

Optimal Robustness Results for Relative Belief Inferences and the Relationship to Prior-Data Conflict*

Luai Al Labadi[†] and Michael Evans[‡]

Abstract. The robustness to the prior of Bayesian inference procedures based on a measure of statistical evidence is considered. These inferences are shown to have optimal properties with respect to robustness. Furthermore, a connection between robustness and prior-data conflict is established. In particular, the inferences are shown to be effectively robust when the choice of prior does not lead to prior-data conflict. When there is prior-data conflict, however, robustness may fail to hold.

MSC 2010 subject classifications: primary 62F15; secondary 62F35.

Keywords: relative belief, robustness, prior-data conflict.

1 Introduction

Robustness to the choice of the prior is an issue of considerable importance in a Bayesian statistical analysis. If an inference is very sensitive to the choice of the prior, then this could be viewed as either a negative for the inference method being used or for the choice of prior. In this paper it is shown that certain inferences are in a sense optimally robust to the choice of the prior. Furthermore, when the sensitivity of the inferences to the prior is measured quantitatively, it is shown that there is an intimate connection between the effective robustness of the inferences and whether or not there is prior-data conflict. So by choice of the inferential methodology and the avoidance of prior-data conflict, robustness of the inferences to the choice of prior is achieved. It is to be noted that, while the results derived here are for specific Bayesian inferences, the optimality of these inferences with respect to robustness implies that the effect of prior-data conflict applies to all Bayesian inferences.

The basic ingredients for a statistical analysis are taken here to be the data x , a statistical model $\{f_\theta : \theta \in \Theta\}$, where each f_θ is a probability density with respect to volume measure μ on the sample space \mathcal{X} , and a proper prior density π with respect to volume measure ν on Θ . Note that volume measure on a discrete set is taken to be counting measure. Furthermore, suppose that interest is in making inferences about the quantity $\psi = \Psi(\theta)$ where $\Psi : \Theta \rightarrow \Psi$ is onto and we don't distinguish between the function and its range to save notation. Throughout the paper the spaces Θ and Ψ

*The authors acknowledge the support of the Natural Sciences and Engineering Research Council of Canada for this research. Also the authors thank the reviewers and an Associate Editor for a number of constructive comments.

[†]Department of Statistical Sciences, University of Toronto, Toronto, ON M5S 3G3, Canada, luai.allabadi@gmail.com

[‡]Department of Statistical Sciences, University of Toronto, Toronto, ON M5S 3G3, Canada

can be thought of as open subsets of Euclidean spaces or more generally as Riemann manifolds with $\dim(\Theta) \geq \dim(\Psi)$.

Let $\pi_\Psi(\cdot|x)$ and π_Ψ denote the posterior and prior densities of ψ where these are both taken with respect to volume measure ν_Ψ on Ψ . It follows that when Ψ is smooth, and ψ is not in the set of volume measure zero given by the image of the critical values of Ψ , then $\pi_\Psi(\psi) = \int_{\Psi^{-1}\{\psi\}} \pi(\theta) J_\Psi(\theta) \nu_{\Psi^{-1}\{\psi\}}(d\theta)$ where $J_\Psi(\theta) = (\det(d\Psi(\theta)) \circ (d\Psi(\theta))^t)^{-1/2}$, $d\Psi$ is the differential of Ψ and $\nu_{\Psi^{-1}\{\psi\}}$ is volume measure on $\Psi^{-1}\{\psi\}$. Also, $\pi_\Psi(\psi|x) = \int_{\Psi^{-1}\{\psi\}} \pi(\theta|x) J_\Psi(\theta) \nu_{\Psi^{-1}\{\psi\}}(d\theta)$ where $\pi(\theta|x) = \pi(\theta) f_\theta(x) / m(x)$, with $m(x) = \int_{\Theta} \pi(\theta) f_\theta(x) \nu(d\theta)$, is the posterior density of θ with respect to ν . Note that m is the prior predictive density of the data with respect to μ . The conditional prior of θ given $\psi = \Psi(\theta)$ has density $\pi(\theta|\psi) = \pi(\theta) J_\Psi(\theta) / \pi_\Psi(\psi)$ with respect to $\nu_{\Psi^{-1}\{\psi\}}$ on the set $\Psi^{-1}\{\psi\}$. The conditional prior predictive density of x is then given by $m(x|\psi) = \int_{\Psi^{-1}\{\psi\}} \pi(\theta|\psi) f_\theta(x) \nu_{\Psi^{-1}\{\psi\}}(d\theta)$. A simple argument, see Baskurt and Evans (2013), gives the Savage–Dickey ratio result that

$$\frac{\pi_\Psi(\psi|x)}{\pi_\Psi(\psi)} = \frac{m(x|\psi)}{m(x)}, \quad (1)$$

which has some use in the developments here. Note that there is no need to coordinatize nuisance parameters as $\lambda = \Lambda(\theta)$, since all nuisance parameters are integrated out via $\pi(\cdot|\psi)$ and, in general, there need not exist a complementing function Λ such that (Ψ, Λ) is 1–1.

The approach taken here is to study robustness to the prior for relative belief inferences for ψ rather than all possible inferences. Relative belief inferences are based on the relative belief ratio defined as

$$RB_\Psi(\psi|x) = \lim_{\delta \rightarrow 0} \frac{\Pi_\Psi(N_\delta(\psi)|x)}{\Pi_\Psi(N_\delta(\psi))} \quad (2)$$

when this limit exists for a sequence of neighborhoods $N_\delta(\psi)$ of ψ converging nicely to ψ (see Rudin (1974)) for the definition of “converging nicely”. Whenever there exist versions of the densities π_Ψ and $\pi_\Psi(\cdot|x)$, of the probability measures Π_Ψ and $\Pi_\Psi(\cdot|x)$ respectively, taken with respect to ν_Ψ , that are continuous at ψ with $\pi_\Psi(\psi) > 0$, then (2) exists and is given by $RB_\Psi(\psi|x) = \pi_\Psi(\psi|x) / \pi_\Psi(\psi)$. A similar statement for the densities of the probability measures M and $M(\cdot|\psi)$ with respect to μ establishes that the relative belief ratio of x , having observed ψ , is given by $m(x|\psi) / m(x)$ and, when both conditions hold, then (1) obtains. Since $RB_\Psi(\psi|x)$ measures the change in belief that ψ is the true value, it is a measure of evidence. Here $RB_\Psi(\psi|x) > 1$ means that there is evidence in favor of ψ being the true value, as belief in ψ has increased after seeing the data, and $RB_\Psi(\psi|x) < 1$ means that there is evidence against ψ being the true value, as belief in ψ has decreased after seeing the data. Section 2 provides some more details concerning relative belief inferences for both estimation and hypothesis assessment but also see Baskurt and Evans (2013) and Evans (2015). The relevant mathematical details for the formulas provided for π_Ψ and $\pi_\Psi(\cdot|x)$ can be found in texts that deal with geometric measure theory such as Federer (1969), Hirsch (1976), Krantz and Parks (2008) and Tjur (1974).

Results in Section 3 establish that relative belief inferences have optimal robustness properties when the marginal prior for ψ is allowed to vary over all possibilities in the class of ϵ -contaminated priors. This generalizes results found in Wasserman (1989), Ruggeri and Wasserman (1993) and de la Horra and Fernandez (1994). Furthermore, an ambiguity concerning the interpretation of the results is resolved. As such this provides further justifications for these inferences.

While inferences may be optimally robust, this does not imply that they are in fact robust. In Section 4 quantitative measures of the sensitivity of relative belief inferences to both the marginal prior of ψ and the conditional prior for θ given $\Psi(\theta) = \psi$ are derived. In Section 5 it is shown that these inferences are indeed robust when the base prior π does not suffer from prior-data conflict. This adds weight to arguments concerning the importance of checking for prior-data conflict before reporting inferences, as prior-data conflict can imply sensitivity of the inferences to the choice of the prior. Prior-data conflict is interpreted as the true value lying in the tails of the prior and consistent methods have been developed for assessing this in Evans and Moshonov (2006) and Evans and Jang (2011b). Methodology for modifying a prior when prior-data conflict is encountered, through the selection of a prior weakly informative with respect to the base prior, is developed in Evans and Jang (2011c). Other approaches to identifying and dealing with prior-data conflict can be found in O'Hagan (2003), Marshall and Spiegelhalter (2007), Dahl et al. (2007), Scheel et al. (2011), Presanis et al. (2013).

Robustness to the prior has been considered by many authors and there are a number of different approaches. Many discussions are concerned with determining the range of values that some characteristic of interest takes when the prior is allowed to vary over some class. Berger (1990), Berger (1994) contain broad reviews of work on this topic and Rios Insua and Ruggeri (2000) is a collection of papers by key contributors. Sivaganesan et al. (1993) develop γ -credible regions for θ that have posterior content at least γ for every prior in an ϵ -contaminated class and have smallest Euclidean volume amongst all such regions and so are optimal. By contrast, as developed in Evans et al. (2006), the relative belief γ -credible region for a general ψ minimizes the prior content, with respect to the base prior, among all γ -credible region for ψ and these regions are shown here to also possess optimal robustness properties. Dey and Birmiwall (1994) consider global robustness measures based upon measures of distance from the posterior distribution. While the concern here is with robustness to the prior, it is also relevant to be concerned with robustness to the likelihood. Given the similarities between likelihood inferences and relative belief inferences, see Evans (2015), the robustness results in Royall and Tsou (2003) suggest similar properties will hold but this topic is not pursued further here. It is to be emphasized that the discussion here is only concerned with proper priors and it is unclear how this relates to the situation where improper priors are used.

Robustness to the prior is studied here via the Gâteaux derivative. Undoubtedly the Fréchet derivative is a more reasonable choice but it comes with some computational costs and so we have used the simpler directional derivative here. Robustness to the prior, as measured by the Fréchet derivative, will be pursued in another study.

2 Relative Belief Inferences

When $RB_{\Psi}(\psi | x) > 1$ this is the factor by which prior belief in the truth of ψ has increased after seeing the data. Clearly the bigger $RB_{\Psi}(\psi | x)$ is, the more evidence there is in favor of ψ while, when $RB_{\Psi}(\psi | x) < 1$, the smaller $RB_{\Psi}(\psi | x)$ is, the more evidence there is against ψ . This leads to a total preference ordering on Ψ , namely, ψ_1 is not preferred to ψ_2 whenever $RB_{\Psi}(\psi_1 | x) \leq RB_{\Psi}(\psi_2 | x)$ since there is at least as much evidence for ψ_2 as there is for ψ_1 . This in turn leads to unambiguous solutions to inference problems.

The best estimate of ψ is the value for which the evidence is greatest, namely,

$$\psi(x) = \arg \sup RB_{\Psi}(\psi | x).$$

Associated with this estimate is a γ -relative belief credible region $C_{\Psi, \gamma}(x) = \{\psi : RB_{\Psi}(\psi | x) \geq c_{\Psi, \gamma}(x)\}$ where $c_{\Psi, \gamma}(x) = \inf\{k : \Pi_{\Psi}(RB_{\Psi}(\psi | x) \leq k | x) \geq 1 - \gamma\}$. Notice that $\psi(x) \in C_{\Psi, \gamma}(x)$ for every $\gamma \in [0, 1]$ and so, for selected γ , the size of $C_{\Psi, \gamma}(x)$ can be taken as a measure of the accuracy of the estimate $\psi(x)$. The interpretation of $RB_{\Psi}(\psi | x)$ as the evidence for ψ , forces the use of the sets $C_{\Psi, \gamma}(x)$ for our credible regions. For if ψ_1 is in such a region and $RB_{\Psi}(\psi_2 | x) \geq RB_{\Psi}(\psi_1 | x)$, then ψ_2 must be in the region as well as there is at least as much evidence for ψ_2 as for ψ_1 . Optimal properties for relative belief credible regions, in the class of all credible regions, have been established in Evans et al. (2006) and Evans and Shakhathreh (2008). As previously mentioned, a relative belief γ -credible region for a general ψ minimizes the prior content and always has posterior content greater than or equal to its prior content. The latter generalizes a result of Piccinato (1984) to any ψ . Optimal properties for the estimator $\psi(x)$ are established in Evans and Jang (2011a).

For the assessment of the hypothesis $H_0 : \Psi(\theta) = \psi_0$, the evidence is given by $RB_{\Psi}(\psi_0 | x)$. One problem that both the relative belief ratio and the Bayes factor share as measures of evidence, is that it is not clear how they should be calibrated. Certainly the bigger $RB_{\Psi}(\psi_0 | x)$ is than 1, the more evidence there is in favor of ψ_0 while the smaller $RB_{\Psi}(\psi_0 | x)$ is than 1, the more evidence we have against ψ_0 . But what exactly does a value of $RB_{\Psi}(\psi_0 | x) = 20$ mean? It would appear to be strong evidence in favor of ψ_0 because beliefs have increased by a factor of 20 after seeing the data. But what if other values of ψ had even larger increases? For example, the discussion in Baskurt and Evans (2013) of the Jeffreys–Lindley paradox makes it clear that the value of a relative belief ratio or a Bayes factor cannot always be interpreted as an indication of the strength of the evidence.

The value $RB_{\Psi}(\psi_0 | x)$ can be calibrated by comparing it to the other possible values $RB_{\Psi}(\cdot | x)$ through its posterior distribution. For example, one possible measure of the strength is

$$\Pi_{\Psi}(RB_{\Psi}(\psi | x) \leq RB_{\Psi}(\psi_0 | x) | x) \tag{3}$$

which is the posterior probability that the true value of ψ has a relative belief ratio no greater than that of the hypothesized value ψ_0 . While (3) may look like a p-value, it has a very different interpretation. For when $RB_{\Psi}(\psi_0 | x) < 1$, so there is evidence against ψ_0 , then a small value for (3) indicates a large posterior probability that the

true value has a relative belief ratio greater than $RB_{\Psi}(\psi_0 | x)$ and so there is strong evidence against ψ_0 . If $RB_{\Psi}(\psi_0 | x) > 1$, so there is evidence in favor of ψ_0 , then a large value for (3) indicates a small posterior probability that the true value has a relative belief ratio greater than $RB_{\Psi}(\psi_0 | x)$ and so there is strong evidence in favor of ψ_0 . Notice that, in the set $\{\psi : RB_{\Psi}(\psi | x) \leq RB_{\Psi}(\psi_0 | x)\}$, the “best” estimate of the true value is given by ψ_0 simply because the evidence for this value is the largest in this set.

Various results have been established in Baskurt and Evans (2013) supporting both $RB_{\Psi}(\psi_0 | x)$, as the measure of the evidence for H_0 , and (3) as a measure of the strength of that evidence. For example, the following simple inequalities are useful in assessing the strength of the evidence, namely, $\Pi_{\Psi}(RB_{\Psi}(\psi | x) = RB_{\Psi}(\psi_0 | x) | x) \leq \Pi_{\Psi}(RB_{\Psi}(\psi | x) \leq RB_{\Psi}(\psi_0 | x) | x) \leq RB_{\Psi}(\psi_0 | x)$. So if $\Pi_{\Psi}(\{RB_{\Psi}(\psi_0 | x)\} | x)$ is large when $RB_{\Psi}(\psi_0 | x) > 1$, then there is strong evidence in favor of ψ_0 while, if $RB_{\Psi}(\psi_0 | x) < 1$ is very small, then there is immediately strong evidence against ψ_0 . Also, in situations where there are only a few possible values of ψ , then $\Pi_{\Psi}(RB_{\Psi}(\psi | x) = RB_{\Psi}(\psi_0 | x) | x)$ can be a more appropriate measure of strength.

It is worth noting that formally $RB_{\Psi}(\cdot | x)$ is an integrated likelihood. The interpretation, however, is quite different as the value $RB_{\Psi}(\psi | x)$ is a measure of the evidence that ψ is the true value and this is not the case for an integrated likelihood which can be multiplied by any positive constant. In effect likelihoods of any variety only give relative measures of evidence when comparing two values. The importance of this distinction is readily seen by noting the difference in the interpretation when $RB_{\Psi}(\psi | x) < 1$ and when $RB_{\Psi}(\psi | x) > 1$. By contrast an integrated likelihood does not provide a clear distinction between having evidence for or against a specific value and this has numerous consequences for the theory of inference. For example, based on a measure of evidence it is possible to consider the bias a prior induces into an analysis as discussed in Baskurt and Evans (2013). Also, the profile likelihood is commonly used for inferences about ψ although it is not a likelihood in general. Piccinato (1984) and Liseo (1996) establish that Bayesian results concerning likelihood regions for θ can be easily generalized to profile likelihood regions for ψ , see Corollary 3. Our concern here, however, is with regions determined using the marginal prior and posterior distributions of ψ and, as noted these can be considered as based on the integrated likelihood. Further properties of inferences based on the integrated likelihood are discussed in Berger et al. (1999).

3 Optimal Robustness with Respect to the Marginal Prior

When interest is in making inferences about $\psi = \Psi(\theta)$, it is reasonable to ask how sensitive the relief belief approach is to the ingredients given by the prior. This entails examining how dependent $\psi(x)$, $C_{\Psi, \gamma}(x)$, $RB_{\Psi}(\psi_0 | x)$ and $\Pi_{\Psi}(RB_{\Psi}(\psi | x) \leq RB_{\Psi}(\psi_0 | x) | x)$ are to changes in the prior, as these four objects represent the essential relative belief inferences.

The full prior π for θ can always be factored as $\pi(\theta) = \pi_{\Psi}(\psi)\pi(\theta | \psi)$. In contrast to other discussions of robustness with respect to the prior, the sensitivity of the inferences

to π_Ψ and the sensitivity of the inferences to $\pi(\cdot | \psi)$ are considered separately, as this leads to more information concerning where the lack of robustness arises when this occurs.

The result (1) implies that $RB_\Psi(\psi | x) = m(x | \psi)/m(x)$. From this it is immediate that $\psi(x) = \arg \sup_\psi RB_\Psi(\psi | x) = \arg \sup_\psi m(x | \psi)$ and so the relative belief estimate is optimally robust to π_Ψ as the estimate has no dependence on the marginal prior. Furthermore, $C_{\Psi, \gamma}(x)$ is of the form $\{\psi : m(x | \psi) \geq k\}$ for some k and so the form of relative belief regions for ψ is optimally robust to π_Ψ . The specific region chosen for the assessment of the accuracy of $\psi(x)$ depends on the posterior and so is not independent of π_Ψ . It is now proved that $C_{\Psi, \gamma}(x)$ has an optimal robustness property among all credible regions for ψ .

Consider ϵ -contaminated priors for θ of the form

$$\Pi_\epsilon = \Pi(\cdot | \psi) \times [(1 - \epsilon)\Pi_\Psi + \epsilon Q], \quad (4)$$

where Q is a probability measure on Ψ and Π is the base prior as described in the Introduction. Note that the conditional prior of θ given $\Psi(\theta) = \psi$ is fixed and independent of ϵ .

To assess the robustness of the posterior content of a set $A \subset \Psi$, it makes sense to look at $\delta(A) = \Pi_\Psi^{upper}(A | x) - \Pi_\Psi^{lower}(A | x)$ where $\Pi_\Psi^{upper}(A | x) = \sup_Q \Pi_\Psi^\epsilon(A | x)$ and $\Pi_\Psi^{lower}(A | x) = \inf_Q \Pi_\Psi^\epsilon(A | x)$ and the supremum/infimum is taken over all probability measures on Ψ . For this let $\epsilon^* = \epsilon/(1 - \epsilon)$ and $r(A) = \sup_{\psi \in A} RB_\Psi(\psi | x) = \sup_{\psi \in A} m(x | \psi)/m(x)$, so $r(\Psi) = RB_\Psi(\psi(x) | x)$ and always one and only one of $r(A)$, $r(A^c)$ equals $r(\Psi)$.

The following known result is needed.

Lemma 1. (Huber (1973)) Let Q denote a probability measure on Ψ . For prior measure $\Pi_\Psi^\epsilon = (1 - \epsilon)\Pi_\Psi + \epsilon Q$ on Ψ and $A \subset \Psi$,

(i) $\Pi_\Psi^{upper}(A | x) = (\Pi_\Psi(A | x) + \epsilon^* r(A))/(1 + \epsilon^* r(A))$,

(ii) $\Pi_\Psi^{lower}(A | x) = \Pi_\Psi(A | x)/(1 + \epsilon^* r(A^c))$,

(iii)

$$\delta(A) = \frac{\Pi_\Psi(A | x) \epsilon^* (r(A^c) - r(A))}{(1 + \epsilon^* r(A))(1 + \epsilon^* r(A^c))} + \frac{\epsilon^* r(A)}{(1 + \epsilon^* r(A))},$$

(iv) $\delta(A^c) = \delta(A)$.

Let $\gamma^*(x) = \Pi_\Psi(C_{\Psi, \gamma}(x) | x)$ be the exact posterior content of the γ -relative belief region. The following result generalizes results found in Wasserman (1989) and de la Horra and Fernandez (1994) who considered robustness to the prior of credible regions for the full parameter θ . In particular, this result applies to arbitrary parameters $\psi = \Psi(\theta)$ and does not require continuity.

Proposition 2. The following hold,

(i) among all sets $A \subset \Psi$ satisfying $\Pi_\Psi(A | x) \leq \gamma^*(x)$ and $r(A) = r(\Psi)$, the set $C_{\Psi, \gamma}(x)$ minimizes $\delta(A)$,

- (ii) among all sets $A \subset \Psi$ satisfying $\Pi_{\Psi}(A|x) \geq \gamma^*(x)$ and $r(A^c) = r(\Psi)$, the set $C_{\Psi,1-\gamma^*(x)}^c(x)$ minimizes $\delta(A)$,
- (iii) when $\gamma^*(x) = \gamma \geq 1/2$ then, among all sets $A \subset \Psi$ satisfying $\Pi_{\Psi}(A|x) = \gamma$, the set $C_{\Psi,\gamma}(x)$ minimizes $\delta(A)$.

Proof. (i) For any set A with $r(A) = r(\Psi)$ then $r(A^c) - r(A) = r(A^c) - r(\Psi) \leq 0$. Therefore,

$$\begin{aligned} \delta(A) &= \frac{\Pi_{\Psi}(A|x)\epsilon^*(r(A^c) - r(\Psi))}{(1 + \epsilon^*r(\Psi))(1 + \epsilon^*r(A^c))} + \frac{\epsilon^*r(\Psi)}{1 + \epsilon^*r(\Psi)} \\ &\geq \frac{\Pi_{\Psi}(C_{\Psi,\gamma}(x)|x)\epsilon^*(r(A^c) - r(\Psi))}{(1 + \epsilon^*r(\Psi))(1 + \epsilon^*r(A^c))} + \frac{\epsilon^*r(\Psi)}{1 + \epsilon^*r(\Psi)}. \end{aligned}$$

Now

$$\frac{r(A^c) - r(\Psi)}{1 + \epsilon^*r(A^c)} \tag{5}$$

is increasing in $r(A^c)$, so we need to show that $A = C_{\Psi,\gamma}(x)$ minimizes $r(A^c)$ among all A satisfying $\Pi_{\Psi}(A|x) \leq \Pi_{\Psi}(C_{\Psi,\gamma}(x)|x)$ and $r(A) = r(\Psi)$. Suppose that $r(A^c) < r(C_{\Psi,\gamma}^c(x))$ and let $B = \{\psi : RB_{\Psi}(\psi|x) > r(A^c)\}$. Note that $r(C_{\Psi,\gamma}^c(x)) \leq \inf_{\psi \in C_{\Psi,\gamma}(x)} RB_{\Psi}(\psi|x)$ and so $C_{\Psi,\gamma}(x) \subset B$, which implies $\Pi_{\Psi}(B|x) > \Pi_{\Psi}(C_{\Psi,\gamma}(x)|x)$ with the strictness of the inequality following from the definition of $C_{\Psi,\gamma}(x)$. But also $B \subset A$ which contradicts $\Pi_{\Psi}(A|x) \leq \Pi_{\Psi}(C_{\Psi,\gamma}(x)|x)$ and so we must have $r(A^c) \geq r(C_{\Psi,\gamma}^c(x))$. This establishes that (5) is minimized by $A = C_{\Psi,\gamma}(x)$.

(ii) Now consider all the sets A with $r(A^c) = r(\Psi)$. Since $\delta(A) = \delta(A^c)$, it is equivalent to minimize $\delta(A^c)$ among all sets A^c satisfying $\Pi_{\Psi}(A^c|x) \leq \Pi_{\Psi}(C_{\Psi,\gamma}^c(x)|x) = 1 - \gamma^*(x)$ and $r(A^c) = r(\Psi)$. By part (i) this is minimized by taking $A^c = C_{\Psi,1-\gamma^*(x)}^c(x)$ and the result is proved.

(iii) The solutions to the optimization problems in parts (i) and (ii), namely, $C_{\Psi,\gamma}(x)$ and $C_{\Psi,1-\gamma}^c(x)$ respectively, both have posterior content equal to γ . As such one of these sets is the solution to the optimization problem stated in (iii). We have that

$$\begin{aligned} \delta(C_{\Psi,\gamma}(x)) - \delta(C_{\Psi,1-\gamma}^c(x)) &= \delta(C_{\Psi,\gamma}(x)) - \delta(C_{\Psi,1-\gamma}(x)) \\ &= \frac{\gamma\epsilon^*(r(C_{\Psi,\gamma}^c(x)) - r(\Psi))}{(1 + \epsilon^*r(\Psi))(1 + \epsilon r(C_{\Psi,\gamma}^c(x)))} - \frac{\gamma\epsilon^*(r(C_{\Psi,1-\gamma}^c(x)) - r(\Psi))}{(1 + \epsilon^*r(\Psi))(1 + \epsilon r(C_{\Psi,1-\gamma}^c(x)))} \\ &= \frac{\gamma\epsilon^*}{(1 + \epsilon^*r(\Psi))} \left\{ \frac{r(C_{\Psi,\gamma}^c(x)) - r(\Psi)}{1 + \epsilon r(C_{\Psi,\gamma}^c(x))} - \frac{r(C_{\Psi,1-\gamma}^c(x)) - r(\Psi)}{1 + \epsilon r(C_{\Psi,1-\gamma}^c(x))} \right\}. \end{aligned}$$

The result follows from this because $C_{\Psi,\gamma}^c(x) \subset C_{\Psi,1-\gamma}^c(x)$, so $r(C_{\Psi,\gamma}^c(x)) \leq r(C_{\Psi,1-\gamma}^c(x))$, and (5) is increasing in $r(A^c)$. \square

It is interesting to consider the statistical meaning of the separate parts of Proposition 2 as the statements create a degree of ambiguity. If a system of credible regions is being used, say $B_{\Psi,\gamma}(x)$, then it makes sense to require that these sets are monotonically increasing in γ and the smallest set $\lim_{\gamma \searrow 0} B_{\Psi,\gamma}(x)$ contains a single point which

is taken as the estimate of ψ . The size of $B_{\Psi,\gamma}(x)$, for some specific γ , can then be taken as an assessment of the accuracy of the estimate where size is measured in some application dependent way. The relative belief regions satisfy this and the estimate, under the assumption of a unique maximizer of $RB_{\Psi}(\cdot|x)$, is $\psi(x)$. So effectively (i) is saying that $C_{\Psi,\gamma}(x)$ is optimally robust with respect to posterior content. Note that we have to exclude sets A with $\Pi_{\Psi}(A|x) > \gamma^*(x)$ because, for example, the set $A = \Psi$ is always optimally robust with respect to content but does not provide a meaningful assessment of the accuracy of the estimate. Given that $\psi(x)$ and the form of $C_{\Psi,\gamma}(x)$ are optimally robust, this further supports the claim that relative belief estimation is optimally robust to the choice of the marginal prior. Note that the sets in (ii) do not satisfy the stated criteria for being a system of credible regions.

Part (iii) indicates that, when there are many sets with posterior content exactly equal to γ , and this is typically true in the continuous case, then $C_{\Psi,\gamma}(x)$ is optimally robust among these sets with respect to content. It makes sense to require $\gamma^*(x) \geq 1/2$ for any credible region as, if $\gamma^*(x) < 1/2$, then there is more belief that the true value is in $C_{\Psi,\gamma}^c(x)$ than in $C_{\Psi,\gamma}(x)$.

It is also interesting to note an immediate consequence of Proposition 2 that is similar to a result in Liseo (1996) developed there in the context of a discussion of likelihood regions. Let $C_{\gamma}(x)$ be a relative belief credible region for θ and note that this is a likelihood region. Put $B_{\Psi,\gamma}(x) = \Psi C_{\gamma}(x)$, $\gamma_*(x) = \Pi_{\Psi}(B_{\Psi,\gamma}(x)|x)$ and note that under weak conditions $B_{\Psi,\gamma}(x)$ is a profile likelihood region for ψ . Denote the profile likelihood by $L_{\Psi}(\cdot|x)$ and put $s(A) = \sup_{\psi \in A} L_{\Psi}(\psi|x)$, so $s(\Psi) = L_{\Psi}(\psi_{profile}(x)|x)$ with $\psi_{profile}(x)$ the profile likelihood estimate of ψ .

Corollary 3. *The following hold,*

- (i) *among all sets $A \subset \Psi$ satisfying $\Pi_{\Psi}(A|x) \leq \gamma_*(x)$ and $s(A) = s(\Psi)$, the set $B_{\Psi,\gamma}(x)$ minimizes $\delta(A)$,*
- (ii) *among all sets $A \subset \Psi$ satisfying $\Pi_{\Psi}(A|x) \geq \gamma_*(x)$ and $s(A^c) = s(\Psi)$, the set $B_{\Psi,1-\gamma_*(x)}^c(x)$ minimizes $\delta(A)$,*
- (iii) *when $\gamma_*(x) = \gamma \geq 1/2$ then, among all sets $A \subset \Psi$ satisfying $\Pi_{\Psi}(A|x) = \gamma$, the set $B_{\Psi,\gamma}(x)$ minimizes $\delta(A)$.*

So profile likelihood regions also have optimal robustness properties. The difference between the two results is that profile likelihood regions and relative belief regions can be quite different. For example, a γ -profile likelihood region need not contain the relative belief estimate $\psi(x)$ and conversely, as the following example demonstrates.

Example 1. *Profile and integrated likelihood regions differ substantially.*

Consider the model-prior combination given in Table 1 with sample space $\mathcal{X} = \{1, 2\}$, parameter space $\Theta = \{0, 1, 2\}$ and $\delta \in [0, 1/2]$. Suppose interest is in $\psi = \Psi(\theta) = I_{\{0,1\}}(\theta)$. For $x = 1$ the profile likelihood is $L(0|1) = 1/3, L(1|1) = 1/2$ independent of δ and so $\psi_{profile}(1) = 1$. The posterior is $\pi(0|1) = \delta/2m_{\delta}(1), \pi(1|1) = (1/2 - \delta)/5m_{\delta}(1), \pi(2|1) = 1/6m_{\delta}(1)$ where $m_{\delta}(1) = 4/15 + 3\delta/10$ and these posterior probabilities converge to 0, 3/8 and 5/8 respectively as $\delta \rightarrow 0$. So the limiting posterior probabilities for ψ are $\pi_{\Psi}(0|1) = 5/8, \pi_{\Psi}(1|1) = 3/8$.

| θ | $f_\theta(1)$ | $f_\theta(2)$ | $\pi(\theta)$ |
|----------|---------------|---------------|----------------|
| 0 | 1/2 | 1/2 | δ |
| 1 | 1/5 | 4/5 | $1/2 - \delta$ |
| 2 | 1/3 | 2/3 | 1/2 |

Table 1: The model and prior in Example 1.

The relative belief ratio for ψ is given by $RB_\Psi(0|x) = 1/3m(1)$, $RB_\Psi(1|x) = (1/5 + 3\delta/5)/m(1)$. From this $\psi(1) = 1$ if and only if $\delta \in [2/9, 1/2]$. In this case, when δ is small and $\gamma = 3/8$, a γ -credible region based on the profile likelihood does not contain the relative belief estimate $\psi(1) = 0$ while the corresponding relative belief region does not contain the profile likelihood estimate. The essential difference here is that the relative belief estimate has the interpretation that it is the value of ψ for which the data have led to the greatest increase in belief and no such interpretation is available for the profile likelihood estimate.

Applying Lemma 1 gives

$$\delta(C_{\Psi,\gamma}(x)) = \frac{\epsilon^* RB_\Psi(\psi(x)\pi x)}{(1 + \epsilon^* RB_\Psi(\psi(x)|x))} \times \left\{ 1 - \frac{\Pi_\Psi(C_{\Psi,\gamma}(x)|x)}{RB_\Psi(\psi(x)|x)} \frac{1 - \inf_{\psi \in C_{\Psi,\gamma}(x)} RB_\Psi(\psi|x)}{(1 + \epsilon^* \inf_{\psi \in C_{\Psi,\gamma}(x)} RB_\Psi(\psi|x))} \right\}$$

and this can be close to 1 when $RB_\Psi(\psi(x)|x)$ is large. So, while $C_{\Psi,\gamma}(x)$ possesses an optimal robustness property with respect to posterior content, this does not imply that the posterior content is necessarily robust. This depends on other aspects of the particular problem which will be discussed.

4 Measuring Robustness Quantitatively

To measure the robustness of an inference to the prior π , when using the ϵ -contaminated class, it is natural to look at Gâteaux derivatives of the relevant quantity at π in various directions Q . The derivative is a measure of the sensitivity of the inference to small changes in the prior and so is local in nature. When the derivative is large for some Q , the inference is highly sensitive to the prior chosen and naturally this is viewed negatively. In this section this behavior of relative belief inferences is analyzed separately for ϵ -contaminated classes for the marginal π_Ψ and the conditional $\pi(\cdot|\psi)$.

4.1 Sensitivity to the Marginal Prior

Consider the family of priors given by (4) but now restricted to those Q that are also absolutely continuous with respect to ν_Ψ on Ψ and let q denote the density of Q . The posterior of ψ based on the contaminated prior is $\Pi_{\epsilon,\Psi}(\cdot|x) = (1-\epsilon_x)\Pi_\Psi(\cdot|x) + \epsilon_x Q(\cdot|x)$ where $\epsilon_x = \epsilon m_Q(x)/[(1-\epsilon)m(x) + \epsilon m_Q(x)]$, $m_Q(x) = \int_\Psi m(x|\psi) Q(d\psi)$ and $Q(A|x) = \int_A (m(x|\psi)/m_Q(x)) Q(d\psi)$. The relative belief ratio for ψ based on a general Π_ϵ equals

$RB_{\epsilon, \Psi}(\psi | x) = (1 - \epsilon_x)RB_{\Psi}(\psi | x) + \epsilon_x RB_{Q, \Psi}(\psi | x)$ and here, using (1), $RB_{Q, \Psi}(\psi | x) = m(x | \psi) / m_Q(x)$ so

$$RB_{\epsilon, \Psi}(\psi | x) = \frac{RB_{\Psi}(\psi | x)}{1 - \epsilon(1 - m_Q(x) / m(x))}. \quad (6)$$

The following result gives the Gâteaux derivative of the relative belief ratio.

Proposition 4. *The Gâteaux derivative of $RB_{\Psi}(\cdot | x)$ at ψ in direction Q equals*

$$RB_{\Psi}(\psi | x) \{1 - m_Q(x) / m(x)\}. \quad (7)$$

Proof. From (6),

$$\lim_{\epsilon \rightarrow 0} \frac{RB_{\Psi}^{\epsilon}(\psi | x) - RB_{\Psi}(\psi | x)}{\epsilon} = RB_{\Psi}(\psi | x) \lim_{\epsilon \rightarrow 0} \left\{ \frac{(1 - m_Q(x) / m(x))}{1 - \epsilon(1 - m_Q(x) / m(x))} \right\}. \quad \square$$

The value of (7) can be large simply because $RB_{\Psi}(\psi | x)$ is large, so it makes more sense to look at the relative change. Therefore, for small ϵ ,

$$\frac{|RB_{\epsilon, \Psi}(\psi | x) - RB_{\Psi}(\psi | x)|}{RB_{\Psi}(\psi | x)} \approx \left| 1 - \frac{m_Q(x)}{m(x)} \right| \epsilon$$

implying a small relative change in $RB_{\Psi}(\psi | x)$ when $m_Q(x) / m(x)$ is not large.

The Gâteaux derivative of the strength of the evidence is now computed.

Proposition 5. *The Gâteaux derivative of $\Pi_{\Psi}(RB_{\Psi}(\psi | x) \leq RB_{\Psi}(\psi_0 | x) | x)$ at ψ_0 in the direction Q is*

$$\frac{m_Q(x)}{m(x)} \left\{ \begin{array}{l} Q(RB_{\Psi}(\psi | x) \leq RB_{\Psi}(\psi_0 | x) | x) - \\ \Pi_{\Psi}(RB_{\Psi}(\psi | x) \leq RB_{\Psi}(\psi_0 | x) | x) \end{array} \right\}.$$

Proof. The strength based on Π_{Ψ}^{ϵ} satisfies $\Pi_{\Psi}^{\epsilon}(RB_{\epsilon, \Psi}(\psi | x) \leq RB_{\epsilon, \Psi}(\psi_0 | x) | x) = (1 - \epsilon_x)\Pi_{\Psi}(RB_{\epsilon, \Psi}(\psi | x) \leq RB_{\epsilon, \Psi}(\psi_0 | x) | x) + \epsilon_x Q(RB_{\epsilon, \Psi}(\psi | x) \leq RB_{\epsilon, \Psi}(\psi_0 | x) | x)$. So, using (6),

$$\begin{aligned} \Pi_{\Psi}^{\epsilon}(RB_{\epsilon, \Psi}(\psi | x) \leq RB_{\epsilon, \Psi}(\psi_0 | x) | x) &= \Pi_{\Psi}(m(x | \psi) \leq m(x | \psi_0) | x) + \\ &\epsilon_x \{Q(m(x | \psi) \leq m(x | \psi_0) | x) - \Pi_{\Psi}(m(x | \psi) \leq m(x | \psi_0) | x)\}. \end{aligned}$$

This implies that

$$\begin{aligned} &\lim_{\epsilon \rightarrow 0} [\Pi_{\Psi}^{\epsilon}(RB_{\Psi}^{\epsilon}(\psi_0 | x) \leq RB_{\Psi}^{\epsilon}(\psi_0 | x) | x) - \Pi_{\Psi}(RB_{\Psi}(\psi | x) \leq RB_{\Psi}(\psi_0 | x) | x)] / \epsilon \\ &= \frac{m_Q(x)}{m(x)} \{Q(m(x | \psi) \leq m(x | \psi_0) | x) - \Pi_{\Psi}(m(x | \psi) \leq m(x | \psi_0) | x)\} \\ &= \frac{m_Q(x)}{m(x)} \left\{ \begin{array}{l} Q(RB_{\Psi}(\psi | x) \leq RB_{\Psi}(\psi_0 | x) | x) - \\ \Pi_{\Psi}(RB_{\Psi}(\psi | x) \leq RB_{\Psi}(\psi_0 | x) | x) \end{array} \right\}. \quad \square \end{aligned}$$

So the strength is robust to choice of the marginal prior π_Ψ whenever $m_Q(x)/m(x)$ is small.

For both the measure of evidence $RB_\Psi(\psi_0 | x)$ and its strength, the ratio $m_Q(x)/m(x)$ plays a key role in determining the robustness. The implications of this are discussed in Section 5. Note that $\sup_Q m_Q(x)/m(x) = RB_\Psi(\psi(x) | x)$ gives the worst case behavior of this ratio.

It is of interest to contrast these results with those for the commonly used MAP inferences which are based on the posterior density $\pi_\Psi(\cdot | x)$.

Proposition 6. *The Gâteaux derivative of the posterior density of ψ in the direction Q at ψ_0 is given by $\{m_Q(x)/m(x)\}\{q(\psi_0 | x) - \pi_\Psi(\psi_0 | x)\}$.*

Proof. Since $\pi_{\epsilon, \Psi}(\psi | x) = (1 - \epsilon_x)\pi_\Psi(\psi | x) + \epsilon_x q(\psi | x)$ it follows that

$$\lim_{\epsilon \rightarrow 0} \frac{\pi_{\epsilon, \Psi}(\psi_0 | x) - \pi_\Psi(\psi_0 | x)}{\epsilon} = \frac{m_Q(x)}{m(x)}(q(\psi_0 | x) - \pi_\Psi(\psi_0 | x)). \quad \square$$

Note that MAP-based inferences implicitly use $\pi_\Psi(\psi_0 | x)$ as a measure of the evidence that ψ_0 is the true value. Comparing this with the relative belief ratio we see that for small ϵ ,

$$\frac{|\pi_{\epsilon, \Psi}(\psi_0 | x) - \pi_\Psi(\psi_0 | x)|}{\pi_\Psi(\psi_0 | x)} \approx \frac{m_Q(x)}{m(x)} \left| 1 - \frac{q(\psi_0 | x)}{\pi_\Psi(\psi_0 | x)} \right| \epsilon$$

and the relative change in $\pi_\Psi(\psi_0 | x)$ is dependent on the ratio of the posteriors as well as $m_Q(x)/m(x)$. So if $\pi_\Psi(\psi_0 | x)$ is small relative to $q(\psi_0 | x)$ we will get a big relative change and this suggests that MAP inferences are much less robust than relative belief inferences. A similar result is obtained for the Bayesian p-value in Evans and Zou (2001). Consider also the following example.

Example 2. *The MAP and relative belief estimates contrasted.*

Suppose x is a sample from a Bernoulli(θ), $\theta \sim \text{beta}(\alpha, \beta)$ with $\alpha \geq 1, \beta \geq 1$ and $\psi = \Psi(\theta) = \theta^p$ for some $p \leq 1$. In this case $\psi(x) = \bar{x}^p$ and the MAP estimate is given by $\psi_{MAP}(x) = (\alpha - p + n\bar{x})^p / (n + \alpha + \beta - p - 1)^p$. Note that the relative belief estimate of ψ is just the appropriate transform of the relative belief estimate of θ . On the other hand the MAP estimate of θ is $(\alpha - 1 + n\bar{x}) / (n + \alpha + \beta - 2)$ and $\psi_{MAP}(x)$ is not the p -th power of this. It is the case, however, that the two estimates are essentially equivalent whenever n is large enough. How large n has to be, however, depends on α, β, p and the data. For example, when $n = 1000, \alpha = 1, \beta = 1, p = 0.1$ and $\bar{x} = 0$, then $\psi(x) = 0$ but $\psi_{MAP}(x) = 0.496$ even though it would be natural to be quite certain that $\psi \approx 0$ in such a case.

It is clear that $\psi_{MAP}(x)$ is much less robust to the prior than $\psi(x)$. For example, as $\alpha \rightarrow \infty$, then $\psi_{MAP}(x) \rightarrow 1$, for any value of \bar{x} . So if the prior were chosen to reflect virtual certainty that $\theta = 1$ by choosing α very large, then $\psi_{MAP}(x)$ will reflect this even when this is contradicted by the data \bar{x} . This issue is discussed more fully in Section 5 where it is seen that prior-data conflict can play a key role in determining robustness to the prior.

4.2 Sensitivity to the Conditional Prior

Consider now priors for θ of the form $\Pi_\epsilon = [(1-\epsilon)\Pi(\cdot|\psi) + \epsilon Q(\cdot|\psi)] \times \Pi_\Psi$ where $Q(\cdot|\psi)$ is a probability measure on $\Psi^{-1}\{\psi\}$ absolutely continuous with respect to $\nu_{\Psi^{-1}\{\psi\}}$ with density $q(\cdot|\psi)$, for each $\psi \in \Psi$. So the marginal prior of ψ is now fixed and the conditional prior of θ is perturbed. The posterior of ψ based on this prior is $\Pi_{\epsilon,\Psi}(\cdot|x) = (1-\epsilon_x)\Pi_\Psi(\cdot|x) + \epsilon_x Q_\Psi(\cdot|x)$ where $Q_\Psi(A|x) = \int_A (m_Q(x|\psi)/m_Q(x)) \Pi_\Psi(d\psi)$, $m_Q(x|\psi) = \int_{\Psi^{-1}\{\psi\}} f_\theta(x) Q(d\theta|\psi)$ and $m_Q(x) = \int_\Psi m_Q(x|\psi) \Pi_\Psi(d\psi)$.

The relative belief ratio for ψ based on Π_ϵ equals $RB_{\epsilon,\Psi}(\psi|x) = (1-\epsilon_x)RB_\Psi(\psi|x) + \epsilon_x RB_{Q,\Psi}(\psi|x)$ where now $RB_{Q,\Psi}(\psi_0|x) = m_Q(x|\psi_0)/m_Q(x)$. This leads to the following result.

Proposition 7. *The Gâteaux derivative of $RB_\Psi(\cdot|x)$ at ψ_0 in the direction Q is $\{m_Q(x)/m(x)\}(RB_{Q,\Psi}(\psi_0|x) - RB_\Psi(\psi_0|x))$.*

Proof. Clearly,

$$\lim_{\epsilon \rightarrow 0} \frac{RB_{\epsilon,\Psi}(\psi_0|x) - RB_\Psi(\psi_0|x)}{\epsilon} = \frac{m_Q(x)}{m(x)} (RB_{Q,\Psi}(\psi_0|x) - RB_\Psi(\psi_0|x)). \quad \square$$

The implications of this result for robustness are discussed in Section 5.

Now consider the robustness of the strength of the evidence.

Proposition 8. *If $RB_\Psi(\cdot|x)$ has a discrete distribution with support containing no limit points, the Gâteaux derivative of $\Pi_\Psi(RB_\Psi(\psi|x) \leq RB_\Psi(\psi_0|x)|x)$ at ψ_0 in the direction Q equals 0. When $RB_\Psi(\cdot|x)$ has a continuous distribution under $\Pi_\Psi(\cdot|x)$ with density $g(\cdot|x)$, the Gâteaux derivative of $\Pi_\Psi(RB_\Psi(\psi|x) \leq RB_\Psi(\psi_0|x)|x)$ at ψ_0 in the direction Q equals $\{m_Q(x)/m(x)\}RB_{Q,\Psi}(\psi_0|x)g(RB_\Psi(\psi_0|x)|x)$.*

Proof. Since

$$\begin{aligned} & \Pi_\Psi(RB_{\epsilon,\Psi}(\psi|x) \leq RB_{\epsilon,\Psi}(\psi_0|x)|x) \\ &= \Pi_\Psi \left(\begin{array}{l} (1-\epsilon_x)RB_\Psi(\psi|x) + \epsilon_x RB_{Q,\Psi}(\psi|x) \\ \leq (1-\epsilon_x)RB_\Psi(\psi_0|x) + \epsilon_x RB_{Q,\Psi}(\psi_0|x) \end{array} \middle| x \right), \end{aligned}$$

then, for all $\epsilon > 0$ such that $\epsilon_x \leq 1$,

$$\begin{aligned} & \Pi_\Psi(RB_{\epsilon,\Psi}(\psi|x) \leq RB_{\epsilon,\Psi}(\psi_0|x)|x) \\ & \leq \Pi_\Psi \left(RB_\Psi(\psi|x) \leq RB_\Psi(\psi_0|x) + \frac{\epsilon_x}{1-\epsilon_x} RB_{Q,\Psi}(\psi_0|x) \middle| x \right) \end{aligned}$$

and for all $\epsilon < 0$,

$$\begin{aligned} & \Pi_\Psi(RB_{\epsilon,\Psi}(\psi|x) \leq RB_{\epsilon,\Psi}(\psi_0|x)|x) \\ & \geq \Pi_\Psi \left(RB_\Psi(\psi|x) \leq RB_\Psi(\psi_0|x) + \frac{\epsilon_x}{1-\epsilon_x} RB_{Q,\Psi}(\psi_0|x) \middle| x \right). \end{aligned}$$

When $RB_{\Psi}(\cdot|x)$ has a discrete distribution with support containing no limit points, then the lower and upper bounds equal $\Pi_{\Psi}(RB_{\Psi}(\psi|x) \leq RB_{\Psi}(\psi_0|x)|x)$ for all ϵ small enough and the result follows. When $RB_{\Psi}(\cdot|x)$ has a continuous distribution with density $g(\cdot|x)$, then

$$\begin{aligned} & \lim_{\epsilon \rightarrow 0} \frac{\Pi_{\Psi}(RB_{\epsilon, \Psi}(\psi_0|x) \leq RB_{\epsilon, \Psi}(\psi_0|x)|x) - \Pi_{\Psi}(RB_{\Psi}(\psi|x) \leq RB_{\Psi}(\psi_0|x)|x)}{\epsilon} \\ &= \{m_Q(x)/m(x)\}RB_{Q, \Psi}(\psi_0|x)g(RB_{\Psi}(\psi_0|x)|x). \quad \square \end{aligned}$$

From this it is seen that in the discrete case the strength is insensitive to local changes in the prior.

Consider the continuous case. When there is strong evidence either for or against ψ_0 , then $RB_{\Psi}(\psi_0|x)$ will be in the right or left tail correspondingly of the posterior distribution of $RB_{\Psi}(\cdot|x)$ and so $g(RB_{\Psi}(\psi_0|x)|x)$ will tend to be small. As such the strength will be robust to small changes in the prior provided $m_Q(x)/m(x)$ is not large. When there is not strong evidence however, then $g(RB_{\Psi}(\psi_0|x)|x)$ could be large and, if $m_Q(x)/m(x)$ is not small, then the strength is not robust. This underscores a recommendation in Baskurt and Evans (2013) that in the continuous case the parameter be discretized when assessing the evidence and its strength. For this, when ψ is real-valued, let $\delta > 0$ be the difference between two ψ values that is deemed to be of practical importance. The prior and posterior distributions of ψ discretized to the intervals $[\psi_0 + (2i - 1)\delta/2, \psi_0 + (2i + 1)\delta/2)$ for $i \in \mathbb{Z}$ are then used to assess the hypothesis corresponds to the interval $[\psi_0 - \delta/2, \psi_0 + \delta/2)$. By Proposition 8 the strength is then insensitive to small changes in the prior.

It is perhaps not surprising that the robustness behavior of the relative belief ratio and its strength is more complicated when considering the effect of the conditional prior than with the marginal prior. The optimality results concerning robustness to the marginal prior underscore this.

5 Robustness and Prior-Data Conflict

The existence of a prior-data conflict means that the data support certain values of $\psi = \Psi(\theta)$ being the true value but the prior places little or no mass there. While various measures can be used to determine whether or not such a conflict has occurred, a logical approach is based on the factorization of the joint probability measure for (θ, x) given by $\Pi \times P_{\theta} = \Pi(\cdot|T) \times M_T \times P(\cdot|T)$, where T is a minimal sufficient statistic, $\Pi(\cdot|T)$ is the posterior probability measure for θ , M_T is the prior predictive probability measure of T and $P(\cdot|T)$ is the conditional probability measure of the data given T . The measure $P(\cdot|T)$ is then available for computing probabilities relevant to checking the model $\{f_{\theta} : \theta \in \Theta\}$, the measure M_T is available for computing probabilities relevant to checking the prior and $\Pi(\cdot|T)$ is the relevant probability measure for computing probabilities for θ . A statistical analysis then proceeds by checking the model, perhaps via a tail probability based on a discrepancy statistic, and then proceeding to check the prior, if the data does not contradict the model. If both the model and prior are not

contradicted by the data, then we can proceed to inference about θ . The logic behind this sequence lies in part with the fact that it makes little sense to check a prior if the model fails. Furthermore, separating the check of the prior from that of the model provides more information in the event of a conflict arising, as it is then possible to identify where the failure lies, namely, with the model or with the prior.

In Evans and Moshonov (2006) this factorization was adhered to and the tail probability

$$M_T(m_T(t) \leq m_T(T(x))) \quad (8)$$

was advocated for checking the prior where m_T is the density of M_T with respect to some support measure. So if (8) is small, then the observed value $T(x)$ of the minimal sufficient statistic lies in the tails of M_T and there is an indication of a prior-data conflict. In Evans and Jang (2011b) the validity of this approach was firmly established by the proof that (8) converges to $\Pi(\pi(\theta) \leq \pi(\theta_{true}))$ under i.i.d. sampling and some additional weak conditions. Furthermore, it was shown how to modify (8) so as to achieve invariance under choice of the minimal sufficient statistic. Also, Evans and Moshonov (2006) argued that (8) should be replaced by $M_T(m_T(t) \leq m_T(T(x)) | U(T(x)))$ for any maximal ancillary $U(T)$ as the variation in T due to $U(T)$ has nothing to do with θ and so reflects nothing about the prior. The tail probability (8) is a check on the full prior and Evans and Moshonov (2006) also developed methods for checking factors of the prior so a failure in the prior could be isolated to a particular aspect.

First, however, consider the case when $\Psi(\theta) = \theta$ and interest is in the robustness of inferences to the whole prior. From the results in Section 4.1, it is seen that the ratio $m_Q(x)/m(x) = m_{Q,T}(T(x))/m_T(T(x))$, where $m_{Q,T}(T(x)) = \int_{\Theta} f_{\theta,T}(T(x)) Q(d\theta)$, plays a key role in determining the local sensitivity in the direction given by Q , of the inferences for given observed data x . This depends on Q and the worst case is given by

$$\sup_Q \frac{m_{Q,T}(T(x))}{m_T(T(x))} = \sup_Q \frac{\int_{\Theta} f_{\theta,T}(T(x)) Q(d\theta)}{m_T(T(x))} = RB(\theta(x) | x)$$

and note that $\theta(x)$ is the maximum likelihood estimator (MLE) in this case as well as the relative belief estimate. Notice that when (8) is small, so there is an indication of a prior-data conflict existing, then $m_T(T(x))$ is relatively small when compared to other values of $m_T(t)$ which are not influenced by the data. This implies that the prior is having a big influence relative to the data and so a lack of robustness can be expected.

This phenomenon is well-illustrated in the following examples where ancillaries play no role because of Basu's theorem.

Example 3. *Location normal model.*

Suppose that $x = (x_1, \dots, x_n)$ is a sample from the $N(\mu, 1)$ distribution with $\mu \sim N(\mu_0, \sigma_0^2)$. Then M_T is given by $T(x) = \bar{x} \sim N(\mu_0, 1/n + \sigma_0^2)$. When Q is the $N(\mu_1, \sigma_1^2)$ distribution, then $M_{Q,T}$ is given by $\bar{x} \sim N(\mu_1, 1/n + \sigma_1^2)$. This implies that

$$\frac{m_{Q,T}(T(x))}{m_T(T(x))} = \sqrt{\frac{1/n + \sigma_0^2}{1/n + \sigma_1^2}} \exp \left\{ -\frac{1}{2} \left[\frac{(1/n + \sigma_1^2)^{-1} (\bar{x} - \mu_1)^2}{(1/n + \sigma_0^2)^{-1} (\bar{x} - \mu_0)^2} \right] \right\} \quad (9)$$

| μ_1 | σ_1^2 | $m_Q(x)/m(x)$ | μ_1 | σ_1^2 | $m_Q(x)/m(x)$ |
|---------|--------------|---------------|---------|--------------|---------------|
| -3.0 | 1 | 0.0065 | 0.5 | 0.5 | 1.3474 |
| -2.0 | 1 | 0.0905 | 0.5 | 1.0 | 1.0000 |
| -1.0 | 1 | 0.4832 | 0.5 | 2.0 | 0.7254 |
| 1.0 | 1 | 0.7917 | 0.5 | 3.0 | 0.5975 |
| 2.0 | 1 | 0.2428 | 0.5 | 50.0 | 0.1488 |
| 3.0 | 1 | 0.0287 | 0.5 | 100.0 | 0.1053 |

Table 2: The ratio $m_Q(x)/m(x)$ in Example 3 when there is no conflict.

| μ_1 | σ_1^2 | $m_Q(x)/m(x)$ | μ_1 | σ_1^2 | $m_Q(x)/m(x)$ |
|---------|--------------|-----------------------|---------|--------------|---------------|
| -3.0 | 1 | 1.88×10^{-8} | 0.5 | 0.5 | 0.0053 |
| -2.0 | 1 | 9.97×10^{-8} | 0.5 | 1.0 | 1.0000 |
| -1.0 | 1 | 2.00×10^{-4} | 0.5 | 2.0 | 14.2070 |
| 1.0 | 1 | 4.90×10^{-1} | 0.5 | 3.0 | 32.5842 |
| 2.0 | 1 | 5.75×10^1 | 0.5 | 50.0 | 58.2823 |
| 3.0 | 1 | 2.61×10^2 | 0.5 | 100.0 | 43.9565 |

Table 3: The ratio $m_Q(x)/m(x)$ in Example 3 when there is conflict.

and, as a function of (μ_1, σ_1^2) this is maximized when $\mu_1 = \bar{x}, \sigma_1^2 = 0$. Notice that this supremum converges to ∞ as $\bar{x} \rightarrow \pm\infty$ and such values correspond to prior-data conflict with respect to the $N(\mu_0, \sigma_0^2)$ prior.

Now, consider a numerical example. A sample of size $n = 20$ was generated from the $N(0, 1)$ distribution obtaining $\bar{x} = 0.2591$. When the base prior is $N(0.5, 1)$ then (8) equals 0.8141 and accordingly there is no indication of any prior-data conflict. Also, $\sup_Q(m_Q(x)/m(x)) = 4.7109$ which seems modest as it describes the worst case robustness behavior. In Table 2 some values of $m_Q(x)/m(x)$ are recorded when Q is a $N(\mu_1, \sigma_1^2)$ distribution for various values of μ_1 and σ_1^2 as these might be expected to be realistic directions in which to perturb the base prior. In all cases the value of $m_Q(x)/m(x)$ is quite modest and the maximum value of (9) is 1.0534. Overall it can be concluded here that the analysis is robust to local perturbations of the prior.

Now consider an example where there is prior-data conflict. In this case a sample of $n = 20$ is generated from a $N(4, 1)$ distribution obtaining $\bar{x} = 4.0867$ and the same base prior is used. The value of (8) is 0.0005 and so there is a strong indication of prior-data conflict. Furthermore, $\sup_Q(m_Q(x)/m(x)) = 2096.85$ which certainly indicates a lack of robustness. In Table 3 some values of $m_Q(x)/m(x)$ are recorded when Q is a $N(\mu_1, \sigma_1^2)$ distribution for various values of μ_1 and σ_1^2 . It is seen that the value of $m_Q(x)/m(x)$ can be relatively large and the maximum value of (9) is 468.86. So it can be concluded that the analysis based on the model, prior and observed data, will not be robust to local perturbations of the prior when there is prior-data conflict.

Example 4. Bernoulli model.

Suppose that $x = (x_1, \dots, x_n)$ is a sample from a Bernoulli(θ) and the prior is $\theta \sim \text{beta}(\alpha_0, \beta_0)$ for some choice of (α_0, β_0) . A minimal sufficient statistic is $T(x) =$

| α_1 | β_1 | $m_Q(x)/m(x)$ | α_1 | β_1 | $m_Q(x)/m(x)$ |
|------------|-----------|---------------|------------|-----------|---------------|
| 20 | 5 | 32647.89 | 5 | 1 | 21523.28 |
| 15 | 5 | 25729.50 | 5 | 25 | 0.12 |
| 10 | 5 | 15010.95 | 5 | 22 | 0.41 |
| 5 | 5 | 3996.37 | 5 | 20 | 1.00 |
| 1 | 5 | 125.87 | 5 | 16 | 6.77 |

Table 4: The ratio $m_Q(x)/m(x)$ in Example 4 when there is conflict.

$\sum_{i=1}^n x_i \sim \text{binomial}(n, \theta)$ and then

$$m_T(t) = \binom{n}{t} \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} \frac{\Gamma(t + \alpha_0)\Gamma(n - t + \beta_0)}{\Gamma(n + \alpha_0 + \beta_0)}.$$

Also,

$$\sup_Q (m_Q(x)/m(x)) = \frac{\Gamma(\alpha_0)\Gamma(\beta_0)}{\Gamma(\alpha_0 + \beta_0)} \frac{\Gamma(n + \alpha_0 + \beta_0)}{\Gamma(t + \alpha_0)\Gamma(n - t + \beta_0)} \bar{x}^{n\bar{x}}(1 - \bar{x})^{n(1-\bar{x})}.$$

To illustrate the relationship between prior-data conflict and robustness, consider a numerical example. Suppose that $\alpha_0 = 5$ and $\beta_0 = 20$. Generating a sample of size $n = 20$ from the Bernoulli(0.25) gave the value $n\bar{x} = 3$. In this case (8) equals 0.7100 and there is no indication of any prior-data conflict. Also, $\sup_Q (m_Q(x)/m(x)) = 1.4211$ which indicates that the inferences will be generally robust to small deviations. If $m_Q(x)/m(x)$ is computed for various Q , where Q is a beta(α_1, β_1), then in all cases it is readily seen that this ratio is quite reasonable in value as indeed it is bounded above by 1.4211.

A sample of $n = 20$ was also generated from a Bernoulli(0.9) with the same prior being used. In this case $n\bar{x} = 17$ and (8) equals 6.2×10^{-6} , so there is a strong indication of prior-data conflict. Also, $\sup_Q (m_Q(x)/m(x)) = 46396.43$ which indicates that the inferences will be generally not be robust to small deviations. Table 4 provides some values of $m_Q(x)/m(x)$ for Q given by a beta(α_1, β_1) for various choices of (α_1, β_1) and there are several large values.

Now consider the case when $\theta = (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$ so the prior factors as $\pi(\theta) = \pi_2(\theta_2 | \theta_1)\pi_1(\theta_1)$. Presumably the conditional prior $\pi_2(\cdot | \theta_1)$ and the marginal prior π_1 are elicited and the goal is inference about some $\psi = \Psi(\theta)$. It is then preferable to check the prior by checking each individual component for prior-data conflict as this leads to more information about where a conflict exists when it does.

In general, it is not clear how to check the individual components but in certain contexts a particular structure holds that allows for this. Suppose that all ancillaries are independent of the minimal sufficient statistic and so can be ignored. The more general situation is covered in Evans and Moshonov (2006).

As discussed in Evans and Moshonov (2006), suppose there is a statistic $V(T)$ such that the marginal distribution of $V(T)$ is dependent only on θ_1 . Such a statistic is

referred to as being ancillary for θ_2 given θ_1 . Naturally we want $V(T)$ to be a maximal ancillary for θ_2 given θ_1 . An appropriate tail probability for checking π_1 is then given by

$$M_{V(T)}(m_{V(T)}(v) \leq m_{V(T)}(V(T(x))), \tag{10}$$

as $M_{V(T)}$ does not depend on $\pi_2(\cdot | \theta_1)$. A natural order is to check π_1 first and then check $\pi_2(\cdot | \theta_1)$ for prior-data conflict, whenever no prior-data conflict is found for π_1 . The appropriate tail probability for checking $\pi_2(\cdot | \theta_1)$ is given by

$$M_T(m_T(t | V(T(x))) \leq m_T(T(x) | V(T(x))) | V(T(x))). \tag{11}$$

Note that this is assessing whether or not $\pi_2(\cdot | \theta_1)$ is a suitable prior for θ_2 among those θ_1 values deemed to be suitable according to the prior π_1 . If (11) were to be used before (10), then it would not be possible to assess if a failure was due to where π_1 was placing the bulk of its mass or was caused by where the conditional priors were placing their mass. Notice that

$$\frac{m_{Q,T}(T(x))}{m_T(T(x))} = \frac{m_{Q,T}(T(x) | V(T(x)))}{m_T(T(x) | V(T(x)))} \frac{m_{Q,V(T)}(V(T(x)))}{m_{V(T)}(V(T(x)))}, \tag{12}$$

so prior-data conflict with either π_1 or $\pi_2(\cdot | \theta_1)$ could lead to large values of the ratio on the left for certain choices of Q . When only the conditional prior of θ_2 given θ_1 is perturbed, then $m_{Q,V(T)}(V(T(x))) = m_{V(T)}(V(T(x)))$.

Letting $f_{\theta_1,V}$ denote the density of V , then

$$\begin{aligned} & \frac{m_{Q,V(T)}(V(T(x)))}{m_{V(T)}(V(T(x)))} \\ &= \int_{\Theta_1} \frac{f_{\theta_1,V}(V(T(x)))}{m_{V(T)}(V(T(x)))} Q_1(d\theta_1) \leq RB_1(\theta_1(V(T(x))) | V(T(x))) \end{aligned}$$

where $RB_1(\cdot | V(T(x)))$ gives the relative belief ratios for θ_1 based on having observed $V(T(x))$. The right-hand side gives the worst-case behavior of the second factor in (12).

Now consider the robustness of relative belief inferences for a general $\psi = \Psi(\theta)$. The following result generalizes Propositions 4 and 7 as we consider a general perturbation to the prior, namely, $\Pi_\epsilon = (1 - \epsilon)\Pi + \epsilon Q$ and the proof is the same as that of Proposition 7.

Proposition 9. *The Gâteaux derivative of $RB_{\Psi}(\cdot | x)$ at ψ in the direction Q is*

$$\{m_Q(x)/m(x)\}(RB_{Q,\Psi}(\psi | x) - RB_{\Psi}(\psi | x)).$$

The factor $RB_{Q,\Psi}(\psi | x) - RB_{\Psi}(\psi | x)$ can be big simply because we choose a prior Q that is very different than Π . For example, $RB_{\Psi}(\psi | x)$ may be big (small) because there is considerable evidence in favor of (against) ψ being the true value and we can choose a prior Q that doesn't (does) place mass near ψ . As such, it makes sense to standardize the derivative by dividing by this factor and this leaves the robustness determined again by $m_Q(x)/m(x)$.

Suppose now that Q and Π have the same marginal for $\zeta = \Xi(\theta)$. Then, $m_Q(x) = m(x) \int_{\Xi} \int_{\Xi^{-1}\{\zeta\}} RB(\theta|x) Q(d\theta|\psi) \Pi_{\Xi}(d\zeta) \leq m(x) \int_{\Xi} RB(\theta_{\zeta}(x)|x) \Pi_{\Xi}(d\zeta)$ where $\theta_{\zeta}(x) = \arg \sup\{RB(\theta|x) : \Xi(\theta) = \zeta\}$. Therefore,

$$\frac{m_Q(x)}{m(x)} \leq \int_{\Xi} RB(\theta_{\zeta}(x)|x) \Pi_{\Xi}(d\zeta) \tag{13}$$

and the right-hand side gives the worst-case behavior of the first factor in (12) when $\Xi(\theta) = \theta_1$, which is related to prior-data conflict with the prior on θ_2 .

The following is a standard example where priors are specified hierarchically.

Example 5. *Location-scale normal model.*

Suppose that $x = (x_1, \dots, x_n)$ is a sample from the $N(\mu, \sigma^2)$ distribution with $\mu|\sigma^2 \sim N(\mu_0, \tau_0^2\sigma^2), \sigma^{-2} \sim \text{gamma}_{rate}(\alpha_0, \beta_0)$ ($X \sim \text{gamma}_{rate}(\alpha_0, \beta_0)$ when the density of X equals $f(x) = \Gamma^{-1}(\alpha_0)\beta_0^{\alpha_0}x^{\alpha_0-1}e^{-\beta_0x}$ when $x > 0$ and 0 otherwise). Then $T(x) = (\bar{x}, \|x - \bar{x}1\|^2)$ is a minimal sufficient statistic for the model. Note that the prior is chosen by eliciting values for $\mu_0, \tau_0^2, \alpha_0, \beta_0$ and so there is interest in how sensitive inferences are to perturbations in each component separately. The posterior distribution of (μ, σ^2) is given by $\mu|\sigma^2, T(x) \sim N(\mu_x, (n + 1/\tau_0^2)^{-1}\sigma^2), \sigma^{-2}|T(x) \sim \text{gamma}_{rate}(\alpha_0 + n/2, \beta(\bar{x}, s^2))$ where $\mu_x = (n + 1/\tau_0^2)^{-1}(n\bar{x} + \mu_0/\tau_0^2)$ and $\beta(\bar{x}, s^2) = \beta_0 + (n - 1)s^2/2 + n(\bar{x} - \mu_0)^2/2(n\tau_0^2 + 1)$ with $s^2 = \|x - \bar{x}1\|^2/(n - 1)$.

Consider first inferences for $\psi = \Psi(\theta) = \sigma^2$ and note that $V(T(x)) = \|x - \bar{x}1\|^2$ is ancillary given ψ and its distribution depends on ψ . Therefore, the prior on σ^2 is checked first using the prior predictive for $V(T(x))$. An easy calculation gives that the prior distribution of $s^2 = V(T(x))/(n - 1)$ is $(\beta_0/\alpha_0)F(n - 1, 2\alpha_0)$ and this specifies (10). While the results of Section 4.1 apply here, consider the behavior of the relative belief ratio $RB_1(\sigma^2|V(T(x)))$ which is based on only observing $V(T(x))$ rather than $T(x)$. By Proposition 4 this has Gâteaux derivative depending on $m_{Q,V(T)}(V(T(x)))/m_{V(T)}(V(T(x)))$. Notice, however, that relative belief ratios accumulate evidence in a simple way. For any statistic $V(T(x))$, then

$$RB_{\Psi}(\psi|T(x)) = \frac{\pi_{\Psi}(\psi|T(x))}{\pi_{\Psi}(\psi)} = \frac{\pi_{\Psi}(\psi|V(T(x)))}{\pi_{\Psi}(\psi)} \frac{\pi_{\Psi}(\psi|T(x))}{\pi_{\Psi}(\psi|V(T(x)))}$$

where the first factor gives the evidence obtained after observing $V(T(x))$ and the second factor gives the evidence obtained after observing $T(x)$ having already observed $V(T(x))$. So $RB_1(\sigma^2|x) = RB_1(\sigma^2|V(T(x)))[RB_1(\sigma^2|T(x))/RB_1(\sigma^2|V(T(x)))]$ with the same interpretation for the factors. As such, a lack of robustness of $RB_1(\sigma^2|V(T(x)))$, which can be connected to prior-data conflict through (10), implies a lack of robustness for $RB_1(\sigma^2|x)$.

When no prior-data conflict is obtained for the prior on σ^2 , then it makes sense to look for prior-data conflict with the prior on μ which is typically the parameter of primary interest. So now consider perturbations to the prior on μ and the relationship to prior-data conflict with this prior. The conditional distribution of $T(x)$ given $V(T(x))$ is given by the conditional prior predictive of \bar{x} given s^2 which is distributed as $\mu_0 +$

| α_1 | β_1 | $\frac{m_{Q,V(T)}(V(T(x)))}{m_{V(T)}(V(T(x)))}$ | α_1 | β_1 | $\frac{m_{Q,V(T)}(V(T(x)))}{m_{V(T)}(V(T(x)))}$ |
|------------|-----------|---|------------|-----------|---|
| 5 | 1 | 0.05 | 1 | 5 | 0.07 |
| 5 | 2 | 0.38 | 2 | 5 | 0.25 |
| 5 | 4 | 0.99 | 4 | 5 | 0.81 |
| 5 | 10 | 0.34 | 10 | 5 | 0.53 |

Table 5: The ratio $m_{Q,V(T)}(V(T(x)))/m_{V(T)}(V(T(x)))$ in Example 5 when there is no conflict with the prior on σ^2 .

$\tilde{\sigma}t_{n+2\alpha_0-1}$ where $\tilde{\sigma}^2 = \{\tau_0^2(n\tau_0^2 + 1)(2\beta_0 + (n - 1)s^2) + 1\} / \{n\tau_0^2(n + 2\alpha_0 - 1)\}$ specifying (11). Furthermore, for (13), $\Xi(\theta) = \sigma^2$ and $\theta_{\sigma^2}(x) = (\bar{x}, \sigma^2)$ with

$$RB((\bar{x}, \sigma^2) | x) = \frac{(n\tau_0^2 + 1)^{\frac{1}{2}}}{\beta_0^{\alpha_0}} \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_0 + n/2)} (\beta(\bar{x}, s^2))^{\alpha_0 + \frac{n}{2}} \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\left\{-\frac{(n - 1)s^2}{2\sigma^2}\right\}$$

and so

$$\int_0^\infty RB((\bar{x}, \sigma^2) | x) \Pi_1(d\sigma^{-2}) = (n\tau_0^2 + 1)^{\frac{1}{2}} \left(\frac{\beta(\bar{x}, s^2)}{\beta_0 + (n - 1)s^2/2}\right)^{\alpha_0 + \frac{n}{2}}.$$

Now consider a number of numerical examples where the base prior is always specified by $\mu_0 = 0, \tau_0^2 = 1, \alpha_0 = 5$ and $\beta_0 = 5$. The behavior of the two factors in (12) is examined when there is no prior-data conflict and when there is.

A sample of size $n = 20$ was generated from the $N(0, 1)$ distribution obtaining $\bar{x} = -0.1066, s^2 = 0.9087$. So there should be no prior-data conflict with the prior on σ^2 . Indeed, (10) equals 0.7626 so there is no indication of any problems with the prior on σ^2 . Values of $m_{Q,V(T)}(V(T(x)))/m_{V(T)}(V(T(x)))$ are recorded in Table 5 when the marginal prior on σ^2 is perturbed by a $\text{gamma}_{rate}(\alpha_1, \beta_1)$ distribution for various values of α_1 and β_1 . In all cases, the ratio is small and indicates robustness to local perturbations of the prior on σ^2 . Note that the worst case behavior, over all possible directions, is given by the maximized relative belief ratio for σ^2 based on $V(T(x))$ which occurs at $\sigma^2 = s^2$ and equals

$$RB_1(s^2 | V(T(x))) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_0 + (n - 1)/2)} \beta_0^{-\alpha_0} e^{-\frac{n-1}{2}(s^2) - \frac{n-1}{2}} \left(\frac{(n - 1)s^2}{2} + \beta_0\right)^{\frac{n-1}{2} + \alpha_0}.$$

In this case $RB_1(s^2 | V(T(x))) = 1.7479$.

Next a sample of size $n = 20$ from the $N(0, 25)$ was generated obtaining $\bar{x} = 0.0950, s^2 = 23.9593$. So there is clearly prior-data conflict with the prior on σ^2 . This is reflected in the value of (10) which equals 0.64×10^{-5} . Table 6 shows that there is a serious lack of robustness. The worst case behavior is given by $RB_1(s^2 | V(T(x))) = 40484.68$.

| α_1 | β_1 | $\frac{m_{Q,V(T)}(V(T(x)))}{m_{V(T)}(V(T(x)))}$ | α_1 | β_1 | $\frac{m_{Q,V(T)}(V(T(x)))}{m_{V(T)}(V(T(x)))}$ |
|------------|-----------|---|------------|-----------|---|
| 5 | 1 | 0.00 | 1 | 5 | 5517.42 |
| 5 | 2 | 0.01 | 2 | 5 | 1245.26 |
| 5 | 4 | 2.34 | 4 | 5 | 13.78 |
| 5 | 10 | 23.51 | 10 | 5 | 0.00 |

Table 6: The ratio $m_{Q,V(T)}(V(T(x)))/m_{V(T)}(V(T(x)))$ in Example 5 when there is conflict with the prior on σ^2 .

| α_1 | β_1 | $\frac{m_{Q,V(T)}(V(T(x)))}{m_{V(T)}(V(T(x)))}$ | α_1 | β_1 | $\frac{m_{Q,V(T)}(V(T(x)))}{m_{V(T)}(V(T(x)))}$ |
|------------|-----------|---|------------|-----------|---|
| 5 | 1 | 0.03 | 1 | 5 | 0.09 |
| 5 | 2 | 0.29 | 2 | 5 | 0.31 |
| 5 | 4 | 0.92 | 4 | 5 | 0.86 |
| 5 | 10 | 0.44 | 10 | 5 | 0.38 |

Table 7: The ratio $m_{Q,V(T)}(V(T(x)))/m_{V(T)}(V(T(x)))$ in Example 5 when there is conflict with the prior on μ but not with the prior on σ^2 .

It is also relevant to consider what happens concerning the robustness of inferences about σ^2 when there is prior-data conflict with the prior on μ but not with the prior on σ^2 . A sample of $n = 20$ was generated from the $N(10, 1)$ distribution obtaining $\bar{x} = 9.7041$, $s^2 = 1.0082$, so there is clearly prior-data conflict with the prior on μ but not with the prior on σ^2 . The value of (10) equals 0.6460 which gives no reason to doubt the relevance of the prior on σ^2 . Table 7 shows that $m_{Q,V(T)}(V(T(x)))/m_{V(T)}(V(T(x)))$ is small and indicates robustness to local perturbations of the prior on σ^2 . The worst case behavior is given by $RB_1(s^2 | V(T(x))) = 1.7218$. This reinforces the claim that the tail probabilities (10) and (11) are measuring different aspects of the data conflicting with the prior.

Now consider perturbations to the prior on μ with the prior on σ^2 fixed. A sample of $n = 20$ was generated from a $N(0, 1)$ obtaining $\bar{x} = -0.1066$, $s^2 = 0.9087$ so there is clearly no prior-data conflict with either component. This is reflected in the value of (11) which equals 0.9150. Table 8 shows that the first factor $m_{Q,T}(T(x) | V(T(x)))/m_T(T(x) | V(T(x)))$ in (12) is small when the conditional prior on μ is perturbed by $N(\mu_1, \tau_1^2)$ priors and thus demonstrates robustness to perturbations in these directions. The worst case behavior is given by $\int_0^\infty RB((\bar{x}, \sigma^2) | x) \Pi_1(d\sigma^{-2}) = 4.6099$ which is comparatively small.

| μ_1 | τ_1^2 | $\frac{m_{Q,T}(T(x) V(T(x)))}{m_T(T(x) V(T(x)))}$ | μ_1 | τ_1^2 | $\frac{m_{Q,T}(T(x) V(T(x)))}{m_T(T(x) V(T(x)))}$ |
|---------|------------|---|---------|------------|---|
| -2 | 1 | 0.17 | 0 | 2 | 0.51 |
| -1 | 1 | 0.66 | 0 | 3 | 0.34 |
| 1 | 1 | 0.54 | 0 | 4 | 0.26 |
| 2 | 1 | 0.12 | 0 | 5 | 0.21 |

Table 8: The ratio $m_{Q,T}(T(x) | V(T(x)))/m_T(T(x) | V(T(x)))$ in Example 5 when there is no conflict with the prior on σ^2 or with the prior on μ .

| μ_1 | τ_1^2 | $\frac{m_{Q,T}(T(x) V(T(x)))}{m_T(T(x) V(T(x)))}$ | μ_1 | τ_1^2 | $\frac{m_{Q,T}(T(x) V(T(x)))}{m_T(T(x) V(T(x)))}$ |
|---------|------------|---|---------|------------|---|
| -2 | 1 | 0.87 | 0 | 2 | 0.51 |
| -1 | 1 | 0.96 | 0 | 3 | 0.34 |
| 1 | 1 | 0.98 | 0 | 4 | 0.26 |
| 2 | 1 | 0.90 | 0 | 5 | 0.21 |

Table 9: The ratio $m_{Q,T}(T(x) | V(T(x)))/m_T(T(x) | V(T(x)))$ in Example 5 when there is conflict with the prior on σ^2 but not with the prior on μ .

| μ_1 | τ_1^2 | $\frac{m_{Q,T}(T(x) V(T(x)))}{m_T(T(x) V(T(x)))}$ | μ_1 | τ_1^2 | $\frac{m_{Q,T}(T(x) V(T(x)))}{m_T(T(x) V(T(x)))}$ |
|---------|------------|---|---------|------------|---|
| -2 | 1 | 0.01 | 0 | 2 | 117,584 |
| -1 | 1 | 0.10 | 0 | 3 | 5,611,980 |
| 1 | 1 | 10.83 | 0 | 4 | 26,012,609 |
| 2 | 1 | 132.09 | 0 | 5 | 55,478,630 |

Table 10: The ratio $m_{Q,T}(T(x) | V(T(x)))/m_T(T(x) | V(T(x)))$ in Example 5 when there is no conflict with the prior on σ^2 but there is with the prior on μ .

Table 9 gives some values of $m_{Q,T}(T(x) | V(T(x)))/m_T(T(x) | V(T(x)))$ when a sample of $n = 20$ was generated from a $N(0, 25)$, obtaining $\bar{x} = 0.0950, s^2 = 23.9593$. So in this case there is prior-data conflict with the prior on σ^2 but not with the prior on μ . The value of (11) equals 0.9150 which gives no indication of prior-data conflict with the prior on μ . The tabulated values also indicate no serious robustness concerns as does $\int_0^\infty RB((\bar{x}, \sigma^2) | x) \Pi_1(d\sigma^{-2}) = 4.5838$. This also reinforces the claim that the tail probabilities (10) and (11) are measuring different aspects of the data conflicting with the prior.

Table 10 gives some values of $m_{Q,T}(T(x) | V(T(x)))/m_T(T(x) | V(T(x)))$ when a sample of $n = 20$ was generated from a $N(10, 1)$ obtaining $\bar{x} = 9.7941, s^2 = 1.0082$. So in this case there is prior-data conflict with the prior on μ but not with the prior on σ^2 . The value of (11) equals 0.1691×10^{-9} which gives a clear indication of prior-data conflict with the prior on μ . In this case the tabulated values indicate a clear lack of robustness with respect to the prior on μ . Also, $\int_0^\infty RB((\bar{x}, \sigma^2) | x) \Pi_1(d\sigma^{-2}) = 8,046,933,962$ indicates that the worst case behavior with respect to robustness is terrible.

Note that $RB(\theta(x) | x)$ and the similar ratios discussed in this section, do not measure robustness directly through the values they assume. Rather, the results of this section show that, when prior-data conflict exists, these ratios tend to be larger than when there is no prior-data conflict. Since these values influence relevant derivatives, this indicates an increased sensitivity to the choice of the prior when such a conflict arises. As such these ratios are not to be interpreted as measures of surprise as discussed in Section 4.7.2 of Berger (1985).

6 Conclusions

Several optimal robustness results have been derived here for relative belief inferences. These results provide support for these inferences for estimation and hypothesis assess-

ment. Even though relative belief inferences have optimal robustness properties with respect to choice of prior, this does not guarantee that they are robust in practice. The issue of practical robustness in a given problem is seen to be connected with whether or not there is prior-data conflict. With no prior-data conflict the inferences are robust to small changes in the prior, at least in the sense measured here. This adds support to the point-of-view that checking for prior-data conflict is an essential aspect of good statistical practice.

It is interesting that the worst case behavior of the measure of sensitivity is associated with the maximized value of a relative belief ratio. The actual maximum value attained is meaningless, however, as there is no way to calibrate this as opposed to calibrating the relative belief ratio at a fixed value via the strength. The relative belief estimate is consistent, however, and the relative belief ratio at this value will, at least in the continuous case, converge to infinity. So large values would seem to be associated with high evidence in favor. What has been shown here is that large values can be associated with prior-data conflict and a lack of robustness rather than providing high evidence. When prior-data conflict is encountered the prior can be modified, following Evans and Jang (2011c), to avoid this. While objections can be raised to taking such a step, it seems necessary if we want to report a valid characterization of the evidence obtained.

References

- Baskurt, Z. and Evans, M. (2013). “Hypothesis assessment and inequalities for Bayes factors and relative belief ratios.” *Bayesian Analysis*, 8(3): 569–590. MR3102226. doi: <http://dx.doi.org/10.1214/13-BA824>. 706, 708, 709, 717
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer, second edition. MR0804611. doi: <http://dx.doi.org/10.1007/978-1-4757-4286-2>. 725
- Berger, J. O. (1990). “Robust Bayesian analysis: sensitivity to the prior.” *Journal of Statistical Planning and Inference*, 25: 303–328. MR1064429. doi: [http://dx.doi.org/10.1016/0378-3758\(90\)90079-A](http://dx.doi.org/10.1016/0378-3758(90)90079-A). 707
- Berger, J. O. (1994). “An overview of robust Bayesian analysis (with discussion).” *Test*, 3: 5–124. MR1293110. doi: <http://dx.doi.org/10.1007/BF02562676>. 707
- Berger, J. O., Liseo, B., and Wolpert, R. L. (1999). “Integrated likelihood methods for eliminating nuisance parameters.” *Statistical Science*, 14(1): 1–28. MR1702200. doi: <http://dx.doi.org/10.1214/ss/1009211803>. 709
- Dahl, F. A., Gasemyr, J., and Natvig, B. (2007). “A robust conflict measure of inconsistencies in Bayesian hierarchical models.” *Scandinavian Journal of Statistics*, 34: 816–828. MR2396940. doi: <http://dx.doi.org/10.1111/j.1467-9469.2007.00560.x>. 707
- de la Horra, J. and Fernandez, C. (1994). “Bayesian analysis under ϵ -contaminated priors: a trade-off between robustness and precision.” *Journal of Statistical Planning*

- and Inference*, 38: 13–30. MR1256845. doi: [http://dx.doi.org/10.1016/0378-3758\(92\)00152-T](http://dx.doi.org/10.1016/0378-3758(92)00152-T). 707, 710
- Dey, D. K. and Birmiwall, L. R. (1994). “Robust Bayesian analysis using divergence measures.” *Statistics and Probability Letters*, 20: 287–294. MR1294607. doi: [http://dx.doi.org/10.1016/0167-7152\(94\)90016-7](http://dx.doi.org/10.1016/0167-7152(94)90016-7). 707
- Evans, M. (2015). *Measuring Statistical Evidence Using Relative Belief*. CRC Press, Taylor & Francis Group. 706, 707
- Evans, M., Guttman, I., and Swartz, T. (2006). “Optimality and computations for relative surprise inferences.” *Canadian Journal of Statistics*, 34(1): 113–129. MR2267713. doi: <http://dx.doi.org/10.1002/cjs.5550340109>. 707, 708
- Evans, M. and Jang, G. H. (2011a). “Inferences from prior-based loss functions.” Technical Report, [arXiv.1104.3258](https://arxiv.org/abs/1104.3258). 708
- Evans, M. and Jang, G. H. (2011b). “A limit result for the prior predictive.” *Statistics and Probability Letters*, 81: 1034–1038. MR2803740. doi: <http://dx.doi.org/10.1016/j.spl.2011.02.025>. 707, 718
- Evans, M. and Jang, G. H. (2011c). “Weak informativity and the information in one prior relative to another.” *Statistical Science*, 26(3): 423–439. MR2917964. doi: <http://dx.doi.org/10.1214/11-STS357>. 707, 726
- Evans, M. and Moshonov, H. (2006). “Checking for prior-data conflict.” *Bayesian Analysis*, 1(4): 893–914. MR2282210. doi: <http://dx.doi.org/10.1016/j.spl.2011.02.025>. 707, 718, 720
- Evans, M. and Shakhatreh, M. (2008). “Optimal properties of some Bayesian inferences.” *Electronic Journal of Statistics*, 2: 1268–1280. MR2471286. doi: <http://dx.doi.org/10.1214/07-EJS126>. 708
- Evans, M. and Zou, T. (2001). “Robustness of relative surprise inferences to choice of prior.” In Chaubey, Y. (ed.), *Recent Advances in Statistical Methods, Proceedings of Statistics 2001 Canada: The 4th Conference in Applied Statistics*, 90–115. Imperial College Press. MR2023214. doi: http://dx.doi.org/10.1142/9781860949531_0008. 715
- Federer, S. G. (1969). *Geometric Measure Theory*. Berlin: Springer-Verlag. 706
- Hirsch, M. W. (1976). *Differential Topology*. New York: Springer-Verlag. MR0448362. 706
- Huber, P. J. (1973). “The use of Choquet capacities in statistics.” *Bulletin of the International Statistical Institute*, 45: 181–191. MR0356322. 710
- Krantz, S. G. and Parks, H. R. (2008). *Geometric Integration Theory*. Boston: Birkhäuser. MR2427002. doi: <http://dx.doi.org/10.1007/978-0-8176-4679-0>. 706
- Liseo, B. (1996). “A note on the concept of robust likelihoods.” *Metron*, LIV: 25–38. 709, 712

- Marshall, E. C. and Spiegelhalter, D. J. (2007). “Identifying outliers in Bayesian hierarchical models: a simulation-based approach.” *Bayesian Analysis*, 2: 409–444. MR2312289. doi: <http://dx.doi.org/10.1214/07-BA218>. 707
- O’Hagan, A. (2003). “HSS model criticism (with discussion).” In *Highly Structured Stochastic Systems*. Oxford University Press. MR2082418. 707
- Piccinato, L. (1984). “A Bayesian property of likelihood sets.” *Statistica*, 2: 197–204. MR0759306. 708, 709
- Presanis, A. M., Ohlssen, D., Spiegelhalter, D. J., and Angelis, D. D. (2013). “Conflict diagnostics in directed acyclic graphs, with applications in Bayesian evidence synthesis.” *Statistical Science*, 28: 376–397. MR3135538. doi: <http://dx.doi.org/10.1214/13-STS426>. 707
- Rios Insua, D. and Ruggeri, F. (2000). *Robust Bayesian Analysis*. Springer-Verlag. MR1795206. doi: <http://dx.doi.org/10.1007/978-1-4612-1306-2>. 707
- Royall, R. M. and Tsou, T. (2003). “Interpreting statistical evidence by using imperfect models: robust adjusted likelihood functions.” *Journal of the Royal Statistical Society. Series B.*, 65(2): 391–404. MR1983754. doi: <http://dx.doi.org/10.1111/1467-9868.00392>. 707
- Rudin, W. (1974). *Real and Complex Analysis*, New York: McGraw-Hill, second edition. MR0344043. 706
- Ruggeri, F. and Wasserman, L. (1993). “Infinitesimal sensitivity of posterior distributions.” *The Canadian Journal of Statistics*, 21(2): 195–203. MR1234761. doi: <http://dx.doi.org/10.2307/3315811>. 707
- Scheel, I., Green, P. J., and Rougier, J. C. (2011). “A graphical diagnostic for identifying influential model choices in Bayesian hierarchical models.” *Scandinavian Journal of Statistics*, 38(3): 529–550. MR2833845. doi: <http://dx.doi.org/10.1111/j.1467-9469.2010.00717.x>. 707
- Sivaganesan, S., Berliner, M., and Berger, J. O. (1993). “Optimal robust credible sets for contaminated priors.” *Statistics and Probability Letters*, 18(5): 383–388. MR1247450. doi: [http://dx.doi.org/10.1016/0167-7152\(93\)90032-E](http://dx.doi.org/10.1016/0167-7152(93)90032-E). 707
- Tjur, T. (1974). *Conditional Probability Distributions*. Institute of Mathematical Statistics, University of Copenhagen. MR0345151. 706
- Wasserman, L. (1989). “A robust Bayesian interpretation of likelihood regions.” *Annals of Statistics*, 17(3): 1387–1393. MR1015159. doi: <http://dx.doi.org/10.1214/aos/1176347277>. 707, 710