Statistica Sinica **30** (2020), 579-599
doi:https://doi.org/10.5705/ss.202017.0265

LOCAL AND GLOBAL ROBUSTNESS WITH CONJUGATE AND SPARSITY PRIORS

Vahed Maroufy and Paul Marriott

University of Texas Health Center at Houston and University of Waterloo

Abstract: This paper studies the sensitivity of posteriors to local and global perturbations of conjugate, shrinkage and sparsity priors. The perturbations are natural, geometrically motivated, and generalize the linear perturbation studied in Gustafson (1996). A geometric approach is also employed for optimizing the sensitivity direction function, which is defined on a convex space with non-trivial boundaries. The robustness of multi-dimensional models with shrinkage and sparsity priors is studied through simulation and through two real data sets; a benign breast disease study, and an adolescent placement study. Our results illustrate that there can exist significant sensitivity of the covariate coefficient estimates to perturbations of the independent weakly informative prior distributions.

Key words and phrases: Bayesian sensitivity, local mixture model, perturbation space, smooth manifold, shrinkage and sparsity priors.

1. Introduction

Statistical analyses are often performed using assumptions which are not directly validated. Hence, there is always interest in investigating the degree to which statistical inference is sensitive to perturbations of the model and data. Specifically, in a Bayesian context, when priors have been chosen, the sensitivity of the posterior to prior choice is an important issue. A rich literature on sensitivity to perturbations of data, prior and sampling distribution exists, see for example: Cook (1986); McCulloch (1989); Lavine (1991); Ruggeri and Wasserman (1993); Blyth (1994); Gustafson (1996); Critchley and Marriott (2004); Linde (2007); Zhu et al. (2007), Zhu, Ibrahim and Tang (2011), Zhu, Ibrahim and Tang (2014) and Kurtek and Bharath (2015).

In this paper we consider both local and global sensitivity with respect to perturbations of a conjugate base prior – still the most common case in practice. We also look at perturbations of weakly informative regularity priors in the multi-dimensional case. Our method is designed to have four important properties. Firstly, the perturbation spaces have a structure that allows the analyst to select the generality of the perturbation in a clear way. Secondly, we want the space to be computationally tractable, hence we focus on convex sets inside linear spaces which have a clear geometric structure. Thirdly, we want, in order to allow for meaningful comparisons, the spaces to be consistent with elicited prior knowledge. Thus if a subject matter expert indicates that a prior moment or quantile has a known value – or if a constraint such as symmetry is appropriate – then all perturbed priors should be consistent with this information. Finally, we are going to consider perturbations which are mixtures over standard families. The motivation here is that mixing allows us to explore if the analyst has been over-precise in the specification of the prior by allowing for unthought of heterogeneity. In general, spaces of mixture models are complex but we build on the work of Maroufy and Marriott (2016a) which shows how discrete mixtures of local mixture models can construct a very flexible, but tractable, space, see §2.1.

Sensitivity analysis with respect to a perturbation of the prior, which is the focus of this paper, is commonly called robustness analysis (Insua and Ruggeri (2000)). In robustness analysis it is customary to choose a base prior model and a plausible class of perturbations. The influence of a perturbation is assessed either locally, or globally, by measuring the change in certain features of the posterior distribution. For instance, Gustafson (1996) studies linear and non-linear model perturbations, and Weiss (1996) uses a multiplicative perturbation to the base prior and specifies the important perturbations using the posterior density of the parameter of interest. Common global measures of influence include divergence functions (Weiss (1996)) and relative sensitivity (Ruggeri and Sivaganesan (2000)).

In local analysis, the rate at which a posterior quantity changes, relative to a change in the prior, quantifies sensitivity (Gustafson (1996); Linde (2007),Berger, Rios Insua and Ruggeri (2000)). Gustafson (1996), which we follow closely, obtains the direction in which a certain posterior expectation has the maximum sensitivity to prior perturbation by considering a mapping from the space of perturbations to the space of posterior expectations. In Linde (2007), the Kullback-Leibler and χ^2-divergence functions are utilized for assessing local sensitivity with respect to a multiplicative perturbation of the base prior or likelihood model. They approximate the local sensitivity using the Fisher information of the mixing parameter under additive and geometric mixing. Non-parametric Fisher information is also used in Kurtek and Bharath (2015) to give a geometric framework for sensitivity analysis with respect to likelihood and prior perturbation.

The approach of this paper to defining the perturbation space extends the linear perturbations studied in Gustafson (1996) in a number of ways. We do not require the same positivity condition, rather use one which is more general and returns naturally normalized distributions. Further, our space is highly tractable, due to its intrinsic linearity and convexity. Finally it is clear, with our formulation, how to remain consistent with prior information which may have been elicited from an expert. The cost associated with our generalization is the existence of boundaries defined by (2.3) in Section 2.1 and having to use the methods developed to work with it. We also can compare our method with the geometric approach of Zhu, Ibrahim and Tang (2011) which uses a manifold based approach. Our, more linear approach, considerably improves interpretability and tractability while sharing a similar underlying geometric foundation.

The paper is organized as follows. In Section 2, the perturbation space is introduced and its properties are studied. Sections 3.1 and 3.2 develop the theory of local and global sensitivity analysis. Section 3.3 describes the geometry of the perturbation parameter space and proposes possible algorithms for quantifying local and global sensitivity. Section 3.4 presents an illustrative example elucidating how experts' knowledge may be incorporated in the prior perturbation. In Section 4 the sensitivity of covariate coefficients in multi-dimensional models with respect to perturbations of weakly informative priors is studied though two real data examples: a benign breast disease study and an adolescent placement study. The proofs are sketched in the Appendix.

2. Perturbation Space

2.1. Theory and geometry

This paper looks at perturbations of a conjugate and sparsity priors which are generated by a mixing mechanism. Specifically, suppose the parameter μ has a prior $\pi_0(\mu;\theta)$, where θ is an, analyst selected, hyperparameter. We perturb by looking at priors which are mixtures over θ, i.e.

$$\int \pi_0(\mu;\theta)dQ(\theta).$$

In this formulation we can think of the distribution the choice of Q as parametrising the perturbation space.

To allow full generality of the mixture structure it is tempting, following Lindsay (1995), to allow Q to be any finite discrete distribution with an unknown number of components. i.e.

$$\int \pi_0(\mu;\theta)dQ(\theta) = \sum_{i=1}^{N} \rho_i \pi_0(\mu;\theta_i), \tag{2.1}$$

where $\rho_i > 0$, and $\sum_{i=1}^{N} \rho_i = 1$. In this case the perturbation space would be 'parameterised' by N, the number of components, $(\theta_1, \ldots, \theta_N)$, the components, and $(\rho_1, \ldots, \rho_N)$, the mixing weights. However, this parameterization has many problems in implementation. Specifically, it is poorly identified and has complex boundaries.

A key example of the identification problems is the case when all the mixing components are close to one another. Here there are many ways of representing essentially the same mixing distribution, see Maroufy and Marriott (2016a). We think of this case, where all mixing components are close to one another, as a *local perturbation* of the prior. Having a single set of closely grouped components – or the much more general situation where Q is any small-variance distribution – is exactly the situation which motivated the design of the local mixture model (LMM), see Marriott (2002), and Anaya-Izquierdo and Marriott (2007a). The intuition is that for a local perturbation all the mixing component distributions will lie close to a low dimensional linear space. This space is spanned by derivatives of the prior and parameterised with a small number of identified parameters.

Definition 1. *For a density function, $f(y;\theta)$, belonging to the exponential family, the local mixture model, of order k, centred at ϑ, is defined as*

$$g_\vartheta(y;\lambda) = f(y;\vartheta) + \sum\nolimits_{j=1}^{k} \lambda_j f^{(j)}(y;\vartheta), \tag{2.2}$$

where $\lambda := (\lambda_1, \ldots, \lambda_k) \in \Lambda_\vartheta$, $f^{(j)}(y;\vartheta) = \partial^j f(y;\theta)/\partial\theta^j \mid_{\theta=\vartheta}$. The parameter space,

$$\Lambda_{\theta_0} = \left\{ \lambda | f(x;\theta_0) + \sum\nolimits_{j=1}^{k} \lambda_j v_j(x;\theta_0) \geq 0, \textit{for all } x \right\} \tag{2.3}$$

is convex with a boundary determined by the non-negativity condition in (2.2).

It is shown in Marriott (2002) and Anaya-Izquierdo and Marriott (2007a) that the local mixture gives an excellent way of parameterising this class of local perturbations by simply approximating expression (2.1) by (2.2). The parameters of (2.2) are well identified, however the 'cost' of this representation is working with the boundary of Λ_ϑ, see Maroufy and Marriott (2015).

Example 1. In Section 3.4 we look at an example where a sample of size $n = 15$ is taken from a normal model $\mathcal{N}(1,1)$, and the base prior, for the mean, is

$\mathcal{N}(2,1)$. For LMMs of order $k=4$, Maroufy and Marriott (2016a) show that for any $\delta > 0$ there exist an interval $I = [\vartheta - \epsilon_1(\delta,\vartheta), \vartheta + \epsilon_2(\delta,\vartheta)]$ such that $\left|\int_I f(y;\theta)dQ - g_\vartheta(y;\boldsymbol{\lambda})\right| < \delta$, for all x. Here I is interpreted as a domain for the mixing distribution Q, for which the LMM model $g_\vartheta(y;\boldsymbol{\lambda})$ can approximate the mixture model $\int_I f(y;\theta)dQ$ up to an arbitrary small error, δ. They also give ways to calculate I for different exponential family distribution. In particular, for a normal distribution the approximation error δ depend only on σ, and for a fixed variance σ they give the interval $I = [\vartheta - 0.6\sigma, \vartheta + 0.6\sigma]$. Then, for the base prior $N(2,1)$ it is easy to show that the $k=4$ local mixture model will cover all perturbations of the prior which are mixtures over $\mathcal{N}(\mu,1)$, when the mixing distribution has support inside $[1.4, 2.6]$, see Maroufy and Marriott (2016a, Example 5).

In Zhu, Ibrahim and Tang (2011, Definition 1) a perturbation manifold is defined to be a triplet consisting of the space of perturbations $\mathcal{M}$, a proper metric closely related to Fisher Information, and the corresponding Levi-Civita connection. Examples of this perturbation space include an additive ϵ-contamination class and can include linear and non-linear perturbation schemes. In our approach we relax the assumption that the perturbation space is a manifold, since we include the boundaries. These can be non-smooth and have singularities, see Maroufy and Marriott (2015). Furthermore, the linear structure on our spaces agree with the, so-called, mixture connection of Amari (1990), which were shown to defined an affine structure in measure space, Marriott (2002). Our perturbation spaces are convex subsets of this affine space. Thus the emphasis is on convex geometry rather than differential geometry.

While local mixtures, and hence local perturbations, have very attractive inferential properties – unlike general mixture models – they are restrictive in the sense that they are only 'local'. This restriction can be completely removed – while still keeping attractive inferential properties – by considering finite mixtures of local mixtures, see for example Maroufy and Marriott (2016a).

Definition 2. *Let θ_l be a set of user selected, and suitably separated, grid-points as defined in* Maroufy and Marriott (2016a)*, then a finite mixture of local mixtures is defined as the convex combination*

$$\sum_{l=1}^{L} \rho_l h(x;\lambda^l,\theta_l),$$

where $\lambda^l \in \Lambda_{\theta_l}$, $\sum_{l=1}^{L} \rho_l = 1$.

In this paper, for simplicity, we mostly consider the single component LMM case but point out that the generalisation of Definition 2 is always possible.

2.2. Prior perturbation

Suppose that the base prior model is $\pi_0(\mu;\theta)$, the probability (density) function of a natural exponential family with the hyper-parameter θ.

Definition 3. *The local perturbed prior model corresponding to* $\pi_0(\mu;\theta)$ *is defined by*

$$\begin{aligned}\pi(\mu,\lambda;\theta) &:= \pi_0(\mu;\theta) + \sum\nolimits_{j=1}^{k} \lambda_j \pi_0^{(j)}(\mu;\theta) \\ &= \pi_0(\mu;\theta)\left\{1 + \sum\nolimits_{j=1}^{k} \lambda_j q_j(\mu,\theta)\right\}, \lambda \in \Lambda_\theta, \qquad (2.4)\end{aligned}$$

where $\lambda = (\lambda_1, \ldots, \lambda_k)$ *is the perturbation parameter vector, and* $q_j(\mu,\theta) = (\pi_0^{(j)}(\mu;\theta))/(\pi_0(\mu;\theta))$ *are polynomials of degree* j.

In Definition 3, π_0 is perturbed linearly, in a way similar to, but distinct from, the linear perturbation

$$\tau(\cdot, \pi_0, u^*) = \pi_0(\cdot) + u^*(\cdot), \quad u^*(\cdot) > 0 \qquad (2.5)$$

studied in Gustafson (1996), but with a different positivity condition, and is, as we shall show, very interpretable for applications. Definition 3 can also be seen as the multiplicative perturbation model $\pi(\mu,\lambda;\theta) = \pi_0(\mu,;\theta)h^*(\mu;\lambda,\theta)$ studied in Linde (2007).

2.3. Sensitivity analyses

We consider two different approaches for evaluating sensitivity with respect to a perturbation: worst case and average. The first, which is the more conventional approach, maximizes an objective function with respect to the multi-dimensional perturbation parameter. The function characterizes the discrepancy between the base model and each feasible perturbed model in the perturbation space. We exploit this approach both for assessing local and global sensitivity. Our local sensitivity (Section 3.1) is derived based on maximizing the directional derivative of the discrepancy function in a similar way to Gustafson (1996). For global sensitivity (Section 3.2) we characterize the discrepancy between the base and perturbed models via two different distance measures; the difference between the posterior means, and the Kullback-Leibler divergence function between the posterior predictive distributions.

For the averaging approach, see Section 4, we treat the perturbation parameters as nuisance parameters and learn about them jointly with the parameters of interest. We use a Markov Chain Monte Carlo method for the estimation procedure which is shown to be efficient and straightforward to implement because of the tractable convex geometry of our perturbation space, Maroufy and Marriott (2016a). This approach is highly efficient for dealing with sensitivity analyses in multi-dimensional models, see Section 4.

3. Measuring Sensitivity

3.1. Directional sensitivity

In this section we study the influence of local perturbations, defined inside the perturbation space, on the posterior mean. Following Gustafson (1996) we obtain the direction of sensitivity using the Fréchet derivative of a mapping between two normed spaces. Throughout the rest of the paper we denote the sampling density and base prior by $f(x;\mu)$ and $\pi_0(\mu;\theta)$, respectively, and $x=(x_1,\ldots,x_n)$ represents the vector of observations.

Lemma 1. *Under the prior perturbation* (2.4)*, the perturbed posterior model is*

$$\pi_p(\mu,\lambda|x;\theta)=\frac{\pi_p^0(\mu|x,\theta)}{\xi(\lambda,\theta)}\left\{1+\sum\nolimits_{j=1}^k\lambda_j q_j(\mu,\theta)\right\}, \tag{3.1}$$

with $\lambda\in\Lambda_\theta$, $\xi(\lambda,\theta):=1+\sum_{j=1}^k\lambda_j E_p^0[q_j(\mu,\theta)]>0$*, where* $\pi_p^0(\mu|x,\theta)$ *and* $E_p^0(\cdot|x)$ *are the posterior density and posterior mean of the base model.*

The following lemma characterizes the l^{th} moment of the perturbed posterior model. Note that, throughout the rest of the paper, for simplicity of exposition, we suppress the explicit dependence of ξ, q_j, π_p^0 and π_p on θ.

Lemma 2. *The moments of the perturbed posterior distribution are given by*

$$E_p(\mu^l|x,\lambda)=\frac{1}{\xi(\lambda)}\left\{E_p^0(\mu^l)+\sum\nolimits_{j=1}^k\lambda_j A_j^l(x)\right\}, \tag{3.2}$$

where $\lambda\in\Lambda_\theta$ *and* $A_j^l(x)=E_p^0(\mu^l q_j(\mu)|x)$.

To quantify the magnitude of the perturbation we exploit the *size function* as defined in Gustafson (1996), i.e., the L^p norm of the ratio u^*/π_0, for $p<\infty$, with respect to the measure induced by π_0. Accordingly, the size function for $u(\mu)=\sum_{j=1}^k\lambda_j\pi_0^{(j)}(\mu;\theta)$ is

$$\text{size}(u)=\left[E_{\pi_0}\left(\left|\sum\nolimits_{j=1}^k\lambda_j q_j(\mu)\right|\right)^p\right]^{1/p},$$

which, (i) is a finite norm and (ii) is invariant with respect to change of the dominating measure and also with respect to any one-to-one transformation of the sample space. Clearly, size(u) is finite if the first $k+p$ moments of $\pi_0(\mu,\theta)$ exist. In addition, property (ii) holds by use of the change of variable formula and the fact that for any one-to-one transformation $m=\nu(\mu)$ we have $\bar{\pi}_0^{(j)}(m,\theta)/\bar{\pi}_0(m,\theta)=\pi_0^{(j)}(\mu,\theta)/\pi_0(\mu,\theta)$.

For a mapping $T:\mathcal{U}\to\mathcal{V}$, where $\mathcal{U}$ and $\mathcal{V}$ are, respectively, the perturbations space normed with size($\cdot$), and the space of posterior expectations normed with their absolute value, the Fréchet derivative at $u_0\in\mathcal{U}$ is defined by the linear functional $\dot{T}(u_0):\mathcal{U}\to\mathcal{V}$ satisfying

$$\|T(u_0+u)-T(u_0)-\dot{T}(u_0)u\|_{\mathcal{V}}=o(\|u\|_{\mathcal{U}}),$$

in which $\dot{T}(u_0)u$ is the rate of change of T at u_0 in direction u. Let $Cov_p^0(\cdot,\cdot)$ be the posterior covariance with respect to the base model. Theorem 1 expresses $\dot{T}(u_0)u$ as a linear function of λ, at $u_0=0$ which corresponds to the base prior model.

Theorem 1. *For the perturbation in Definition* 3, $\dot{T}(0)u$ *is a linear function of the perturbation parameter* λ *obtained by the following equation*

$$\varphi(\lambda)=\sum\nolimits_{j=1}^{k}\lambda_j Cov_p^0\left(\mu,q_j(\mu)\right),\lambda\in\Lambda_\theta. \tag{3.3}$$

Section 3.3 illustrates how this function is maximized with respect to the perturbation parameter.

3.2. Global sensitivity

Here we use two commonly applied measures of sensitivity – the posterior mean difference and Kullback-Leibler divergence function – for assessing the global influence of a prior perturbation on the posterior mean and on prediction, respectively. The following theorem characterizes the difference between the posterior mean of the base and perturbed models as a function of λ.

Theorem 2. *Let* $\Psi(\lambda)=E_p(\mu|x,\lambda)-E_p^0(\mu|x)$ *represent the difference between the posterior expectations, then*

$$\Psi(\lambda)=\frac{1}{\xi(\lambda)}\varphi(\lambda),\quad \lambda\in\Lambda_\theta. \tag{3.4}$$

The function in (3.4) behaves in a intuitively natural way, for as $\lambda\to0$ we have $\xi(\lambda)\to1$, and consequently $\Psi(\lambda)$ behaves locally in a similar way to $\varphi(\lambda)$.

To assess the influence of the prior perturbation on prediction, we also quantify the change as measured by the divergence in the posterior predictive distribution. As an illustrative example, suppose the sampling distribution and the base prior model are respectively $\mathcal{N}(\mu, \sigma^2)$ and $\mathcal{N}(\theta, \sigma_0^2)$. The posterior predictive distribution for the base model is $\mathcal{N}(\mu_\pi, \sigma_\pi^2 + \sigma^2)$, where

$$\mu_\pi = \frac{\theta\sigma^2 + n\sigma_0^2\bar{x}}{n\sigma_0^2 + \sigma^2}, \quad \sigma_\pi^2 = \frac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2}.$$

Lemma 3. *The posterior predictive distribution for the perturbed model is*

$$g_p(y) = \frac{1}{\xi(\lambda)} \left\{ g_p^0(y) + \Gamma \sum\nolimits_{j=1}^{k} \lambda_j E^\star[q_j(\mu)] \right\} \tag{3.5}$$

in which, $g_p^0(y)$ is the posterior predictive density for the base model, Γ is a function of $(y, x, n, \theta_0, \sigma_0^2, \sigma^2)$ and $E^\star(\cdot)$ is the expectation with respect to a normal distribution.

For probability measures P_0 and P_1, with the same support, S, and densities $g_p^0(\cdot)$ and $g_p(\cdot)$, respectively, the Kullback-Leibler divergence functional is defined by,

$$D_{KL}(P_0, P_1) = \int_S \log \left[\frac{g_p^0(y)}{g_p(y)} \right] g_p^0(y) dy. \tag{3.6}$$

Theorem 3. *Kullback-Leibler divergence between $g_p^0(\cdot)$ and $g_p(\cdot)$, as a function of $\lambda \in \Lambda_\theta$ is*

$$\begin{aligned} D_{KL}(\lambda) = &\int_S \log \left[g_p^0(y) \right] g_p^0(y) dy + \log[\xi(\lambda)] \\ &- \int_S \log \left(g_p^0(y) + \Gamma \sum\nolimits_{j=1}^{k} \lambda_j E^\star[q_j(\mu)] \right) g_p^0(y) dy. \end{aligned} \tag{3.7}$$

In Section 3.3 we discuss the maximization process of the above discrepancy functions with respect to the perturbation parameter.

3.3. Optimising and estimating the perturbation parameter

Throughout the rest of this paper we let $k = 4$, as it gives a perturbation space which is flexible enough for our analysis and, as has been illustrated in Marriott (2002) , simply increasing the order of local mixture models does not significantly increase flexibility. Nevertheless, all the results and algorithms can be generalized to higher dimensions. This section outlines the theoretical framework for obtaining λ in the maximization method.

To obtain the values of λ which finds the most sensitive local and global perturbations, as described in Section 1, we apply an optimization approach to the functions (3.3), (3.4) and (3.7). We have that $\varphi(\lambda)$ is a linear function of λ on the convex space Λ_θ which represents the directional derivative of the mapping T at $\lambda = 0$. Thus, for obtaining the maximum direction of sensitivity, called the worst local sensitivity direction in Gustafson (1996), we need to maximize $\varphi(\lambda)$ over all the possible directions at $\lambda = 0$, but restricted by the boundary of Λ_θ. However, $\Psi(\lambda)$ and $D_{KL}(\lambda)$ are smooth non-linear objective functions on the convex space Λ_θ, for which we propose a suitable gradient based constraint optimization method that exploits the geometry of the parameter space. By Definition 1, for a fixed known θ, the space Λ_θ is a non-empty convex subspace in $\mathbb{R}^k$ with its boundary obtained by the following infinite set of hyperplanes

$$\mathcal{H} = \left\{ \lambda \middle| 1 + \sum\nolimits_{j=1}^{k} \lambda_j q_j(\mu) = 0; \mu \in \mathbb{R} \right\}.$$

Lemma 4 describes the boundary of Λ_θ as a smooth immersed manifold which can have self intersections, see Maroufy and Marriott (2016b) for proof.

Lemma 4. *The boundary of Λ_θ is a the union of smooth manifolds in $\mathbb{R}^4$.*

In addition, the interior of Λ_θ, which guarantees positivity of $\pi(\mu, \lambda; \theta)$ for all $\mu \in \mathbb{R}$, can be characterized by the necessary and sufficient positivity conditions on general polynomials of degree four. Comprehensive necessary and sufficient conditions are given in Barnard and Child (1936) and Bandy (1966).

Lemma 5. *The function $\varphi(\lambda)$ attains its maximum value at the gradient direction $\nabla\varphi$ if it is feasible; otherwise, the maximum direction is the direction of the orthogonal projection of $\nabla\varphi$ onto the boundary plane corresponding to $\lambda_4 = 0$.*

$D_{KL}(\lambda)$ and $\Psi(\lambda)$ are smooth functions which can achieve their maximum either in the interior or on the boundary of Λ_θ. Therefore, optimization shall be implemented in two steps: searching the interior using a regular Newton-Raphson algorithm, and then searching the boundary using a generalized form of Newton-Raphson algorithm for smooth manifolds (Shub (1986)).

3.4. An illustrative example

As an illustrative example, suppose the sampling distribution and the base prior model are respectively $\mathcal{N}(\mu, \sigma^2)$, and $\mathcal{N}(\theta, \sigma_0^2)$ and both σ and σ_0 are considered known. We use this example to illustrate both directional and global sensitivity. We also show how we can exploit the interpretability of our perturbation space to adapt the perturbation to be consistent with certain types of prior knowledge in the optimization approach.

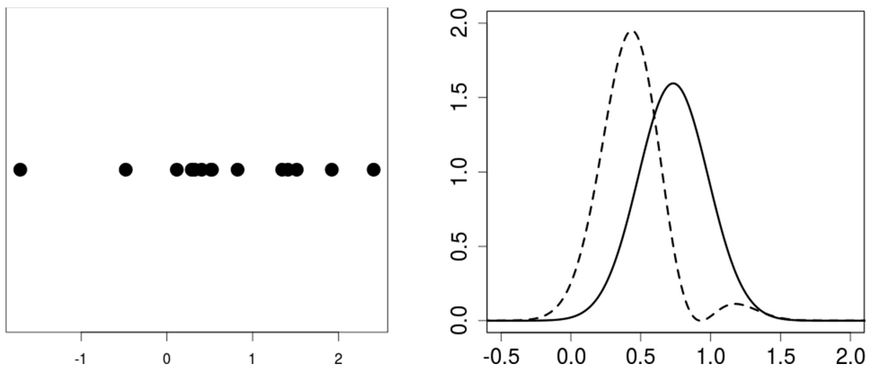

Figure 1. This shows, respectively, plots for the sample, and the posterior densities of the based (solid) and perturbed (dashed) model corresponding to $\hat{\lambda}_\Psi$.

We can calculated the relative difference between the Bayes estimates by

$$d = \frac{|E_p^0(\mu) - E_p^{\hat{\lambda}}(\mu)|}{std_p^0(\mu)}$$

in which $E_p^0(\mu)$ and $E_p^{\hat{\lambda}}(\mu)$ are the Bayes estimates with respect to the base and perturbed models, respectively, and $std_p^0(\mu)$ is the posterior standard deviation under the base model.

Example 2. A sample of size $n = 15$ is taken from a normal $\mathcal{N}(1, 1)$ distribution, and the base prior for the mean is $\mathcal{N}(2, 1)$. The estimate $\hat{\lambda}_\Psi = (-0.323,\ 1.44,\ -0.218,\ 0.441)$ is obtained from minimizing $\Psi(\lambda)$, defined in Theorem 2, and the corresponding relative discrepancy in Bayes estimate is $d = 1.19$; that is, the resulted change in posterior expectation is 119% of the posterior standard deviation of the base model. The corresponding density plots of both models are given in Figure 1.

For a directional analysis, we obtained the unit vector $\hat{\lambda}_\varphi$ which maximizes the directional derivative $\varphi(\lambda)$. Figure 2 shows the posterior density displacement corresponding to the perturbation parameter $\lambda_\alpha = \alpha\hat{\lambda}_\varphi$ for different values of $\alpha > 0$, as well as the boundary point λ_b in direction of $\hat{\lambda}_\varphi$. The corresponding relative differences in posterior expectation are $d = 0.09, 0.15, 0.3, 0.55$. Hence, additionally to obtaining the worst direction, these values suggest how far one can perturb the base prior along the worst direction so that relative discrepancy in posterior mean estimation is less than, say 50%.

We can extend this analysis by looking at cases where we might want to leave unchanged some aspect of the prior specification. For example we might only

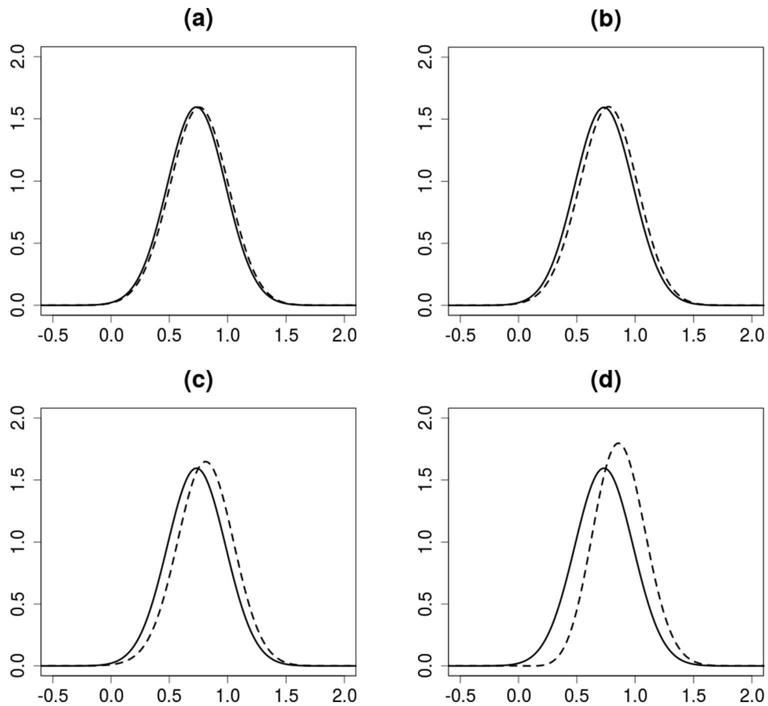

Figure 2. Posterior density displacement corresponding to $\lambda = \alpha\hat{\lambda}_\varphi$ for $\alpha = (a)0.1, (b)0.15, (c)0.25$ and (d) the boundary point at the maximum direction.

want to look at perturbations which leave the prior mean unchanged, or do not add prior skewness. These intuitive restrictions can be accommodated into our model simply by restricting the moment structure of the perturbed prior model. The central moments of the perturbed prior model, in Definition 3, are linearly related to the perturbation parameter λ. Specifically, for the normal model the mean, second and third central moments are

$$\bar{\mu}_\pi = \theta + \lambda_1, \qquad \bar{\mu}_\pi^{(2)} = \sigma^2 + 2\lambda_2 - \lambda_1^2,$$
$$\bar{\mu}_\pi^{(3)} = 6\lambda_3 + 2\lambda_1^3 - 6\lambda_1\lambda_2. \tag{3.8}$$

Clearly, λ_1 modifies the mean value, (λ_1, λ_2) adjust variance, and $(\lambda_1, \lambda_2, \lambda_3)$ adds skewness to the normal base model. Assuming $\lambda_1 = 0$, guarantees the perturbed model has its mean unchanged, and restricting $\lambda_1 = \lambda_3 = 0$ returns a symmetric perturbed model with same mean as the base prior model.

Example 3. For concreteness we look at perturbations where the prior mean is left invariant. By implementing the restriction $\lambda_1 = 0$, we are restricting the perturbation space Λ_θ to a lower dimensional space defined by the intersection of Λ_θ

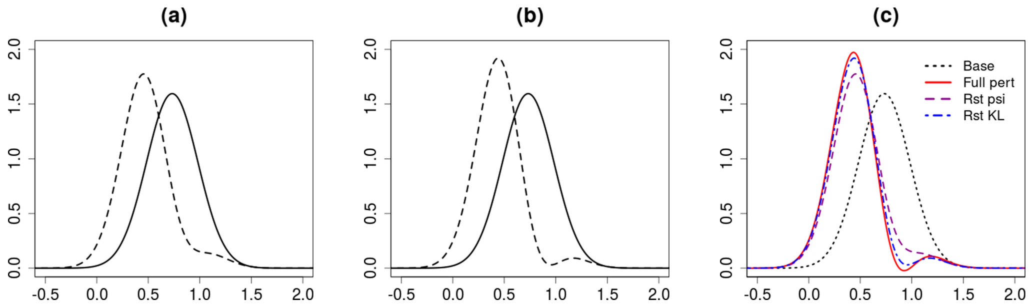

Figure 3. Panels (a),(b) correspond to $\hat{\lambda}_\Psi$ and $\hat{\lambda}_D$, under $\lambda_1 = 0$, respectively, including the base (solid) and perturbed posterior (dashed). Panel (c) presents posterior densities of based model (Base), and perturbed models for $\hat{\lambda}_\Psi$ (Rst psi) and $\hat{\lambda}_D$ (Rst KL) under $\lambda_1 = 0$, and the full perturbed posterior model (Full pert) from Example 2.

with the hyperplane defined by $\lambda_1 = 0$. The resultant space is a three dimensional convex space. Then we can find the most effective local and global perturbations in the new space using similar theoretical methods used for the earlier case. The estimate $\hat{\lambda}_D = (1.821, -0.011, 0.482)$ and $\hat{\lambda}_\Psi = (1.836, 0.016, 0.481)$ are obtained from maximizing $D_{KL}(\lambda)$ and minimizing $\Psi(\lambda)$, respectively. The corresponding relative discrepancies in the Bayes estimate are respectively $d = 1.19, 1.2$; that is, the resultant changes in posterior expectation are respectively 119% and 120% of the posterior standard deviation of the base model. Also, the corresponding posterior distributions are plotted in Figure 3.

4. Sensitivity in Multi-dimensional Models

In this section we apply our perturbation and sensitivity analysis to multi-dimensional models; specifically looking at both shrinkage and sparsity priors. In these cases we show how studying the space of mixtures of these priors gives insight into the sensitivity to prior choice.

In the shrinkage case of ridge regression, Sec. 4.1.1, we use the local mixture methods of Definition 3 where we mix over the location parameter. These result in the same expansions as used in the previous examples.

In the case of sparsity priors, Sec. 4.1.2, we follow the approach of Griffin and Brown (2010, 2017), and focus on scale mixtures of normal priors, as well as looking at other possible mixture structures. These can have exact closed form structure (for example the Laplace prior), have a local mixture structure, or be finite mixtures of local mixtures as describe in Definition 2.

In general we think of both shrinkage and sparsity priors as being families of priors indexed by the level of regularity required by the analyst. This common structure adds a little complexity to a sensitivity analysis as we do not want to conflate perturbations which just change the amount of regularity with other forms of perturbation.

4.1. Methodology: logistic regression

To show the methodology explicitly we focus on logistic regression, but emphasis that this is not a constraint and the methods can be applied to any generalised linear model.

Consider a logistic regression model $\text{logit}(p) = X\beta$ where X is a $n \times p$ design matrix and $\beta = (\beta_1, \ldots, \beta_p)^T$ is a vector of unknown covariate coefficients. The maximum likelihood method is a common approach for estimating β via maximizing

$$\ell(\beta) = \sum_{i=1}^{n} y_i(x_i\beta) - \log\left\{1 + \exp\left(x_i\beta\right)\right\}, \tag{4.1}$$

where x_i is the i^{th} row of X. Correspondingly, ridge regression parameters are obtained by maximizing

$$\ell_r(\beta) = \ell(\beta) - \gamma \sum_{j=1}^{p} \beta_j^2 \tag{4.2}$$

where γ is the regularization parameter and is often determined through cross validation techniques, Le Cessie and Van Houwelingen (1992). This problem can also be seen as a Bayesian inference problem with weakly informative priors

$$\pi(\beta_j) \propto \mathrm{e}^{-\beta_j^2/2};$$

i.e an independent set of standard normal prior models on $\beta_1, \ldots, \beta_p$.

4.1.1. Ridge prior

Let the base prior for β_j be

$$\pi(\beta_j) \propto \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\beta_j^2/(2\sigma^2)}$$

where σ^2 presents the minimum variation in the prior. Here σ is a measure of the degree of regularity in this problem. Then the perturbed prior is

$$\pi(\beta_j, \lambda^j) \propto \mathrm{e}^{-\beta_j^2/(2\sigma^2)} \left(1 + \sum_{k=1}^{4} \lambda_k^j q_k(\beta_j)\right),$$

where

$$q_1(\beta_j) = \frac{\beta_j}{\sigma^2},\ q_2(\beta_j) = \frac{\beta_j^2}{\sigma^4} - \frac{1}{\sigma^2},\ q_3(\beta_j) = \frac{\beta_j^3}{\sigma^6} - \frac{3\beta_j}{\sigma^4},\ q_4(\beta_j) = \frac{\beta_j^4}{\sigma^8} - \frac{6\beta_j^2}{\sigma^6} + \frac{3}{\sigma^4}.$$

The perturbed prior is not symmetric and has more variation than the minimum variance σ^2, due to extra flexibility added by λ^j parameters. Therefore, we can extend (perturb) the model (4.2) to

$$\ell_r(\beta, \underline{\lambda}) = \ell(\beta) - \gamma \sum_{j=1}^{p} \frac{\beta_j^2}{2} + \gamma \sum_{j=1}^{p} \log\left(1 + \sum_{k=1}^{4} \lambda_k^j q_k(\beta_j)\right), \tag{4.3}$$

where $\underline{\lambda} = [\lambda_k^j]$ is a $4 \times p$ perturbation parameter matrix.

Instead of optimisation over the perturbation space we average over it via MCMC methods. One reason for doing this is to allow the data to be more involved in the selection of the regularity prior. In this case the variance term σ^2, in the expansion above, gives a lower bound, but the local mixing can increase this if the data indicates this is desirable. We would therefore recommend setting σ at the lower end of what might be of interest and let the sensitivity analysis point out if inflation is needed.

4.1.2. Sparsity priors

There are a number of ways of modelling sparsity with prior distributions which are mixtures of different kinds. For example a two component finite mixture of a discrete point mass at $\beta = 0$ and a continuous component, typically normal. This the 'spike-and-slab' model of Mitchell and Beauchamp (1988). A completely continuous alternative is double exponential or Laplace prior of Park and Casella (2008). Griffin and Brown (2010) generalise this approach by looking at scaled mixture of mean zero normals where

$$\beta_i | \Psi_i \sim N(\beta_i | 0, \Psi_i), \Psi_i \sim G. \tag{4.4}$$

If the distribution G is selected to be gamma then this results in a two dimensional family which includes the Laplace prior as a special case.

Approximating (4.4) as a local mixture model is more subtle than in the case considered above since it requires an expansion around the delta function at $\beta = 0$. In fact it requires the methods of Definition 2. The space of mixtures is approximated by a finite mixture, one of whose components is the delta function at $\beta = 0$ and the remaining components of the form

$$f(\beta; \sigma_i, \lambda_i) := \mathrm{e}^{-\beta^2/\sigma_i^2}\left(1 + \lambda_{i1}\frac{\beta - \sigma_i^2}{\sigma_i^3} + \lambda_{2i}\frac{(2\sigma_i^4 - 5\sigma_i^2\beta^2 + \beta^4)}{\sigma_i^6}\right)$$

for a set of user selected σ_i. The parameter space of the λ_i is determined by the positivity of the expression as above. We will discuss this approach in more detail in the discussion section. For concreteness, in the data examples below, we use a specific path through this mixture space by using Griffin and Brown (2010) normal-gamma family.

4.2. Data examples

Example 4 (Benign Breast Disease study). The study is focused on examining the risk factors associated with benign breast disease, and comprises 200 observations, including 50 women who were diagnosed as having benign breast disease and 150 age matched controls (Pastides et al. (1985)). The covariates considered here are: age of the subject at the interview (AGMT), highest grade in School (HIGD), weight of the Subject (WT), age at last menstrual period (AGLP), with coefficients $\beta_1, \beta_2, \beta_3, \beta_4$, respectively. To assess the sensitivity of the covariate coefficients to weakly informative prior models, we compare their estimates under the base and the perturbed models. For the perturbed model we are computing the estimates by marginalising over the perturbation space.

Table 1 presents the estimated parameters and their standard deviation for the covariate coefficients as well as the absolute relative difference of the estimates between the two models, defined as

$$\text{r-bias} = \frac{|\hat{\beta}^{base} - \hat{\beta}^{ptb}|}{sd(\hat{\beta}^{base})}.$$

The results illustrate significant difference between the parameter estimates between the two models, the largest of which is that reported for β_3 at 570% of the standard deviation. Hence, simply selecting weekly informative priors here can cause significant bias in the coefficient estimates.

Example 5 (Adolescent Placement Study). In this example we examine sensitivity of the regression parameters in a study on aftercare placement for 508 psychiatrically hospitalized adolescents based on the data presented in Fontanella, Early and Phillips (2008), also see Hosmer, Lemeshow and Sturdivant (2013, Sec. 1.6.4). The outcome variable has four categories: 0 = Outpatient, 1 = Day Treatment, 2 = Intermediate Residential and 3 = Residential. For the purpose of our modelling we combine the outcome variable into two categories 0 = Outpatient or Day Treatment and 1 = Intermediate Residential or Residential. The predictors considered are Age, Race, Gender, Length of Hospitalization and State of Custody.

Table 1. The first two rows are the estimates and standard deviations of the coefficients based on the base model. The next two represent the same numbers marginalised over the perturbed model. The term r-base is the relative absolute difference between the estimates of the two models. The final two represent the results of using the normal-gamma shrinkage prior.

parameter	β_1	β_2	β_3	β_4
Estimate base model	0.004	0.098	−0.023	0.0143
Standard deviation base model	0.010	0.047	0.001	0.007
Estimate perturbed model	−0.007	0.021	−0.015	0.028
Standard deviation perturbed model	0.005	0.029	0.002	0.008
r-bias	1.225	1.618	5.723	2.082
Shrinkage mean	−0.006	0.064	−0.030	0.055
Standard deviation	0.020	0.056	0.007	0.033

Table 2. The first two rows are the estimates and standard deviations of the coefficients based on the base model. The next two represent the same numbers marginalised over the perturbed model. The term r-base is the relative absolute difference between the estimates of the two models. The final two represent the results of using the normal-gamma shrinkage prior.

parameter	β_1	β_2	β_3	β_4	β_5
Estimate base model	−0.176	0.318	0.250	0.074	3.165
Standard deviation base model	0.014	0.221	0.225	0.011	0.251
Estimate perturbed model	−0.184	0.304	0.272	0.082	3.248
Standard deviation perturbed model	0.022	0.242	0.254	0.0135	0.272
r-bias	0.585	0.063	0.096	0.691	0.331
Shrinkage mean	0.073	−0.534	0.411	0.081	3.377
Standard deviation	0.084	0.285	0.281	0.013	0.312

Unlike the previous study, Table 2 does not illustrate any significant difference for the covariate coefficients. Here the likelihood dominates the prior information as a result of a relatively large sample size so that the results are reasonably stable with respect to the defined prior perturbations.

5. Discussion

This paper has looked at the sensitivity of Bayesian inference to perturbations of the prior. It takes a geometric approach throughout. The key idea is to think of mixing over priors as defining a plausible perturbation space. Within this space – or the approximation of it generated by local mixing – we used both gradient and averaging approaches to explore this space.

The spaces defined by local mixture models have some nice properties such as convexity, low-dimensionality and identifiability, Marriott (2002); Anaya-Izquierdo and Marriott (2007b) but do come with boundaries which can be complex, Maroufy and Marriott (2015). In this paper we have explored some ways of exploiting the advantages and dealing with the boundaries.

The space of scale mixture of normal (Section 4.1.2) is a very interesting one from the point of view of local mixtures since it contains a singular limiting point which is the delta function at $\beta = 0$. Thus it naturally includes the 'slab-and-spike' priors of Mitchell and Beauchamp (1988). The extreme point of this space are the unmixed models, $N(0, \sigma^2)$, including the singular case. It would be a great utility to find good, finite dimensional, approximations to this space. Using the methods of finite mixtures of LMM define in Definition 2 may be one way of doing this. The key open problem is to find optimal ways of placing the grid-points σ_i balancing the dimensionality of the approximating space with the approximation error.

Acknowledgment

This work is partly supported by NSERC discovery grant 'Computational Information Geometry and Model Uncertainty'.

Appendix: Proofs

Lemma A1.

$$\pi_p(\mu|x, \lambda) = \frac{\pi(\mu, \lambda) f(x;\mu)}{g(x, \lambda)}, \tag{A.1}$$

where

$$\begin{aligned} g(x, \lambda) &= \int \pi(\mu, \lambda; \theta) f(x; \mu) d\mu \\ &= \int f(x;\mu)\pi_0(\mu;\theta) d\mu \\ &\quad + \sum_{j=1}^{k} \lambda_j \int q_j(\mu, \theta) f(x;\mu)\pi_0(\mu;\theta) d\mu \\ &= g(x) \left\{ 1 + \sum_{j=1}^{k} \lambda_j E_p^0[q_j(\mu, \theta)] \right\}. \end{aligned} \tag{A.2}$$

Since $f(x;\mu)\pi_0(\mu;\theta) = g(x)\pi_p^0(\mu|x,\theta)$ *and* $g(x) = \int f(x;\mu)\pi_0 \ (\mu;\theta) d\mu$ *where,* $g(x)$ *is the marginal density of sample in the base model. Hence,*

$$\pi_p(\mu,\lambda|x;\theta) = \frac{f(x;\mu)\pi_0(\mu;\theta)\left\{1+\sum_{j=1}^k \lambda_j q_j(\mu,\theta)\right\}}{g(x)\left\{1+\sum_{j=1}^k \lambda_j E_p^0[q_j(\mu,\theta)]\right\}}$$
$$= \frac{\pi_p^0(\mu|x,\theta)}{\xi(\lambda,\theta)}\left\{1+\sum\nolimits_{j=1}^k \lambda_j q_j(\mu,\theta)\right\}, \lambda \in \Lambda_\theta$$

with $\xi(\lambda,\theta) = 1+\sum_{j=1}^k \lambda_j E_p^0[q_j(\mu,\theta)]$.
Also $\xi(\lambda,\theta) > 0$, *since* $1+\sum_{j=1}^k \lambda_j q_j(\mu,\theta) > 0$, *for all* $\mu \in \mathbb{R}$ *and* $\lambda \in \Lambda_\theta$, *and* $\xi(\lambda,\theta) = E_p^0[1+\sum_{j=1}^k \lambda_j q_j(\mu,\theta)]$.

Lemma A2. *Result follows by direct calculation and using the fact that,*

$$A_j^l(x) := \int \mu^l q_j(\mu)\pi_{post}^0(\mu|x)d\mu = E_p^0[\mu^l q_j(\mu)]. \tag{A.3}$$

Theorem A1. *Substitute* $u^*(\cdot)$ *by* $u(\cdot)$ *in* Gustafson (1996, Result 8).

Theorem A2. *By direct calculation and use of align* (A.3)

Lemma A3.

$$g_p(y) = \int f(y;\mu)\pi_p(\mu,\lambda|x)d\mu \tag{A.4}$$

is the convolution of $\mathcal{N}(\mu,\sigma^2)$ *and* $\mathcal{N}(\mu_\pi,\sigma_\pi^2)$. *Since,*

$$\frac{(y-\mu)^2}{\sigma^2}+\frac{(\mu-\mu_\pi)^2}{\sigma_\pi^2} = \frac{\left(\mu-(\sigma_\pi^2 y+\sigma^2\mu_\pi)/(\sigma^2+\sigma_\pi^2)\right)^2}{(\sigma^2\sigma_\pi^2)/(\sigma^2+\sigma_\pi^2)}+\frac{(y-\mu_\pi)^2}{\sigma^2+\sigma_\pi^2}$$

hence, the posterior predictive distribution for base model is $\mathcal{N}(\mu_\pi,\sigma_\pi^2+\sigma^2)$ *and* (3.5) *is obtained by direct calculation, where,*

$$\Gamma = \frac{1}{\sqrt{2\pi(\sigma_\pi^2+\sigma^2)}}\exp\left\{-\frac{(y-\mu_\pi)^2}{2(\sigma_\pi^2+\sigma^2)}\right\}$$

and $E^\star(\cdot)$ *is expectation with respect to* μ *according to the following normal distribution*

$$\mathcal{N}\left(\frac{\sigma_\pi^2 y+\sigma^2\mu_\pi}{\sigma_\pi^2+\sigma^2}, \frac{\sigma_\pi^2\sigma^2}{\sigma_\pi^2+\sigma^2}\right).$$

Theorem A3. *Use of Lemma 3 and direct calculation finishes the proof.*

Lemma A4. *Implied by direct application of the implicit function theorem,* Rudin (1976, p.225), *see* Maroufy and Marriott (2016b).

Lemma A5. $\nabla\varphi = (a_1,a_2,a_3,a_4)$, *is a vector originated at* $\lambda = 0$, *where* $a_j = Cov_p^0(\mu, q_j(\mu))$. *If it is feasible then clearly gives the maximum direction. However, if it is not feasible then* $a_4 \le 0$ *since the condition* $a_4 > 0$ *is necessary* *for feasibility. Thus, the direction of the orthogonal projection of $\nabla\varphi$ onto the boundary plane corresponding to $\lambda_4 = 0$ is the closest we get to a maximum and feasible direction.*

References

Amari, S. I. (1990). *Differential-Geometrical Methods in Statisitics: Lecture Notes in Statisitics*, 2nd Edition. Springer-Verlag: New York.

Anaya-Izquierdo, K. and Marriott, P. (2007a). Local mixture models of exponential families. *Bernoulli* **13**, 623–640.

Anaya-Izquierdo, K. and Marriott, P. (2007b). Local mixture of the exponential distribution. *Annals of the Institute of Statistical Mathematics.* **59**, 111–134.

Bandy, M. L. (1966). A theorem on positive quartic forms. *The American Mathematical Monthly* **73**, 864–866.

Barnard, S. and Child, J. M. (1936). *Higher Algebra*. Macimillian and Co.

Berger, J. O., Rios Insua, D. and Ruggeri, F. (2000). Bayesian robustness. *Robust Bayesian Analysis*, 1–32.

Blyth, S. (1994). Local divergence and association. *Biometrika* **81**, 579–584.

Cook, D. R. (1986). Assessment of local influence. *Journal of the Royal Statistical Society. Series B. (Statistical Methodology)* **48**, 133–169.

Critchley, F. and Marriott, P. (2004). Data-informed influence analysis. *Biometrika* **91**, 125–140.

Fontanella, C. A., Early, T. J. and Phillips, G. (2008). Need or availability? modeling aftercare decisions for psychiatrically hospitalized adolescents. *Children and Youth Services Review* **30**, 758–773.

Griffin, J. and Brown, P. (2017). Hierarchical shrinkage priors for regression models. *Bayesian Analysis* **12**, 135–159.

Griffin, J. E. and Brown, P. J. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis* **5**, 171–188.

Gustafson, P. (1996). Local sensitivity of posterior expectations. *The Annals of Statistics* **24**, 174–195.

Hosmer, D. W., Lemeshow, S. and Sturdivant, R. X. (2013). *Applied logistic regression* **398**, John Wiley & Sons.

Insua, D. R. and Ruggeri, F. (2000). *Robust Bayesian Analysis*. Springer: New York.

Kurtek, S. and Bharath, K. (2015). Bayesian sensitivity analysis with the fisher–rao metric. *Biometrika* **102**, 601–616.

Lavine, M. (1991). The prior and the likelihood. *Journal of the American Statistical Assiciation* **86**, 396–399.

Le Cessie, S. and Van Houwelingen, J. C. (1992). Ridge estimators in logistic regression. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **41**, 191–201.

Linde, V. D. A. (2007). Local influence on posterior distributions under multiplicative models of perturbation. *Bayesian Analysis* **2**, 319–332.

Lindsay, B. G. (1995). *Mixture Models: Theory, Geometry and Applications*. Inst of Mathematical Statistics.

Maroufy, V. and Marriott, P. (2015). Computing boundaries in local mixture models. *Geometric*

Science of Information, Lecture Notes in Computer Science, 577–585.

Maroufy, V. and Marriott, P. (2016a). Mixture models: Building a parameter space. *Statistics and Computing* **27**, 591–597.

Maroufy, V. and Marriott, P. (2016b). On the effect of generalising frailty assumptions in survival analysis. *Journal of the Amercan Statistical Association*, in submision.

Marriott, P. (2002). On the local geometry of mixture models. *Biometrika* **89**, 77–93.

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

Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association* **83**, 1023–1032.

Park, T. and Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association* **103**, 681–686.

Pastides, H., Kelsey, J. L., Holford, T. R. and Livolsi, V. A. (1985). An epidemiologic study of fibrocystic breast disease with reference to ductal epithelial atypia. *American journal of epidemiology* **121**, 440–447.

Rudin, W. (1976). *Principles of Mathematical Analysis*, 3rd Edition. McGraw-Hill: New York.

Ruggeri, F. and Sivaganesan, S. (2000). On a global sensitivity measure for bayesian inference. *Sankhya* **62**, 110–127.

Ruggeri, F. and Wasserman, L. (1993). Infinitesimal sensitivity of posterior distributions. *The Canadian Journal of Statistics* **21**, 195–203.

Shub, M. (1986). Some remarks on dynamical systems and numerical analysis. In Proc. *VII ELAM* (Edited by L. Lara-Carrero and J. Lewowicz), 69–92. Equinoccio, U. Simon Bolivar, Caracas.

Weiss, R. (1996). An approach to Bayesian sensitivity analysis. *Journal of the Royal Statistical Society. Series B (Statistical Methodological)* **58**, 739–770.

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

Zhu, H., Ibrahim, J. G. and Tang, N. (2011). Bayesian influence analysis: a geometric approach. *Biometrika* **98**, 307–323.

Zhu, H., Ibrahim, J. G. and Tang, N. (2014). Bayesian sensitivity analysis of statistical models with missing data. *Statistica Sinica* **24**, 871.

Department of Biotatistics and Data Science, School of Public Health, University of Texas Health Center at Houston, Houston, Tx, USA.

E-mail: vahed.maroufy@uth.tmc.edu

Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

E-mail: pmarriott@uwaterloo.ca

(Received June 2017; accepted May 2018)