Bayesian Analysis (2007) **2**, Number 2, pp. 319–332

Local Influence on Posterior Distributions under Multiplicative Modes of Perturbation

Angelika van der Linde*

Abstract. Any unperturbed and perturbed posterior density can formally be linked by a mixture. Many divergences between the unperturbed and perturbed posterior density - global measures of influence of the perturbation - are then essentially determined by the Fisher information with respect to the mixing parameter evaluated at the unperturbed density. It is investigated which aspect of change this Fisher information - commonly interpreted as local measure of influence - captures in assessing influence of the perturbation. Under multiplicative modes of perturbation it is nicely interpretable as unperturbed posterior variance of the (log-)perturbation function.

Keywords: Bayesian robustness, case deletion

1 Introduction

Bayesian sensitivity analysis is concerned with the impact of a modification of the prior or the likelihood on the posterior. The analysis of the effect of changes in the prior is often addressed as Bayesian robustness analysis. A collection of papers discussing the main issues and tools of Bayesian sensitivity studies was edited by Insua and Ruggeri (2000). One common approach is to define a null model and a mode of perturbation. Then either globally the difference between the posterior in the null model and a perturbed posterior is assessed or locally - with a parameter defining the degree of perturbation - the rate of change is determined by a derivative evaluated in the null model. A review of approaches to global and local robustness analysis is provided by Sivaganesan (2000) (ch.5 in the aforementioned book), more details and a critical discussion of local sensitivity analysis are given by Gustafson (2000).

A frequently used global measure of influence of a model component is the (Kullback-Leibler) divergence of perturbed and unperturbed posterior densities, a related local measure is its second derivative with respect to the perturbation parameter (e.g. McCulloch (1989); Lavine (1992); Geisser (2000)). This is also the general set-up referred to in this paper.

If the perturbation can be represented by a change of a parameter, the divergence can be approximated using a second order Taylor expansion of the perturbed posterior density. For many divergences it is then essentially governed by the Fisher information which therefore is used as local measure of influence under the specified parameterization. For example McCulloch (1989) analyzed the effect of weights on error variances in regression. Here the null model corresponds to a weight equal to one and the degree of

*University of Bremen, Germany, http://www.math.uni-bremen.de/~avdl/

perturbation is quantified by the weight itself which therefore has been considered as an intrinsic parameter of the perturbation. This paper focuses on modes of perturbation of the likelihood or the prior (or both) such that the posterior is modified multiplicatively. Such multiplicative modes of perturbation introduced by Weiss (1996) comprise case deletion, modelling outlying observations and change of the prior distribution. Multiplicative modes of perturbation often describe a qualitative change like an alternative prior or the omission of a case. A formal perturbation parameter describing the amount of perturbation can then be defined as a mixing parameter. Dey and Birmiwal (1994) studied a null model with a prior p_0, a mode of perturbation (alternative prior) p_1 and perturbed priors $p_\lambda = (1-\lambda)p_0 + \lambda p_1$ (an additive contamination class of priors) or $\tilde{p}_\lambda = c(\lambda)p_0^{1-\lambda}p_1^\lambda$ (a geometric contamination class of priors). In such cases again the Fisher information with respect to the mixture can be used as local measure of influence. In this paper the results of Dey and Birmiwal (1994) for priors are generalized to additive and geometric contamination classes of models (that is joint densities of observations and parameters) induced by multiplicative modes of perturbation of a null model, particularly the sampling model.

Gustafson (2000) points to two potential problems in analyses to assess the local influence of the sampling model: (i) Model fit may be more important than influence. This is an important point if for example alternative distributional assumptions for (conditionally) i.i.d. observations are discussed. But there are set-ups of interest like case deletion where the argument does not apply. (ii) The (interpretation of) parameters may not be invariant under a change of the sampling distribution. Again this is true but not always a point of concern. For instance the mean in regression is a parameter of interest for many error distributions, or a future observation (from the null sampling distribution) is a parameter of interest in a predictive approach under case deletion. Hence, although the technical results obtained in this paper are applicable to all formally multiplicative modes of perturbation, one has to be cautious about interpretations. In this paper special emphasis is given to the more involved details of perturbations of the sampling distribution corresponding to case weights and case deletion.

A main contribution of this work is to provide a simple unifying framework for a branch of Bayesian sensitivity analysis referring to multiplicative modes of perturbation which comprise changes in both the likelihood and the prior. In this way previous results (Dey and Birmiwal (1994); Peng and Dey (1995); Millar and Stewart (2005)) are generalized and integrated in one approach. It is demonstrated that the Fisher information is interpretable as (unperturbed) posterior variance of the (log-)perturbation function. Thus it assesses (a posteriori) the change in the 'prior input' rather than the 'posterior output' due to the mode of perturbation. Furthermore it is shown that the χ^2-divergence is proportional to the Fisher information when an additive mixture is used giving further support to recommendations to prefer the χ^2-divergence to the Kullback-Leibler divergence if a choice is to be made. Thus also a new link between global and local Bayesian sensitivity analyses is established under multiplicative modes of perturbation and another interpretation of the χ^2-divergence is provided.

The paper is organized as follows: In section 2 the formal set-up is defined and the main results are proven. Application of the local measure of influence to case weights and case deletion are discussed in detail. In section 3 the ideas are exemplified for likelihood and prior perturbations, and section 4 concludes with a brief discussion.

2 Measures of sensitivity for multiplicative modes of perturbation

2.1 General results

Assume that in a null model the sampling density for observations y_i, $i = 1...n$, attains values $p_0(y_i|\theta)$ and that the prior density for parameters θ is given by $p_0(\theta)$ such that for the posterior density $p_0(\theta|y_d) = p_0(y_d|\theta)p_0(\theta)/p_0(y_d)$ holds, where $y_d = (y_1, ..., y_n)$ denotes the data, $p_0(y_d|\theta) = \prod_{i=1}^{n} p_0(y_i|\theta)$ and $p_0(y_d) = \int p_0(y_d, \theta)d\theta$. A multiplicative mode of perturbation of the null model $p_0(y_d, \theta)$ is given by $p_1(y_d, \theta) = p_0(y_d, \theta)h^*(\theta)$ with h^* denoting the perturbation function. It yields the posterior

$$p_1(\theta|y_d) = \frac{p_0(\theta|y_d)h^*(\theta)}{E^0_{\theta|y_d}[h^*(\theta)]}, \tag{1}$$

where the superscript indicates with respect to which posterior density the expectation (and similarly in the sequel a variance) is taken. Equation (1) results from perturbations of the likelihood or the prior. For example, $h^*(\theta) = p_0(y_i|y_{d\setminus i})/p_0(y_i|\theta)$ with $p_0(y_i|y_{d\setminus i}) = E^0_{\theta|y_{d\setminus i}}[p(y_i|\theta)]$ and $y_{d\setminus i}$ denoting the data without y_i corresponds to the deletion of the i-th case. (This particular choice of h^* to represent case deletion is justified in section 2.2.1.) Similarly $h^*(\theta) = p_1(\theta)/p_0(\theta)$ corresponds to a change of the prior. An *additive mixture*

$$p_\lambda(y_d, \theta) = (1 - \lambda)p_0(y_d, \theta) + \lambda p_1(y_d, \theta) \tag{2}$$

represents a parametric weighting scheme for the perturbation of the model even if initially it is qualitative.

Let ζ be a parameter of interest, either the full parameter, $\zeta = \theta$, a partial parameter, $\zeta = \tau$ if $\theta = (\tau, \rho)$, or a future observation from the unperturbed model, $\zeta = \tilde{y}$. The influence of the perturbation can be assessed by the difference of the posterior densities $p_0(\zeta|y_d)$ and $p_1(\zeta|y_d)$ either graphically (Weiss and Cook (1992)) or formally using a summary measure of the unperturbed posterior distribution of $p_1(\zeta|y_d)/p_0(\zeta|y_d)$. Such a summary measure is provided by a divergence, and a frequently used family of divergences between densities p and q is given by $\{D_\phi(p, q) = \int \phi[q(u)/p(u)]p(u)du \mid \phi\, convex,\ \phi(1) = 0\}$. $\phi_{KL}(x) = -\ln(x)$ yields the (directed) Kullback-Leibler divergence, $\phi_{\chi^2}(x) = (x - 1)^2$ yields the χ^2-divergence. According to (1)

$$D_\phi(p_0(\theta|y_d), p_1(\theta|y_d)) = E^0_{\theta|y_d}[\phi(\frac{p_1(\theta|y_d)}{p_0(\theta|y_d)})] = E^0_{\theta|y_d}[\phi(\frac{h^*(\theta)}{E^0_{\theta|y_d}[h^*(\theta)]})]. \tag{3}$$

For λ denoting a mixing parameter $\lambda_0 = 0$ corresponds to p_0, $\lambda = 1$ to p_1. It is well known (e.g. Blyth (1994)) that

$$D_\phi(p_{\lambda_0}(\zeta|y_d), p_\lambda(\zeta|y_d)) \approx \frac{\phi''(1)}{2} I_{\zeta|y_d}(\lambda_0)(\lambda - \lambda_0)^2 \tag{4}$$

with the Fisher information

$$I_{\zeta|y_d}(\lambda_0) = var^0_{\zeta|y_d}[\frac{d}{d\lambda}\ln(p_\lambda(\zeta|y_d))|_{\lambda=\lambda_0}]$$

for $\zeta \in \{\theta, \tau, \tilde{y}\}$. Hence

$$D_\phi(p_0(\zeta|y_d), p_1(\zeta|y_d)) \approx \frac{\phi''(1)}{2} I_{\zeta|y_d}(0). \tag{6}$$

Note that for the χ^2-divergence $\phi''_{\chi^2}(1) = 2$ and for the directed Kullback-Leibler divergence $\phi''_{KL}(1) = 1$. For all functions ϕ the divergence $D_\phi(p_0(\zeta|y_d), p_1(\zeta|y_d))$ is essentially determined by the Fisher information $I_{\zeta|y_d}(0)$. It is therefore common practice to use $I_{\zeta|y_d}(0)$ as an omnibus local measure of influence. Under multiplicative modes of perturbation where $p_1(y_d, \theta)/p_0(y_d, \theta) = h^*(\theta)$ define

$$h^*(\zeta) := \quad \tfrac{p_1(y_d,\zeta)}{p_0(y_d,\zeta)} = \langle \begin{array}{ccc} h^*(\theta) & if & \zeta = \theta \\ E^0_{\rho|\tau,y_d}[h^*(\theta)] & if & \zeta = \tau,\ \theta = (\tau, \rho) \\ E^0_{\theta|\tilde{y},y_d}[h^*(\theta)] & if & \zeta = \tilde{y} \end{array} .$$

The following representation is then immediate.

Theorem 1

The Fisher information $I^a_{\zeta|y_d}(0)$ with respect to the mixing parameter in an additive mixture is equal to the unperturbed posterior variance of the perturbation function $h^*(\zeta)$,

$$I^a_{\zeta|y_d}(0) = var^0_{\zeta|y_d}[h^*(\zeta)]. \tag{8}$$

Proof:

$$\begin{aligned} I^a_{\zeta|y_d}(0) &= var^0_{\zeta|y_d}[\frac{d}{d\lambda}\ln(p_\lambda(\zeta|y_d))|_{\lambda=0}] \\ &= var^0_{\zeta|y_d}[\frac{d}{d\lambda}\ln(p_\lambda(y_d,\zeta))|_{\lambda=0}] \\ &= var^0_{\zeta|y_d}[-1 + \frac{p_1(y_d,\zeta)}{p_0(y_d,\zeta)}] \\ &= var^0_{\zeta|y_d}[h^*(\zeta)]. \end{aligned} \tag{9}$$

Thus the Fisher information measures (a posteriori) the amount of initially introduced perturbation. Theorem 1 generalizes theorem 3.1 of Dey and Birmiwal (1994) who obtained this result (with a different proof) for a perturbation of the prior distribution. The representation of $I^a_{\zeta|y_d}(0)$ can be extended to a representation of the χ^2- divergence.

Corollary 1

$$D_{\phi_{\chi^2}}(p_0(\zeta|y_d), p_1(\zeta|y_d)) = I^a_{\zeta|y_d}(0)(\frac{p_0(y_d)}{p_1(y_d)})^2. \tag{10}$$

Proof:

The ratio of the posterior density values can be re-written as

$$h(\zeta) := \frac{p_1(\zeta|y_d)}{p_0(\zeta|y_d)} = \frac{p_1(y_d,\zeta)}{p_0(y_d,\zeta)}\frac{p_0(y_d)}{p_1(y_d)} = h^*(\zeta)\frac{p_0(y_d)}{p_1(y_d)} \tag{11}$$

and hence

$$D_{\phi_{\chi^2}}(p_0(\zeta|y_d), p_1(\zeta|y_d)) = var^0_{\zeta|y_d}[h(\zeta)] = var^0_{\zeta|y_d}[h^*(\zeta)](\frac{p_0(y_d)}{p_1(y_d)})^2.$$

Thus the χ^2- divergence is proportional to the Fisher information. In comparison to the Taylor approximation (6) the corollary describes the approximation error. In particular for case deletion represented by $h^*(\theta) = p_0(y_i|y_{d\setminus i})/p_0(y_i|\theta)$ one has $p_1(y_d) = \int p_1(y_d|\theta)p_0(\theta)d\theta = \int p_0(y_{d\setminus i}|\theta)p_0(y_i|y_{d\setminus i})p_0(\theta)d\theta = p_0(y_i|y_{d\setminus i})p_0(y_{d\setminus i}) = p_0(y_d)$ and hence

$$D_{\phi_{\chi^2}}(p_0(\zeta|y_d), p_1(\zeta|y_d)) = I^a_{\zeta|y_d}(0) \tag{13}$$

exactly.

The proportionality factor $p_0(y_d)/p_1(y_d)$ in equation (11) is the inverse normalizing constant of the perturbed posterior, $p_1(y_d)/p_0(y_d) = E^0_{\zeta|y_d}[h^*(\zeta)]$, yielding as in (3)

$$D_{\phi_{\chi^2}}(p_0(\zeta|y_d), p_1(\zeta|y_d)) = var^0_{\zeta|y_d}[h^*(\zeta)]/(E^0_{\zeta|y_d}[h^*(\zeta)])^2. \tag{14}$$

Thus the χ^2- divergence is a squared coefficient of variation of the perturbation function $h^*(\zeta)$. These properties of the χ^2- divergence support preference of it to other divergences like the Kullback-Leibler divergence (see the discussion by Weiss (1996)).

For a *geometric mixture*

$$\tilde{p}_\lambda(y_d,\theta) = c(\lambda)p_0^{1-\lambda}(y_d,\theta)p_1^\lambda(y_d,\theta)$$

a result similar to theorem 1 is obtained generalizing theorem 3.2 of Dey and Birmiwal (1994) to perturbations of the likelihood only or of likelihood and prior simultaneously.

Theorem 2

The Fisher information $I^g_{\theta|y_d}(0)$ with respect to the mixing parameter in a geometric mixture is equal to the unperturbed posterior variance of the log-perturbation function $\ln(h^*(\theta))$,

$$I^g_{\theta|y_d}(0) = var^0_{\theta|y_d}[\ln(h^*(\theta))]. \tag{16}$$

Proof:

$$\begin{aligned} I^g_{\theta|y_d}(0) &= var^0_{\theta|y_d}[\frac{d}{d\lambda}\ln(\tilde{p}_\lambda(\theta|y_d))|_{\lambda=0}] \\ &= var^0_{\theta|y_d}[-\ln(p_0(y_d,\theta)) + \ln(p_1(y_d,\theta))] \\ &= var^0_{\theta|y_d}[\ln(h^*(\theta))]. \end{aligned} \tag{17}$$

The representation can be extended to $\zeta \in \{\tau, \tilde{y}\}$.

Corollary 2

a)

$$I^g_{\tau|y_d}(0) = var^0_{\tau|y_d}[E^0_{\rho|\tau,y_d}\{\ln(h^*(\theta))\}]. \tag{18}$$

b)

$$I^g_{\tilde{y}|y_d}(0) = var^0_{\tilde{y}|y_d}[E^0_{\theta|\tilde{y},y_d}\{\ln(h^*(\theta))\}]. \tag{19}$$

Proof:

a) By definition

$$I^g_{\tau|y_d}(0) = var^0_{\tau|y_d}[\frac{d}{d\lambda}\ln(\tilde{p}_\lambda(\tau|y_d))|_{\lambda=0}].$$

With $\theta = (\tau, \rho)$

$$\frac{d}{d\lambda}\ln(\tilde{p}_\lambda(\tau|y_d)) = \frac{1}{\tilde{p}_\lambda(\tau|y_d)}\frac{d}{d\lambda}\tilde{p}_\lambda(\tau|y_d) = \frac{1}{\tilde{p}_\lambda(\tau|y_d)}\frac{d}{d\lambda}\int \tilde{p}_\lambda(\tau,\rho|y_d)d\rho.$$

Assuming that differentiation and integration can be interchanged

$$\begin{aligned} \frac{d}{d\lambda}\ln(\tilde{p}_\lambda(\tau|y_d)) &= \frac{1}{\tilde{p}_\lambda(\tau|y_d)}\int \frac{d}{d\lambda}\tilde{p}_\lambda(\tau,\rho|y_d)d\rho \\ &= \frac{1}{\tilde{p}_\lambda(\tau|y_d)}\int \tilde{p}_\lambda(\tau,\rho|y_d)\frac{d}{d\lambda}\ln(\tilde{p}_\lambda(\tau,\rho|y_d))d\rho \\ &= \int \tilde{p}_\lambda(\rho|\tau,y_d)\frac{d}{d\lambda}\ln(\tilde{p}_\lambda(\tau,\rho|y_d))d\rho. \end{aligned} \tag{22}$$

Hence

$$
\begin{aligned}
& var^0_{\tau|y_d}[\frac{d}{d\lambda}\ln(\tilde{p}_\lambda(\tau|y_d))|_{\lambda=0}] \\
= \quad & var^0_{\tau|y_d}[E^0_{\rho|\tau,y_d}\{\frac{d}{d\lambda}\ln(\tilde{p}_\lambda(\tau,\rho|y_d))|_{\lambda=0}\}] \\
= \quad & var^0_{\tau|y_d}[E^0_{\rho|\tau,y_d}\{\frac{d}{d\lambda}\ln(\tilde{p}_\lambda(\theta,y_d))|_{\lambda=0}\}] \qquad (23) \\
= \quad & var^0_{\tau|y_d}[E^0_{\rho|\tau,y_d}\{\frac{d}{d\lambda}\ln(h^*(\theta))\}]
\end{aligned}
$$

where the last equation is obtained as in the proof of theorem 2.

b) Extending θ to $\tilde{\theta} = (\tilde{y}, \theta)$ and considering partial parameters $\tilde{y}$ and θ one obtains similarly

$$
\begin{aligned}
& var^0_{\tilde{y}|y_d}[\frac{d}{d\lambda}\ln(\tilde{p}_\lambda(\tilde{y}|y_d))|_{\lambda=0}] \\
= \quad & var^0_{\tilde{y}|y_d}[E^0_{\theta|\tilde{y},y_d}\{\frac{d}{d\lambda}\ln(\tilde{p}_\lambda(y_d,\tilde{y},\theta))|_{\lambda=0}\}] \qquad (24) \\
= \quad & var^0_{\tilde{y}|y_d}[E^0_{\theta|\tilde{y},y_d}\{\ln(\frac{p_1(y_d,\tilde{y},\theta)}{p_0(y_d,\tilde{y},\theta)})\}].
\end{aligned}
$$

As $\tilde{y}$ is a future observation from the unperturbed model $p_1(y_d, \tilde{y}, \theta) = p_1(y_d, \theta)p_0(\tilde{y}|\theta)$ and $p_0(y_d, \tilde{y}, \theta) = p_0(y_d, \theta)p_0(\tilde{y}|\theta)$ such that $p_1(y_d, \tilde{y}, \theta)/p_0(y_d, \tilde{y}, \theta) = h^*(\theta)$.

The assessments of local influence due to an additive or geometric mixture thus only differ in the (log-)scale of $h^*(\theta)$.

2.2 Sensitivity to case weights and case deletion

Perturbations of the likelihood by case weights or case deletion are of special interest. They are discussed in detail in this section because they involve some subtle issues.

2.2.1 Invariance to non-informative observations

Conventionally case deletion is represented by the perturbation function $h^{**}(\theta) = 1/p_0(y_i|\theta)$. However, $p_1(y_d|\theta) = p_0(y_{d\setminus i}|\theta)$ is not a proper density for y_d and hence an additive mixture $p_\lambda(y_d, \theta) = [(1-\lambda)p_0(y_d|\theta) + \lambda p_1(y_d|\theta)]p_0(\theta)$ is not well defined. The problem can be fixed easily using an arbitrary density q, say, such that $q(y_i)$ is not informative about θ. The posterior $p_1(\theta|y_d)$ resulting from $p_1(y_d|\theta) = p_0(y_{d\setminus i}|\theta)q(y_i)$ equals $p_0(\theta|y_{d\setminus i})$ for all such densities q and according to (3) the global measure of influence $D_\phi(p_0(\zeta|y_d), p_1(\zeta|y_d))$ is the same for all q. The posterior $p_\lambda(\theta|y_d)$ induced by the additive mixture though is *not* invariant to the choice of q and neither is the Fisher information, the local measure of influence: For $h^*_q(\theta) := q(y_i)/p_0(y_i|\theta)$ obviously $var^0_{\theta|y_d}[h^*_q(\theta)] = q^2(y_i) var^0_{\theta|y_d}[h^{**}(\theta)]$. At first sight this lack of invariance seems to be counterintuitive but only if one intuitively expects $p_\lambda(\theta|y_d) = (1-\lambda)p_0(\theta|y_d) + \lambda p_1(\theta|y_d)$. This equation does not hold in general, as the right hand side does not depend on $q(y_i)$ whereas the left hand side does. The equation *is* satisfied only for the choice $q(y_i) = p_0(y_i|y_{d\setminus i})$. It corresponds to the perturbation function $h^*(\theta) = p_0(y_i|y_{d\setminus i})/p_0(y_i|\theta)$ which is therefore recommended to define the mode of perturbation for case deletion in an additive mixture of likelihoods.

A problem of invariance does not occur with a geometric mixture

$$\begin{aligned}\tilde{p}_\lambda(y_d,\theta) &= c(\lambda)p_0^{1-\lambda}(y_d,\theta)p_1^\lambda(y_d,\theta) \\ &= c(\lambda)p_0^{1-\lambda}(y_d|\theta)p_0^\lambda(y_{d\setminus i}|\theta)q^\lambda(y_i)p_0(\theta) \\ &= c(\lambda)p_0(y_{d\setminus i}|\theta)p_0^{1-\lambda}(y_i|\theta)q^\lambda(y_i)p_0(\theta), \end{aligned} \tag{25}$$

where $\tilde{p}_\lambda(\theta|y_d)$ does not depend on the choice of q. Therefore the local measure of influence

$$var^0_{\theta|y_d}[\ln(h^*_q(\theta))] = var^0_{\theta|y_d}[\ln(p_0(y_i|\theta))] \tag{26}$$

is independent of q, too.

2.2.2 Geometric mixtures and case weights

With $h^{**}(\theta) = 1/p_0(y_i|\theta)$ the geometric mixture represents the familiar idea of case weights $\omega = 1-\lambda$,

$$\pi_\omega(y_d|\theta) := p_0(y_{d\setminus i}|\theta)p_0^\omega(y_i|\theta) \xrightarrow{\omega\to 0} p_0(y_{d\setminus i}|\theta). \tag{27}$$

A decreasing weight for a single case eventually corresponds to a constant density ($\equiv 1$) in y_i respectively case deletion.

2.2.3 Case deletion in geometric mixtures

Case weights for the Normal distribution and more generally in scaled exponential families yield a perturbation scheme for overdispersion and - in the limit - case deletion and have therefore been used as convenient parameters. A Bayesian application of these ideas was investigated by Millar and Stewart (2005).

Case weights induce the mixture of sampling densities $\pi_\omega(y_d|\theta)$ with $\omega = 1$ corresponding to the null model. In general, the normalizing constant $c_\pi(\omega,\theta)$ of $\pi_\omega(y_d|\theta)$ - if it exists - depends also on ω. Millar and Stewart (2005) suggest to neglect the normalizing constant and to base a local sensitivity analysis on the 'weighted likelihood posterior'

$$\pi_\omega(\theta|y_d) = \pi_\omega(y_d|\theta)p_0(\theta) \Big/ \int \pi_\omega(y_d|\theta)p_0(\theta)d\theta \tag{28}$$

(provided the integral in the denominator is finite). The local influence of case deletion is then assessed as second derivative of the Kullback-Leibler divergence between $\pi_\omega(\theta|y_d)$ and $p_0(\theta|y_d)$ evaluated at $\omega = 1$. It turns out to be $var^0_{\theta|y_d}[\ln(p_0(y_i|\theta))]$ as in (26). The coincidence results from the fact that (despite the technical problems with normalizing constants) Millar and Stewart (2005) implicitly work out a short cut of a Bayesian analysis based on the geometric mixture (25) : In order to avoid constant densities refer again to $h^*_q(\theta) = q(y_i)/p_0(y_i|\theta)$. Then

$$\tilde{p}_\lambda(y_d, \theta) = c(\lambda)\pi_\omega(y_d|\theta)q^\lambda(y_i)p_0(\theta)$$

yields

$$\tilde{p}_\lambda(y_d) = c(\lambda)q^\lambda(y_i)\int \pi_\omega(y_d|\theta)p_0(\theta)d\theta$$

and hence

$$\tilde{p}_\lambda(\theta|y_d) = \pi_\omega(\theta|y_d).$$

Thus the results obtained by Millar and Stewart (2005) correspond to special choices of multiplicative modes of perturbation with a geometric mixture. Their results are justified here in a different Bayesian set-up that technically avoids constant densities. In this set-up the normalizing constant of $\pi_\omega(y_d|\theta)q^\lambda(y_i)$ as a function of y_d becomes part of $\tilde{p}_\lambda(\theta) \neq p_0(\theta)$. The existence of a normalizing constant $c(\lambda)$ in the geometric mixture has still to be assumed, though.

2.2.4 Case weights: intrinsic or formal parameters ?

In robustness analyses of regression models (Cook (2004)) case weights indicating a large variance of a response variable have been considered. Let for example $p(y_j|\mu,\omega)$ be obtained from $Y_j|\mu \sim N(\mu, \sigma^2/\omega)$ with σ^2 known, and assume independent observations. Define the null model by $\omega = 1$ for all observations, that is $p_0(y_j|\mu) = p(y_j|\mu, 1)$ for all j. Let further p_1 correspond to a single weighted observation y_i such that $p_1(y_j|\mu,\omega) = p(y_j|\mu, 1)$ for $j \neq i$ and $p_1(y_i|\mu,\omega) = p(y_i|\mu,\omega)$. The local measure of influence induced by the direct parameterization with case weights ω is given by

$$\begin{aligned}
& \frac{d^2}{d\omega^2} D_{\phi_{KL}}(p_0(\mu|y_d), p_1(\mu|y_d,\omega))|_{\omega=1} \\
= \;& var^0_{\mu|y_d}[\frac{d}{d\omega}\ln(p_1(y_d|\mu,\omega))|_{\omega=1}] && (32)\\
= \;& var^0_{\mu|y_d}[\frac{(y_i-\mu)^2}{2\sigma^2}] && (33)\\
= \;& var^0_{\mu|y_d}[\ln(p_0(y_i|\theta))] && (34)
\end{aligned}$$

as in (26). Whether case weights are intrinsic or artificial parameters has been controversial (cp. the contributions by Loynes and Lawrence to the discussion of Cook's paper (Cook (2004))). A case weight seems to be an intrinsic perturbation parameter whenever it is a scaling parameter with the mode of perturbation corresponding to case deletion. This approach was investigated by McCulloch (1989) and extended by Lavine

(1992) both using the (directed) Kullback-Leibler divergence. The present analysis emphasizes that a case weight in scaled families and more generally is a formal geometric mixing parameter.

2.2.5 Fixed case weights

Whenever the case weight ω is used as the parameter according to which the derivative is formed to obtain a local measure of influence it describes the amount of deviation from the unperturbed likelihood (for instance as over- or underdispersion) with the mode of perturbation corresponding to case deletion. In contrast, a fixed value ω defining a certain degree of over- or underdispersion as the mode of perturbation for Gaussian distributions as above, the amount of deviation from the null model in the direction of that fixed over- or underdispersed density may be represented by a mixing parameter in a geometric mixture. The influence of the fixed over- or underdispersion can then be assessed evaluating for example $\frac{d^2}{d\lambda^2} D_{\phi_{KL}}(p_0(\mu|y_d), \tilde{p}_\lambda(\mu|y_d,\omega))|_{\lambda=0}$ with $\tilde{p}_\lambda(y_d,\mu,\omega) = c(\lambda)p_0^{1-\lambda}(y_d|\mu)p_1^\lambda(y_d|\mu,\omega)p_0(\mu)$. The local measure of influence

$$var^0_{\mu|y_d}[\ln \frac{p_{N(\mu,\sigma^2/\omega)}(y_i)}{p_{N(\mu,\sigma^2)}(y_i)}] = var^0_{\mu|y_d}[(1-\omega)(\frac{(y_i-\mu)^2}{2\sigma^2})]. \tag{35}$$

differs from the one in equation (33) in that the amount of over- or underdispersion is taken into account by the factor $(1-\omega)$.

3 Examples

In order to illustrate and compare different modes of perturbation yielding measures of local influence, two simple examples with additive mixtures are given. The first example focuses on changes of the likelihood, in the second example changes of the prior are considered.

More examples with modifications of the prior can be found in the paper by Dey and Birmiwal (1994). As case deletion represented by a mode of perturbation with the non-informative (predictive) density yields equality (13) of the Fisher information to the χ^2-divergence, the computationally more complex examples given by Peng and Dey (1995) for the χ^2-divergence as global measure of influence also fit in here. Examples for the effect of case deletion using a geometric mixture as discussed in section 2.2.3 are presented by Millar and Stewart (2005).

Computational issues in the calculation of the Fisher information of the form $var^0_{\zeta|y_d}[E^0_{\theta|\zeta,y_d}\{h^*(\theta)\}]$ or $var^0_{\zeta|y_d}[E^0_{\theta|\zeta,y_d}\{\ln(h^*(\theta))\}]$ are not special within a Bayesian analysis as ‘only’ (conditional) posterior means and variances of analytically known functions of the parameters are required. Sampling from a conditional posterior distribution may not be immediate in a complex model but no extra effort is to be made for a local sensitivity analysis of the type discussed in this paper.

3.1 Example 1: Stack loss data

The famous stack loss data have been analyzed with several regression procedures (e.g. Spiegelhalter et al. (1996)) and are known to comprise at least three 'outliers'. The response variable Y is 'stack loss', the three covariables are X_1 = 'air flow', X_2 = 'temperature', X_3 = 'concentration'. A standard conjugate Bayesian regression analysis was applied to this data set and used as null model. More precisely, it was assumed that independently $Y_i|\beta, \sigma^2 \sim N(x_i^T\beta, \sigma^2)$, $i = 1, ..., 21$, $x_i^T = (1, x_{1i}, x_{2i}, x_{3i})$, and the priors were $\beta|\sigma^2 \sim N(0, \nu\sigma^2 I_4)$ with $\nu = 1000$ and $\sigma^2 \sim Inv - \Gamma(0.001, 0.001)$. The posterior estimates are similar to those given in previous analyses under slightly different assumptions: $E(\beta^T|y_d) = (-39.39, 0.717, 1.293, -0.158)$, $SD(\beta^T|y_d) = (11.23, 0.13, 0.35, 0.15)$; $E(\sigma|y_d) = 3.04$, $SD(\sigma|y_d) = 0.52$; the response $\tilde{Y}$ at $\tilde{x}^T = (1, 60, 20, 85)$, which is within the range of but not exactly comprised by the experimental design, follows a posterior Student-t-distribution with mean 16.01, standard deviation 0.37 and 21 degrees of freedom. Using the posterior probability of a large residual, $|y_i - x_i^T\beta| > 2.5\sigma$, as an indicator of 'outlyingness' of the i-th observation, the observations $y_{21}, y_4, y_3, y_1, y_{17}$ (in this order) were suggested for an assessment of influence.

3.1.1 Case deletion

To analyze case deletion $h^*(\theta) = p_0(y_i|y_{d\setminus i})/p_0(y_i|\theta)$ was used. Calculation of the Fisher information (equal to the χ^2-divergence here) for each case deleted in turn indicated that observation y_{21} is by far most influential for all parameters. The four remaining observations y_4, y_3, y_1, y_{17} are (in this order) the next influential observations for θ and are split into the subsets $\{y_1, y_{17}\}$, $\{y_3\}$ and $\{y_4\}$ as being influential for β, σ^{-2}, $\tilde{y}$ respectively.

3.1.2 Sequence of fixed case weights

The sequence of weights $\omega^{-1} \in \{2, 5, 10, 100, 1000, 10000\}$ was tried for each individual case $y_i \in \{y_{21}, y_4, y_3, y_1\}$, and the χ^2-divergences were evaluated for each fixed ω. For each parameter $\zeta \in \{\theta, \beta, \sigma^{-2}, \tilde{y}\}$ and each case the χ^2-divergence increases with ω^{-1} and approximates the value of the χ^2-divergence obtained with case deletion at $\omega^{-1} = 100$. For higher weights ω^{-1} the χ^2-divergences stabilize.
As mentioned before the χ^2-divergence is based on the ratio of posterior densities whereas the Fisher information is based on the ratio of joint densities reflecting directly the perturbation introduced in the likelihood or prior. For case weights the ratio of marginal densities does not equal 1, and therefore the Fisher information assesses 'prior input' rather than 'posterior output'. For example, the sequence of weights applied to the cases y_{21}, y_4, y_3 simultaneously yields for β again an increasing χ^2-divergence in ω^{-1}, but the Fisher information first increases, then peaks at $\omega^{-1} = 10$ and decreases for $\omega^{-1} \geq 100$. The peak thus indicates the maximum induced perturbation.

3.1.3 More robust distributional assumptions

In order to account for outlying observations alternative distributional assumptions like the t_4- or double exponential instead of the Normal distribution have been suggested for the stack loss data. First the t_d-distribution for all observations was assessed for $d \in \{4, 10, 15, 20, 40\}$ and, as expected, the Fisher information for $\zeta \in \{\theta, \beta, \sigma^{-2}\}$ decreased rapidly as the degrees of freedom increased and the t_d-distribution approximated the Normal distribution. In comparison the assumption of a double exponential distribution for all response variables turns out to be more influential in terms of the Fisher information for all parameters than the t_4-distribution. The same statement holds if the alternative distribution is assumed for Y_{21}, Y_4, Y_3 only.

3.2 Example 2: Casino data

In his introductory text on Bayesian Statistics Bolstad (2004) discusses several prior assumptions for the success rate θ in a sequence of Bernoulli trials to exemplify robustness issues. The data represent the amount of support for a casino by a town's population. Out of $n = 100$ inhabitants $r = 26$ voted for the casino. The prior distributions under consideration are a) the reference prior $Beta(0.5, 0.5)$, b) the uniform prior $Beta(1, 1)$, c) a conjugate informative prior $Beta(4.8, 19.4)$ and d) a subjective informative prior

$$p(\theta) = \frac{1}{0.7} \langle \begin{array}{cc} 20\theta & 0 \leq \theta \leq 0.1 \\ 2 & 0.1 \leq \theta \leq 0.3 \\ 5 - 10\theta & 0.3 \leq \theta \leq 0.5 \\ 0 & 0.5 \leq \theta \leq 1 \end{array} . \tag{36}$$

In this case the Fisher information assesses the induced initial perturbation as the posterior variance of the ratio of prior densities. The $Beta(0.5, 0.5)-$prior is chosen as reference. A first comparison of $Beta(\alpha, \alpha)$-priors with $\alpha \in \{0.6, 0.7, 0.8, 0.9, 1\}$ confirmed that the Fisher information for θ increases with α. In a second comparison of the priors b)-d) to the reference prior the Fisher information points to prior c) as the most influential perturbation followed by prior d). The posterior densities though are rather similar (cp. Bolstad (2004), fig.10.1 and fig.10.2), and the difference seen in the Fisher information is compensated in the χ^2-divergence by the ratio of marginal densities, that is by a high (unperturbed) posterior mean of the perturbation function according to eq. (14).

4 Discussion

The Fisher information $I_{\zeta|y_d}(0)$ refines a global analysis of influence based on a (χ^2-) divergence. It points to that mode of perturbation that introduces the largest deviation in the ingredients of a Bayesian analysis, likelihood and prior. It does not directly assess the effect of the mode of perturbation on the posterior distribution of a parameter of interest but is counterbalanced by the impact on the marginal densities.

Hence the measure of global and local influence usually differ. Only if case deletion is represented by the predictive density in an additive mixture, the two measures coincide. In this case the Fisher information might offer a computational alternative to the χ^2-divergence and vice versa. In the examples the local measure of influence is applied only in comparisons of several modes of perturbation, and thus calibration does not matter. However, in order to assess the magnitude and hence relevance of a potentially influential mode of perturbation some calibration is needed. The calibration initially suggested by McCulloch (1989) has been referred to by many authors.

Applied to case deletion the additive mixture (with non-informative observations) preserves integrability to one and thus avoids the problem of possibly non-existing normalizing constants occurring with geometric mixtures. Therefore it might be preferred. An additive mixture of posterior densities can be induced by either an additive mixture of sampling densities or an additive mixture of priors. Thus it is always in this sense coherent and an additive mixture of sampling densities corresponds to conventional models for robustness.

References

Blyth, S. (1994). "Local Divergence and Assoiation." *Biometrika*, 81: 579–584.

Bolstad, W. (2004). *Introduction to Bayesian Statistics*. New York: Wiley.

Cook, D. (2004). "Assessment of Local Influence." *Journal of the Royal Statistical Society, Series B*, 48: 133–169. With discussion.

Dey, D. and Birmiwal, L. (1994). "Robust Bayesian Analysis Using Divergence Measures." *Statistics and Probability Letters*, 20: 287–294.

Geisser, S. (2000). "Bayesian Perturbation Diagnostics and Robustness." In Insua, D. R. and Ruggeri, F. (eds.), *Robust Bayesian Analysis*, 289–301. New York: Springer.

Gustafson, P. (2000). "Local Robustness in Bayesian Analysis." In Insua, D. R. and Ruggeri, F. (eds.), *Robust Bayesian Analysis*, 71–88. New York: Springer.

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

Lavine, M. (1992). "Local Predictive Influence in Bayesian Linear Models with Conjugate Priors." *Communications in Statistics - Simulation*, 21: 269–283.

McCulloch, R. (1989). "Local Model Influence." *Journal of the American Statistical Association*, 84: 473–478.

Millar, R. and Stewart, W. (2005). "Detection of Locally Influential Observations in Bayesian Models." In *Research Report 2005:2*. Auckland, New Zealand: Dept. of Statistics. Available from: http://www.stat.auckland.ac.nz/publications.

Peng, F. and Dey, D. (1995). "Bayesian Analysis of Outlier Problems Using Divergence Measures." *Canadian Journal of Statistics*, 23: 199–213.

Sivaganesan, S. (2000). "Global and Local Robustness Approaches: Uses and Limitations." In Insua, D. R. and Ruggeri, F. (eds.), *Robust Bayesian Analysis*, 89–108. New York: Springer.

Spiegelhalter, D., Thomas, A., Best, N., and W.Gilks (1996). *BUGS Examples, Version 0.5*, volume 1. Cambridge Medical Research Council Biostatistics Unit.

Weiss, R. (1996). "An Approach to Bayesian Sensitivity Analysis." *Journal of the Royal Statistical Society, Series B*, 58: 739–750.

Weiss, R. and Cook, D. (1992). "A Graphical Case Statistic for Assessing Posterior Influence." *Biometrika*, 79: 51–55.

Acknowledgments

The hints and suggestions of anonymous referees helped to considerably improve a first draft of the paper.
This work was stimulated during a visit at the Department of Statistics, University of Auckland, New Zealand supported by DFG-grant GZ:447 NSL-111/1/05.