Bayesian Analysis (2015) **10**, Number 2, pp. 321–349

Sensitivity Analysis for Bayesian Hierarchical Models

Małgorzata Roos*, Thiago G. Martins†, Leonhard Held‡, and Håvard Rue§

Abstract. Prior sensitivity examination plays an important role in applied Bayesian analyses. This is especially true for Bayesian hierarchical models, where interpretability of the parameters within deeper layers in the hierarchy becomes challenging. In addition, lack of information together with identifiability issues may imply that the prior distributions for such models have an undesired influence on the posterior inference. Despite its importance, informal approaches to prior sensitivity analysis are currently used. They require repetitive re-fits of the model with ad-hoc modified base prior parameter values. Other formal approaches to prior sensitivity analysis suffer from a lack of popularity in practice, mainly due to their high computational cost and absence of software implementation. We propose a novel formal approach to prior sensitivity analysis, which is fast and accurate. It quantifies sensitivity without the need for a model re-fit. Through a series of examples we show how our approach can be used to detect high prior sensitivities of some parameters as well as identifiability issues in possibly over-parametrized Bayesian hierarchical models.

Keywords: Base prior, formal local sensitivity measure, Bayesian robustness, calibration, Hellinger distance, Bayesian hierarchical models, identifiability, overparametrisation.

1 Introduction

Today applied statisticians have a wealth of both frequentist and Bayesian procedures at their disposal. The prominent feature of the latter approach is its ability to incorporate prior knowledge in the analysis. This feature, however, is both a benefit and a challenge. A Bayesian model is said to be sensitive and non-robust with respect to the prior distribution if its posterior distribution dramatically changes when the base prior parameter values are altered slightly. Recently, implementation of the hierarchical framework has led to the development of increasingly intricate models. Unfortunately, their complexity makes an extensive elicitation of the base prior for each hierarchy layer practically impossible. Instead, base priors tend to be determined in a rather casual fashion and without appropriate reflection, so arbitrary and inaccurately specified parameter values may arise. At the same time, due to model complexity, the impact

*Division of Biostatistics, Institute for Social and Preventive Medicine, University of Zurich, Zurich, Switzerland, mroos@ifspm.uzh.ch

†Department of Mathematical Sciences, Norwegian University for Science and Technology, Trondheim, Norway, guerrera@math.ntnu.no

‡Division of Biostatistics, Institute for Social and Preventive Medicine, University of Zurich, Zurich, Switzerland, leonhard.held@ifspm.uzh.ch

§Department of Mathematical Sciences, Norwegian University for Science and Technology, Trondheim, Norway, hrue@math.ntnu.no

 DOI: 10.1214/14-BA909

of possibly misspecified prior parameter values on outputs is unclear. In addition, the adequacy of the sample size needed for a reliable estimation of each layer is, in fact, unknown. It can happen that models are overparametrized (Carlin and Louis, 1998) and it may not be obvious to decide which parameters are well identified by the data and which are not (Dawid, 1979; Gelfand and Sahu, 1999). Hence, development of complex Bayesian models without any prior robustness diagnostics may be problematic. In order to ensure reliable and robust results, it is crucial to verify how sensitive the resulting posteriors are for each prior input.

The relevance of sensitivity and uncertainty analyses for exploring complex models has been highly emphasized in the literature. Oakley and O'Hagan (2004) and Saltelli et al. (2008) applied it successfully in the context of frequentist analysis. In Bayesian statistics the general sensitivity concept treated by Geisser (1992), Clarke and Gustafson (1998), Millar and Stewart (2007) and Zhu et al. (2011) involves broad issues like influential observations, uncertainty of the sampling model and prior inadequacy.

1.1 Bayesian formal sensitivity analysis

Sensitivity to the prior parameter specification is a crucial part of the general sensitivity setting (Berger et al., 2000). Ríos Insua et al. (2000) and Ruggeri (2008) argue that inappropriate prior parameter specifications can lead to distorted findings for both influential observations and uncertainty of the sampling model. To date, two approaches to sensitivity analysis can be distinguished: the global and the local one. The global approach considers the class of all priors compatible with the elicited prior information and computes the range of the posteriors as the prior varies over the class. This range is typically found by determining the "extreme" priors in the class that yield maximally distant posteriors, without explicitly carrying out the analysis for every prior in the class. In contrast, the local sensitivity approach is interested in the rate of change in posterior with respect to changes in the prior, and usually uses differential calculus to approximate it. Despite its desirability the global approach is impractical in the Bayesian hierarchical framework whereas the local one, recommended by Gustafson (2000) and Sivaganesan (2000), is the method of choice.

The local sensitivity approach routinely applied in complex Bayesian hierarchical models can determine which model components are hard to learn from the data. Elevated worst-case sensitivity estimates quickly identify priors that may need more careful attention. Since at the stage of model construction, the analyst might be interested in specific directions on the hyperparameter space (Kadane, 1992), there is a strong need for investigation of the circular sensitivity around particular base prior parameter values.

Gustafson (2000) distinguishes a variety of frameworks for local Bayesian robustness investigation. They differ with respect to the posterior results (distribution or summaries), by prior perturbations (geometric or parametric), whether the worst-case sensitivity is measured in the absolute or relative sense, and by the discrepancy measure used. In particular, McCulloch (1989) following Cook (1986) approximated prior worst-case robustness by the principal eigenvalue of an appropriate infinitesimal ratio. This approach has been further refined by Zhu et al. (2011).

1.2 Informal approaches and dedicated software

Surprisingly, despite considerable theoretical advances in formal sensitivity analysis, it is barely used in every-day practice. In the few cases when the extent of prior robustness is assessed, brute force and informal approaches are used instead. An informal technique consists of repetitive fits of the model with ad hoc modified prior inputs. If the posteriors subjectively do not differ much, non-sensitivity and robustness are claimed. The main drawback of such an approach is that it requires several re-fits of the model, which may be extremely time consuming. The informal approach gives no guidance about how the input modifications should be performed and how differences in the results should be judged in a standardized way. Consequently, in order to guarantee reproducibility of Bayesian robustness considerations, the use of a formal sensitivity approach is required.

Although the need and importance of a formal prior robustness investigation have been ubiquitously approved, its lack of popularity in practice seems mainly due to the non-existence of such a facility in current Bayesian programs (Ruggeri, 2008). Hence, Berger et al. (2000) and Lesaffre and Lawson (2012) stress that a development of a formal robustness methodology, which is feasible, fairly quick, operating with low extra computing effort and provided by default in a dedicated software, is strongly required. Furthermore, in order to become widely used, its compatibility with the Markov chain Monte Carlo (Gilks et al., 1996) framework is beneficial.

1.3 Scope of paper

In this paper we suggest the use of a Bayesian formal ϵ-local sensitivity, which can be conveniently applied to Bayesian hierarchical models. The novelty of our approach hinges on the choice of an epsilon grid, a set of base prior parameter specifications modified in a standardized way. The typical worst-case sensitivity estimates provided by the infinitesimal local approach are complemented by circular sensitivity summaries. Our approach guarantees a nearly instantaneous sensitivity assessment without any need for a model re-fit. Prior and posterior Hessian approximations are not required any more. Instead, the base prior specification and the corresponding marginal posterior density alone are used. Because our ϵ-local sensitivity approach operates without much additional computational effort, it is a convenient measure for an every-day use.

The remainder of this article is organized as follows: Section 2 defines the ϵ-local sensitivity measure and its calibration with respect to the unit-variance normal distribution. In Section 3, an illustrative and comparative example is discussed. General approaches for fast numerical computation and epsilon grid search are presented in Section 4. Although our local robustness approach is generally applicable, its performance for a range of applications with increasing complexity and several latent models is presented in Section 5 and in the Supplementary Material. In these examples we show how to use the proposed methodology in practice to identify sensitive parameters. Some concluding remarks are given in Section 6. Appendices A-C provide proofs and discuss computational issues.

2 Local sensitivity

2.1 Definition

Given a scalar θ, we denote by $\pi_{\boldsymbol{\gamma}_0}(\theta)$ and $\pi_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y})$ the base prior density with parameter values fixed at $\boldsymbol{\gamma}_0$ and the resulting marginal posterior density for θ, respectively. The corresponding entities indexed by $\boldsymbol{\gamma}$ originate from a prior distribution with arbitrary parameter values denoted by $\boldsymbol{\gamma}$.

We define the ϵ-local circular sensitivity $S^c_{\boldsymbol{\gamma}_0}(\epsilon)$ as the set of ratios

$$S^c_{\boldsymbol{\gamma}_0}(\epsilon) = \left\{ \frac{\mathrm{d}(\pi_{\boldsymbol{\gamma}}(\theta|\boldsymbol{y}), \pi_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y}))}{\epsilon}, \text{ for } \boldsymbol{\gamma} \in \mathrm{G}_{\boldsymbol{\gamma}_0}(\epsilon) \right\}, \tag{1}$$

with the grid $\mathrm{G}_{\boldsymbol{\gamma}_0}(\epsilon)$ of parameter values defined by

$$\mathrm{G}_{\boldsymbol{\gamma}_0}(\epsilon) = \{\boldsymbol{\gamma} : \mathrm{d}(\pi_{\boldsymbol{\gamma}}(\theta), \pi_{\boldsymbol{\gamma}_0}(\theta)) = \epsilon\}, \tag{2}$$

where $\mathrm{d}(\cdot, \cdot)$ denotes a convenient discrepancy measure between two densities. We can also consider $\mathrm{G}_{\boldsymbol{\gamma}_0}(\epsilon)$ as a contour line around $\boldsymbol{\gamma}_0$. In our definition, the distributional assumption of the prior $\pi_{\boldsymbol{\gamma}}(\theta)$ for one particular component θ of the Bayesian hierarchical model is held fixed and only its parameter values $\boldsymbol{\gamma}$ are allowed to vary.

In practice, we use a fixed small ϵ for sensitivity evaluation instead of its infinitesimal approximation. We suggest detailed exploration of the local geometry implied by $\mathrm{d}(\cdot, \cdot)$ in the space of prior distributions. This is done by a numerical search for the prior parameter value grid, $\mathrm{G}_{\boldsymbol{\gamma}_0}(\epsilon)$, with center set at $\boldsymbol{\gamma}_0$ and the distance value fixed to ϵ as in Equation (2). Our circular approach naturally adjusts for possible non-orthogonalities of the prior parametrisation as it examines all directions in the space of prior parameter values on equal footing.

Circular sensitivity can be conveniently summarized by a single number. For example, the worst-case sensitivity $S_{\boldsymbol{\gamma}_0}(\epsilon)$ is defined to be the maximum of the circular sensitivity $S^c_{\boldsymbol{\gamma}_0}(\epsilon)$

$$S_{\boldsymbol{\gamma}_0}(\epsilon) = \max\left\{S^c_{\boldsymbol{\gamma}_0}(\epsilon)\right\} = \max_{\boldsymbol{\gamma} \in \mathrm{G}_{\boldsymbol{\gamma}_0}(\epsilon)} \frac{\mathrm{d}(\pi_{\boldsymbol{\gamma}}(\theta|\boldsymbol{y}), \pi_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y}))}{\epsilon}. \tag{3}$$

In this paper, we mainly concentrate on the worst-case sensitivity, though alternative estimates such as mean, median or minimum could be also reported.

For complex Bayesian hierarchical models the sensitivity of each model component θ is addressed separately. Computations are conducted according to Equations (1)–(3). The only input required for sensitivity estimation is the base prior distribution specification $\pi_{\boldsymbol{\gamma}_0}(\theta)$ and the corresponding marginal posterior density $\pi_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y})$. As a first step, the worst-case robustness $S_{\boldsymbol{\gamma}_0}(\epsilon)$ is checked. A high value indicates that a particular prior has to be investigated with more care. It could be caused by a variety of reasons. One, a misplaced prior distribution leading to a prior-data conflict (Box, 1980; Evans and Moshonov, 2006). Two, an inappropriate prior caused by misspecified

prior parameter values. Three, an insufficient sample size at the hierarchy level under consideration. At this step any other circular sensitivity summary such as mean, median or minimum can be applied.

In the second step, the circular sensitivity $S^c_{\gamma_0}(\epsilon)$ in all directions around γ_0 can be referred to. Circular sensitivity can be easily depicted, as will be shown later. The circular sensitivity plots indicate directions in which the most pronounced sensitivity value modification was found. The plot's shape depends on the prior distribution, the base prior parameter specification and the model assumed. The choice of the base prior parameter values at the model construction stage can be conveniently guided by circular sensitivity and its summary values.

As our approach uses posterior and prior densities directly, a convenient discrepancy measure $\mathrm{d}(\cdot,\cdot)$ to quantify the discrepancy between two distributions is required (Gustafson, 2000). One possible choice is the ϕ-divergence, also called f-divergence, between two densities π_0 and π_1. It is defined as

$$\mathrm{D}_\phi(\pi_1,\pi_0) = \int \pi_1(\theta)\phi\Big(\frac{\pi_0(\theta)}{\pi_1(\theta)}\Big)d\theta, \tag{4}$$

where ϕ is a smooth convex function (Amari, 1990; Amari and Nagaoka, 2000). The Kullback-Leibler divergence and Hellinger distance are special cases of the ϕ-divergence: with $\phi_{KL}(x) = x\log(x)$ for the Kullback-Leibler divergence and $\phi_H(x) = (\sqrt{x}-1)^2/2$ for the Hellinger distance (Dey and Birmiwal, 1994). Robert (1996) found that the Kullback-Leibler divergence and the Hellinger distance can frequently be used interchangeably but that the Hellinger distance is more natural as a true distribution distance.

Our preference for the Hellinger distance (Le Cam, 1986) is motivated by convenience. The Hellinger distance is advantageous given marginal posterior distributions and prior distributions provided numerically and attaining nonzero values only on a finite discrete set of points (Roos and Held, 2011). It is a symmetric measure of discrepancy between two densities π_0 and π_1:

$$\begin{aligned}
\mathrm{H}(\pi_1,\pi_0) &= \sqrt{\frac{1}{2}\int_{-\infty}^{\infty}\Big\{\sqrt{\pi_1(\theta)}-\sqrt{\pi_0(\theta)}\Big\}^2 d\theta} \\
&= \sqrt{\frac{1}{2}\int_{-\infty}^{\infty}\Big\{\pi_1(\theta)-2\sqrt{\pi_1(\theta)\pi_0(\theta)}+\pi_0(\theta)\Big\}d\theta} \\
&= \sqrt{\frac{1}{2}\Big(2-2\int_{-\infty}^{\infty}\sqrt{\pi_1(\theta)\pi_0(\theta)}d\theta\Big)} \\
&= \sqrt{1-\mathrm{BC}(\pi_1,\pi_0)}.
\end{aligned}$$

Here, the Bhattacharyya coefficient $\mathrm{BC}(\pi_1,\pi_0) = \int_{-\infty}^{\infty}\sqrt{\pi_1(\theta)\pi_0(\theta)}d\theta$ measures the affinity of both densities (Bhattacharyya, 1943). 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; Roos and Held, 2011).

We assume throughout that the Gaussian distribution is parametrized by mean μ and precision λ and the gamma distribution with shape α and rate β parameters leading to expectation α/β and variance α/β^2. For both distributions the Hellinger distance between densities with differing parameter values can be computed analytically.

Rao (1945) and Dawid (1977) discussed the direct correspondence of the Bhattacharyya coefficient and the Fisher information matrix. In the context of differential geometry, Amari (1990) stated that both the Hellinger distance and the Bhattacharyya distance are directly related to the Riemannian distance. Indeed, the Hellinger distance introduces a non-Euclidean geometry on the space of probability distributions. As an example, consider the gamma prior assumed in Section 5 for the precision of the structured intrinsic conditional autoregressive component. Figure 1 shows contour plots of $\mathrm{G}_{\boldsymbol{\gamma}_0}(\epsilon)$ in Equation (2) based on the Hellinger distance with center set at $\boldsymbol{\gamma}_0 = (\alpha_0, \beta_0) = (1, 0.34)$. Equal scaling of x and y-axes highlights that the contours tend to be ellipses rather than circles in Euclidean geometry.

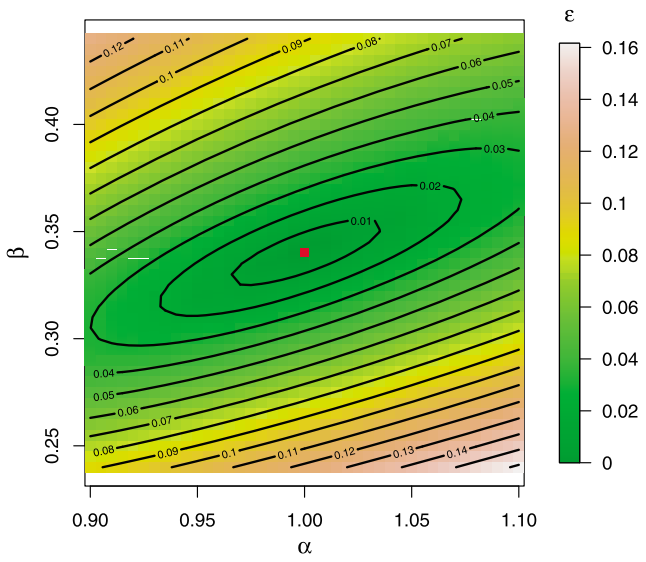

Figure 1: Contour plots of $\mathrm{G}_{\boldsymbol{\gamma}_0}(\epsilon)$ for gamma distribution with center $\boldsymbol{\gamma}_0 = (\alpha_0, \beta_0) = (1, 0.34)$ based on the Hellinger distance.

2.2 Calibration and interpretation

Calibration of differences between two distributions has the following advantage: for a particular reference distribution, the experimenter can assess the relevance of the discrepancy in terms of the natural parameter of the benchmark. Various calibrations have been suggested by McCulloch (1989), Dey and Birmiwal (1994), Goutis and Robert (1998) and Roos and Held (2011). Although ϕ-divergence and Kullback-Leibler divergence have been discussed in the literature, the calibration of the Hellinger distance with

respect to the unit-variance normal distribution, derived in Lemma 2 in Appendix B4, seems to be new. It states that for a particular Hellinger distance value h and

$$\mu(h) = \sqrt{-8\log(1-h^2)}$$

the Hellinger distance between two unit-variance normal distributions with shifted locations satisfies

$$\mathrm{H}(\mathrm{N}(\mu(h),1),\mathrm{N}(0,1)) = h.$$

Note that $\mu(h)$ is the calibration of the Hellinger distance h, as h between any two densities is the same as that between N(0,1) and N($\mu(h)$,1). Given the Hellinger distance h between any two densities, we can quantify discrepancies between them, in terms of the differences in mean from 0 to $\mu(h)$ for normal distribution with unit standard deviation.

Usually, it is expected that the sensitivity estimates in Equations (1) and (3) attain values smaller than 1. This indicates that changes in the marginal posteriors are smaller than changes in priors. Sensitivity close to 1 denotes that differences in marginal posteriors and priors are comparable. However, in practice both the circular $S^c_{\boldsymbol{\gamma}_0}(\epsilon)$ and the worst-case $S_{\boldsymbol{\gamma}_0}(\epsilon)$ sensitivity estimates can attain values larger than 1 leading to so-called super-sensitivity (Plummer, 2001). This possibility has been already attested by McCulloch (1989), Clarke and Gustafson (1998), Pérez et al. (2006) and Zhu et al. (2011). Müller (2012) truncates at 1 to remove excessively large sensitivity values, but we prefer to report the unmodified sensitivity estimates.

In order to get an impression about the relevance of sensitivity values, we calibrate both the numerator and the denominator of the sensitivity measures defined in Equations (1) and (3). Interestingly, this ratio of calibrated Hellinger distances can be conveniently approximated by the ratio of Hellinger distances involved in the sensitivity estimates themselves as

$$\frac{\mu\big(\mathrm{H}(\pi_{\boldsymbol{\gamma}}(\theta|\boldsymbol{y}),\pi_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y}))\big)}{\mu(\epsilon)} \approx \frac{\mathrm{H}(\pi_{\boldsymbol{\gamma}}(\theta|\boldsymbol{y}),\pi_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y}))}{\epsilon}. \tag{5}$$

To see this, apply the following property of the logarithm to the numerator and denominator of the left hand side: $\log(1-x) \approx -x$, for $-1 \leq x < 1$, with $x = h^2$. Therefore, the sensitivity estimates obtained in applications can be directly interpreted as an approximation of the ratio of calibrated Hellinger distances with respect to the unit-variance normal distribution. Although a particular choice of ϵ anchors our calibration, the above observation offers an option to interpret the sensitivity magnitude independently of any particular ϵ value used for the grid computation. Apart of that, the use of the ratio of calibrations leads to its applicability for the whole range of sensitivity values including small and super-sensitivities.

We will show in the illustrative example in Section 3 that calibration with respect to the unit-variance normal distribution converts sensitivity values obtained under the Kullback-Leibler divergence and the Hellinger distance and helps to compare them on an equal footing. Therefore, it is very useful for a unified understanding of sensitivity estimates obtained by different approaches.

3 An illustrative and comparative conjugate example

For illustration, we consider an analytical conjugate example which was also discussed by McCulloch (1989), Zhu et al. (2011) and Box (1980). On the basis of this example we will compare the infinitesimal local sensitivity method based on the Kullback-Leibler divergence with the ϵ-local sensitivity approach suggested in Section 2.1. The latter approach will be illustrated for both the Kullback-Leibler divergence and the Hellinger distance. The infinitesimal local sensitivity approach based on the Kullback-Leibler divergence was originally developed by McCulloch (1989). It is a special case of the more general methodology suggested by Zhu et al. (2011) when constrained to the Kullback-Leibler divergence and parametric perturbations of the base prior. Due to the explicit analytical formulae in the conjugate example, we can provide a software-independent comparison of the methods.

Assume that observations $y_1, y_2, \ldots, y_n$ are realisations of iid $\mathrm{N}(m, \kappa^{-1})$ random variables, where κ is fixed and $m \sim \mathrm{N}(\mu, \lambda^{-1})$. Due to conjugacy the posterior distribution is also normal with $m \,|\, \boldsymbol{y} \sim \mathrm{N}\big((n\kappa\bar{y} + \lambda\mu)/(n\kappa + \lambda), (n\kappa + \lambda)^{-1}\big)$. Following Box (1980), we assume that $\kappa = 1$, $n = 4$, $\bar{y} = 76$ and $\hat{\sigma}^2 = 3.33$. We are interested in the local sensitivity estimate of $\theta = m$ at the base prior parameter specification $\boldsymbol{\gamma}_0 = (\mu_0, \lambda_0) = (70, 0.5)$.

3.1 Infinitesimal local sensitivity based on the Kullback-Leibler divergence

In the following we describe the approach proposed by McCulloch (1989) and Zhu et al. (2011). Recall the definition of the Kullback-Leibler divergence $D_{\mathrm{KL}}(\pi_0||\pi)$ given in Equation (4) for an appropriate choice of $\phi_{KL}(x) = x\log(x)$. It is an unbounded, non-symmetric and directed measure, where the distribution having density π_0 is assumed to be the true probability distribution and the distribution leading to density π is treated as an arbitrary one. Now apply the Taylor expansion to the numerator and denominator of the worst-case infinitesimal local sensitivity S at $\boldsymbol{\gamma}_0$ defined in Equation (6) and approximate it by the principal eigenvalue of the ratio of two matrices

$$S = \lim_{\boldsymbol{\gamma}\to\boldsymbol{\gamma}_0} \frac{\mathrm{D}_{\mathrm{KL}}(\pi_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y})||\pi_{\boldsymbol{\gamma}}(\theta|\boldsymbol{y}))}{\mathrm{D}_{\mathrm{KL}}(\pi_{\boldsymbol{\gamma}_0}(\theta)||\pi_{\boldsymbol{\gamma}}(\theta))} \approx \max_{\text{eigenvalue}} \{(\mathbf{D}^2 i(\boldsymbol{\gamma}_0))^{-1}\mathbf{D}^2 d(\boldsymbol{\gamma}_0)\}. \tag{6}$$

Here $d(\boldsymbol{\gamma}) = \mathrm{D}_{\mathrm{KL}}(\pi_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y})||\pi_{\boldsymbol{\gamma}}(\theta|\boldsymbol{y}))$, $i(\boldsymbol{\gamma}) = \mathrm{D}_{\mathrm{KL}}(\pi_{\boldsymbol{\gamma}_0}(\theta)||\pi_{\boldsymbol{\gamma}}(\theta))$ and $\mathbf{D}^2 f(\boldsymbol{x})$ denotes the second derivative matrix of the function f evaluated at $\boldsymbol{x}$.

In our conjugate example with unknown $\theta = m$, the analytical formulae in Appendix B2 for the eigenvalue analysis of the matrix $(\mathbf{D}^2 i(\boldsymbol{\gamma}_0))^{-1}\mathbf{D}^2 d(\boldsymbol{\gamma}_0)$ can be applied. The resulting worst-case infinitesimal local sensitivity, which is the principal eigenvalue, is equal to 3.284 and the corresponding principal eigenvectors lie on the line λ = -2.677μ. The second eigenvalue is equal to 4e-04 with the corresponding eigenvector lying on the line $\lambda = 0.093\mu$. Both eigenvectors are indicated by red and green lines in Figures 2 and 3.

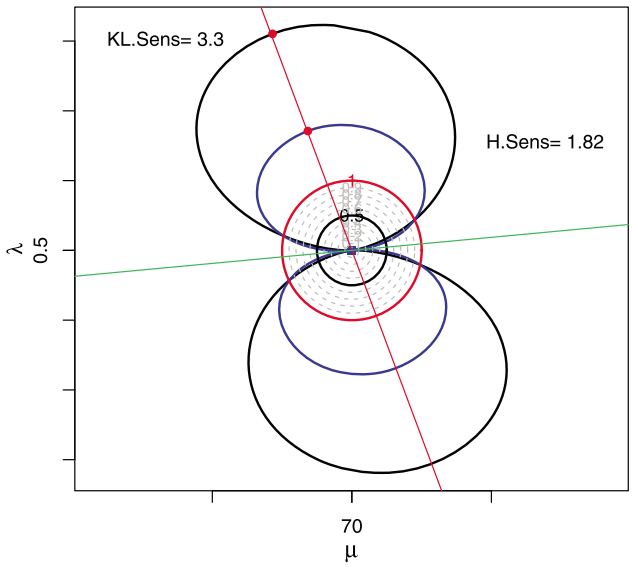

Figure 2: Comparison of local sensitivity methods in the plane spanned by parameters μ and λ centered at $\boldsymbol{\gamma}_0 = (\mu_0, \lambda_0) = (70, 0.5)$. Red and green lines: eigenvectors provided by the infinitesimal local approach based on the Kullback-Leibler divergence with the principal eigenvalue (worst-case sensitivity) 3.284. Black curve: circular sensitivity in each direction provided by the ϵ-local approach for ϵ_0 = 5e-05 based on the Kullback-Leibler divergence. Blue curve: circular ϵ-local sensitivity based on the Hellinger distance for $\epsilon_0 = 0.00354$. In both cases the choice of ϵ_0 corresponds to a shift in means by 0.01 with respect to the unit-variance normal distribution. The worst-case sensitivities provided by the ϵ-local approaches are indicated by a red dot. The scale for sensitivity values (0.1, 0.2,. . ., 0.9, 1) is indicated by grey circles. Sensitivity equal to 0.5 is marked by a black circle, whereas the value 1 is marked by a red one.

For a vector $\boldsymbol{\theta}$ the method is usually applied to the whole parameter vector of the base priors. Both matrices $\mathbf{D}^2 d(\boldsymbol{\gamma}_0)$ and $\mathbf{D}^2 i(\boldsymbol{\gamma}_0)$ for numerator and denominator are usually estimated by the corresponding Fisher information matrices. In practice, for a large cardinality of $\boldsymbol{\theta}$ it may be difficult to determine the direction indicated by the loadings of the principal eigenvector in the multidimensional space. Only the worst-case infinitesimal local sensitivity is provided by the principal eigenvalue. Usually, the matrix $(\mathbf{D}^2 i(\boldsymbol{\gamma}_0))^{-1}\mathbf{D}^2 d(\boldsymbol{\gamma}_0)$ is non-symmetric. Consequently, complex eigenvalues are possible. One may use analytical formulae to compute the Fisher information matrices under base prior and posterior or draw MCMC samples for their approximation.

Equation (6) is an infinitesimal local sensitivity measure expressed by a ratio. Incorporation of the denominator in such a measure is crucial. McCulloch (1989) concentrates on sensitivity of posterior distributions to parametric changes in the base prior parameter values quantified by the Kullback-Leibler divergence. In contrast, the methodology

suggested by Zhu et al. (2011) accounts for three sources of sensitivity: data, model and prior. Moreover, it is capable of handling additive contamination class perturbations of the form

$$\omega(t) = p(\theta) + t(g(\theta) - p(\theta)),$$

with $p(\theta)$ and $g(\theta)$ denoting the base prior and the contaminating distributions, respectively. Such perturbations are conveniently examined in the context of differential geometry. Comparison of sensitivity estimates produced by different discrepancy measures is necessary. Ibrahim et al. (2011) use a parametric bootstrap to indicate sensitivity 'significance' in the context of survival models.

3.2 Epsilon local sensitivity based on the Kullback-Leibler divergence

The first step for the ϵ-local sensitivity computation consists of finding the epsilon grid $\mathrm{G}_{\boldsymbol{\gamma}_0}(\epsilon)$ with respect to the chosen discrepancy measure and the base prior distribution by Equation (2). For a normal prior and the Kullback-Leibler divergence, the epsilon grid $\mathrm{G}_{\boldsymbol{\gamma}_0}(\epsilon)$ is specified by analytical formulae (16) and (17) derived in Appendix B3. Additionally, given a specific ϵ value, the circular ϵ-local sensitivity $S^c_{\boldsymbol{\gamma}_0}(\epsilon)$ defined in Equation (1) can also be computed analytically for each $\boldsymbol{\gamma} = (\mu, \lambda) \in \mathrm{G}_{\boldsymbol{\gamma}_0}(\epsilon)$.

In particular, for $\epsilon_0 = 5\text{e-}05$, the resulting circular sensitivity estimates $S^c_{\boldsymbol{\gamma}_0}(\epsilon_0)$ at $\boldsymbol{\gamma}_0 = (\mu_0, \lambda_0) = (70, 0.5)$ are shown in Figure 2 and the worst-case sensitivity $S_{\boldsymbol{\gamma}_0}(\epsilon_0) = 3.304$ is indicated by the red dot on the black curve. Note that this plot operates in three dimensions as it indicates the values of the circular sensitivity in each direction around $\boldsymbol{\gamma}_0$. The scale for sensitivity values (0.1, 0.2,. . ., 0.9, 1) is indicated by grey circles. Sensitivity equal to 0.5 is marked by a black circle, whereas the value 1 is marked by a red one. The worst-case sensitivity found by the ϵ-local approach lies on the red line, which corresponds to the principal eigenvector given by the infinitesimal local sensitivity approximation. Estimates of both worst-case sensitivity values based on the Kullback-Leibler divergence are comparable.

In case of the ϵ-local sensitivity it is interesting to look at its calibration (see Lemma 1 in Appendix B3). In particular, the choice of $\epsilon_0 = 5\text{e-}05$ for the Kullback-Leibler divergence corresponds to a shift by 0.01 in means of the unit-variance normal distribution. Following Equation (18), the calibrated sensitivity values are obtained by taking the square root of the circular ϵ-local sensitivity estimates based on the Kullback-Leibler divergence (Figure 3). The calibrated worst-case sensitivity equals 1.82 as compared to the calibrated infinitesimal local sensitivity 1.81.

There are two objections towards the use of the Kullback-Leibler divergence as a discrepancy measure. First, for complex Bayesian hierarchical models prior elicitation at different hierarchy levels is frequently impossible. Therefore, it is quite premature to assume a non-elicited base distribution to be a true one, as the directed Kullback-Leibler divergence actually does. Second, the unboundedness of the Kullback-Leibler divergence makes it less suitable for dealing with numerical representations of marginal posterior distributions (Roos and Held, 2011). To overcome both objections the use

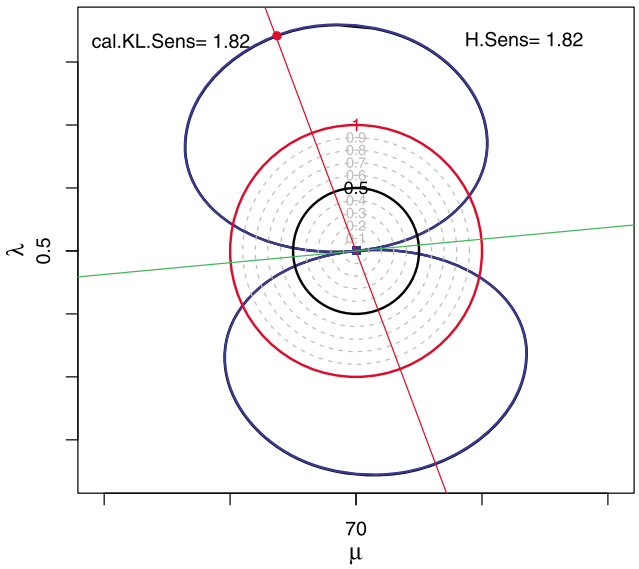

Figure 3: Comparison of calibrated local sensitivity estimates from Figure 2 centered at $\boldsymbol{\gamma}_0 = (\mu_0, \lambda_0) = (70, 0.5)$. Red and green lines: eigenvectors provided by the infinitesimal local approach based on the Kullback-Leibler divergence with the calibrated principal eigenvalue (worst-case infinitesimal local sensitivity) $\sqrt{3.284} = 1.81$. Black curve: circular sensitivity in each direction provided by the ϵ-local approach for $\epsilon_0 =$ 5e-05 based on the Kullback-Leibler divergence. Blue curve: circular ϵ-local sensitivity based on the Hellinger distance for $\epsilon_0 = 0.00354$. In both cases the choice of ϵ_0 corresponds to a shift in means by 0.01 with respect to the unit-variance normal distribution. The worst-case sensitivities provided by the ϵ-local approaches are indicated by a red dot. Both black and blue curves match well. Scale for sensitivity values (0.1, 0.2,. . ., 0.9, 1) is indicated by grey circles. Sensitivity equal to 0.5 is marked by a black circle, whereas the value 1 is marked by a red one.

of a symmetric and bounded discrepancy measure might be more appropriate. In the next subsection we discuss the ϵ-local sensitivity based on the Hellinger distance, which exhibits both requested properties.

3.3 Epsilon local sensitivity based on the Hellinger distance

In the case of the normal prior and the Hellinger distance, the epsilon grid is provided by the analytical Formulae (20) and (21) in Appendix B4. In particular, for $\epsilon_0 = 0.00354$ at $\boldsymbol{\gamma}_0 = (\mu_0, \lambda_0) = (70, 0.5)$, circular sensitivity estimates $S^c_{\boldsymbol{\gamma}_0}(\epsilon_0)$ are computed as shown by the blue curve in Figure 2. The worst-case sensitivity $S_{\boldsymbol{\gamma}_0}(\epsilon_0)$ indicated by the red dot on the blue curve attains a value of 1.823. The position of the worst-case sensitivity in the polar plot indicates that the marginal posterior is affected more by changes in the

precision λ than in the expectation μ. Further computational aspects of this example are discussed in Appendix C5. Notably, the ϵ-local sensitivity estimates are stable for a wide range of ϵ value choices.

We apply the calibration of the Hellinger distance with respect to the unit-variance normal distribution discussed in Section 2.2 and Appendix B4. The use of $\epsilon_0 = 0.00354$ for the epsilon grid search corresponds to a unit-variance normal distribution with mean equal to 0.01. Furthermore, by Formula (5), the circular ϵ-local sensitivity estimates $S^c_{\gamma_0}(\epsilon_0)$ approximate directly the corresponding sensitivities with respect to the unit-variance normal distribution. Calibration revealed super-sensitivity for m, as the mean change in the unit-variance normal distributions for the posteriors is 182.3% of the mean change in the unit-variance normal distributions for priors.

The calibration facilitates the comparison of sensitivity estimates when different discrepancy measures are used. Although the ϵ-local sensitivity estimates based on the Kullback-Leibler divergence and the Hellinger distance in Figure 2 differ, the corresponding ϵ-local sensitivity estimates calibrated with respect to the unit-variance normal distribution in Figure 3 agree well. Both worst-case ϵ-local sensitivity estimates lie on the red principal eigenvector line provided by the infinitesimal local approach.

The ϵ-local sensitivity approach concentrates only on the parametric sensitivity to the base prior parameter values for one model component θ. It is easier and less general than the methodology developed by Zhu et al. (2011). However, it is capable of providing not only the worst-case sensitivity estimate but also other descriptive statistics of circular sensitivity estimates and circular plots. For example, a circular plot with the worst-case ϵ-local sensitivity 0.88, median 0.86 and minimum 0.71 conveys different information than the values 0.83, 0.4 and 0.03. Careful investigation of each model component may provide valuable suggestions at the stage of model building.

Interestingly, the ϵ-local sensitivity approach suggests that the frequently used ad-hoc methodology (changing base prior parameter values and re-fitting the model) can easily be turned into a formal one. In fact, it suggests not to use arbitrary parameter values but standardized ones contained in the epsilon grid. When doing so, for a small epsilon value the ϵ-local sensitivity estimates will be close to the infinitesimal local ones.

4 Numerical computation

We still need to address two vital topics dealing with the instantaneous computation of posterior density for differing prior parameter values and computation of the sensitivity measure itself. The previous illustrative example is quite optimistic, as both the marginal posterior distribution and the epsilon grid are accessible analytically. In practice, however, the estimates of the marginal posteriors $(\theta^{(j)}, \tilde{\pi}_{\gamma_0}(\theta^{(j)}|\boldsymbol{y}))$ are obtained numerically on a finite set $j = 1, \ldots, J$ corresponding to a non-negligible probability mass. In addition, the epsilon grid for any arbitrary prior distribution based on the Hellinger distance cannot be computed analytically. Below we concentrate on the first issue, while the general methodology needed for the ϵ-grid search will be explained in the next subsection.

4.1 Fast computation

In general, given the marginal posterior density $\pi_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y})$ computed for the base prior $\pi_{\boldsymbol{\gamma}_0}(\theta)$, the marginal posterior density $\pi_{\boldsymbol{\gamma}}(\theta|\boldsymbol{y})$ for the prior $\pi_{\boldsymbol{\gamma}}(\theta)$ with a new parameter specification $\boldsymbol{\gamma}$ instead of $\boldsymbol{\gamma}_0$ can be computed instantaneously by

$$\pi_{\boldsymbol{\gamma}}(\theta|\boldsymbol{y}) \propto \frac{\pi_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y})}{\pi_{\boldsymbol{\gamma}_0}(\theta)}\pi_{\boldsymbol{\gamma}}(\theta) \tag{7}$$

as proved in Appendix A. It is an extremely general and useful formula, which is the basis of, among others, importance sampling (Gelman et al., 2004) and the Laplace approximation (Tierney and Kadane, 1986). Formula (7) applied to an estimate of the marginal posterior distribution at the base prior $\tilde{\pi}_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y})$, provided by any tool capable of supplying the marginal posterior density, gives

$$\tilde{\pi}_{\boldsymbol{\gamma}}(\theta|\boldsymbol{y}) \propto \frac{\tilde{\pi}_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y})}{\pi_{\boldsymbol{\gamma}_0}(\theta)}\pi_{\boldsymbol{\gamma}}(\theta).$$

This general observation allows for an instantaneous computation of the Hellinger distance between two marginal posteriors $\tilde{\pi}_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y})$ and $\tilde{\pi}_{\boldsymbol{\gamma}}(\theta|\boldsymbol{y})$, arising from two slightly shifted prior parameter values $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma} \in \mathrm{G}_{\boldsymbol{\gamma}_0}(\epsilon)$. Note that

$$\mathrm{H}(\tilde{\pi}_{\boldsymbol{\gamma}}(\theta|\boldsymbol{y}), \tilde{\pi}_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y})) = \sqrt{1 - \mathrm{BC}(\tilde{\pi}_{\boldsymbol{\gamma}}(\theta|\boldsymbol{y}), \tilde{\pi}_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y}))},$$

with

$$\mathrm{BC}(\tilde{\pi}_{\boldsymbol{\gamma}}(\theta|\boldsymbol{y}), \tilde{\pi}_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y})) \approx \int \sqrt{\tilde{\pi}_{\boldsymbol{\gamma}}(\theta|\boldsymbol{y})\tilde{\pi}_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y})}d\theta$$

leads directly to the circular sensitivity estimates $S^c_{\boldsymbol{\gamma}_0}(\epsilon)$ and the worst-case sensitivity $S_{\boldsymbol{\gamma}_0}(\epsilon)$ in Equations (1) and (3). In general, for the computation of $\mathrm{BC}(\tilde{\pi}_{\boldsymbol{\gamma}}(\theta|\boldsymbol{y}), \tilde{\pi}_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y}))$ numerical integration is used. Although the true $\mathrm{BC}(\pi_{\boldsymbol{\gamma}}(\theta|\boldsymbol{y}), \pi_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y}))$ is restricted to the unit interval, the numerical approximation occasionally gives values slightly larger than one making truncation at 1 necessary. In such a case, minimal circular ϵ-local sensitivity estimates equal 0 and non-smooth circular sensitivity plots would indicate this numerical inaccuracy.

The above approach to instantaneous computation of the posterior for differing prior parameter specifications makes the model re-fit unnecessary. Similar computations can be carried out within any framework supporting marginal posterior density $\pi_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y})$ estimation and capable of providing $(\theta^{(j)}, \tilde{\pi}_{\boldsymbol{\gamma}_0}(\theta^{(j)}|\boldsymbol{y}))$, for $j = 1, \ldots, J$, numerically. We recommend, however, that for marginal posterior densities of precisions their log-transformed representations are used.

4.2 Grid search

For a fixed, small ϵ the search for the grid $\mathrm{G}_{\boldsymbol{\gamma}_0}(\epsilon)$ defined in Equation (2) requires exploration of the geometry around the prior parameter values $\boldsymbol{\gamma}_0$ in the space of distributions $\pi_{\boldsymbol{\gamma}}(\theta)$ (Figure 1). The goal is to find the set of prior parameter specifications $\boldsymbol{\gamma}$ such

that Hellinger distance between $\pi_{\boldsymbol{\gamma}}(\theta)$ and the base prior $\pi_{\boldsymbol{\gamma}_0}(\theta)$ is equal to ϵ fulfilling $\mathrm{G}_{\boldsymbol{\gamma}_0}(\epsilon) = \{\boldsymbol{\gamma} : \mathrm{H}(\pi_{\boldsymbol{\gamma}}(\theta), \pi_{\boldsymbol{\gamma}_0}(\theta)) - \epsilon = 0\}$. In order to find an ϵ grid for a base prior distribution in, say, two dimensions, a suitable transformation of the Cartesian coordinates $(\gamma^{(1)}, \gamma^{(2)})$ to the polar coordinates (ϕ, r) is used, where ϕ and r denote the angle in radians and modulus, respectively. For sake of stability of the algorithm $\log(r) = z$ is considered. We use a scaling factor $(\exp(z)\cos(\phi), \exp(z)\sin(\phi))$ which transforms the base prior parameter values $(\gamma_0^{(1)}, \gamma_0^{(2)})$ into an ϵ-distant pair $(\gamma^{(1)}, \gamma^{(2)})$ by finding the roots of the analytical equation numerically. Further details on the back-transformation used can be found in Appendix C3.

The polar coordinates approach guarantees that each direction is treated equally as the angles ϕ run through an equidistant grid on the $[-\pi, \pi]$ interval. It also implies a natural ordering of the grid points. This polar approach is applied to both normal and gamma priors used in applications in Section 5 and in the Supplementary Material but could easily be applied to any other two-parameter prior distribution of interest or even extended to higher dimensions.

5 Application: Disease mapping

In this section we demonstrate the use of the circular and worst-case ϵ-local sensitivity for a Bayesian hierarchical model. Additional examples for data sets with increasing hierarchical model complexity are discussed in the Supplementary Material. Marginal posterior densities necessary for the ϵ-local sensitivity computation have been contributed by `inla` (Appendix C).

The analysis of spatial variation of lip cancer in Scotland is based on a non-conjugate hierarchical model (Breslow and Clayton, 1993). Here, we consider observed (y) and expected (e) cases of lip cancer registered between 1975 and 1980 in each of the $n = 56$ counties in Scotland. We include an intercept α; a covariable x denoting the proportion of individuals who are employed in agriculture, fishing or forestry scaled by 1/10; a known offset $\log e$; and spatial components $\boldsymbol{v}$ and $\boldsymbol{u}$.

The unstructured region specific random effects in the vector $\boldsymbol{v}$ consist of independent Gaussian distributed random variables $v_i \overset{iid}{\sim} \mathrm{N}(0, \tau_{\boldsymbol{v}}^{-1})$ with precision $\tau_{\boldsymbol{v}}$, for which the gamma prior is assumed. A more involved structured intrinsic conditional autoregressive model (Besag et al., 1991) in component $\boldsymbol{u}$ assumes that conditions for neighbouring random effects tend to be similar. The Gaussian random field $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$ is defined as

$$u_i | u_j, i \neq j, \tau_{\boldsymbol{u}} \sim \mathrm{N}\left(\frac{1}{n_i}\sum_{i\sim j} u_j, \frac{1}{n_i \tau_{\boldsymbol{u}}}\right),$$

where $i \sim j$ indicates that two random effects i and j are neighbours and n_i is the number of neighbouring entities of the ith object. In order to guarantee the identifiability of the intercept, a sum-to-zero constraint on each connected component is used. For the precision $\tau_{\boldsymbol{u}}$ a gamma prior is assumed.

Let y_i be realisations of $Y_i \,|\, \mu_i \overset{ind}{\sim} \text{Po}(\mu_i),\ i = 1, \ldots, n$ and consider the following six models:

$$\begin{aligned}
\log \mu_i &= \log e_i + \alpha + v_i && (8)\\
\log \mu_i &= \log e_i + \alpha + u_i && (9)\\
\log \mu_i &= \log e_i + \alpha + v_i + u_i && (10)\\
\log \mu_i &= \log e_i + \alpha + \beta x + v_i && (11)\\
\log \mu_i &= \log e_i + \alpha + \beta x + u_i && (12)\\
\log \mu_i &= \log e_i + \alpha + \beta x + v_i + u_i. && (13)
\end{aligned}$$

Table 1: Worst-case sensitivity estimates for model components α, β, $\tau_{\boldsymbol{v}}$ and $\tau_{\boldsymbol{u}}$ in Section 5 for $\epsilon_0 = 0.00354$.

Model	α	β	$\tau_{\boldsymbol{v}}$	$\tau_{\boldsymbol{u}}$
(8)	0.004		0.244	
(9)	0.002			0.245
(10)	0.002		1.587	0.274
(11)	0.005	0.004	0.237	
(12)	0.004	0.004		0.268
(13)	0.005	0.005	1.568	0.355

As Fong et al. (2010, Supplementary Material) provided a very careful probabilistic elicitation of prior values, we have adopted their choices here. For the intercept α and the regression coefficient of the covariate β, we assumed normal priors with base prior parameter specification set at $\boldsymbol{\gamma}_0^{(\alpha),(\beta)} = (0, 0.001)$. For the unstructured and structured components we assumed gamma priors for $\tau_{\boldsymbol{v}}$ and $\tau_{\boldsymbol{u}}$ with base prior parameter values set to $\boldsymbol{\gamma}_0^{(\boldsymbol{v})} = (1, 0.14)$ and $\boldsymbol{\gamma}_0^{(\boldsymbol{u})} = (1, 0.34)$, respectively. Table 1 reports the worst-case sensitivities for each component in all six models for the base prior parameter values defined above. As an example, consider the sensitivity values in the last row of Table 1. For the structured component $\boldsymbol{u}$, a worst-case sensitivity of 0.355 was found. This value means that the mean change in the unit-variance normal distributions for posteriors is 35.5% of the mean change of unit-variance normal distributions in the prior. In contrast, super-sensitivity of 1.568 for the unstructured component $\boldsymbol{v}$ shows that the mean change in the unit-variance normal distributions for posteriors is 156.8% of the mean change in the unit-variance normal distributions for priors. Figure 4 shows for precisions $\boldsymbol{v}$ and $\boldsymbol{u}$ in model (13) the polar circular sensitivity plot centered at $\boldsymbol{\gamma}_0$ and rolled out on the line. Note that the left panel plots operate in three dimensions as they indicate the values of the circular sensitivity in each direction around $\boldsymbol{\gamma}_0^{(\boldsymbol{v})}$ (top) and $\boldsymbol{\gamma}_0^{(\boldsymbol{u})}$ (bottom).

Fong et al. (2010, Supplementary Material) prefer to always include both the unstructured $\boldsymbol{v}$ and the structured $\boldsymbol{u}$ components together. They argue that since the $\boldsymbol{u}$

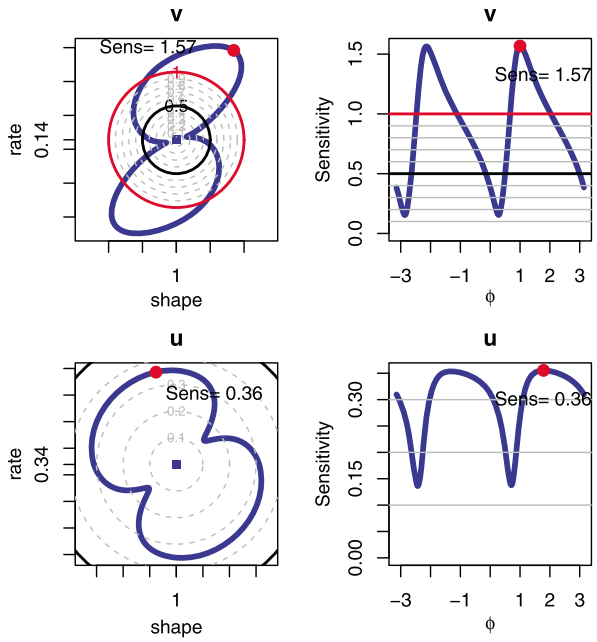

Figure 4: Polar sensitivity plots (left) centered at $\boldsymbol{\gamma}_0$ and sensitivity plots rolled out on the line (right) in each polar direction of the circular $S^c_{\boldsymbol{\gamma}_0}(\epsilon_0)$ together with the worst-case sensitivity $S_{\boldsymbol{\gamma}_0}(\epsilon_0)$ (red dot) for $\tau_{\boldsymbol{v}}$ at $\boldsymbol{\gamma}_0^{(\boldsymbol{v})}$ (top) and $\tau_{\boldsymbol{u}}$ at $\boldsymbol{\gamma}_0^{(\boldsymbol{u})}$ (bottom) components in model (13) in Section 5 for $\epsilon_0 = 0.00354$. The scale for sensitivity values (0.1, 0.2,. . ., 0.9, 1) is indicated by grey circles and lines, respectively. Sensitivity equal to 0.5 is marked by a black colour, whereas the sensitivity value 1 is marked by a red one.

model contains only a single parameter to govern both the spatial extent of dependence and the strength of this dependence, there is no place for pure randomness to be accommodated. Their models (0.3), (0.5), (0.4) and (0.6) are complemented with worst-case sensitivity estimates in rows corresponding to Equations (8), (10), (11) and (13) in Table 1. We conclude that for the lip cancer in Scotland data, having both $\boldsymbol{v}$ and $\boldsymbol{u}$ components at the same time in the model leads to super-sensitive marginal posteriors for the precision of the unstructured component with respect to the base prior parameter values choice. Apparently, a simultaneous inclusion of both latent components ends in an overparametrized model and the unstructured component becomes nearly non-identifiable (Eberly and Carlin, 2000).

This example illustrates that, ideally, the local sensitivity estimates should be investigated prior to the release of the final model. We suggest that the sensitivity values should be openly disclosed to the readers. In cases where a model with an included super-sensitivity is released, its justification or all measures aiming for its impact mitigation should be discussed.

6 Discussion

We introduced and utilized a new formal local robustness measure, which was able to automatically handle both circular and worst-case sensitivities in complex Bayesian hierarchical models. It hinged on two essential ingredients. Firstly, on an appropriately generated grid, which provided a well standardized way for modification of the base prior parameter values. Secondly, on an instantaneous computation of the marginal posterior density for priors with parameter values contained in the grid without any model re-fit.

It is a formal local robustness approach which dispenses with Taylor-expansion approximation and infinitesimal asymptotics. Instead, the circular sensitivity is computed for each polar direction chosen equidistantly on $[-\pi, \pi]$ around $\boldsymbol{\gamma}_0$. It can be both plotted and summarized by a single number. Whereas infinitesimal methods restrict their output to the worst-case value, the deeper insight offered by the circular sensitivity and its versatile summaries seems to be new.

We provided a careful and extensive investigation of the properties of the introduced ϵ-local sensitivity and comparison with the infinitesimal local sensitivity on conjugate examples, where estimation of exact analytical sensitivity estimates was possible. Due to limited space we reported mostly on the worst-case values. However, for all of the examples, polar plots of the circular sensitivity estimates and other summaries were available for more careful investigation.

As expected, a strong influence of the sample size on the prior sensitivity estimates emerged (Appendix C and Supplementary Material). Indeed, we observed that our measure automatically adjusts for increasing sample size by returning smaller prior sensitivity estimates. The choice of the ϵ for the grid search did not have much influence on sensitivity estimates but anchored calibration in terms of unit-variance normal distribution with shifted means. Our novel calibration use gave rise to a convenient interpretation of sensitivity estimates independent of the actual ϵ choice. It also provided a device to compare the sensitivity estimates obtained under both the Kullback-Leibler divergence and the Hellinger distance assumptions.

We identified several model components and base parameter values requiring more careful attention. Sensitivity estimates in Section 5 indicated clear identifiability problems when both unstructured and structured models were included in the model simultaneously. We believe that inclusion of both latent models at the same time leads to an overparametrized model.

One possible drawback of our approach is that we investigated ϵ-local sensitivity for each model component separately while keeping all other model component prior parameters fixed at their base values. It can happen, however, that a model is insensitive to changes in only one input at time, while being sensitive to simultaneous changes in more than one input. We believe, however, that ϵ-local sensitivity for each model component separately is actually what we are able to interpret in practice.

Another possible drawback of our approach is that it hinges on the choice of ϵ for the grid search surrounding the base prior parameter values. Fortunately, we were able to show in Appendix C and in the Supplementary Material that sensitivity estimates stay

numerically stable over a wide range of ϵ values leaving much freedom for its choice. We stress, however, that it is essential to apply the same ϵ for all model components for which local sensitivity is examined in order to provide well standardized robustness comparisons.

MCMC methodology is often not directly compatible with many of the robust Bayesian techniques that had been developed (Berger et al., 2000). Lesaffre and Lawson (2012) admit that a routine use of sensitivity procedures in MCMC cannot be afforded due to a substantial computational burden. In contrast, our fast ϵ-local sensitivity estimation technique has a potential to be implemented without much extra cost by any framework capable of estimating of marginal posterior densities. In particular, when the marginal posterior densities are obtained from MCMC samples followed by Rao-Blackwellization (Gelfand and Smith, 1990) or a dynamic exploration of posteriors by means of importance sampling (Narasimhan, 2005), our approach can be directly used.

The formal ϵ-local sensitivity measure gave a novel, reasonable, easy to interpret and useful piece of information about the marginal posterior distribution sensitivity to base prior parameter values. Its use was not restricted to conjugate examples but was easily extended to complex Bayesian hierarchical models giving new insight into the identifiability of model components given the data at hand. Additionally, we were able to determine in applications which model components were hard to learn from the data and identified several base prior parameter specifications requiring more careful attention. Therefore, we believe that thanks to our formal sensitivity measure checking for local robustness in complex Bayesian hierarchical models will become a part of routine statistical practice.

Appendix A (Proof)

Proof of the general Formula (7) used for instantaneous computation of the slightly shifted posterior in Section 4.1

Denote by $\boldsymbol{\gamma}_0$ the base prior parameter values and by $\boldsymbol{\gamma}$ the shifted prior parameter values. Reformulate

$$\pi_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y}) = \frac{\pi(\boldsymbol{y}|\theta)\pi_{\boldsymbol{\gamma}_0}(\theta)}{\pi_{\boldsymbol{\gamma}_0}(\boldsymbol{y})}$$

to obtain

$$\pi(\boldsymbol{y}|\theta) = \frac{\pi_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y})\pi_{\boldsymbol{\gamma}_0}(\boldsymbol{y})}{\pi_{\boldsymbol{\gamma}_0}(\theta)}.$$

Consequently,

$$\pi_{\boldsymbol{\gamma}}(\theta|\boldsymbol{y}) = \frac{\pi(\boldsymbol{y}|\theta)\pi_{\boldsymbol{\gamma}}(\theta)}{\pi_{\boldsymbol{\gamma}}(\boldsymbol{y})} = \frac{\pi_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y})\pi_{\boldsymbol{\gamma}_0}(\boldsymbol{y})\pi_{\boldsymbol{\gamma}}(\theta)}{\pi_{\boldsymbol{\gamma}}(\boldsymbol{y})\pi_{\boldsymbol{\gamma}_0}(\theta)}$$

can be rewritten as

$$\pi_{\boldsymbol{\gamma}}(\theta|\boldsymbol{y}) = \frac{c(\boldsymbol{y})\pi_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y})\pi_{\boldsymbol{\gamma}}(\theta)}{\pi_{\boldsymbol{\gamma}_0}(\theta)} \propto \frac{\pi_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y})\pi_{\boldsymbol{\gamma}}(\theta)}{\pi_{\boldsymbol{\gamma}_0}(\theta)},$$

with $c(\boldsymbol{y}) = \pi_{\boldsymbol{\gamma}_0}(\boldsymbol{y})/\pi_{\boldsymbol{\gamma}}(\boldsymbol{y})$. □

Appendix B (Sensitivity)

B1 Additional literature about sensitivity

Saltelli et al. (2004) and Cacuci et al. (2005) are prominent advocates for uncertainty analyses with respect to influential observations and the sampling distribution in complex frequentist models. In Bayesian statistics an additional source of uncertainty due to prior inadequacy has to be accounted for. The general sensitivity concept studied by Lavine (1992), Geisser (1993) and Gustafson and Wasserman (1995) involves consideration of influential observations, uncertainty of the sampling distribution and prior inadequacy at the same time. Zhu et al. (2007) develop the differential geometric approach to deal with this topic via geometric contamination of the prior distribution. The earlier advances by Van der Linde (2007) become special cases of this general methodology. The robustness analysis deals with the sensitivity of the results with respect to the prior distribution only. Weiss and Cook (1992) suggested a graphical approach for assessing posterior influence. In particular, the local approach is discussed by Gustafson (1996). Other advances to local Bayesian robustness can be found in Kass et al. (1989) and Weiss (1996).

B2 Formulae for the infinitesimal local sensitivity based on the Kullback-Leibler divergence in Section 3.1

Denote by $\pi_{\boldsymbol{\gamma}_0}$ and $\pi_{\boldsymbol{\gamma}}$ the densities of the $\mathrm{N}(\mu_0, \lambda_0^{-1})$ (true) and $\mathrm{N}(\mu, \lambda^{-1})$ (arbitrary) distributions, respectively. The value of the Kullback-Leibler divergence

$$\mathrm{D}_{\mathrm{KL}}(\pi_{\boldsymbol{\gamma}_0} || \pi_{\boldsymbol{\gamma}}) = \int \pi_{\boldsymbol{\gamma}_0}(\theta) \log\Big(\frac{\pi_{\boldsymbol{\gamma}_0}(\theta)}{\pi_{\boldsymbol{\gamma}}(\theta)}\Big) d\theta$$

between both distributions reads as follows:

$$\mathrm{D}_{\mathrm{KL}}(\pi_{\boldsymbol{\gamma}_0} || \pi_{\boldsymbol{\gamma}}) = \lambda(\mu_0 - \mu)^2/2 + \Big(\lambda/\lambda_0 - 1 - \log(\lambda/\lambda_0)\Big)/2. \tag{14}$$

For a fixed value of $\boldsymbol{\gamma}_0 = (\mu_0, \lambda_0)$ it is a function with respect to two arbitrary parameters $(\mu, \lambda) = \boldsymbol{\gamma}$. In the illustrative example in Section 3 the model component of interest is a scalar $\theta = m$. Equation (14) leads to an explicit expression for $i(\boldsymbol{\gamma}) = \mathrm{D}_{\mathrm{KL}}(\pi_{\boldsymbol{\gamma}_0}(m) || \pi_{\boldsymbol{\gamma}}(m))$. Due to conjugacy, Equation (14) can be re-used for analytical computation of the Kullback-Leibler divergence between two normal posteriors with parameters $(n\kappa\bar{y} + \lambda_0\mu_0)/(n\kappa + \lambda_0)$, $n\kappa + \lambda_0$, $(n\kappa\bar{y} + \lambda\mu)/(n\kappa + \lambda)$, $n\kappa + \lambda$ instead of μ_0, λ_0, μ and λ, respectively. This provides an analytical expression for $d(\boldsymbol{\gamma}) = \mathrm{D}_{\mathrm{KL}}(\pi_{\boldsymbol{\gamma}_0}(m|\boldsymbol{y}) || \pi_{\boldsymbol{\gamma}}(m|\boldsymbol{y}))$. Additionally, the entries of the matrices $\mathbf{D}^2 i(\boldsymbol{\gamma}_0)$ and $\mathbf{D}^2 d(\boldsymbol{\gamma}_0)$ can be derived analytically by differentiating functions $i(\boldsymbol{\gamma})$ and $d(\boldsymbol{\gamma})$ twice with respect to the parameters μ and λ and evaluating the result at $\boldsymbol{\gamma}_0 = (\mu_0, \lambda_0)$. This leads to

$$\mathbf{D}^2 i(\boldsymbol{\gamma}_0) = \begin{pmatrix} \lambda_0 & 0 \\ 0 & (2\lambda_0^2)^{-1} \end{pmatrix}$$

and

$$\mathbf{D}^2 d(\boldsymbol{\gamma}_0) = \begin{pmatrix} \lambda_0^2/(n\kappa + \lambda_0) & n\kappa\lambda_0(\mu_0 - \bar{y})/(n\kappa + \lambda_0)^2 \\ n\kappa\lambda_0(\mu_0 - \bar{y})/(n\kappa + \lambda_0)^2 & (2(n\kappa + \lambda_0)^2)^{-1} + (n\kappa(\mu_0 - \bar{y}))^2/(n\kappa + \lambda_0)^3 \end{pmatrix}.$$

The observation that $\mathbf{D}^2 i(\boldsymbol{\gamma}_0)$ and $\mathbf{D}^2 d(\boldsymbol{\gamma}_0)$ are Fisher information matrices for the prior and posterior families of distributions evaluated at $\boldsymbol{\gamma}_0$ leads to an identical analytical finding. The resulting matrix $(\mathbf{D}^2 i(\boldsymbol{\gamma}_0))^{-1}\mathbf{D}^2 d(\boldsymbol{\gamma}_0) = \mathbf{A}$ has the following entries:

$$\begin{aligned} a_{11} &= \lambda_0/(n\kappa + \lambda_0), \\ a_{12} &= n\kappa(\mu_0 - \bar{y})/(n\kappa + \lambda_0)^2, \\ a_{21} &= 2n\kappa\lambda_0^3(\mu_0 - \bar{y})/(n\kappa + \lambda_0)^2, \\ a_{22} &= \lambda_0^2/(n\kappa + \lambda_0)^2 + 2\lambda_0^2(n\kappa(\mu_0 - \bar{y}))^2/(n\kappa + \lambda_0)^3. \end{aligned}$$

The eigenvalues can be also computed analytically by using Equation (15) with entries from matrix $\mathbf{A}$:

$$\text{eigenvalue}^{\max}, \text{eigenvalue}^{\min} = \Big(a_{11} + a_{22} + / - \sqrt{(a_{11} - a_{22})^2 + 4a_{21}a_{12}}\Big)/2. \quad (15)$$

B3 Formulae for ϵ-local sensitivity based on the Kullback-Leibler divergence in Section 3.2

Epsilon grid based on the Kullback-Leibler divergence

For normal prior distribution, following equation (14), the search for the epsilon grid

$$\mathrm{G}_{\boldsymbol{\gamma}_0}(\epsilon) = \{\boldsymbol{\gamma} : \mathrm{D}_{\mathrm{KL}}(\pi_{\boldsymbol{\gamma}_0} || \pi_{\boldsymbol{\gamma}}) = \epsilon\}$$

corresponds to finding roots of

$$\mathrm{D}_{\mathrm{KL}}(\pi_{\boldsymbol{\gamma}_0} || \pi_{\boldsymbol{\gamma}}) - \epsilon = 0,$$

which is equivalent to finding roots μ of

$$\lambda(\mu_0 - \mu)^2/2 + \Big(\lambda/\lambda_0 - 1 - \log(\lambda/\lambda_0)\Big)/2 - \epsilon = 0$$

for λ, μ_0 and λ_0 kept fixed. For normal prior the roots themselves

$$\mu^{\min}, \mu^{\max} = \mu_0 - / + \sqrt{-(\lambda/\lambda_0 - 1 - \log(\lambda/\lambda_0) - 2\epsilon)/\lambda} \quad (16)$$

and an approximate range of λ, in which roots exist,

$$\lambda^{\min}, \lambda^{\max} = \lambda_0\big(1 + \epsilon - / + \sqrt{\epsilon(2 + \epsilon)}\big) \quad (17)$$

can be found analytically.

Calibration of the Kullback-Leibler divergence with respect to the normal distribution

Calibration of the Kullback-Leibler divergence with respect to the normal distribution has been already discussed by Goutis and Robert (1998). For sake of completeness, we provide the formulae explicitly below. They follow directly from Equation (14).

Lemma 1. *Calibration of the Kullback-Leibler divergence k between two normal densities N(0,1) and N(μ,1), respectively, can be computed as follows:*

$$k(\mu) = \mu^2/2$$

$$\mu(k) = \sqrt{2k}$$

Proof. Assume $\mu_0 = 0$, $\mu_1 = \mu$ with precisions $\lambda_0 = \lambda_1 = 1$ in Equation (14) to obtain $k(\mu) = k = \mu^2/2$. Solve this equality with respect to μ to obtain $\mu(k) = \sqrt{2k}$. □

Note that $\mu(k)$ is the calibration of the Kullback-Leibler divergence k with respect to the unit-variance normal distribution, as k between any two densities is the same as that between N(0,1) and N($\mu(k)$,1). If the Kullback-Leibler divergence k between any two densities is known, then we can quantify discrepancies between them in terms of the differences in mean from 0 to $\mu(k)$ for normal distribution with unit standard deviation. In particular, if we are interested in a mean shift $\mu = 0.01$, then the value of the ϵ for the epsilon grid based on the Kullback-Leibler divergence can be fixed by $k(\mu) = k(0.01) = \epsilon_0 =$ 5e-05. In such a case, the choice of $\epsilon_0 =$ 5e-05 corresponds to the shift in means by 0.01 for unit-variance normal distribution as $\mu(\epsilon_0) = \sqrt{2\epsilon_0} = 0.01$. Moreover,

$$\frac{\mu\big(\mathrm{D_{KL}}(\pi_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y})||\pi_{\boldsymbol{\gamma}}(\theta|\boldsymbol{y}))\big)}{\mu(\epsilon)} = \sqrt{\frac{\mathrm{D_{KL}}(\pi_{\boldsymbol{\gamma}_0}(\theta|\boldsymbol{y})||\pi_{\boldsymbol{\gamma}}(\theta|\boldsymbol{y}))}{\epsilon}}. \tag{18}$$

Consequently, the square root of the ϵ-local sensitivity based on the Kullback-Leibler divergence corresponds to the ϵ-local sensitivity in terms of unit-variance normal distributions.

B4 Formulae for ϵ-local sensitivity based on the Hellinger distance in Section 3.3

Hellinger distance between two normal distributions

Denote by $\pi_{\boldsymbol{\gamma}_0}$ and $\pi_{\boldsymbol{\gamma}}$ the densities of the $\mathrm{N}(\mu_0, \lambda_0^{-1})$ and $\mathrm{N}(\mu, \lambda^{-1})$ distributions, respectively. The value of the Hellinger distance

$$\mathrm{H}(\pi_{\boldsymbol{\gamma}}, \pi_{\boldsymbol{\gamma}_0}) = \sqrt{\frac{1}{2}\int_{-\infty}^{\infty}\Big\{\sqrt{\pi_{\boldsymbol{\gamma}}(\theta)} - \sqrt{\pi_{\boldsymbol{\gamma}_0}(\theta)}\Big\}^2 d\theta}$$

between both distributions reads as follows:

$$\mathrm{H}(\pi_{\boldsymbol{\gamma}}, \pi_{\boldsymbol{\gamma}_0}) = \sqrt{1 - \sqrt{\frac{2\sqrt{\lambda_0\lambda}}{\lambda_0 + \lambda}}\exp\Big(-\frac{\lambda_0\lambda(\mu_0 - \mu)^2}{4(\lambda_0 + \lambda)}\Big)}. \tag{19}$$

Epsilon grid based on the Hellinger distance

For normal prior distribution, following equation (19), the search for the epsilon grid

$$\mathrm{G}_{\gamma_0}(\epsilon) = \{\boldsymbol{\gamma} : \mathrm{H}(\pi_{\boldsymbol{\gamma}}, \pi_{\boldsymbol{\gamma}_0}) = \epsilon\}$$

corresponds to finding roots of

$$\mathrm{H}(\pi_{\boldsymbol{\gamma}}, \pi_{\boldsymbol{\gamma}_0}) - \epsilon = 0,$$

which is equivalent to finding roots μ of

$$\sqrt{\frac{2\sqrt{\lambda_0\lambda}}{\lambda_0+\lambda}}\exp\Big(-\frac{\lambda_0\lambda(\mu_0-\mu)^2}{4(\lambda_0+\lambda)}\Big) - (1-\epsilon^2) = 0$$

for λ, μ_0 and λ_0 kept fixed. For normal prior not only the range of λ, in which roots exist,

$$\lambda^{\min}, \lambda^{\max} = \lambda_0\big(1 - / + \sqrt{1-(1-\epsilon^2)^4}\big)^2(1-\epsilon^2)^{-4} \tag{20}$$

but also roots themselves

$$\mu^{\min}, \mu^{\max} = \mu_0 - / + \sqrt{-\frac{4(\lambda_0+\lambda)}{\lambda_0\lambda}\log\Big((1-\epsilon^2)\sqrt{\frac{\lambda_0+\lambda}{2\sqrt{\lambda_0\lambda}}}\Big)} \tag{21}$$

can be found analytically.

Given the grid $\mathrm{G}_{\gamma_0}(\epsilon)$ specified by equations (20) and (21) the circular ϵ-local sensitivity $S^c_{\gamma_0}(\epsilon)$ defined in Equation (1) based on the Hellinger distance can be obtained analytically. The Hellinger distance between two priors can be directly computed analytically using Equation (19). Due to conjugacy, Equation (19) can be used again for analytical computation of the Hellinger distance between two posteriors with parameters $(n\kappa\bar{y} + \lambda_0\mu_0)/(n\kappa + \lambda_0)$, $n\kappa + \lambda_0$, $(n\kappa\bar{y} + \lambda\mu)/(n\kappa + \lambda)$, $n\kappa + \lambda$ instead of μ_0, λ_0, μ and λ, respectively.

Calibration of the Hellinger distance with respect to the normal distribution

Lemma 2. *Calibration of the Hellinger distance h between two normal densities N(0,1) and N(μ,1), respectively, can be computed as follows:*

$$h(\mu) = \sqrt{1-\exp(-\mu^2/8)}$$
$$\mu(h) = \sqrt{-8\log(1-h^2)}$$

Proof. Assume $\mu_0 = 0$, $\mu_1 = \mu$ and precisions $\lambda_0 = \lambda_1 = 1$ in Equation (19) to obtain $h(\mu) = h = \sqrt{1-\exp(-\mu^2/8)}$. Solve this equality with respect to μ to obtain $\mu(h) = \mu = \sqrt{-8\log(1-h^2)}$. □

Table 2: Normal calibration of the Hellinger distance. Left: equal $\mu = 1$ increment, Right: equal $h = 0.1$ increment.

$h = h(\mu)$	$\mu(h) = \mu$	$h = h(\mu)$	$\mu(h) = \mu$
0.000	0.000	0.000	0.000
0.343	1.000	0.100	0.284
0.627	2.000	0.200	0.571
0.822	3.000	0.300	0.869
0.930	4.000	0.400	1.181
0.978	5.000	0.500	1.517
0.994	6.000	0.600	1.890
0.999	7.000	0.700	2.321
1.000	8.000	0.800	2.859
1.000	9.000	0.900	3.645

Table 2 demonstrates the values obtained by calibration. In particular, if we are interested in a mean shift $\mu = 0.01$, then the value of the ϵ for the epsilon grid based on the Hellinger distance can be fixed by $h(\mu) = h(0.01) = \epsilon_0 = 0.00354$. By Equation (5) the ϵ-local sensitivity based on the Hellinger distance approximates the calibrated ratio with respect to the unit-variance normal distribution directly.

Appendix C (Computations)

C1 `R-INLA` review

A wide range of Bayesian problems are covered by the latent Gaussian models framework and therefore effectively handled by INLA (Rue et al., 2009). An `R` package (`http://www.r-inla.org`) called INLA serves as an interface to the `inla` program. Its usage is similar to the familiar user-friendly `glm` function in `R`. The `inla` program allows the user to conveniently perform approximate Bayesian inference in latent Gaussian models. It is a fast and very versatile program, providing full Bayesian analysis of generalized linear mixed models (Fong et al., 2010; Martins et al., 2013). Computationally expensive models on high-dimensional data within stochastic partial differential equations (SPDEs) framework (Lindgren et al., 2011) can be tackled by `inla` as well. As output marginal posterior densities of all parameters in the model together with summary characteristics are offered by default. Although `inla` provides diagnostics for outlying observations via the conditional predictive ordinate (CPO) (Pettit, 1990; Geisser, 1993) default prior sensitivity diagnostics are still missing. Here, we closed this gap and provided a ready to use `priorSens` package in `R` facilitating the use of sensitivity measure described here.

C2 The `priorSens` package in `R`

We developed a handy package `priorSens` for routine, every-day sensitivity computation. It can be used with practically no extra programming effort needed for default sensitivity investigations within `inla` or to obtain standardized grids for any alternative robustness considerations. Currently, normal and gamma priors are supported by the package covering already a great amount of possible models. Further extensions to support other priors are possible. The `priorSens` package can be obtained upon request from the authors. We plan to include it as a default option in the `R-INLA` framework. All sensitivity estimates presented in Section 5 and in the Supplementary Material were evaluated by the `priorSens` package in `R`. The required input for the ϵ-local sensitivity to be computed are the base prior distribution $\pi_{\boldsymbol{\gamma}_0}(\theta)$ with parameter specification $\boldsymbol{\gamma}_0$ and the corresponding marginal posterior density $(\theta^{(j)}, \tilde{\pi}_{\boldsymbol{\gamma}_0}(\theta^{(j)}|\boldsymbol{y}))$ obtained numerically on a finite set $j = 1, \ldots, J$.

C3 Back-transformation for grid search

As discussed in Section 4.2 the scaling factor $(\exp(z)\cos(\phi), \exp(z)\sin(\phi))$ is transformed back to Cartesian coordinates using

$$\left[\gamma_0^{(1)} + r\cos(\phi)c^x(\phi)\right]$$

and

$$\left[\gamma_0^{(2)} + r\sin(\phi)c^y(\phi)\right],$$

where

$$c^x(\phi) = \begin{cases} r^*(0) & \text{if } \phi \in [-\pi/2, \pi/2], \\ r^*(\pi) & \text{if } \phi \in [\pi/2, -\pi/2] \end{cases}$$

and

$$c^y(\phi) = \begin{cases} r^*(\pi/2) & \text{if } \phi \in [0, \pi], \\ r^*(-\pi/2) & \text{if } \phi \in [\pi, 0], \end{cases}$$

with $r^*(\delta)$, for radian values $\delta = -\pi/2, 0, \pi/2, \pi$, denoting the modulus values obtained at δ angles during a pre-exploration of the polar coordinate space. The factors $c^x(\phi)$ and $c^y(\phi)$ are necessary to scale the problem so that r is close to 1 across different prior distributions. This practice standardizes the task of computing $\mathrm{G}_{\boldsymbol{\gamma}_0}(\epsilon)$, which makes the numerical algorithm more stable and generally applicable. The `priorSens` package relies on the above general approach for epsilon grid search.

C4 Epsilon local sensitivity in `R-INLA`

Our fast general ϵ-local sensitivity methodology based on the Hellinger distance can be implemented without much extra cost by any technique capable of computing marginal posterior distributions, in particular, by the `R-INLA` framework. In practice, however, two settings for the cardinality of the grid $\mathrm{G}_{\boldsymbol{\gamma}_0}(\epsilon)$ and the value of ϵ have to be fixed (Section 4.2). In applications shown below we consider 400 polar directions and use one particular $\epsilon_0 = 0.00354$ for the grid search, which corresponds to a unit-variance normal

distribution with mean equal to 0.01 with respect to the Hellinger distance (Section 2.2 and Appendix B4). The continued illustrative example in Appendix C5 and several examples in the Supplementary Material indicate that the exact sensitivity estimates stay stable over a wide range of ϵ values. Therefore, there is some room for a tolerable ϵ choice.

Our `inla` computations were run for default settings: simplified Laplace strategy for the marginal posterior approximation and central composite design (CCD) for integrating out the hyperparameters. In general, for precisions log-transformed marginal posterior densities were used. In addition, for precisions of latent components in `R-INLA` such as unstructured ("iid"), structured ("ICAR"), the latent Gaussian random walk of the first ("rw1") and the second ("rw2") order an appropriate tuning for their marginal posteriors provided by the function `inla.hyperpar()` was utilized.

C5 Epsilon local sensitivity based on the Hellinger distance (continued)

Reconsider the model and data introduced in the illustrative example in Section 3. We studied the ϵ-local sensitivity values, when the exact analytical marginal posterior density estimate was replaced by the approximate one provided by `inla`. The worst-case sensitivity $\mathrm{S}^{\texttt{inla}}_{\boldsymbol{\gamma}_0}(\epsilon_0) = 1.82$ for the `inla`-driven approach and the analytically computed estimate $\mathrm{S}^{\text{exact}}_{\boldsymbol{\gamma}_0}(\epsilon_0) = 1.82$ agreed perfectly well. Absolute and relative error ranges of the `inla`-driven circular sensitivity estimates with respect to the analytical ones were equal to (-2e-04, 0.0015) and (-0.006, 0.006), respectively.

In addition, we varied the grid epsilon values and computed both the corresponding analytical and `inla`-driven sensitivities. Table 3 shows good agreement of exact and `inla`-driven sensitivity estimates for the small sample at hand. Besides, we assumed in the example a larger sample size $n = 100$ instead of $n = 4$ keeping all other model parameters fixed. We expected less sensitivity to the prior parameter values when much evidence is provided by the data. Again exact and `inla`-driven sensitivity estimates agreed very well. Lower estimates of sensitivity obtained for $n = 100$ in Table 3 indicated that our measure reacts reasonably to increased sample size. Interestingly, even if the prior is clearly displaced but there is enough evidence in the data ($n = 100$), sensitivity can be small.

Table 3: Worst-case sensitivity estimates for mean as a function of ϵ for $n = 4$ and $n = 100$ in the example in Section 3 and Appendix C5 at $\boldsymbol{\gamma}_0 = (\mu_0, \lambda_0) = (70, 0.5)$.

ϵ	$\mathrm{S}^{\text{exact}}_{\boldsymbol{\gamma}_0}(\epsilon), n = 4$	$\mathrm{S}^{\texttt{inla}}_{\boldsymbol{\gamma}_0}(\epsilon), n = 4$	$\mathrm{S}^{\text{exact}}_{\boldsymbol{\gamma}_0}(\epsilon), n = 100$	$\mathrm{S}^{\texttt{inla}}_{\boldsymbol{\gamma}_0}(\epsilon), n = 100$
0.0001	1.81	1.81	0.43	0.43
0.0005	1.81	1.82	0.43	0.43
0.0010	1.82	1.82	0.43	0.43
0.0050	1.83	1.83	0.43	0.43
0.0100	1.84	1.84	0.44	0.44

For several additional examples discussed in the Supplementary Material we once again detected an excellent agreement of the `inla`-driven and exact sensitivity estimates. As expected, a strong influence of the sample size on the prior sensitivity estimates emerged. Indeed, we observed that our measure automatically adjusts for increasing sample size by returning smaller prior sensitivity estimates.

Supplementary Material

Supplementary Material: Sensitivity Analysis for Bayesian Hierarchical Models (DOI: 10.1214/14-BA909SUPP; .pdf).

References

Amari, S. (1990). *Differential-Geometrical Methods in Statistics. 2nd Edition. Lecture Notes in Statistics*, volume 28. Springer-Verlag. 325, 326

Amari, S. and Nagaoka, H. (2000). *Methods of Information Geometry*. Oxford University Press. 325

Berger, J. O., Ríos Insua, D., and Ruggeri, F. (2000). "Bayesian Robustness." In Ríos Insua, D. and Ruggeri, F. (eds.), *Robust Bayesian Analysis*, 1–32. Springer-Verlag. 322, 323, 338

Besag, J., York, J., and Mollié, A. (1991). "Bayesian image restoration, with two applications in spatial statistics." *Annals of the Institute of Statistical Mathematics*, 43(1): 1–59. 334

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. 325

Box, G. (1980). "Sampling and Bayes' inference in scientific modelling and robustness." *Journal of the Royal Statistical Society, Series A*, 143(4): 383–430. 324, 328

Breslow, N. E. and Clayton, D. G. (1993). "Approximate inference in generalized linear mixed models." *Journal of the American Statistical Association*, 88(421): 9–25. 334

Cacuci, D. G., Ionescu-Bujor, M., and Navon, I. M. (2005). *Sensitivity and Uncertainty Analysis. Volume II, Applications to Large-Scale Systems*. Chapman & Hall. 339

Carlin, B. and Louis, T. (1998). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall/CRC. 322

Clarke, B. and Gustafson, P. (1998). "On the overall sensitivity of the posterior distribution to its inputs." *Journal of Statistical Planning and Inference*, 71(1-2): 137–150. 322, 327

Cook, R. (1986). "Assessment of local influence." *Journal of the Royal Statistical Society, Series B*, 48(2): 133–169. 322

Dawid, A. (1977). "Further comments on some comments on a paper by Bradley Efron." *The Annals of Statistics*, 5(6): 1249. 326

— (1979). "Conditional independence in statistical theory." *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(1): 1–31. 322

Dey, D. and Birmiwal, L. (1994). "Robust Bayesian analysis using divergence measures." *Statistics & Probability Letters*, 20(4): 287–294. 325, 326

Eberly, L. and Carlin, B. (2000). "Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models." *Statistics in Medicine*, 19(17-18): 2279–2294. 336

Evans, M. and Moshonov, H. (2006). "Checking for prior-data conflict." *Bayesian Analysis*, 4(1): 893–914. 324

Fong, Y., Rue, H., and Wakefield, J. (2010). "Bayesian inference for generalized linear mixed models." *Biostatistics*, 11(3): 397–412. 335, 343

Geisser, S. (1992). "Bayesian perturbation diagnostics and robustness." In Goel, P. and Iyengar, N. (eds.), *Bayesian Analysis in Statistics and Econometrics*, 289–301. Springer-Verlag. 322

— (1993). *Predictive Inference: An Introduction*. Chapman & Hall, Inc. 339, 343

Gelfand, A. and Sahu, S. (1999). "Identifiability, improper priors, and Gibbs sampling for Generalized Linear Models." *Journal of the American Statistical Association*, 94(445): 247–253. 322

Gelfand, A. and Smith, A. (1990). "Sampling-based approaches to calculating marginal densities." *Journal of the American Statistical Association*, 85(410): 398–409. 338

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004). *Bayesian Data Analysis. 2nd Edition*. Chapman & Hall/CRC. 333

Gilks, W., Richardson, S., and Spiegelhalter, D. (1996). *Markov Chain Monte Carlo*. Chapman & Hall. 323

Goutis, C. and Robert, C. (1998). "Model choice in generalised linear models: A Bayesian approach via Kullback-Leibler projections." *Biometrika*, 85(1): 29–37. 326, 340

Gustafson, P. (1996). "Local sensitivity of inferences to prior marginals." *Journal of the American Statistical Association*, 91(434): 774–781. 339

— (2000). "Local robustness in Bayesian analysis." In Ríos Insua, D. and Ruggeri, F. (eds.), *Robust Bayesian Analysis*, 71–88. Springer-Verlag. 322, 325

Gustafson, P. and Wasserman, L. (1995). "Local sensitivity diagnostics for Bayesian inference." *The Annals of Statistics*, 23(6): 2153–2167. 339

Ibrahim, J., Zhu, H., and Tang, N. (2011). "Bayesian local influence for survival models." *Lifetime Data Analysis*, 17(1): 43–70. 330

Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press. 325

Kadane, J. (1992). "Comments to: "Bayesian perturbation diagnostics and robustness" by S. Geisser." In Goel, P. and Iyengar, N. (eds.), *Bayesian Analysis in Statistics and Econometrics*, 298–300. Springer-Verlag. 322

Kass, R., Tierney, L., and Kadane, J. (1989). "Approximate methods for assessing influence and sensitivity in Bayesian analysis." *Biometrika*, 76(4): 663–674. 339

Lavine, M. (1992). "Local predictive influence in Bayesian linear models with conjugate priors." *Communications in Statistics - Simulation and Computation*, 21(1): 269–283. 339

Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag. 325

Lesaffre, E. and Lawson, A. (2012). *Bayesian Biostatistics*. John Wiley & Sons. 323, 338

Lindgren, F., Rue, H., and Lindström, J. (2011). "An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic differential equation approach." *Journal of the Royal Statistical Society, Series B.*, 73(4): 423–498. 343

Martins, T., Simpson, D., Lindgren, F., and Rue, H. (2013). "Bayesian computing with INLA: new features." *Computational Statistics & Data Analysis*, 67: 68–83. 343

McCulloch, R. (1989). "Local model influence." *Journal of the American Statistical Association*, 84(406): 473–478. 322, 326, 327, 328, 329

Millar, R. and Stewart, W. (2007). "Assessment of locally influential observations in Bayesian models." *Bayesian Analysis*, 2(2): 365–384. 322

Müller, U. (2012). "Measuring prior sensitivity and prior informativeness in large Bayesian models." *Journal of Monetary Economics*, 59(6): 581–597. 327

Narasimhan, B. (2005). "Lisp-Stat to Java to R." *Journal of Statistical Software*, 13(4): 1–10. 338

Oakley, J. and O'Hagan, A. (2004). "Probabilistic sensitivity analysis of complex models: a Bayesian approach." *Journal of the Royal Statistical Society, Series B*, 66(3): 751–769. 322

Pérez, C., Martín, J., and Rufo, M. (2006). "MCMC-based local parametric sensitivity estimation." *Computational Statistics & Data Analysis*, 51(2): 823–835. 327

Pettit, L. (1990). "The conditional predictive ordinate for the normal distribution." *Journal of the Royal Statistical Society, Series B*, 52(1): 175–184. 343

Plummer, M. (2001). "Local sensitivity in Bayesian graphical models." URL http://www-ice.iarc.fr/~martyn/papers/sensitivity.ps 327

Rao, C. (1945). "Information and the accuracy attainable in the estimation of statistical parameters." *Bulletin of the Calcutta Mathematical Society*, 37: 81–91. 326

Ríos Insua, D., Ruggeri, F., and Martín, J. (2000). "Bayesian Sensitivity Analysis." In Saltelli, A., Chan, K., and Scott, E. M. (eds.), *Sensitivity Analysis*, 225–244. John Wiley & Sons. 322

Robert, C. (1996). "Intrinsic losses." *Theory and Decision*, 40(2): 191–214. 325

Roos, M. and Held, L. (2011). "Sensitivity analysis in Bayesian generalized linear mixed models for binary data." *Bayesian Analysis*, 6(2): 259–278. 325, 326, 330

Rue, H., Martino, S., and 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. 343

Ruggeri, F. (2008). "Bayesian Robustness." *Forum: Robustness Analysis, In: European Working Group "Multiple Criteria Decision Aiding"*, 3(17). 322, 323

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global Sensitivity Analysis. The Primer*. John Wiley & Sons. 322

Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M. (2004). *Sensitivity Analysis in Practice. A Guide to Assessing Scientific Models*. John Wiley & Sons. 339

Sivaganesan, S. (2000). "Global and local robustness approaches: uses and limitations." In Ríos Insua, D. and Ruggeri, F. (eds.), *Robust Bayesian Analysis*, 89–108. Springer-Verlag. 322

Tierney, L. and Kadane, J. (1986). "Accurate approximations for posterior moments and marginal densities." *Journal of the American Statistical Association*, 81(393): 82–86. 333

Van der Linde, A. (2007). "Local influence on posterior distributions under multiplicative modes of perturbation." *Bayesian Analysis*, 2(2): 319–332. 339

Weiss, R. (1996). "An approach to Bayesian sensitivity analysis." *Journal of the Royal Statistical Society, Series B.*, 58(4): 739–750. 339

Weiss, R. and Cook, R. (1992). "A graphical case statistic for assessing posterior influence." *Biometrika*, 79(1): 51–55. 339

Zhu, H., Ibrahim, J., Lee, S., and Zhang, H. (2007). "Perturbation selection and influence measures in local influence analysis." *The Annals of Statistics*, 35(6): 2565–2588. 339

Zhu, H., Ibrahim, J., and Tang, N. (2011). "Bayesian influence analysis: a geometric approach." *Biometrika*, 98(2): 307–323. 322, 327, 328, 330, 332

Acknowledgments

The authors are grateful to Finn Lindgren, Daniel Simpson, Daniel Sabanés Bové and Andrea Riebler for enlightening discussions, to Diego Morosoli for acquiring references and to Isaac Gravestock for looking over the final version of the manuscript. They thank two Editors and two anonymous reviewers for their valuable comments and suggestions on an earlier version of the manuscript.