{1/2 + \sqrt{q(1-q)}}}.$$

Table 2 shows the desired calibration $q(h)$ for a range of h values. It tells us that the Hellinger distance h between two distributions is the same as that between a $\text{Ber}(1/2)$

Table 2: Calibration of the Hellinger distance.

$h = h(q)$	$q(h) = q$
0.000	0.500
0.035	0.550
0.071	0.600
0.108	0.650
0.145	0.700
0.185	0.750
0.227	0.800
0.272	0.850
0.325	0.900
0.391	0.950
0.541	1.000

and a $\text{Ber}(q(h))$ distribution. This latter difference is one that we can easier interpret. For simplicity, we denote it a $q(h) - 1/2$ distance in Bernoulli probability scale. In our application we never observe Hellinger distances larger than 0.541. For larger distances this definition would need appropriate amendment.

We now provide the calibration for the sensitivity measure S as introduced in (3):

$$C(S, p) = q(S \times h(p)). \tag{5}$$

Note that if $S > 1$ then $C(S, p) > p$ and if $S < 1$ then $C(S, p) < p$ and if $S = 1$ then $C(S, p) = p$. The value of $C(S, p)$ can be interpreted as follows. Given a pair of priors whose Hellinger distance is calibrated by p, then the Hellinger distance between the corresponding posteriors is calibrated by $C(S, p)$.

4.2. Application to gamma and half-normal hyperpriors

We first explain the general approach described above for gamma hyperpriors. Alternatively, half-normal priors will be considered.

It can be shown that for any two gamma densities with values $\boldsymbol{\theta}_1 = (a_{11}, a_{12})$ and $\boldsymbol{\theta}_2 = (a_{21}, a_{22})$

$$H(pri(\boldsymbol{\theta}_1), pri(\boldsymbol{\theta}_2)) = \sqrt{1 - \Gamma\Big(\frac{a_{11}+a_{21}}{2}\Big)\sqrt{\frac{a_{12}^{a_{11}} a_{22}^{a_{21}}}{\Gamma(a_{11})\Gamma(a_{21})(\frac{a_{12}+a_{22}}{2})^{a_{11}+a_{21}}}}}. \tag{6}$$

For $\boldsymbol{\theta}_0 = (1, 0.622)$ and the two alternatives $\boldsymbol{\theta}_m = (1, 0.311)$ and $\boldsymbol{\theta}_p = (1, 0.933)$ we have $H(pri(\boldsymbol{\theta}_0), pri(\boldsymbol{\theta}_m)) = 0.239$ and $H(pri(\boldsymbol{\theta}_0), pri(\boldsymbol{\theta}_p)) = 0.142$. For example, because $q(0.239) = 0.814$ (see Table 2) this means that using a $\mathrm{G}(1, 0.311)$ instead of a $\mathrm{G}(1, 0.622)$ prior corresponds to a distance of $0.814 - 1/2 = 0.314$ in Bernoulli probability scale.

The Hellinger distances between the prior distributions can be now compared with the distances of the corresponding posterior distribution, see Section 5.3. Our investigations benefit from the ability of INLA to provide the marginal posterior distributions of every desirable quantity in the model. The `inla` function called `inla.dmarginal()` evaluates the posterior densities at a fixed grid of values. The Hellinger distance between two posterior densities say $post(\boldsymbol{\theta}_0)$ and $post(\boldsymbol{\theta})$ can now be evaluated numerically at a finite set of integration points k

$$H(post(\boldsymbol{\theta}_0), post(\boldsymbol{\theta})) = \sqrt{1 - \sum_k \sqrt{post(\boldsymbol{\theta}_0)(k) post(\boldsymbol{\theta})(k)} \Delta_k},$$

with area weights Δ_k provided by the trapezoidal rule.

The Hellinger distance between two half-normal $\mathrm{HN}(\kappa)$ priors with values κ_1 and κ_2 can be computed according to the following formula

$$H(pri(\kappa_1), pri(\kappa_2)) = \sqrt{1 - \frac{(\kappa_1 \kappa_2)^{1/4}}{\sqrt{\frac{\kappa_1+\kappa_2}{2}}}}.$$

Note that due to the invariance of the Hellinger distance, this is in fact a special case of Equation (6) with $\boldsymbol{\theta}_1 = (a_{11}, a_{12}) = (1/2, \kappa_1/2)$ and $\boldsymbol{\theta}_2 = (a_{21}, a_{22}) = (1/2, \kappa_2/2)$ (see Section 2.2 for a justification). For the default choice $\kappa_0 = 0.01$ and the alternative values $\kappa_m = 0.0269$ and $\kappa_p = 0.00563$, respectively, the Hellinger distances for a prior change from $\mathrm{HN}(\kappa_0)$ to $\mathrm{HN}(\kappa_m)$ is 0.239 and from $\mathrm{HN}(\kappa_0)$ to $\mathrm{HN}(\kappa_p)$ is 0.142. As intended, our choice of κ_m and κ_p leads to the same Hellinger distances as for the gamma priors just below Equation (6).

4.3. Sensitivity on the prior grid

Following suggestions by Geisser (1993) we also investigated sensitivity of the posterior mean estimates for both large and small perturbations of the gamma and half-normal hyperpriors. In this way we can evaluate both macro- and micro-sensitivity. Gamma priors are more challenging in this context as changes in two parameters have to be examined. The following 3×3 grid of prior values has been considered: $\mathrm{G}(a_1, a_2)$, where $a_1 \in \{0.5,\ 1,\ 1.5\}$ and $a_2 \in \{0.311,\ 0.622,\ 0.933\}$. Consequently, there are 9 pairs of values (a_1, a_2) called central. For each central pair 9 models with slightly perturbed prior values $(a_1 + \epsilon_1, a_2 + \epsilon_2)$, where $\epsilon_1, \epsilon_2 \in \{-0.01,\ 0,\ 0.01\}$ have been fitted.

A similar idea can be applied to the half-normal priors. We consider three centers $0.0269, 0.01, 0.00563$ for the κ value. For each central value 3 models with slightly perturbed prior values $\kappa+\epsilon$ are examined. Their choice is as follows: $\epsilon_m \in \{-0.002, 0, 0.002\}$ for $\kappa_m = 0.0269$, $\epsilon_0 \in \{-0.0007, 0, 0.0007\}$ for $\kappa_0 = 0.01$ and $\epsilon_p \in \{-0.0004, 0, 0.0004\}$ for $\kappa_p = 0.00563$.

The Hellinger distance for the prior densities between centers of the same prior type is always larger than 0.1. On the other hand the Hellinger distance between the central priors and the micro-perturbed priors is never lager than 0.02.

5. Results

We first demonstrate the use of the model choice criteria for the default choice $\boldsymbol{\theta}_0 = (1, 0.622)$ of the gamma prior of the random effects precision. Next, we investigate the sensitivity of the estimates of the best fitting model to the alternative gamma and half-normal hyperprior specifications.

5.1. Model choice

All model choice criteria considered favor model B3, see Table 3. This finding is in agreement with the analysis by Lee and Nelder (1996). For the other choices of

Table 3: Comparison of DIC, LML, $\overline{\text{LCPO}}$, $\overline{\text{BS}}$ and AUC for different models for gamma hyperpriors with value $\boldsymbol{\theta}_0 = (1, 0.622)$.

Model	DIC	LML	$\overline{\text{LCPO}}$	$\overline{\text{BS}}$	AUC
A	394.3	-211.5	0.548	0.185	0.790
B1	392.3	-210.8	0.545	0.184	0.790
B2	388.0	-209.9	0.540	0.182	0.796
B3	384.1	-207.5	0.534	0.180	0.801
B4	393.8	-211.3	0.547	0.185	0.791
C	397.5		0.556	0.188	0.786

the gamma and half-normal prior (see Sections 4.2 and 4.3) model B3 was also the best. We therefore conduct our sensitivity analysis in Sections 5.2 to 5.4 exclusively for model B3. We note that the LML for model C is equal to -204.8. However, it is not comparable with LMLs from other models as there are separate intercepts fitted for each of the three experiments. Note also that if we assumed that the probability of the mating success is equal to the overall prevalence $\hat{p} = 0.525$ of mating success in the data set, then the Brier score would be equal to $\hat{p}(1 - \hat{p}) = 0.249$, a natural reference value (Schmid and Griffith, 2005). The values of $\overline{\text{BS}}$ are substantially better than this threshold.

5.2. Sensitivity of the fixed effects

We now investigate sensitivity of the marginal posterior distributions of the fixed effects in model B3 when the values of the gamma priors of all five hyperparameters are changed at the same time as described in Section 4.2. Inspection of the Hellinger distances and their calibrations (columns H_1, H_2, $q(H_1)$ and $q(H_2)$ in Table 4) reveals that the posterior distributions of $\beta(fW)$ and $\beta(WW)$ vary the most which is in agreement with the marginal posterior densities shown in Figure 1. Interestingly, the Hellinger distances for half-normal hyperpriors (columns H_3 and H_4 in Table 4) are smaller than for gamma hyperpriors (H_1 and H_2). The columns $q(H_3)$ and $q(H_4)$

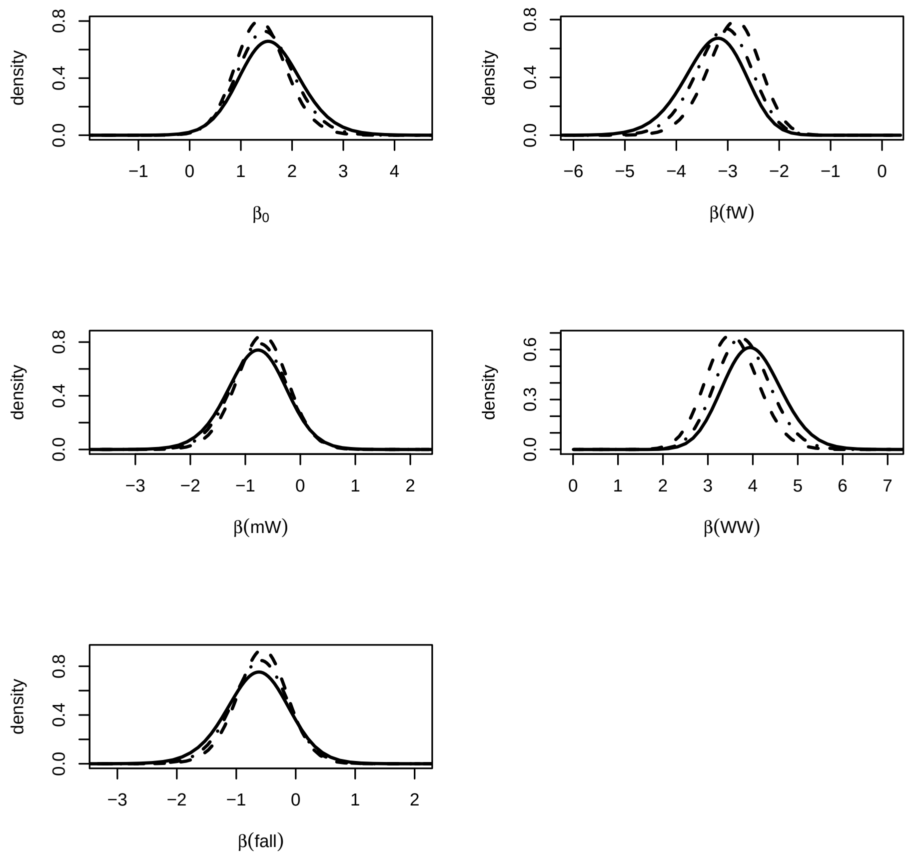

Figure 1: Marginal posterior densities of the fixed effects for $\boldsymbol{\theta}_0$ =(1,0.622) (solid), $\boldsymbol{\theta}_m$ =(1,0.311) (dashed) and $\boldsymbol{\theta}_p$ =(1,0.933) (dot-dashed) gamma prior values for all precision parameters in model B3.

Table 4: Hellinger distance and its calibration $q(H)$ between the posterior distributions of fixed effects in model B3 for changes of the values of all hyperpriors at the same time: $H_1 = H(post(\boldsymbol{\theta}_0), post(\boldsymbol{\theta}_m))$, $H_2 = H(post(\boldsymbol{\theta}_0), post(\boldsymbol{\theta}_p))$ for the gamma prior values $\boldsymbol{\theta}_0$ =(1,0.622), $\boldsymbol{\theta}_m$ =(1,0.311), $\boldsymbol{\theta}_p$ =(1,0.933). $H_3 = H(post(\kappa_0), post(\kappa_m))$, $H_4 = H(post(\kappa_0), post(\kappa_p))$ for the half-normal prior values κ_0 =0.01, κ_m =0.0269 and κ_p =0.00563.

Variable	H_1	H_2	H_3	H_4	$q(H_1)$	$q(H_2)$	$q(H_3)$	$q(H_4)$
β_0	0.042	0.030	0.005	0.005	0.559	0.543	0.507	0.507
$\beta(fW)$	0.077	0.055	0.003	0.005	0.608	0.578	0.505	0.507
$\beta(mW)$	0.029	0.021	0.003	0.003	0.542	0.530	0.504	0.505
$\beta(WW)$	0.082	0.059	0.006	0.007	0.615	0.583	0.508	0.510
$\beta(fall)$	0.030	0.022	0.005	0.005	0.542	0.530	0.507	0.507

show that changing the value of κ results in posterior changes which are at most 0.01 in Bernoulli probability scale. Thanks to the calibration of the Hellinger distance we conclude that our assumptions on the values of the gamma and half-normal hyperpriors influence the estimates of all fixed effects only to some minor extent.

5.3. Sensitivity of the random effects precisions

We now want to investigate the sensitivity of the five precision parameters in model B3 for the same prior changes as in Section 5.2, applied to each precision in turn with all other priors kept fixed at the default choice $\boldsymbol{\theta}_0$. Figure 2 gives the marginal posterior distributions while Table 5 lists the sensitivities and their calibrations (with respect to 0.8) computed according to Equations (3) and (5), respectively.

We describe the use of the sensitivity calibration at one particular example of $S =$ 0.483 in column S_1 in Table 5 obtained for τ_{M3} for the change from $\boldsymbol{\theta}_0$ to $\boldsymbol{\theta}_m$. Let $p = 0.8$, then following the formula in Equation (5) $C(S, 0.8) = q(S \times h(0.8)) =$ $q(0.483 \times 0.227) = q(0.109) = 0.652$ which can be found in the corresponding entry in column $C(S_1)$. This means that two priors whose difference is comparable with the difference between an event having probability 0.5 and 0.8 correspond to posteriors whose difference is comparable with the difference between an event having probability

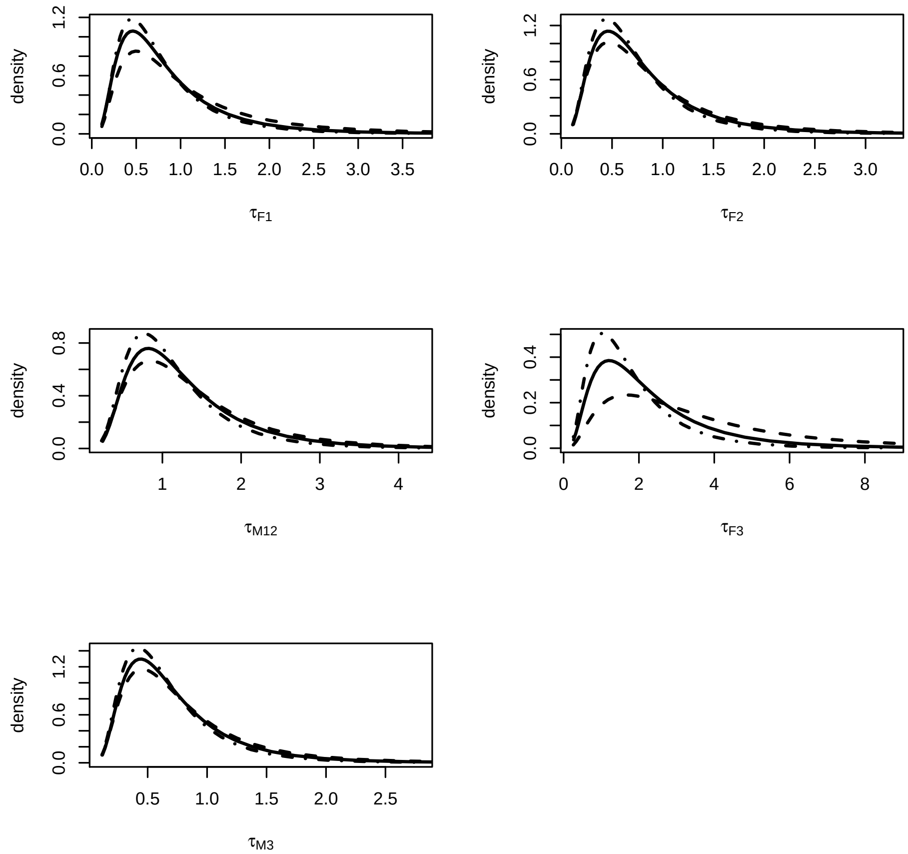

Figure 2: Marginal posterior densities of the precisions in model B3 for individual changes of the gamma prior values $\boldsymbol{\theta}_0$ =(1,0.622) (solid), $\boldsymbol{\theta}_m$ =(1,0.311) (dashed) and $\boldsymbol{\theta}_p$ =(1,0.933) (dot-dashed).

Table 5: Sensitivity and its calibration $C(S, 0.8)$ of the precisions in the model B3 for individual changes of the values of the hyperpriors: $S_1 = S(\boldsymbol{\theta}_0, \boldsymbol{\theta}_m)$, $S_2 = S(\boldsymbol{\theta}_0, \boldsymbol{\theta}_p)$ for the gamma prior values $\boldsymbol{\theta}_0$ =(1,0.622), $\boldsymbol{\theta}_m$ =(1,0.311), $\boldsymbol{\theta}_p$ =(1,0.933). $S_3 = S(\kappa_0, \kappa_m)$, $S_4 = S(\kappa_0, \kappa_p)$ for the half-normal prior values κ_0 =0.01, κ_m =0.0269 and κ_p =0.00563.

Variable	S_1	S_2	S_3	S_4	$C(S_1)$	$C(S_2)$	$C(S_3)$	$C(S_4)$
τ_{F1}	0.625	0.747	0.331	0.557	0.695	0.731	0.605	0.675
τ_{F2}	0.564	0.793	0.327	0.549	0.677	0.744	0.604	0.673
τ_{M12}	0.566	0.748	0.329	0.553	0.678	0.731	0.605	0.674
τ_{F3}	1.017	1.104	0.669	1.119	0.804	0.826	0.708	0.830
τ_{M3}	0.483	0.654	0.321	0.538	0.652	0.704	0.602	0.669

0.5 and 0.652.

Remarkably, the values of sensitivities and calibrations are mostly smaller for half-normal than the corresponding ones for gamma priors. The largest sensitivity for τ_{F3} in Table 5 in columns S_1 and S_2 is confirmed by the discrepancy of the marginal posterior densities in Figure 2. Calibrations provided in the columns $C(S_1)$ to $C(S_4)$ for gamma and half-normal priors indicate that the posterior distribution of τ_{F3} is the most sensitive to alterations of the values of hyperpriors. Information introduced by the prior is practically retained without any change on the posterior. We have not gained much information about τ_{F3} after taking into account the observed data.

5.4. Sensitivity on the prior grid

Figures 3 and 4 show boxplots of the posterior mean estimates of the fixed effects and the precision parameters, respectively, for both gamma and half-normal priors. As in Section 5.2 the posterior mean estimates of $\beta(fW)$ and $\beta(WW)$ vary the most when the values of the gamma hyperpriors are altered on the grid. Moreover, the posterior mean estimates of the precision of the hyperparameter τ_{F3} vary the most on the grid of values for gamma and half-normal priors which confirms the findings in Section 5.3. Macro-sensitivity for half-normal priors is clearly less than for gamma priors. Interestingly, the posterior mean estimates of the most sensitive parameters disagree the most for the

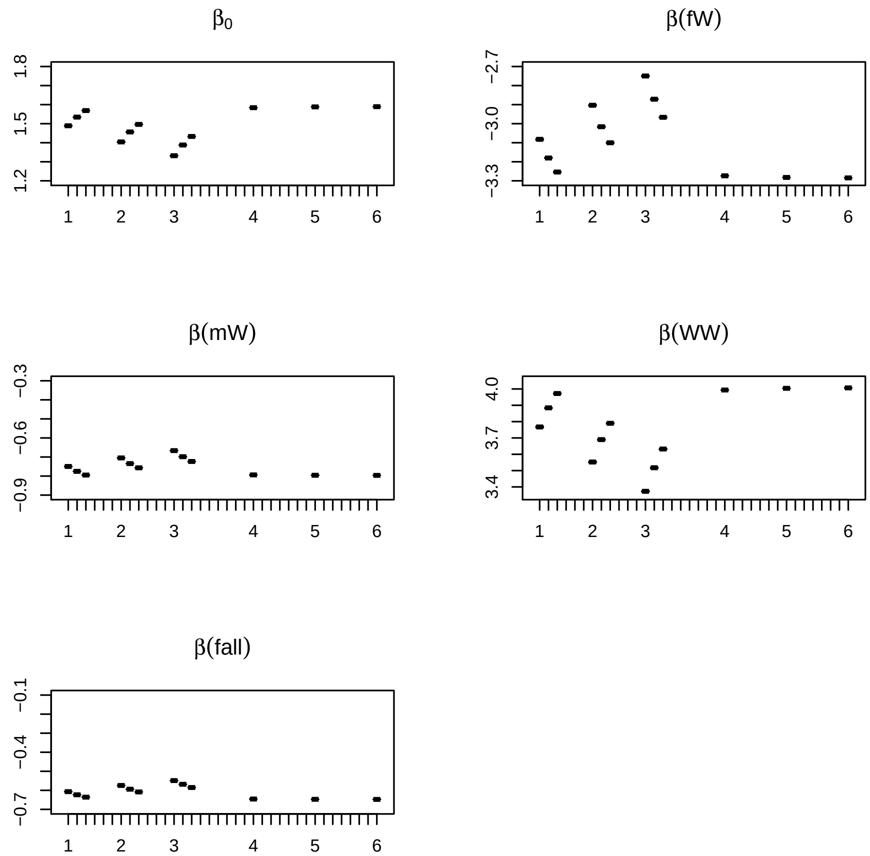

Figure 3: Posterior mean estimates of the fixed parameters in model B3 for different values (a_1, a_2) of the gamma and κ of the half-normal prior of the hyperparameters. The ordering of the center parameters from left to right on the x–axis is as follows: gamma (a_1, a_2) 1: $(0.5, 0.311)$, $(0.5, 0.622)$, $(0.5, 0.933)$, 2: $(1, 0.311)$, $(1, 0.622)$, $(1, 0.933)$, 3: $(1.5, 0.311)$, $(1.5, 0.622)$, $(1.5, 0.933)$ and half-normal κ 4: 0.0269, 5: 0.01, 6: 0.00563.

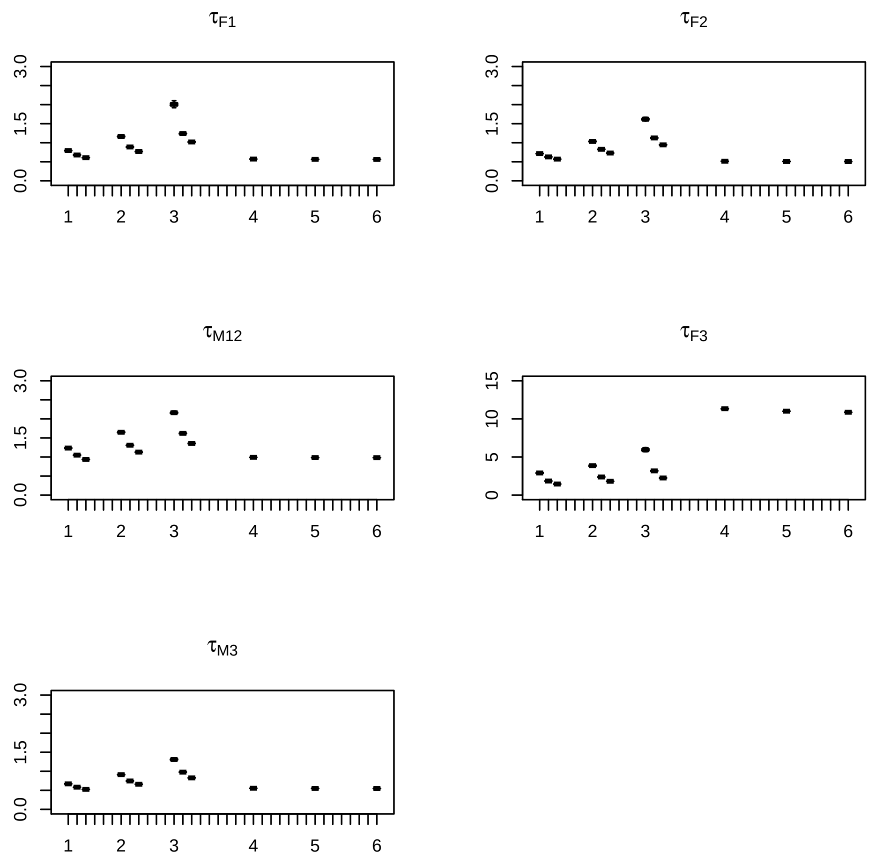

Figure 4: Posterior mean estimates of the precision parameters in model B3 for different values (a_1, a_2) of the gamma and κ of the half-normal prior of the hyperparameters. The ordering of the center parameters from left to right on the x–axis is as follows: gamma (a_1, a_2) 1: $(0.5, 0.311)$, $(0.5, 0.622)$, $(0.5, 0.933)$, 2: $(1, 0.311)$, $(1, 0.622)$, $(1, 0.933)$, 3: $(1.5, 0.311)$, $(1.5, 0.622)$, $(1.5, 0.933)$ and half-normal κ 4: 0.0269, 5: 0.01, 6: 0.00563. Note that for τ_{F3} the scale is different.

gamma and half-normal hyperprior assumptions.

6. Discussion

In this paper we have proposed a novel sensitivity measure for Bayesian hierarchical models and re-visited the salamander mating data to illustrate the methodology. First, we applied various model choice criteria to arrive at the best model which turned out to be model B3. Next, we conducted a sensitivity analysis of this model. The calibration of the Hellinger distance enabled us to discuss the relevance of the discrepancies of the estimates of the fixed and random effects for varying hyperprior choices. It turned to be a useful tool for judging the extend of the discrepancies of the estimates due to the hyperprior values alterations. This measure can be useful for practical statisticians to assess the amount of sensitivity with respect to prior assumptions. It can be extended easily to other models and its application is not constrained to binary GLMMs. The sensitivity measure based on the Hellinger distance developed above might be of some help to furnish the visual impression with a measure of their relevance for example if posterior densities have been estimated using MCMC and Rao-Blackwellization (Gelfand and Smith, 1990). We complemented the sensitivity analysis by investigating the posterior mean estimates of the fixed effects and the random effects precisions on the grid of macro- and micro-perturbed parameters of the gamma and half-normal prior distributions.

As far as the fixed effects are considered, we observed that the marginal posterior densities and posterior mean estimates for $\beta(fW)$ and $\beta(WW)$ react the most to changes of the hyperprior values. Moreover, we found that the marginal posterior distributions of the random effects precisions are very sensitive to prior alterations. The estimates of τ_{F3} were the most sensitive as the information introduced by the prior is practically transferred to the posterior without any change. The binary data in the salamander mating experiment apparently do not carry much information with respect to τ_{F3}.

Our work enabled the comparison of the sensitivity performance in model B3 under two different distributional assumptions for gamma and half-normal hyperpriors. It suggests that half-normal priors lead to less sensitive estimates and give a more stable performance. Consequently, we agree with Gelman (2006) and Lunn et al. (2009b) that one should be careful with gamma priors.

Interestingly, for the fixed and random effects which are relatively unsensitive to the changes of the hyperprior values the results from both prior assumptions (gamma and half-normal) are close to each other. However, the estimates which are highly sensitive to the choice of the hyperprior values within one particular prior specification disagree the most for different distributional hyperprior assumptions. For these model components the estimates can differ considerably depending on the distributional assumption of the prior of the hyperparameters. Therefore, we agree with Box and Tiao (1973) that the precise choice of the prior may be crucial and the models may be not robust to the choice of the prior.

Acknowledgements

The authors are grateful to Håvard Rue and Andrea Riebler for valuable discussions about INLA and to Michaela Paul for her support dealing with LaTeX and Sweave. They thank two anonymous reviewers for their valuable comments and suggestions which led to a considerable improvement of the manuscript.

Conflict of Interest: None declared.

References

Besag, J., Green, P., Higdon, D., Mengersen, K., 1995. Bayesian computation and stochastic systems. Statistical Science 10 (1), 3–66.

Bhattacharyya, A., 1943. On a measure of divergence between two statistical populations defined by their probability distributions. Bulletin of the Calcutta Mathematical Society 35, 99–109.

Box, G., Tiao, G., 1973. Bayesian Inference in Statistical Analysis. Addison–Wesley Publishing Company.

Breslow, N. E., Clayton, D. G., 1993. Approximate inference in generalized linear mixed models. Journal of the American Statistical Association 88 (421), 9–25.

Browne, W., Draper, D., 2006. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. Bayesian Analysis 1 (3), 473–514.

Chan, J. S. K., Kuk, A. Y. C., 1997. Maximum likelihood estimation for probit-linear mixed models with correlated random effects. Biometrics 53 (1), 86–97.

Fong, Y., Rue, H., Wakefield, J., 2010. Bayesian inference for generalized linear mixed models. Biostatistics 11 (3), 397–412.

Geisser, S., 1993. Predictive Inference: An Introduction. Chapman & Hall, Inc.

Gelfand, A., Smith, A., 1990. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association 85 (410), 398–409.

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models (Comment on Article by Browne and Draper). Bayesian Analysis 1 (3), 515–534.

Gneiting, T., Raftery, A. E., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102 (477), 359–378.

Held, L., Schrödle, B., Rue, H., 2010. Posterior and cross-validatory predictive checks: A comparison of MCMC and INLA. In: Kneib, T., Tutz, G. (Eds.), Statistical Modelling and Regression Structures – Festschrift in Honour of Ludwig Fahrmeir. Springer, pp. 91–110.

Jeffreys, H., 1961. Theory of Probability. Oxford University Press.

Karim, M. R., Zeger, S. L., 1992. Generalized linear models with random effects; salamander mating revisited. Biometrics 48 (2), 631–644.

Kass, R., Natarajan, R., 2006. A default conjugate prior for variance components in Generalized Linear Mixed Models (Comment on Article by Browne and Draper). Bayesian Analysis 1 (3), 535–542.

Lambert, P., 2006. (Comment on Article by Browne and Draper). Bayesian Analysis 1 (3), 543–546.

Le Cam, L., 1986. Asymptotic Methods in Statistical Decision Theory. Springer-Verlag.

Lee, Y., Nelder, J. A., 1996. Hierarchical generalized linear models. Journal of the Royal Statistical Society, Series B. 58 (4), 619–678.

Lunn, D., Spiegelhalter, D., Thomas, A., Best, N., 2009a. The BUGS project: Evolution, critique and future directions. Statistics in Medicine 28 (25), 3049–3067.

Lunn, D., Spiegelhalter, D., Thomas, A., Best, N., 2009b. Rejoinder to commentaries on "The BUGS project: Evolution, critique and future directions". Statistics in Medicine 28 (25), 3081–3082.

McCullagh, P., Nelder, J. A., 1989. Generalized Linear Models. Second Edition. Chapman & Hall/CRC.

McCulloch, R., 1989. Local model influence. Journal of the American Statistical Association 84 (406), 473–478.

Narasimhan, B., 2005. Lisp-Stat to Java to R. Journal of Statistical Software 13 (4), 1–10.

Pepe, M., 2003. The Statistical Evaluation of Medical Tests for Classifition and Prediction. Oxford University Press.

Plummer, M., 2008. Penalized loss function for Bayesian model comparison. Biostatistics 9 (3), 523–539.

Rue, H., Held, L., 2005. Gaussian Markov Random Fields. Theory and Applications. Chapman & Hall/CRC.

Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society, Series B. 71 (2), 319–392.

Schmid, C. H., Griffith, J. L., 2005. Multivariate classification rules: Calibration and discrimination. In: Armitage, P., Colton, T. (Eds.), Encyclopedia of Biostatistics, Second Edition. John Wiley & Sons, pp. 3491–3497.

Spiegelhalter, D., Best, N., Carlin, B., van der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B. 64 (4), 583–616.

Stone, M., 1977. Asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. Journal of the Royal Statistical Society, Series B. 39 (1), 44–47.

Verrell, P. A., Arnold, S. J., 1989. Behavioral observations of sexual isolation among allopatric populations of the mountain dusky salamander, Desmognathus Ochrophaeus. Evolution 43 (4), 745–755.

Wakefield, J., 2009. Comments on "The BUGS project: Evolution, critique and future directions". Statistics in Medicine 28 (25), 3079–3080.

Zeger, S., Karim, M., 1991. Generalized linear models with random effects: A Gibbs sampling approach. Journal of the American Statistical Association 86, 79–86.