

Working Paper 93-26
Statistics and Econometrics Series 20
October, 1993

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

**ON BAYESIAN ROBUSTNESS:
AN ASYMPTOTIC APPROACH**

Daniel Peña and Ruben H. Zamar*

Abstract

This paper presents a new asymptotic approach to study the robustness of Bayesian inference to changes on the prior distribution. We study the robustness of the posterior density score function when the uncertainty about the prior distribution has been restated as a problem of uncertainty about the model parametrization. Classical robustness tools, such as the influence function and the maximum bias function, are defined for uniparametric models and calculated for the location case. Possible extensions to other models are also briefly discussed.

Key Words

Gross error sensitivity, influence function, maximum bias curve, prior robustness.

*Peña, Departamento de Estadística y Econometría. Universidad Carlos III de Madrid; Zamar, Department of Statistics, UBC, Vancouver, Canadá. Research partially supported by DGICYT grant PB90-0266, Spain, and NSERC and ONR Contract N00014-84-84-C-0169.

On Bayesian Robustness: An Asymptotic Approach.

By

Daniel Peña

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid - Spain

and

Ruben H. Zamar

Department of Statistics - UBC
Vancouver - Canada

SUMMARY

This paper presents a new asymptotic approach to study the robustness of Bayesian inference to changes on the prior distribution. We study the robustness of the posterior density score function when the uncertainty about the prior distribution has been restated as a problem of uncertainty about the model parametrization. Classical robustness tools, such as the influence function and the maximum bias function, are defined for uniparametric models and calculated for the location case. Possible extensions to other models are also briefly discussed.

AMS 1980 subject classifications. Primary 62F15; secondary 62F35

Key Words: gross error sensitivity, influence function, maximum bias curve, prior robustness.

Research partially supported by DGICYT grant PB90-0266, Spain, and NSERC and ONR Contract N00014-84-C-0169

1. INTRODUCTION

The robustness of Bayesian analysis to possible misspecification of the prior distribution has been extensively studied. Berger (1984,1990) and Kadane (1984) give a detailed review of the subject. See also Walley (1990) and Pericchi and Walley (1991). The main idea in this approach is to consider a whole set of prior distributions \mathcal{P} instead of a single prior distribution Π alone. Then one focuses on a certain measure of interest (e.g. the posterior risk) and calculates its maximum and infimum as Π ranges over \mathcal{P} . A widely used class of prior distribution is the ϵ -contamination class given by

$$\mathcal{P} = \{ (1 - \epsilon)\Pi_0 + \epsilon Q, Q \in \mathcal{Q} \}, \quad (1)$$

where $0 < \epsilon < 1$ is given, and \mathcal{Q} depends on the given setup. This type of contamination family has been used by many researchers including Huber (1981), in the classical robustness field, and Box and Tiao (1968) in the Bayesian field. More recently Marazzi (1985), Berger and Berliner (1986), Walley (1990) and Moreno and Cano (1991), among others, used model (1) to represent contamination situations.

In general (see for instance Moreno and Cano, 1991) we have that the robustness of the Bayesian inference is analyzed by looking at the infimum and the supremum of

$$\rho(\Pi) = \frac{\int a(\theta)d\Pi(\theta)}{\int b(\theta)d\Pi(\theta)},$$

as Π ranges over \mathcal{P} . For instance, if $b(\theta) = f(x|\theta)$, the likelihood density, and $a(\theta) = \theta f(x|\theta)$, $\rho(\Pi)$ becomes the posterior mean. If $a(\theta) = L(\theta, d(x))f(x|\theta)$ and $b(\theta) = f(x|\theta)$, where $L(\theta, d(x))$ is a loss function for the decision rule $d(x)$, $\rho(\Pi)$ becomes the posterior risk. Finally, if $L(\theta, d(x)) = I_C(\theta)$ where $I_C(\theta)$ is the indicator function for the set C and $b(\theta)$ is as before, then $\rho(\Pi)$ becomes the posterior probability of the set C (a credible set). Although a formal definition of Bayesian robustness is not given, all the literature on this subject shows that the robustness of the Bayesian inference is highly dependent on the observed data.

On the other hand, there is also an extensive literature on classical robustness. A nice presentation of the main topics and ideas underlying this subject can be found in Hampel et al. (1986). The classical robustness theory basically relies on some asymptotic tools to measure the degree of robustness of a certain procedure, namely: the influence function, the gross-error-sensitivity and the maximum bias function. The asymptotic theory provides a simple setup to study the sensitivity of different statistical procedures to changes on the model's assumptions. One would expect, of course, that the findings from this "asymptotic lab" will be validated by Monte Carlo simulation and practical examples.

No comparable asymptotic developments have occurred, so far, in the Bayesian field. This may be due in part to the fact that a major component of Bayesian procedures, namely the prior distribution, does not seem to survive the limiting process.

In this paper we present an asymptotic approach to study the robustness properties of Bayesian inference under changes in the assumed prior distribution. In this approach, the uncertainty about the prior distribution is expressed as a “dual” uncertainty about the model parametrization.

The rest of the paper is organized as follows. In Section 2 we show that the usual approach of keeping the parameter fixed and varying the prior distribution leads to the conclusion that Bayesian inference can not be robust, by classical robustness measures. In Section 3 we derive an asymptotic representation for the changes on the Bayesian posterior density. This representation is obtained by transferring the uncertainty on the prior distribution to a “dual” uncertainty on the parametrization of the model. This is then used to define the Bayesian counterparts for the influence function, the gross-error-sensitivity and the maximum bias curve. This asymptotic approach allows the derivation of general conclusions independently from any given sample situation. In particular, one can state under which conditions and in which sense a given Bayesian procedure will be robust (robustness being measured now by standard robustness tools). In Section 4 we apply the general methodology of Section 3 to the simple location model. Finally, in Section 5, we briefly discuss some possible extensions to the multiple parameter case and to the study of changes on the likelihood part of the model.

2. ROBUSTNESS TO CHANGES ON THE PRIOR

Suppose that a statistician is interested in making inferences on a parameter θ . His prior assumptions are that θ follows a distribution Π_0 and that the sample X_1, X_2, \dots, X_n are i.i.d. random variables with common distribution $G(x|\theta)$. Suppose that $G(x|\theta)$ is differentiable and has a density $g(x|\theta)$. If Π_0 has also a density, p_0 , the posterior distribution for θ under this model is given by

$$\mu_n(\theta) = k_0 p_0(\theta) g(x_1|\theta) \dots g(x_n|\theta), \quad (2)$$

where $k_0 = (\int g(x_1|\theta) \dots g(x_n|\theta) p_0(\theta) d\theta)^{-1}$ does not depend on θ .

If the pair $[\Pi_0, G]$ represents the statistician’s ideal model only approximately (e.g. Π_0 might represent only a crude approximation for his actual prior beliefs), he may also be interested in knowing how changes on the prior density p_0 and/or the likelihood density g can affect the inferences on θ . That is, what would happen if he used the alternative model $[\Pi, F]$ – with prior distribution Π and likelihood distribution F – which is close to $[\Pi_0, G]$ (in some metric) and therefore might also approximately represent his prior beliefs.

The posterior distribution for θ under (Π, F) is given by

$$v_n(\theta) = k_1 p(\theta) f(x_1|\theta) \dots f(x_n|\theta), \quad (3)$$

where $p = \Pi'$, $f = F'$ and k_1 does not depend on θ .

A natural way to study the robustness of the Bayesian procedure is to consider a class \mathcal{C} of models (Π, F) and to study the changes on the posterior densities as (Π, F) ranges on \mathcal{C} .

It is well known that changes on the tails of a density $u(t)$ can be better visualized by looking at the score function,

$$\hat{u}(t) = \frac{d}{dt} \log u(t) = \frac{u'(t)}{u(t)},$$

rather than the density itself. For instance, a plot of the standard normal and Cauchy densities does not reveal the striking tails differences as well as a plot of the the corresponding score functions. Therefore, we will use the posterior score (the score function for the posterior density) to measure the sensitivity to changes on the priors. This approach has also the advantage that the proportionality constants k_0 and k_1 can be safely ignored.

Therefore, the changes on the posterior densities (when going from (2) to (3)) will be measured by the *finite sample discrepancy function*,

$$\Delta_n(\theta|\mathcal{D}) = \frac{1}{n} [\hat{r}_n(\theta) - \hat{u}_n(\theta)], \quad (4)$$

where

$$\mathcal{D} = \{(\Pi_0, G), (\Pi, F)\}. \quad (5)$$

To study the effect of changes on the prior distribution let us first assume that $F = G$ and Π is in \mathcal{P} given by (1), with fixed $0 < \epsilon < 1$ and Q ranging over a certain class \mathcal{Q} of distribution functions. In this case

$$\Delta_n(\theta|\Pi, \Pi_0) = \frac{1}{n} \left[\frac{p'_0(\theta)}{p_0(\theta)} - \frac{p'(\theta)}{p(\theta)} \right]. \quad (6)$$

We now give the following definitions which are the Bayesian counterparts for similar concepts defined in classical robustness.

Definition 1: The finite sample influence function with respect to changes on the prior distribution is defined as

$$\text{IF}_n(Q|\Pi_0, \theta) = \lim_{\epsilon \rightarrow 0} \frac{\Delta_n(\theta|(1-\epsilon)\Pi_0 + \epsilon Q, \Pi_0)}{\epsilon}, \quad (7)$$

provided that the limit exists.

Definition 2: The finite sample sensitivity to changes on the prior distribution is defined as

$$\text{ES}(\Pi_0, \theta) = \sup_{Q \in \mathcal{Q}} |\text{IF}_n(Q|\Pi_0, \theta)|. \quad (8)$$

where the class \mathcal{Q} must be specified in accordance with the given setup.

We have the following theorem.

Theorem 1: The finite sample influence function $\text{IF}_n(Q|\Pi_0, \theta)$ is given by

$$\text{IF}_n(Q|\Pi_0, \theta) = \frac{1}{n} \frac{q(\theta)}{p_0(\theta)} [\hat{p}_0(\theta) - \hat{q}(\theta)], \quad (9)$$

where $q = Q'$.

Proof: Since $\Pi(\theta) = \Pi_\epsilon(\theta) = (1 - \epsilon)\Pi_0(\theta) + \epsilon Q(\theta)$,

$$p(\theta) = p_\epsilon(\theta) = (1 - \epsilon)p_0(\theta) + \epsilon q(\theta),$$

and so,

$$\begin{aligned} \text{IF}_n(Q|\Pi_0, \theta) &= \frac{d}{d\epsilon} \Delta_n(\theta|(1 - \epsilon)\Pi_0 + \epsilon Q, \Pi_0)|_{\epsilon=0} = -\frac{1}{n} \left[\frac{q'(\theta) - p_0'(\theta)}{p_0(\theta)} - \frac{q(\theta) - p_0(\theta)}{p_0(\theta)} \frac{p_0'(\theta)}{p_0(\theta)} \right] \\ &= \frac{1}{n} \frac{q(\theta)}{p_0(\theta)} [\hat{p}_0(\theta) - \hat{q}(\theta)]. \end{aligned}$$

Definition 3: The Bayesian inference based on Π_0 is said to be robust to the prior if the finite sample influence function (7) is bounded.

Since $\text{IF}_n(Q|\Pi_0, \theta)$ depends on the derivative $q'(\theta)$ of $q(\theta)$, it will be typically unbounded for most “reasonable” families \mathcal{Q} . For instance, in the case of the location model, for any given θ it is easy to construct a density q which is symmetric, unimodal and continuous and for which $|q'(\theta)|$ is arbitrarily large.

If the unboundedness of the empirical influence function with respect to changes on the prior distribution, $\text{IF}_n(Q|\Pi_0, \theta)$, were to be taken as an indication of the lack of robustness of Π_0 and if $\text{IF}_n(Q|\Pi_0, \theta)$ is unbounded for all possible Π_0 , then one would logically conclude that robust Bayesian inference, in this sense, is impossible. It becomes apparent, then, that the comparison of the scores for the posterior densities (2) and (3) is not a convenient way to introduce the classical robustness ideas and tools into the Bayesian setup.

An alternative approach, which has the additional advantage of allowing the use of the asymptotic theory, is presented in the next section.

3. A NEW DEFINITION OF BAYESIAN ROBUSTNESS

Suppose that the two pairs $[\Pi_0(\theta), G(x|\theta)]$ and $[\Pi(\theta), F(x|\theta)]$ approximately represent the statistician’s prior beliefs. Notice that, in this formulation, all the uncertainty resides on the shape of the

prior and likelihood distributions while the parametrization for the model – e.g. that θ is the mean of the random variable X – is assumed “known”.

The uncertainty problem in the above paragraph can be formulated as follows. Let us introduce the transformation

$$\tau = T^{-1}(\theta) = \Pi_0^{-1}(\Pi(\theta)), \quad (10)$$

where, as usual, $\Pi_0^{-1}(u) = \inf\{t : \Pi_0(t) \geq u\}$. Using this transformation, the two pairs in the above paragraph can be written as $[\Pi_0(\theta), G(x|\theta)]$ and $[\Pi_0(\tau), F(x|T(\tau))]$. Observe that, as before, these two pairs represent the statistician prior uncertainty, but now the initial uncertainty on the shape of the prior distribution Π has been replaced by a “dual” uncertainty on the parametrization of the model. Consequently, the prior distribution is now kept fixed and equal to Π_0 for both models.

For instance, consider the following simple example. The first model has likelihood $g(x|\theta) = (1/\theta) \exp(-x/\theta)$ and exponential prior with mean $E_{\Pi_0}(\theta) = 1$, whereas the second model has the same likelihood (i.e. $f(x|\theta) = (1/\theta) \exp(-x/\theta)$) and exponential prior with mean $E_{\Pi}(\theta) = \log(2)$. In this case,

$$\tau = T^{-1}(\theta) = \theta \log(2),$$

and the second model has now prior $p(\tau) = \exp(-\tau)$ and likelihood $f(x|\tau) = (\tau/\log(2)) \exp(-x \log(2)/\tau)$. Note that in the reparametrized second model the parameter τ is the median of X and the prior for this parameter is identical to the prior for the mean in the first model. That is, going from one prior distribution to another while keeping the mean of X as the location parameter is “equivalent” to going from one location parameter to another – from the mean to the median of X , in this case – and keeping the same prior distribution.

One would expect, of course, that if Π is close to Π_0 (in some sense), then θ and $\tau = T^{-1}(\theta)$ will also be close, and the posterior distribution for θ , given by (2), will be close to the posterior distribution for τ , given by

$$d_n(\tau) = k_1 p_0(\tau) f(x_1|T(\tau)) \dots f(x_n|T(\tau)). \quad (11)$$

In such case, the statistician inferences would be approximately the same for the intended parameter θ and for a parameter τ which is close to the intended one.

The corresponding score function $\hat{d}_n(\tau)$ (see (11)) is given by

$$\begin{aligned} \hat{d}_n(\tau) &= \hat{p}_0(\tau) + \sum \frac{\partial}{\partial T(\tau)} \log f(x|T(\tau)) \frac{d}{d\tau} T(\tau) \\ &= \hat{p}_0(\tau) + \sum \hat{f}(x_i|T(\tau)) \frac{p_0(\tau)}{p(T(\tau))}. \end{aligned} \quad (12)$$

Observe that if, for example, Π_0 is a $N(\mu, \sigma^2)$ and Π is a mixture $(1 - \epsilon)N(\mu, \sigma^2) + \epsilon N(\mu, k\sigma^2)$ (with $k \leq 1$), then

$$\tau = T^{-1}(\theta) = \sigma \Phi^{-1} \left[(1 - \epsilon) \Phi \left(\frac{\theta - \mu}{\sigma} \right) + \epsilon \Phi \left(\frac{\theta - \mu}{k\sigma} \right) \right] + \mu,$$

and so, if $\theta = \mu + \delta$, $\delta > 0$, then

$$\mu + \frac{\delta}{k} \leq \tau \leq \mu + \delta.$$

When $k = 1$ then $\theta = \tau$, otherwise the new parameter τ is a shrinkage of θ over the prior mean μ . The amount of shrinkage depends on the “contaminating” distribution.

We will compare the asymptotic posterior scores before and after the transformation T^{-1} is applied. In the sequel, we will use θ as a dummy argument for all the posterior score functions involved. More precisely, we will study the asymptotic value, $\Delta(\theta)$, of the discrepancy measure

$$\Delta_n(\theta) = \frac{1}{n} \left[\hat{d}_n(\theta) - \hat{\mu}_n(\theta) \right] = \frac{1}{n} \sum_1^n \left[\frac{p_0(\theta)}{p(T(\theta))} \hat{f}(x_i | T(\theta)) - \hat{g}(x_i | \theta) \right],$$

where \hat{d}_n is given by (12) and $\hat{g}(x | \theta) = (\partial / \partial t) \log g(x | t) |_{t=\theta}$.

Suppose now that the “true” mechanism that generates the sequence X_1, X_2, \dots is the following: given θ_0 , the random variables X_1, X_2, \dots are i.i.d. with common distribution $H_{\theta_0}(x)$. By the Strong Law of Large Numbers, for each θ ,

$$\Delta(\theta | \mathcal{D}) = E_{H_{\theta_0}} \left\{ \frac{p_0(\theta)}{p(T(\theta))} \hat{f}(x | T(\theta)) - \hat{g}(x | \theta) \right\}, \quad (13)$$

where

$$\mathcal{D} = \{(\Pi_0, G), (\Pi, F, H_{\theta_0})\}.$$

As before we will first assume that the likelihood part of the model has been correctly specified, that is, $G(x | \theta) = F(x | \theta)$, and for simplicity, we will take $H_{\theta_0}(x) = F(x | \theta_0)$. The prior distribution Π belongs to \mathcal{P} given by (1), with fixed $0 < \epsilon < 0.5$ and “contaminating” distribution Q ranging over \mathcal{Q} . This last set must be specified in accordance with the situation at hand.

To reflect the dependency of Δ on Π and Π_0 we will use the notation $\Delta(\theta | \Pi, \Pi_0)$. Then we introduce the following definitions.

Definition 4: The maximum posterior score change with respect to changes on the prior distribution, $MC(\Pi_0, \theta)$, is defined as

$$MC(\Pi_0, \theta) = \sup_{\Pi \in \mathcal{P}} |\Delta(\theta | \Pi, \Pi_0)|. \quad (14)$$

Observe that $MC(\Pi_0, \theta)$ is a global robustness measure, similar in nature to the maximum bias curve used in classical robustness.

Definition 5: The influence function with respect to changes on the prior distribution is defined as

$$\text{IF}(Q|\Pi_0, \theta) = \lim_{\epsilon \rightarrow 0} \frac{\Delta(\theta|(1-\epsilon)\Pi_0 + \epsilon Q, \Pi_0)}{\epsilon}, \quad (15)$$

provided that the limit exists.

Observe that $\text{IF}(Q|\Pi_0, \theta)$ is the directional derivative of $\Delta(\theta|\Pi, \Pi_0)$ in the direction of Q , that is,

$$\text{IF}(Q|\Pi_0, \theta) = \frac{d}{d\epsilon} \Delta(\theta|(1-\epsilon)\Pi_0 + \epsilon Q, \Pi_0) |_{\epsilon=0} \quad (16)$$

Definition 6: The sensitivity to changes on the prior distribution is defined as

$$S(\Pi_0, \theta) = \sup_{Q \in \mathcal{F}} |\text{IF}(Q|\Pi_0, \theta)|. \quad (17)$$

Now we can formally define b-robustness to changes on the prior, where as in Hampel et al. (1986) “b” stands for “bias” (to differentiate this concept from that of variance-robustness).

Definition 7: The Bayesian inference based on Π_0 is said to be b-robust with respect to the prior if the influence function (15) is bounded (or equivalently, if $S(\Pi_0, \theta)$ is finite).

The IF and the S are infinitesimal concepts, similar in nature to Hampel’s influence function and gross-error-sensitivity. They give the infinitesimal rate of change of $\Delta(\theta|\Pi, \Pi_0)$ when the assumed model (Π_0, F) is perturbed in the direction of $((1-\epsilon)\Pi_0 + \epsilon Q, F)$ (in the case of the IF) and in the worst possible direction (in the case of S).

One would expect that, paralleling results in classical robustness, under regularity conditions

$$S(\Pi_0, \theta) = \frac{\partial}{\partial \epsilon} MC(\epsilon|\Pi_0, \theta) |_{\epsilon=0}.$$

In such case, likewise in the classical setup, the left hand side above would constitute an “easier” way to compute the right hand side.

Let

$$A(\theta) = - \int_{-\infty}^{\infty} \hat{f}(x|\theta) f(x|\theta_0) dx, \quad (18)$$

and then we can write

$$\Delta(\theta|\Pi, \Pi_0) = A(\theta) - \frac{p_0(\theta)}{p(T(\theta))} A(T(\theta)). \quad (19)$$

Moreover, let

$$\hat{f}'_0(x|\theta) = \frac{d}{dt} \hat{f}_0(x|t) |_{t=\theta} \quad \text{and} \quad B(\theta) = A'(\theta) = \int_{-\infty}^{\infty} \hat{f}'(x|\theta) f(x|\theta_0) dx, \quad (20)$$

where the last equality holds only if differentiation under the integral sign is allowed. Then we have the following result.

Theorem 2: Under regularity assumptions,

$$\text{IF}(Q|\Pi_0, \theta) = \frac{q(\theta) - p_0(\theta)}{p_0(\theta)} + \frac{Q(\theta) - \Pi_0(\theta)}{p_0(\theta)} [B(\theta) - \hat{p}_0(\theta)A(\theta)]. \quad (21)$$

Proof: Let $h(\epsilon) = T(\theta) = [(1 - \epsilon)\Pi_0 + \epsilon Q]^{-1} \Pi_0(\theta)$. Then,

$$\Pi_0(\theta) = (1 - \epsilon)\Pi_0(h(\epsilon)) + \epsilon Q(h(\epsilon)).$$

Differentiating both sides with respect to ϵ at $\epsilon = 0$, we obtain,

$$0 = -\Pi_0'(h(0)) + p_0(h(0))h'(0) + Q(h(0)),$$

Since $h(0) = \theta$ it follows that

$$\frac{d}{d\epsilon} T(\theta) |_{\epsilon=0} = h'(0) = \frac{\Pi_0(\theta) - Q(\theta)}{p_0(\theta)}.$$

In addition,

$$\begin{aligned} \frac{d}{d\epsilon} \left(\frac{p_0(\theta)}{p(T(\theta))} \right) |_{\epsilon=0} &= -\frac{p_0(\theta)}{p^2(T(\theta))} \left[-p_0'(T(\theta)) + (1 - \epsilon)p_0'(T(\theta)) \frac{d}{d\epsilon} (T(\theta)) |_{\epsilon=0} \right. \\ &\quad \left. + q(T(\theta)) + \epsilon q'(T(\theta)) \frac{d}{d\epsilon} (T(\theta)) |_{\epsilon=0} \right] |_{\epsilon=0} \\ &= -\frac{1}{p_0(\theta)} \left[q(T(\theta)) - p_0(\theta) + p_0'(\theta) \frac{\Pi_0(\theta) - Q(\theta)}{p_0(\theta)} \right] \\ &= \frac{p_0(\theta) - q(\theta)}{p_0(\theta)} - \hat{p}_0(\theta) \frac{\Pi_0(\theta) - Q(\theta)}{p_0(\theta)}. \end{aligned}$$

Therefore, from (19)

$$\begin{aligned} \text{IF}(Q|\Pi_0, \theta) &= \frac{d}{d\epsilon} \Delta(\theta|(1 - \epsilon)\Pi_0 + \epsilon Q, \Pi_0) |_{\epsilon=0} = -\frac{d}{d\epsilon} \left(\frac{p_0(\theta)}{p(T(\theta))} A(T(\theta)) \right) |_{\epsilon=0} \\ &= -\frac{d}{d\epsilon} \left(\frac{p_0(\theta)}{p(T(\theta))} \right) |_{\epsilon=0} A(\theta) - \frac{d}{d\epsilon} (A(T(\theta))) |_{\epsilon=0} \\ &= \left[\frac{q(\theta) - p_0(\theta)}{p_0(\theta)} - \hat{p}_0(\theta) \frac{Q(\theta) - \Pi_0(\theta)}{p_0(\theta)} \right] A(\theta) + \frac{Q(\theta) - \Pi_0(\theta)}{p_0(\theta)} B(\theta). \end{aligned}$$

Note that this theorem shows that the influence function (21) is very large if $p_0(\theta)$ is negligibly small when compared with $q(\theta)$. In particular, it is equal to infinity when $p_0(\theta) = 0$, and therefore, a

p_0 with compact support is automatically non-robust by this criterion. On the other hand, if $p_0(\theta) > 0$ for all θ , $IF(Q|\Pi_0, \theta)$ is bounded (for all Q !) and so, boundedness of (21) does not seem to be a good criterion to define the robustness of p_0 . This is not necessarily surprising in view of analogous results for regression estimates in the classical robustness setup, where boundedness of the influence function does not necessarily imply robustness of the regression estimate (see for instance Martin et al. (1989)).

Given this limitation, we suggest to use the sensitivity (17) – which is given by the supremum of (21) when Q ranges over a certain class to determine the robustness merits of two prior distributions over a given interval of values of θ ; such interval will be normally decided in accordance with the objectives at hand. In the next section we will compute the influence function for the location model and compare the robustness properties of the normal, double exponential and Cauchy priors, over different parameter ranges.

Finally, the somewhat concurrent results of Sections 2 and 3 can be interpreted as follows: while robust-to-the-prior Bayesian inference for a specific parameter cannot be achieved (by the criterion defined in Section 2), robust-to-the-prior Bayesian inference for a class of parameters is possible (by the criterion defined in Section 3).

4. THE LOCATION MODEL

Let us assume that θ is a location parameter and $F(x|\theta) = F_0(x - \theta)$. We make the following assumptions on Π_0 and F_0 : (i) Π_0 has a positive and unimodal density $p_0(\theta)$ which is symmetric about a central value θ_0 ; (ii) F_0 has a positive and unimodal density f_0 which is symmetric about zero; (iii) $\psi_0(x) = -f'_0(x)/f_0(x)$ is well defined almost everywhere (set $\psi_0(x) = 0$ where it is not defined); and (iv) $B(\theta) = A'(\theta) \geq 0$, $\forall \theta \geq 0$. Observe that ψ_0 is odd and $\psi_0(x) \geq 0$ for $x \geq 0$. Our choice for Q is given by

$$Q = \{ Q : Q(\theta_0) = 1/2, \text{ and } q(\theta) \text{ is unimodal and symmetric about } \theta_0 \}. \quad (22)$$

To simplify the notations we will take $\theta_0 = 0$.

Remark: By symmetry of Π_0 and Q , $T(-\theta) = -T(\theta)$, for all θ , and $T(\theta) \geq 0$, for $\theta \geq 0$. Since in addition f_0 is symmetric, $A(-\theta) = -A(\theta)$, for all θ , and $A(\theta) \geq 0$ for $\theta \geq 0$. Consequently,

$$\begin{aligned} \Delta(-\theta|\Pi, \Pi_0) &= A(-\theta) - \frac{p_0(-\theta)}{p(T(-\theta))} A(T(-\theta)) \\ &= -A(\theta) + \frac{p_0(\theta)}{p(T(\theta))} A(T(\theta)) \\ &= -\Delta(\theta|\Pi, \Pi_0), \end{aligned}$$

and we can concentrate on the case $\theta \geq 0$.

Notice that by the assumptions on f_0 , $A(t) = -\int_{-\infty}^{\infty} \psi_0(x - \theta) f_0(x) dx > 0$ for all $\theta > 0$. In addition, if $\psi'_0(x) \geq 0$ for all x and differentiation can be performed under the integral sign, then $B(\theta) = A'(\theta) = \int_{-\infty}^{\infty} \psi'_0(x - \theta) f_0(x) dx \geq 0$ for all θ .

If

$$C(\theta) = B(\theta) - \hat{p}_0(\theta)A(\theta),$$

then

$$IF(Q|\Pi_0, \theta) = \frac{q(\theta) - p_0(\theta)}{p_0(\theta)} A(\theta) + \frac{Q(\theta) - \Pi_0(\theta)}{p_0(\theta)} C(\theta).$$

The functions h_1 and h_2 defined below are needed for the statement and the proof of Theorem 3. For any fixed $\theta > 0$ let

$$\begin{aligned} h_1(u, v) &= \left| \frac{u - p_0(\theta)}{p_0(\theta)} A(\theta) + \frac{v - \Pi_0(\theta)}{p_0(\theta)} C(\theta) \right|, \text{ and} \\ h_2(\alpha, q) &= \frac{p_0(\theta)}{(1 - \epsilon)p_0(\alpha) + \epsilon q(\alpha)} A(\alpha). \end{aligned}$$

Observe that $h_1(q(\theta), Q(\theta)) = |IF(\Pi|\Pi_0, \theta)|$ and $A(\theta) - h_2(T(\theta), q) = \Delta(\theta|Q, \Pi_0)$.

Theorem 3: Suppose that $B(\theta) > 0$. Then,

$$\underline{T}(\theta) = \inf_{Q \in \mathcal{Q}} T(\theta) = \Pi_0^{-1} \left(\frac{\Pi_0(\theta) - \epsilon}{1 - \epsilon} \right), \quad (23)$$

and, if $\Pi_0(\theta) - .5\epsilon < 1 - \epsilon$ (the “regular” case)

$$\bar{T}(\theta) = \sup_{Q \in \mathcal{Q}} T(\theta) = \Pi_0^{-1} \left(\frac{\Pi_0(\theta) - (\epsilon/2)}{1 - \epsilon} \right), \quad (24)$$

and $\bar{T}(\theta) = \infty$, otherwise. Furthermore, in the regular case,

$$MC(\Pi_0, \theta) = \max \left\{ \frac{A(\bar{T}(\theta))}{(1 - \epsilon)p_0(\bar{T}(\theta))}, \frac{A(\underline{T}(\theta))}{(1 - \epsilon)p_0(\underline{T}(\theta)) + [\epsilon/(2\underline{T}(\theta))]} \right\}, \quad (25)$$

and $MC(\Pi_0, \theta) = \infty$, otherwise. Moreover,

$$S(\Pi_0, \theta) = \max \left\{ h_1(0, .5), h_1(0, 1), h_1\left(\frac{1}{2\theta}, 1\right) \right\}. \quad (26)$$

Proof: Let $\theta > 0$. For any fixed $q(\theta) = y \geq 0$,

$$|\text{IF}(Q|\Pi_0, \theta)| = h_1(y, Q(\theta))$$

is the absolute value of a linear function of $Q(\theta)$. Since $Q(\theta)$ takes all the values on $[\cdot 5 + y\theta, 1]$ as Q ranges over

$$\mathcal{Q}_y = \{ Q \in \mathcal{Q}, q(\theta) = y \},$$

it follows that

$$\sup_{Q \in \mathcal{Q}_y} |\text{IF}(Q|\Pi_0, \theta)| = \max\{ h_1(y, .5 + \theta y), h_1(y, 1) \}.$$

Equation (26) holds now because $h_1(y, .5 + \theta y)$ and $h_1(y, 1)$ are the absolute value of linear functions of y which reach their maximum and minimum over $[0, 1/2\theta]$ at either $y = 0$ or $y = 1/(2\theta)$.

To prove (25) we first notice that $|\Delta(\theta|Q, \Pi_0)|$ is maximized (in q) when

$$h_2(T(\theta), q) = \frac{p_0(\theta)}{(1 - \epsilon)p_0(T(\theta)) + \epsilon q(T(\theta))} A(T(\theta))$$

takes either its maximum or its minimum value. Second, since by assumption $B(\theta) \geq 0$, one follows that

$$\frac{\partial}{\partial \alpha} h_2(\alpha, q) \geq 0.$$

So, maximizing or minimizing $h_2(T(\theta), q)$ (for a fixed q) is equivalent to maximizing or minimizing $T(\theta)$. Now, the minimum of $T(\theta)$ (as Q ranges over \mathcal{Q}) is given by the smallest value of t for which the equation

$$(1 - \epsilon)\Pi_0(t) + \epsilon Q(t) = \Pi_0(\theta) \quad (27)$$

holds, for some Q . That is,

$$\underline{T}(\theta) = \inf_{Q \in \mathcal{Q}} \{t : (1 - \epsilon)\Pi_0(t) + \epsilon Q(t) = \Pi_0(\theta)\}.$$

Since,

$$(1 - \epsilon)\Pi_0(t) + \epsilon Q(t) \leq (1 - \epsilon)\Pi_0(t) + \epsilon, \quad \forall Q, \forall t,$$

then

$$\underline{T}(\theta) \geq \inf_{Q \in \mathcal{Q}} \{t : (1 - \epsilon)\Pi_0(t) + \epsilon = \Pi_0(\theta)\} = \Pi_0^{-1} \left(\frac{\Pi_0(\theta) - \epsilon}{1 - \epsilon} \right).$$

Equation (23) follows now because (27) is satisfied with

$$t = \Pi_0^{-1} \left(\frac{\Pi_0(\theta) - \epsilon}{1 - \epsilon} \right),$$

and Q uniform on $[-\underline{T}(\theta), \underline{T}(\theta)]$. Analogously, one notices that the maximum value of $T(\theta)$ is given by the largest value of t for which equation (27) holds, for some Q . Since

$$(1 - \epsilon)\Pi_0(t) + \epsilon Q(t) \geq (1 - \epsilon)\Pi_0(t) + \epsilon.5, \quad \forall Q, \forall t,$$

we have

$$\bar{T}(\theta) \leq \Pi_0^{-1} \left(\frac{\Pi_0(\theta) - .5\epsilon}{1 - \epsilon} \right),$$

provided that $\Pi_0(\theta) - .5\epsilon \leq 1 - \epsilon$. Otherwise, it clearly follows that $\bar{T}(\theta) = \infty$. Equation (24) follows now because (27) is satisfied with

$$t = \Pi_0^{-1} \left(\frac{\Pi_0(\theta) - .5\epsilon}{1 - \epsilon} \right),$$

and Q uniform on $[-K, K]$, with $K = \infty$. Finally, $q(\theta) = .5\theta$ and $q(\theta) = 0$ are the largest and smallest possible values for $q(\theta)$, respectively, and therefore (25) follows.

Table 1 gives the values of $S(\Pi_0, \theta)$ when Π_0 is the standard normal, double exponential and Cauchy distributions, for several values of θ . In general, one notices that $S(\Pi_0, \theta)$ is relatively small in places where Π_0 concentrates a relatively large mass. For instance, the double exponential has the smallest sensitivity for $0 \leq \theta \leq 0.25$ and $0.82 \leq \theta \leq 6.42$. The standard normal has the smallest sensitivity (but fairly close to the double exponential) in the range $0.26 \leq \theta \leq 0.81$. Finally the Cauchy dominates the other two when $\theta \geq 6.43$.

The conclusion emerging from these comparisons is the following: other things equal, the double exponential should be preferred over the Normal and Cauchy distributions to achieve a good level of b-robustness.

5. SOME POSSIBLE EXTENSIONS AND CONCLUDING REMARKS

This paper concentrates on the univariate case, and particularly on the location model. However, the procedures presented in Section 2 can be easily applied to other single parameter situations such as the Binomial, Poisson and exponential models, etc.

The extension to the multiparameter case is also straightforward for the case of independent priors. For general priors, we can define the multivariate analogue of the transformation (10) as follows: let $\theta = (\theta_1, \dots, \theta_k)$ be the vector of parameters and Π_0 and Π be two possible multivariate priors on θ with joint densities p_0 and p , respectively. Let us decompose p_0 as

$$p_0(\theta_1, \dots, \theta_k) = p_{01}(\theta_1) p_{02}(\theta_2|\theta_1) \dots p_{0k}(\theta_k|\theta_{k-1}, \dots, \theta_1) \quad (28)$$

and analogously for p . Let us call Π_{0i} and Π_i , ($i = 1, \dots, k$), the corresponding cumulative distribution functions. Then, it can be shown using for instance the Roseblatt's transformation (see Mardia et al. (1979), pp. 36-37) that the new vector of parameters $\tau = (\tau_1, \dots, \tau_k)$ defined as

$$\tau_i = \Pi_{0i}^{-1}(\Pi_i(\theta)), \quad i = 1, \dots, k, \quad (29)$$

has joint distribution Π_0 when θ has joint distribution Π . This result allows the straightforward generalization of our procedures to study the robustness properties of multivariate priors.

The discrepancy measure (13) can be easily differentiated with respect to the likelihood part of the Bayesian model by setting

$$f(x|\theta) = (1 - \delta)g(x|\theta) + \delta l(x|\theta).$$

In fact, it is not difficult to verify that in this case,

$$\frac{d}{d\delta} \Delta(\theta|\mathcal{T})|_{\delta=0} = \frac{l(x|\theta)}{g(x|\theta)} [\hat{l}(x|\theta) - \hat{g}(x|\theta)].$$

This result can be used together with (16) to study the joint sensitivity of the Bayesian inference to changes on both, the prior and likelihood parts of the model.

The asymptotic nature of the present approach carries with it the disadvantage of being an approximation to any given real situation, and its finite sample relevance should be further investigated. On the other hand, it has the advantage of being relatively simple and general, in the sense of allowing the comparison of different Bayesian models independently of the particular sampling situation.

In summary, the present approach seems to offer a way to study the robustness properties of Bayesian procedures. It also allows the use of many tools, initially developed in the classical robustness setup, to the Bayesian field. One would hope, of course, that this connection will lay the ground for further cross-fertilizations between these two important areas of statistical research.

REFERENCES

- Berger, J.O. (1984). The robust Bayesian viewpoint. *In Robustness of Bayesian Analysis* (J. Kadane Ed.). North Holland, Amsterdam.
- Berger, J.O. (1990). Robust Bayesian analysis: sensitivity to the prior. *Journal of Statistical Planning and Inference*, **25**, 303-328.
- Berger, J.O. and Berliner, L.M. (1986). Robust Bayes and empirical Bayes analysis with ϵ -contaminated priors. *Ann. of Statist.* **14** 461-486.
- Box, G.E.P. and Tiao, G.C. (1968). A Bayesian approach to some outlier problems. *Biometrika*, **75** , 651-659.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics*. J. Wiley and Sons. New York.
- Huber, P.J. (1981). *Robust Statistics*. Wiley, N. York.
- Kadane, J.B. (1984). *Robustness of Bayesian Analysis*. North Holland, Amsterdam.
- Marazzi, A. (1985). On constrained minimization of the Bayes risk for the linear model. *Statistics and Decision*, **3**, 277-296.
- Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979). *Multivariate Analysis*. Academic Press. New York.
- Martin, R.D., Yohai, V.J. and Zamar, R.H. (1989). Min-max bias robust regression. *Ann. of Statistics* **17**, 1608- 1630.
- Moreno, E. and Cano, J.A. (1991). Robust Bayesian analysis with ϵ -contamination partially known. *Journal of the Royal Statistical Society B*, **53**, 143-155.
- Pericchi, L.R. and Walley, P. (1991). Robust Bayesian credible interval and prior ignorance. *International Statistical Review*. **58**. 1-23.
- Walley, P. (1990). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall. London.

θ	Normal	Double Exponential	Cauchy
0.00	2.51	2.00	3.14
0.25	2.15	2.13	2.99
0.50	2.02	2.28	3.40
0.75	2.17	2.45	4.20
1.00	3.82	2.65	5.28
1.50	12.37	4.29	9.11
2.00	46.20	7.15	16.39
2.50	206.74	11.22	26.01
3.00	1128.14	16.93	37.97
4.00	63512.33	35.95	68.96
5.00	8286946.00	71.61	108.93
6.00	∞	138.70	158.67
7.00	∞	262.28	217.84

Table 1