ELSEVIER

Statistics & Probability Letters 36 (1997) 205–212

STATISTICS & PROBABILITY LETTERS

A simple diagnostic tool for local prior sensitivity

Daniel Peña [a], Ruben Zamar [b]

[a] *Department of Statistics and Econometrics, Universidad Carlos III, de Madrid, Spain*
[b] *Department of Statistics, University of British Columbia, 333-6356 Agricultural Road, Vancouver, BC, Canada V6T 1Z2*

Received 1 November 1996; received in revised form 1 February 1997

Abstract

This paper presents a simple diagnostic tool to assess the sensitivity of the posterior mode in the presence of an infinitesimal contamination in the prior distribution. The proposed diagnostic measure is easy to compute and can be used as a first step in judging the robustness of the Bayesian inference. The procedure is illustrated in the estimation of the mean of a normal distribution. Some extensions of this diagnostic measure to the multivariate case and credibility intervals are briefly discussed. © 1997 Elsevier Science B.V.

Keywords: Bayesian robustness; Influence function; Mixture of distributions

1. Introduction

As eliciting prior distributions is not an easy task and a prior distribution is needed in Bayesian inference, it is not surprising that reference priors and robustness to the prior distribution are two important lines of Bayesian research. Robust Bayesian Analysis includes not only the study of the prior distribution but the whole process of inference. Berger (1994) presents an overview of this topic and gives many references.

The standard approach in prior robustness is to consider a whole set of prior distributions, instead of a single one, and study the range of a certain measure of interest when the prior varies over this class. Some references to this field are Berger (1984, 1990, 1994), Cuevas and Sanz (1988), Moreno and Cano (1991), Delampady and Dey (1994), Moreno and Pericchi (1993), Pericchi and Walley (1991), Wasserman (1992) and Peña and Zamar (1995). Gustafson et al. (1995) study the local sensitivity of general functionals of the prior using several distances between distribution, obtaining interesting results, and Gustafson and Wasserman (1995) investigate diagnostics for small prior changes over a k-dimensional parameter space. Recently, Gustafson (1996) investigates the local sensitivity of posterior expectations.

We are interested in deriving a simple (preliminary) sensitivity analysis tool and concentrate on a single (although central) feature of the posterior distribution, namely the posterior density mode.

We have chosen the posterior mode by the following reasons: (a) The mode is one of the most sensitive features of the posterior distribution. We are more likely to pick up potential sensitivity problems by looking at the mode than at other location measures (e.g. the mean); (b) Simplicity. We could look at other location and dispersion summaries but this would make the sensitivity analysis more complex. (c) The mode is the

PII S0167-7152(97)00067-9

most likely value for the parameter, according to the Bayesian analysis. The diagnostic tool we derive is the posterior mode influence function (PMIF), that is obtained by computing the directional (Gateaux) derivative of the posterior density mode on the direction of a "contaminating" prior, normalized by the standard deviation of the posterior distribution. This function shows the effect of a small degree of uncertainty in the likelihood of some values of the prior domain. As a very small amount of contamination cannot be regarded as a change in the prior opinion, if it produces a large change in the posterior mode we can conclude that the Bayesian inference is sensitive to the prior specification.

The PMIF can be easily obtained by taking advantage of the fact that the posterior density mode, $\widehat{\theta}$, under mild regularity conditions, satisfies the equation

$$\frac{\partial}{\partial\theta} \ln p(\widehat{\theta}|y) = 0.$$

The rest of the paper in organized as follows. Section 2 develops the basic theory. Section 3 applies it to study the sensitivity of the estimation of the mean in the normal case. Section 4 discusses some possible extensions of the procedure to multivariate problems and credibility intervals. Section 5 includes some final remarks.

2. The sensitivity of the posterior mode

Suppose that we are interested in a parameter θ. We have some prior distribution, $\pi_0(\theta)$, and we observe a random sample $x=(x_1,\dots,x_n)$ from the distribution $f(x|\theta)$. Then, the posterior distribution of θ is given by

$$p_0(\theta|x) = k\pi_0(\theta)\Pi f(x_i|\theta). \tag{1}$$

Under mild regularity assumptions, the mode of this posterior distribution satisfies the equation

$$\frac{\partial \log p_0(\theta|x)}{\partial\theta} = \frac{\pi_0'(\theta)}{\pi_0(\theta)} + \sum \frac{f'(x_i|\theta)}{f(x_i|\theta)} = 0, \tag{2}$$

where

$$f'(x_i|\theta) = \frac{\partial f(x_i|\theta)}{\partial\theta}.$$

Suppose now that instead of the single prior $\pi_0(\theta)$ we consider the class of ε-contaminated prior distributions defined by

$$\pi(\theta) = (1-\varepsilon)\pi_0(\theta) + \varepsilon q(\theta), \tag{3}$$

where $0<\varepsilon<1$ and $q \in \mathcal{Q}$, where $\mathcal{Q}$ is a class of contaminating distributions. Then, the new posterior distribution is given by

$$p(\theta|x) = \lambda(x)p_0(\theta|x) + (1-\lambda(x))q(\theta|x), \tag{4}$$

where $p_0(\theta|x)$ and $q(\theta|x)$ are the posterior distributions obtained from the priors $\pi_0(\theta)$ and $q(\theta)$ and

$$\lambda(x) = \frac{(1-\varepsilon)m(x|\pi_0)}{m(x|\pi)}, \tag{5}$$

where $m(x|\pi_0)$ is the marginal distribution obtained from π_0:

$$m(x|\pi_0) = \int f(x|\theta)\pi_0(\theta)\,\mathrm{d}\theta, \tag{6}$$

and $m(x|\pi)$ is the marginal distribution obtained from π:

$$m(x|\pi)=\int f(x|\theta)\pi(\theta)\,\mathrm{d}\theta. \tag{7}$$

To study the sensitivity of the posterior density $p(\theta|x)$ when the prior moves away from π_0 in the direction of q, we focus on the mode of $p(\theta|x)$ which satisfies the equation

$$G(\theta(\varepsilon),\varepsilon)=\tilde{\pi}(\theta(\varepsilon))+\psi_n(\theta(\varepsilon))=0 \tag{8}$$

where $\tilde{\pi}(\theta)=\pi'(\theta)/\pi(\theta)$ is the score function of the prior and $\psi_n(\theta)=\sum f'(x_i|\theta)/f(x_i|\theta)$ is the score of the likelihood. Then, for the general prior (3)

$$\tilde{\pi}(\theta)=\frac{(1-\varepsilon)\pi_0'(\theta)+\varepsilon q'(\theta)}{(1-\varepsilon)\pi_0(\theta)+\varepsilon q(\theta)}. \tag{9}$$

Let $\widehat{\theta}_0$ and $\widehat{\theta}$ be the mode of the posterior densities $p_0(\theta|x)$ and $p(\theta|x)$ respectively, that are obtained from the corresponding prior densities $\pi_0(\theta)$ and $\pi(\theta)$. Under regularity assumptions on q and π_0, the derivative of $\widehat{\theta}(\varepsilon)$ with respect to ε at $\varepsilon=0$ is obtained from (8) as follows:

$$\left(\frac{\partial G(\widehat{\theta}(\varepsilon),\varepsilon)}{\partial\varepsilon}\right)_{\varepsilon=0}=\frac{-\pi_0'(\widehat{\theta}_0)+\pi_0''(\widehat{\theta}_0)\left(\dfrac{\mathrm{d}\widehat{\theta}(\varepsilon)}{\mathrm{d}\varepsilon}\right)_{\varepsilon=0}+q'(\widehat{\theta}_0)}{\pi_0(\widehat{\theta}_0)}$$
$$-\frac{\left[-\pi_0(\widehat{\theta}_0)+\pi_0'(\widehat{\theta}_0)\left[\dfrac{\mathrm{d}\widehat{\theta}(\varepsilon)}{\mathrm{d}\varepsilon}\right]_{\varepsilon=0}+q(\widehat{\theta}_0)\right]\pi_0'(\widehat{\theta}_0)}{[\pi_0(\widehat{\theta}_0)]^2}+\left(\frac{\mathrm{d}\widehat{\theta}(\varepsilon)}{\mathrm{d}\varepsilon}\right)_{\varepsilon=0}\psi_n'(\widehat{\theta}_0)=0.$$

Dropping the argument $\widehat{\theta}_0$ to simplify the notation and denoting

$$\dot{\theta}_0=\left(\frac{\mathrm{d}\widehat{\theta}(\varepsilon)}{\mathrm{d}\varepsilon}\right)_{\varepsilon=0} \tag{10}$$

we get

$$\dot{\theta}_0=\frac{q\pi_0'-q'\pi_0}{\pi_0^2\left[\dfrac{\pi_0''}{\pi_0}-\left[\dfrac{\pi_0'}{\pi_0}\right]^2+\psi_n'\right]}. \tag{11}$$

Observe that when n is large the leading term in the denominator in (11) is ψ_n' which is of order n and negative. Suppose that π_0 and q are both unimodal. Then, if $q'>0$ and $\pi_0'<0$ it follows that $\dot{\theta}_0>0$, as one would expect. Also, if $q'<0$ and $\pi_0'>0$ then $\dot{\theta}_0<0$. Eq. (11) can be rewritten as

$$\dot{\theta}_0=-\frac{\mathrm{d}}{\mathrm{d}\theta}\left[\frac{q(\theta)}{\pi_0(\theta)}\right]_{\theta=\widehat{\theta}_0}\frac{1}{\mathrm{O}(n)}, \tag{12}$$

where

$$\mathrm{O}(n)=\left(\frac{\pi_0'}{\pi_0}\right)^2-\frac{\pi_0''}{\pi_0}-\psi_n',$$

is positive for large n. Note that, $\dot{\theta}_0 \to 0$ when $n \to \infty$, which is consistent with the well known result that the prior is expected to have small influence when the sample size is large.

Two difficulties with the interpretation of (11) are: (1) $\dot{\theta}_0$ depends on the measurement units of the data, and (2) it needs to be interpreted in comparison to the spread of the posterior distribution. A possible way to overcome these problems is to standardize this measure. Therefore, we define the posterior mode influence function (PMIF) as

$$\mathrm{PMIF}(\pi_0, q) \equiv \frac{\dot{\theta}_0}{\mathrm{DT}(p_0(\theta|x))},$$

where $\mathrm{DT}(p_0(\theta|x))$ is the posterior standard deviation under the central prior, π_0. Typically, PMIF will be of order $\mathrm{O}(\sqrt{n})$ and one may then consider stabilizing this measure by multiplying it by $\sqrt{n}$. However, this normalization is not used in this paper.

One may consider using as a diagnostic tool the supremum of $\mathrm{PMIF}(\pi_0, q)$ over a given class $\mathcal{Q}$,

$$\mathrm{PMS} \equiv \sup_{q \in \mathcal{Q}} \mathrm{PMF}(\pi_0, q)$$

which will be called the posterior mode sensitivity (PMS). However, the PMS may diverge to infinity due to the effect of some unrealistic sequence of prior distributions in $\mathcal{Q}$. Consequently, we recommend the direct use of the PMIF to better understand the effect of different types of prior uncertainty.

The choice of the "direction" q depends on the situation at hand. For instance, in the case of a normal prior illustrated in Section 3, we consider $q = \mathrm{N}(\mu_1, \delta^2)$ for different values of μ_1 and δ^2. The reasons for this choice are the computational simplicity and the interpretability of the results in terms of "change in the mean" and "change in the variance" (which are easy to communicate). Another possibility is the use of shifted and rescaled Student-t distributions with several degrees of freedom. The Student-t approach is, of course, more general and retains the interpretability features of the normal approach but loses in terms of computational simplicity.

Of course, here we are only looking at a particular feature of the posterior distribution and a more global analysis could be done by comparing the posterior distributions $p(\theta|x)$ and $p_0(\theta|x)$ themselves. This comparison can be made, for instance, by using the Kullblack–Leibler divergence or any other relevant measure of distance. See Guftanson et al. (1995). However, the study of the mode of the posterior distribution appears to be a simple and natural way to judge the sensitivity of the inference to local perturbations in the prior distribution.

To simplify the presentation of this section we have avoided the statement of the assumption under which our derivations are rigorous. However, it is clear that if the mode is well defined under the initial prior, and π_0, q and f are differentiable our derivation are justified. They may also be justified under milder assumptions, but the previous conditions cover most cases of practical importance.

3. An application to the estimation of the mean of a normal distribution

To illustrate the diagnostic tool presented in the previous section, suppose that we want to estimate the mean of a normal population, θ, and let us assume that $\pi_0(\theta) \sim \mathrm{N}(\mu_0, \sigma_0^2)$, $x/\theta \sim \mathrm{N}(\theta, \sigma^2)$, and , for simplicity, σ^2 is known. Then, the posterior mode, $\widehat{\theta}_0$, is given by

$$\widehat{\theta}_0 = \frac{\sigma^2 \mu_0 + n\sigma_0^2 \bar{x}}{\sigma^2 + n\sigma_0^2} \tag{13}$$

and the standard deviation of the posterior distribution is

$$\sigma_p = \sigma\sigma_0 \Big/ \sqrt{n\sigma_0^2 + \sigma^2}. \tag{14}$$

Because of the symmetry we assume without loss of generality that $\bar{x} > \mu_0 \geqslant 0$ and therefore $\widehat{\theta}_0 > 0$.

Suppose now that the prior distribution is $\pi(\theta) = (1-\varepsilon)\pi_0(\theta) + \varepsilon q(\theta)$, with $q(\theta) = \mathrm{N}(\mu_1, \delta^2)$. Then, as $\psi_n = n(\bar{x} - \widehat{\theta}_0)/\sigma^2$, $\psi_n' = -n/\sigma^2$, $\pi_0' = -[(\widehat{\theta}_0 - \mu_0)/\sigma_0^2]\pi_0$, $\pi_0'' = [((\widehat{\theta}_0 - \mu_0)/\sigma_0^2)^2 - 1/\sigma_0^2]\pi_0$, and $q' = [(\mu_0 - \widehat{\theta}_0)/\delta^2]q$, from (8) we obtain

$$\mathrm{PMIF}(\mu_1, \delta) = \frac{\sigma_p}{\delta}\left[d_1 - \frac{\sigma_0}{\delta}d_2\right]\exp\{0.5(d_1^2 - d_2^2)\}, \tag{15}$$

where $d_1 = (\widehat{\theta}_0 - \mu_0)/\sigma_0$ and $d_2 = (\widehat{\theta}_0 - \mu_1)/\delta$. Not surprisingly, the PMIF is directly proportional to the posterior standard deviation which converges to zero when $n \to \infty$. This is consistent with the stable estimation property of Bayesian procedures (Savage, 1963). Moreover, the PMIF is inversely proportional to the variance of the contaminating distribution: a flat contamination can hardly affect the posterior mode. The PMFI increases with $d_1 = n\sigma_0(\bar{x} - \mu_0)/(\sigma^2 + n\sigma_0^2)$, that is, the posterior mode is less robust when there is a big discrepancy between the prior mean and the sample mean. Finally, the PMFI decreases when d_2 is large. Therefore, the posterior mode is more robust when there is a big discrepancy between the modes of the posterior and contaminating distributions.

In order to better understand the combined effect of μ_1 and δ we consider the following three cases.

Case 1: $\mu_1 = \widehat{\theta}_0$. One notices in this case that $|\mathrm{PMIF}(\widehat{\theta}_0, \delta)| \to \infty$ as $\delta \to 0$, and the sign is determined by that of d_1. Therefore, the posterior mode tends to move towards the prior mean, and it is most sensitive to a point mass contamination at $\widehat{\theta}_0$ (provided that $d_1 \neq 0$, i.e. $\widehat{\theta}_0 \neq \mu_0$). The practical conclusion from this result is that small uncertainty in the value of $\pi_0(\theta)$ far from $\widehat{\theta}_0$ has less effect than small uncertainty on values around $\widehat{\theta}_0$.

Case 2: $(\widehat{\theta}_0 - \mu_0)(\widehat{\theta}_0 - \mu_1) > 0$. In this case the sign of PMIF is opposite to the sign of $(\widehat{\theta}_0 - \mu_0)$ for values of δ smaller than δ_0, where

$$\delta_0^2 = \sigma_0^2 \frac{(\widehat{\theta}_0 - \mu_1)}{(\widehat{\theta}_0 - \mu_0)},$$

and it has the same sign of $(\widehat{\theta}_0 - \mu_0)$ for values of δ larger than δ_0. The practical conclusion from this result is that relatively spiky contaminations $(\delta < \delta_0)$ moves the posterior mode away from the sample mean, and relatively flat ones $(\delta > \delta_0)$ moves $\widehat{\theta}$ towards the sample mean. The two values of maximum influence are given by

$$\delta_\pm^2 = \frac{\sigma_0^2}{2}\frac{d_3}{d_1}[(d_1d_3 + 3) \pm \sqrt{(d_1d_3 + 3)^2 - 4d_1d_3}], \tag{16}$$

where $d_3 = (\widehat{\theta}_0 - \mu_1)/\sigma_0$. The minus (plus) sign produces the largest displacement towards (away from) the sample mean.

Case 3: $(\widehat{\theta}_0 - \mu_0)(\widehat{\theta}_0 - \mu_1) < 0$. In this case the sign of PMIF is determined by that of d_1. Therefore, the contamination always moves the posterior mode towards the sample mean. The maximum value is given by (16) with the plus sign (this is the only positive root since $d_1d_3 < 0$). The practical conclusion from this result is that any type of small uncertainty will move the posterior mode closer to the sample mean.

3.1. Symmetric contamination

If the uncertainty only concerns the tails of the prior distribution one can make a even simpler analysis by restricting attention to the symmetric normal mixture model ($\mu_1 = \mu_0$). In this case, using (15) and the relation $\delta d_2 = \sigma_0 d_1$ one finds that

$$\mathrm{PMIF}(\delta) = \frac{\sigma_p}{\delta} d_1 \left[1 - \frac{\sigma_0^2}{\delta^2}\right] \exp\left\{0.5 d_1^2 \left(1 - \frac{\sigma_0^2}{\delta^2}\right)\right\}. \tag{17}$$

Assuming, without loss of generality, that $d_1 > 0$, the PMFI is positive when $\delta > \sigma_0$, and negative otherwise. In other words, a small increase in the prior variance moves the posterior mode away from the prior mean and towards the sample mean, as one may expect.

The PMFI has a positive and a negative maximum values achieved at

$$\delta_\pm^2 = \frac{\sigma_0^2}{2}[(d_1^2 + 3) \pm \sqrt{(d_1^2 + 3)^2 - 4d_1^2}], \tag{18}$$

where the plus (minus) sign produces the positive (negative) maximum. Observe that when d_1 is small, $\delta_+^2 \approx 3\sigma_0^2$. and $\delta_-^2 \approx 0$. This means that when the data and the prior agree, the most damaging symmetric contamination has a variance that is three times that of the prior.

On the other hand, when d_1 is large, that is, the data and the prior are not consistent, it is easy to see that when $d_1 \to \infty$, $\delta_-^2 \to \sigma_0^2$ and $\delta_+^2 \to \infty$. Therefore when d_1 is large the posterior mode can only be moved towards the sample mean and the most damaging symmetric contamination has a large variance close to $d_1^2 \sigma_0^2$.

Although the PMS is always infinite for all σ_0, we note that the PMIF goes to zero (infinity) when σ_0 goes to zero (infinity). A non-informative prior leads to a very non-robust posterior mode whereas a strong prior belief produces the most robust situation.

4. Some possible extension

The simple tools presented in the previous section can be generalized in two directions. The first, and most obvious one, is to consider the case of a vector parameter case. The second, is to consider the sensitivity of credibility intervals.

4.1. Vector parameters

For the multivariate case, let θ be a $k \times 1$ vector of parameters. Then the posterior density is given by (1) where now x_i and θ are vectors. We will consider a family of multivariate prior distributions

$$\pi(\theta) = (1 - \varepsilon)\pi_0(\theta) + \varepsilon q(\theta), \tag{19}$$

where $0 < \varepsilon < 1$ and $q \in \mathscr{Q}$, where $\mathscr{Q}$ is a class of multivariate contaminating distributions. The analog of (2) is

$$\nabla \log p(\theta|x) = \nabla \log \pi(\theta) + \sum \nabla \log f(x_i|\theta) = 0, \tag{20}$$

where $\nabla h(t)$ is the gradient of h. Letting

$$\dot{\theta}_0^{\mathrm{T}} = \left[\frac{\partial \theta_1}{\partial \varepsilon}, \ldots, \frac{\partial \theta_k}{\partial \varepsilon}\right]_{\varepsilon=0} \tag{21}$$

we obtain the following generalization of (8)

$$\dot{\theta}_0^{\mathrm{T}} = (\pi_0 \boldsymbol{H} - (\nabla\pi_0)(\nabla\pi_0)^{\mathrm{T}} + \pi_0^2 \Psi_n')^{-1}(q\nabla\pi_0 - \pi_0\nabla q) \tag{22}$$

provided that the inverse exists. In this equation $\boldsymbol{H}$ is the Hessian matrix for π_0, given by

$$\boldsymbol{H} = \{h_{ij}\} = \left\{\frac{\partial^2\pi(\theta)}{\partial\theta_i\partial\theta_j}\right\}_{\theta=\widehat{\theta}_0} \tag{23}$$

Ψ_n' is the Hessian matrix for the log-likelihood function and $(\nabla\pi_0)$ and (∇q) are column vectors representing the gradients of π_0 and q evaluated at $\theta = \widehat{\theta}_0$.

We define the PMIF in the vector case as

$$\mathrm{PMIF}(\pi_0, q) = \Sigma_0^{-1/2}\dot{\theta}_0, \tag{24}$$

where Σ_0 is the posterior covariance matrix of θ under π_0.

4.2. Credibility region

Following the notation in (6) and (7), let $A = m(x|\pi_0)$, $B = m(x|q)$ and c_1 and c_2 be defined by the equations

$$\frac{\alpha}{2} = \int_{-\infty}^{c_1} p(t|x)\,\mathrm{d}t = \int_{c_2}^{\infty} p(t|x)\,\mathrm{d}t \tag{25}$$

where $0<\alpha<1/2$ and $p(t|x)$ is given by (4). It is not difficult to see that

$$\dot{c}_1 = \dot{c}_1(q) = \left(\frac{\mathrm{d}c_1(\varepsilon)}{\mathrm{d}\varepsilon}\right)_{\varepsilon=0} = \frac{\frac{\alpha}{2}B - A\int_{-\infty}^{c_1} q(t)f(x|t)\,\mathrm{d}t}{A\pi_0(c_1^*)f(x|c_1^*)} \tag{26}$$

where c_1^* is defined by the first equality in (25) with $\varepsilon = 0$. Analogously,

$$\dot{c}_2 = \dot{c}_2(q) = \frac{\frac{\alpha}{2}B + A\int_{c_2}^{\infty} q(t)f(x|t)\,\mathrm{d}t}{A\pi_0(c_2^*)f(x|c_2^*)} \tag{27}$$

where c_2^* is given by the second equality in (25) with $\varepsilon = 0$.

Clearly, the sensitivity of the length of the credibility region can be measured by $\dot{c}_2 - \dot{c}_1$. In the particular case of symmetry of the posterior density under π_0, we have that

$$\dot{c}_2 - \dot{c}_1 = \frac{A[\int_{-\infty}^{c_1^*} q(t)f(x|t)\,\mathrm{d}t + \int_{c_2^*}^{\infty} q(t)f(x|t)\,\mathrm{d}t] - B}{Ap(c_1^*)f(x|c_1^*)}.$$

Notice that in our approach the coverage probability is kept constant and we study the changes in the credibility intervals due to the contamination on the prior. The alternative approach of keeping the interval boundaries fixed and studying the changes in the coverage probabilities have been considered by several authors (see, for instance, De la Horra and Fernández (1994)). Both approaches are complementary. One can easily imagine situations where the coverage probability changes very little which the extremes of the credible intervals are greatly affected by changes an the prior distribution and vice versa.

5. Concluding remarks

Assessing robustness in Bayesian inference requires consideration to the prior and to the likelihood. In this paper we have presented a single diagnostic statistic, the posterior mode influence function PMIF, for studying the sensitivity of the estimation to local changes in the prior distribution. The statistic is very simple to compute, and provides a first step in the analysis of robustness. If the PMIF is large, the inference is not robust to the prior. On the other hand, if PMIF is small, further studies should be made to assess the sensitivity of other characteristics of interest in the posterior distribution to changes in the prior and/or the likelihood.

As shown in Section 4, these simple tools can be easily generalized to cover other more complicated situations, such as the vector parameter case and credibility interval (and regions). A more complete study of these problems will be the subject of further research.

Acknowledgements

This research has been partially supported by DGISYT under grant PB93-0232, Spain and NSERC, Canada. Thanks to the referee for his/her helpful comments.

References

Berger, J.O., 1984. The robust Bayesian view point. In: J. Kadene (Ed.), Robustness in Bayesian Analysis. North-Holland, Amsterdam.

Berger, J.O., 1990. Robust Bayesian analysis: sensitivity to the prior. J. Statist. Plann. Inference 25, 303–328.

Berger, J.O., 1994. An overview of robust Bayesian analysis (with discussion). Test, 3, 5–124.

Cuevas, A., Sanz, P., 1988. On differentiability properties of Bayes Operators. In: Bernardo, J.M. et al. (Eds.), Bayesian Statistics, vol. 3. Oxford University Press, Oxford, pp. 569–577.

Delampady, M., Dey, D., 1994. Bayesian Robustness for multiparameter problems. J. Statist. Plann. Inference 23(6), 2153–2167.

De la Horra, J., Fernández, C., 1994. Bayesian robustness of credible regions in the presence of nuisance parameters. Commun. Statist. 23(3), 689–699.

Gustafson, P., Srinivasan, C., Wasserman, L., 1995. Local sensitivity analysis. In: Berger et al. (Eds.), Bayesian Statistics, vol. 5. Oxford University Press, Oxford.

Gustafson, P., Wasserman, L., 1995. Local sensitivity diagnostics for Bayesian inference. Ann. Statist. 23(6), 2153–2167.

Moreno, E., Cano, J.A., 1991. Robust Bayesian analysis with e-contamination partially known. J. Roy. Statist. Soc. B 53, 143–155.

Moreno, E., Pericchi, L.R., 1993. Bayesian robustness for hierarchical e-contamination models. J. Statist. Plann. Inference 37, 159–167.

Peña, D., Zamar, R., 1995. On Bayesian robustness: an asymptotic approach. In: Rieder, H. (Ed.), Robust Statistics, Data Analysis and Computer Intensive Methods, In honor of Peter Huber. Springer Lecture Notes in Statistics, Springer, Berlin.

Pericchi, L.R., Walley, P., 1991. Robust Bayesian credible interval and prior ignorance. Int. Statist. Rev. 58, 1–23.

Savage, L.J. et al., 1962. The Foundations of Statistical Inference. Methuen, London.

Wasserman, L., 1992. Recent methodological advances in robust Bayesian inference. In: Bernardo, J.M. et al. (Eds.), Bayesian Statistics, vol. 4. Oxford University Press, Oxford, pp. 483–502.