BAYESIAN ROBUSTNESS WITH RESPECT TO THE LIKELIHOOD

by

N. D. Shyamalkumar
Purdue University

Technical Report #96-23

Department of Statistics
Purdue University
West Lafayette, IN USA

June 1996

BAYESIAN ROBUSTNESS WITH RESPECT TO THE LIKELIHOOD

N. D. Shyamalkumar

Department Of Statistics, Purdue University, U.S.A.*

Abstract

In this paper we approach the problem of studying sensitivity of posterior quantities to uncertainty in the likelihood. The papers of Lavine (1991) and Pericchi & Perez (1994) are among those which have considered this problem. We suggest an approach which applies in situations where the uncertainty in the likelihood can be modeled as a class which admits a Choquet type representation. An example which falls in the above category would be when the uncertainty can be modeled as an ϵ-contamination class with, say, the base model being normal with unknown location and we allow the contaminants to belong to the class of all symmetric unimodal location models. The mathematical problem reduces to evaluation of extrema of the expectation of a real valued function of n variables, the joint law of the n variables being allowed to vary over all iid laws. Our approach is to get lower bounds for the infimum by evaluating the infimum over more tractable larger classes. These nested classes are those formed by the lower dimensional marginals of N-exchangeable laws, the classes decreasing with increasing N. We show that these infimums do approach the infimum we sought out to evaluate.

KEY WORDS: Sensitivity, ϵ-contamination, Symmetric unimodal.

*Research supported by Frederick N. Andrews fellowship and partially supported by the National Science Foundation, Grants DMS-8923071 and DMS-9303556.

1 Introduction

In the theory of Bayesian Robustness, one studies the effect of the imprecisions in the elicited quantities on the posterior measures of interest. Most of the papers in this area deal with imprecisions solely involving the prior, the primary reason being that the prior is perceived as the weakest link in the Bayesian approach; see Berger (1985). But it is equally important to study the sensitivity of the posterior quantities with respect to the likelihood.

Box & Tiao (1962) and Pericchi & Perez (1994) study robustness with respect to the likelihood, but both are in some sense parametric. Box & Tiao (1962) embeds the normal family in a larger parametric class by the addition of a new parameter. Pericchi & Perez (1994) study robustness by finding the variation in the posterior quantities when the likelihood takes one of a few of the standard functional forms.

Much different from the above approaches is that of Lavine (1991), where for the first time, the class of models considered was nonparametric. Actually, DeRobertis (1978) does mention a non-parametric neighborhood for likelihoods, but very briefly. By non-parametric we mean that the parameter space is infinite dimensional. Also Lavine treats both the prior and likelihood in one stroke. Let $\{P_\theta : \theta \in \Theta\}$ be the initial choice for the model. Lavine looks at the prior , say π, as the law of the random variable $(\boldsymbol{\theta}, \mathbf{P})$, where $\boldsymbol{\theta}$ takes values in Θ and $\mathbf{P}$ takes values in $N(\boldsymbol{\theta})$, where $N(\theta)$ is a neighborhood of the probability P_θ. For example, if the model is known to be $\{P_\theta : \theta \in \Theta\}$, then the usual prior, say π_0, can be considered as giving $\mathbf{P}$ the degenerate conditional distribution at P_θ given $\boldsymbol{\theta} = \theta$. The law of $(\boldsymbol{\theta}, \mathbf{P})$ is allowed to vary in a class, which is defined by first selecting the neighborhoods $\{N(\theta) : \theta \in \Theta\}$. The conditional law of $\mathbf{P}$ given $\boldsymbol{\theta}$ is allowed to vary over all laws which assign probability one to $N(\boldsymbol{\theta})$. The marginal law of $\boldsymbol{\theta}$ is allowed to vary in any standard classes that is used in the study of Bayesian Robustness with respect to the prior. Lavine observes that whenever the neighborhoods do not bound densities away from both 0 and ∞, the bounds achieved are trivial. Hence they suggest using density bounded neighborhoods.

As he notes, the extrema are attained at degenerate conditionals. In this case, there is a sort of model at which the extremum is attained. But these models might violate some reasonable conditions that one might expect any plausible model to satisfy. It seems that this inclusion of implausible models does affect the results. However we can avoid this, at least in some cases.

In this paper we deal with a restricted class of problems, unlike Lavine(1991). Even though it might be possible to study variations in the prior, along with that of the likelihood, we restrict ourselves to just variations in the likelihood. The class of problems that we consider includes, for example, the problem when the uncertainty can be modeled as an ϵ-contamination class with, say, the base model being normal with unknown location, and the contaminants are allowed to belong to the class of all symmetric unimodal location models. Note that, in this example, the densities are not bounded away from ∞.

The main obstacle that one faces, in solving the extremal problem with likelihood robustness involving classes of likelihoods as discussed above, is the problem of obtaining the extrema of the expectation of a real valued function defined on $\mathbb{R}^n$ over the class of all probability measures on $\mathbb{R}^n$ which make the coordinate mappings i.i.d.. This problem cannot be simplified, as any such probability measure is an extreme point of this subset of probability measures. What we do is find a bigger class whose extreme points are simple enough to facilitate the evaluation of the extrema, and small enough for the extrema to be good approximation to the extrema we sought, in the first place. For this, we use the class of finite exchangeable probability measures. This method is also useful in other problems, apart from those arising from likelihood robustness considerations. One such, in the area of prior robustness, will be given as an example.

2 Notations

Throughout this paper we shall be working with finite dimensional Euclidean spaces for the sake of mathematical simplicity though the results hold in the generality of standard Borel spaces.

Let $(\mathcal{R}^n, \mathcal{B}^n)$ denote the n-dimensional Euclidean space with its associated Borel σ field. By $\mathcal{R}_n^\star$ we shall denote the set of all probabilities on $(\mathcal{R}^n, \mathcal{B}^n)$ and let $\mathcal{I}_n$ denote the subset of all probabilities on $(\mathcal{R}^n, \mathcal{B}^n)$ which make the coordinate mappings i.i.d.. By $\mathcal{P}_n$ we denote the subset of all product probabilities. We shall endow $\mathcal{R}_n^\star$ with its Borel σ field with respect to the weak star topology, denoted by $\mathcal{B}_n^\star$. Note that this is also the same σ field as generated by mappings of the type $P \longmapsto P(A)$ where $A \in \mathcal{B}^n$. It is easy to check that $\mathcal{I}_n, \mathcal{P}_n$ are indeed measurable with respect to this σ field.

By S_n we denote the permutation group on the first n natural numbers. For $\underset{\sim}{x} \in \mathcal{R}^n$ we denote by $\underset{\sim}{x}_\sigma$, for $\sigma \in S_n$, the element $(x_{\sigma(1)}, \ldots., x_{\sigma(n)})$ of $\mathcal{R}^n$. For ψ a real valued function on $(\mathcal{R}^n, \mathcal{B}^n)$ we denote by $\tilde{\psi}_N$, for $N \geq n$, the function on $(\mathcal{R}^N, \mathcal{B}^N)$ defined by $\tilde{\psi}_N(\underset{\sim}{x}) = \sum_{\sigma \in S_N} \psi^e(\underset{\sim}{x}_\sigma)$, where ψ^e is the function on $(\mathcal{R}^N, \mathcal{B}^N)$ defined by $\psi^e(\underset{\sim}{x}) = \psi((x_1, \ldots.., x_n))$. In the case when $N = n$ we shall write $\tilde{\psi}$ instead of $\tilde{\psi}_N$.

We shall call a $P \in \mathcal{R}_n^\star$ n-exchangeable if the coordinate mappings form an n-exchangeable sequence. Recall that $X_1, \ldots, X_n$, n real valued random variables, is said to be an n-exchangeable sequence if the law of $(X_1, \ldots, X_n)$ is same as that of $(X_{\sigma(1)}, \ldots, X_{\sigma(n)})$ for all $\sigma \in S_n$ Let $\mathcal{E}_n$ denote the set of all n-exchangeable probabilities on $(\mathcal{R}^n, \mathcal{B}^n)$. It is easy to check the measurability of $\mathcal{E}_n$. For the theory of exchangeability, see Aldous(1985).

3 Main Result

We start by stating some preliminary results.

Lemma 3.1 *Let $\mathcal{Q}$ be a measurable subset of $\mathcal{R}_n^\star$ and ϕ, ψ be two bounded real valued functions on $(\mathcal{R}^n, \mathcal{B}^n)$. Also assume that $\phi \geq 0$. Then*

$$\sup_{P \in \mathcal{Q}} \frac{\int \psi dP}{\int \phi dP} = \sup_{\mu \in S(\mathcal{Q})} \frac{\int \int \psi dP d\mu(P)}{\int \int \phi dP d\mu(P)}$$

where $S(\mathcal{Q}) = \{\mu : \mu \in \mathcal{R}_n^{\star\star} \text{ and } \mu(\mathcal{Q}) = 1\}$. The notation $\mathcal{R}_n^{\star\star}$ is represents the space of probability measures on $\mathcal{R}^\star$..

The above lemma is essentially the same as given in Sivaganesan and Berger (1989).

Lemma 3.2 *Let $H_{\underset{\sim}{y}}$ represent the probability measure associated with n draws without replacement from the set $\{y_1, ..., y_n\}$. Then*

$$\mathcal{E}_n = \{P : P = \int H_{\underset{\sim}{y}} dQ(\underset{\sim}{y}), Q \in \mathcal{R}_n^\star\}.$$

Proof. Follows from the definition of $\mathcal{E}_n$. □

Lemma 3.3 *Let ψ and ϕ be two bounded real valued functions on $(\mathcal{R}^n, \mathcal{B}^n)$ such that $\phi \geq 0$. Then*

$$\sup_{P \in \mathcal{E}_n} \frac{\int \psi dP}{\int \phi dP} = \sup_{\underset{\sim}{x} \in \mathcal{R}^n} \frac{\tilde{\psi}(\underset{\sim}{x})}{\tilde{\phi}(\underset{\sim}{x})}.$$

Proof. Follows as an easy application of the above lemmas. □

Our main result is based on a theorem of Diaconis & Freedman (1980) which we state below. But before that we need a few more notations.

For a $P \in \mathcal{E}_n$ and $k \leq n$, we denote the k-dimensional marginal of P by P_k. For a $\mu \in \mathcal{R}_n^{\star\star}$, we denote by P_μ its barycentre. The notation $P_{\mu k}$ represents the k-dimensional marginal of P_μ, when it is well defined. The variation distance $\|P - Q\|$ is defined as usual:

$$\|P - Q\| = 2 \cdot \sup_{A \in \mathcal{B}^n} |P(A) - Q(A)|.$$

Theorem 3.1 *Let $P \in \mathcal{E}_n$. Then there exists a probability $\mu \in \mathcal{R}_n^{\star\star}$ satisfying $\mu(\mathcal{I}_n) = 1$ such that*

$$\|P_k - P_{\mu k}\| \le 2\beta(n,k), \quad \forall k \le n$$

where $\beta(n,k) = 1 - n^{-k} n!/(n-k)!$. The probability μ depends on n and P but not on k.

It is observed, in Diaconis and Freedman (1980), that $\beta(n,k) \le k(k-1)/2n$. See also Freedman (1977). It was also shown in their paper that the bound in the theorem is sharp. The above theorem was proved for an arbitrary measurable space and, in this case note that $\mathcal{B}_n^\star$ is defined in the alternative way, as mentioned before. Now we are ready to state and prove the main result.

Theorem 3.2 *Let ψ and ϕ be two $[0,1]$ valued functions on $(\mathcal{R}^n, \mathcal{B}^n)$. Assume further that ϕ is bounded below by $c > 0$. Then, for $N \ge n$,*

$$0 \le \sup_{\underset{\sim}{x} \in \mathcal{R}^N} \frac{\tilde{\psi}_N(\underset{\sim}{x})}{\tilde{\phi}_N(\underset{\sim}{x})} - \sup_{P \in \mathcal{I}_n} \frac{\int \psi dP}{\int \phi dP} \le 2\frac{\beta(N,n)}{c^2}.$$

PROOF. Let

$$\sup_{P \in \mathcal{E}_N} \frac{\int \psi^e dP}{\int \phi^e dP} \ge q \ge 0.$$

Then we have

$$\sup_{P \in \mathcal{E}_N} \int (\psi^e - q\phi^e) dP \ge 0.$$

Fix $P \in \mathcal{E}_N$. By the previous theorem we have the existence of a $\mu \in \mathcal{R}_N^{\star\star}$ satisfying $\mu(\mathcal{I}_N) = 1$ such that

$$\left|\int (\psi - q\phi)(dP_n - dP_{\mu n})\right| \le 2\|\psi - q\phi\|_\infty \beta(N,n) \le 2\beta(N,n)/c.$$

Note that $\int (\psi^e - q\phi^e) dP = \int (\psi - q\phi) dP_n$. Hence

$$\begin{aligned} 0 &\le \sup_{P \in \mathcal{E}_N} \int (\psi^e - q\phi^e) dP \\ &\le 2\beta(N,n)/c + \sup_{\mu \in \mathcal{R}_N^{\star\star}} \int (\psi^e - q\phi^e) dP_{\mu n} \\ &\le \sup_{\mu \in \mathcal{R}_N^{\star\star}} \int (\psi - (q - 2\beta(N,n)/c^2)\phi) dP_{\mu n}, \end{aligned}$$

which implies

$$\sup_{\mu \in \mathcal{R}_N^{**}} \frac{\int \psi dP_{\mu n}}{\int \phi dP_{\mu n}} \geq q - 2\beta(N,n)/c^2.$$

But the L.H.S. is the same as

$$\sup_{P \in \mathcal{I}_n} \frac{\int \psi dP}{\int \phi dP}$$

by Lemma 3.1. By Lemma 3.3 it follows that

$$\sup_{P \in \mathcal{E}_N} \frac{\int \psi^e dP}{\int \phi^e dP} = \sup_{\underset{\sim}{x} \in \mathcal{R}^N} \frac{\tilde{\psi}_N(\underset{\sim}{x})}{\tilde{\phi}_N(\underset{\sim}{x})}.$$

Hence the proof. □

4 Application to Bayesian Robustness

Suppose that $Y_1, Y_2, ..., Y_n$ are i.i.d. real valued random variables with common law P_θ, where θ is the unknown parameter taking values in Θ. Although one might feel confident in P_θ, it could be possible to give a suitable class of laws, say $\mathcal{F}_\theta$, such that $P_\theta \in \mathcal{F}_\theta$. We assume that the class admits a Choquet type representation and that the set of extreme points of $\mathcal{F}_\theta$ can be indexed by a real valued parameter, i.e. $\mathcal{E}(\mathcal{F}_\theta) = \{Q_\alpha^\theta : \alpha \in \mathbb{R}\}$. In other words, if $P_\theta \in \mathcal{F}_\theta$, then there exists an $M \in \mathbb{R}^*$ such that

$$P_\theta = \int Q_\alpha^\theta dM(\alpha).$$

We shall assume further that the class $\mathcal{F}_\theta$ is dominated by a σ-finite measure μ for all values of θ. The density of Q_α^θ w.r.t. μ will be denoted by q_α^θ. Denote, by $\mathcal{F}$, the set of all families of probability measures, such that if $\{P_\theta\} \in \mathcal{F}$ then there exists an $M \in \mathbb{R}^*$ such that

$$P_\theta = \int Q_\alpha^\theta dM(\alpha) \ \ \forall\, \theta \in \Theta.$$

Note that $\mathcal{F}$ is a strict subset of $\prod_\theta \mathcal{F}_\theta$. This is not too small a subset as far as statistical applications go, as the above class includes all scale and location families when the minimal assumptions are satisfied. For example, when we restrict attention to the class of all symmetric unimodal location densities, we see that $\mathcal{F}_\theta$, given by

$$\mathcal{F}_\theta = \{P : P = \int Unif_\eta(\cdot - \theta) dM(\eta), M \text{a probability}\},$$

where $Unif_\eta$ denotes the symmetric uniform probability with scale parameter η and median 0, is precisely the family of all uniform distributions symmetric about θ.

The problem that we study is the following. We suppose that the statistician knows that the set of all plausible models is $\mathcal{F}$, although he cannot ascertain which particular element is true. He is interested in knowing how much the posterior inferences will vary under this uncertainty. We assume that the prior on θ, say π, is known. We may also be able to handle uncertainty in the prior in a simultaneous fashion, but at this preliminary stage we will not entertain such uncertainty. We assume that the statistician works with squared error loss. If the statistician is interested in a parametric function, say $g(\theta)$, then the estimate after observing n data points $y_1, y_2, \ldots, y_n$ under model $\mathbf{P}$ is

$$E_{\mathbf{P}}(g(\theta)|\mathbf{y}) = \frac{\int g(\theta) \prod_{i=1}^n p_\theta(y_i)\pi(d\theta)}{\int \prod_{i=1}^n p_\theta(y_i)\pi(d\theta)},$$

where π is the prior and $p_\theta(\cdot)$ is the μ-density of P_θ. Now, if we assume that $\mathbf{P} \in \mathcal{F}$, then the there exists a measure M such that

$$P_\theta = \int Q_\alpha^\theta dM(\alpha) \ \ \forall\, \theta \in \Theta.$$

We can write the above posterior expectation as

$$E_{\mathbf{P}}(g(\theta)|\mathbf{y}) = \frac{\int \cdots \int g(\theta) \prod_{i=1}^n q_{\alpha_i}^\theta(y_i)\pi(d\theta) \prod_{i=1}^n M(d\alpha_i)}{\int \cdots \int \prod_{i=1}^n q_{\alpha_i}^\theta(y_i)\pi(d\theta) \prod_{i=1}^n M(d\alpha_i)}.$$

In the above, if we define

$$G_n^{\mathbf{y}}(\alpha_1, \ldots, \alpha_n) = \int g(\theta) \prod_{i=1}^n q_{\alpha_i}^\theta(y_i)\pi(d\theta)$$

and

$$G_d^{\mathbf{y}}(\alpha_1, \ldots, \alpha_n) = \int \prod_{i=1}^n q_{\alpha_i}^\theta(y_i)\pi(d\theta),$$

then

$$E_{\mathbf{P}}(g(\theta)|\mathbf{y}) = \frac{\int \cdots \int G_n^{\mathbf{y}}(\alpha_1, \ldots, \alpha_n) \prod_{i=1}^n M(d\alpha_i)}{\int \cdots \int G_d^{\mathbf{y}}(\alpha_1, \ldots, \alpha_n) \prod_{i=1}^n M(d\alpha_i)}.$$

Hence

$$\sup_{\mathbf{P}\in\mathcal{F}} E_{\mathbf{P}}(g(\theta)|\mathbf{y}) = \sup_M \frac{\int \cdots \int G_n^{\mathbf{y}}(\alpha_1, \ldots, \alpha_n) \prod_{i=1}^n M(d\alpha_i)}{\int \cdots \int G_d^{\mathbf{y}}(\alpha_1, \ldots, \alpha_n) \prod_{i=1}^n M(d\alpha_i)},$$

and similarly for the infimum. This implies that that the extremal problem can be written in the form of finding extremum over the class of all i.i.d. laws, and hence the results of the previous section are very pertinent. In what follows we show some examples where this method of finding the extremum is applied.

In some cases, as the following example shows, the method gives perfect answers.

Example 1: Let us suppose that X_1, X_2 are two i.i.d. observations from a symmetric unimodal density with unknown form and location parameter. Let the prior on the location parameter be the uniform distribution on (-1,1). We try to find bounds for the posterior probability of the interval (0.0,0.5), after having observed the values 0.25 and 1.0. It follows from the Khintchine representation (Feller (1965)) that the extremum problem can be formulated as finding the extremum, as P varies over $\mathbb{R}^*$, of the quantity

$$\frac{\int I_{[0,0.5]}(\theta)U_1(\theta)U_{\alpha_1}(x_1-\theta)U_{\alpha_2}(x_2-\theta)\ d\theta\ dP(\alpha_1)\ dP(\alpha_2)}{\int U_1(\theta)U_{\alpha_1}(x_1-\theta)U_{\alpha_2}(x_2-\theta)\ d\theta\ dP(\alpha_1)\ dP(\alpha_2)},$$

where $(x_1, x_2) = (0.25, 1.0)$. Here, and in what follows, U_β will represent the density of the symmetric uniform density on the interval $(-\beta, \beta)$. The problem of finding the infimum is trivial, as it is zero, which is seen to be attained at many symmetric uniform distributions, uniform on (-0.50,0.50) being one such. The problem of finding the supremum is not trivial, as we shall see. A lower bound for the supremum is obtained by varying over all degenerate P. This quantity is

$$\sup_{\alpha\ \in\ \mathbf{R}^+} \frac{\int I_{[0,0.5]}(\theta)U_1(\theta)U_\alpha(0.25-\theta)U_\alpha(1.0-\theta)\ d\theta,}{\int U_1(\theta)U_\alpha(0.25-\theta)U_\alpha(1.0-\theta)\ d\theta}$$

which can be easily seen to be 0.5. Similarly, varying over all independent, but not necessarily identical models yields an upper bound for the supremum, which turns out to be 1.0. Next we try to use the method of the previous section, that is we first find the supremum over the set of all 2-exchangeable models. This leads to the evaluation of the following supremum:

$$\sup_{(\alpha_1,\alpha_2)\in\mathbf{R}^+\times\mathbf{R}^+} \frac{\int I_{[0,0.5]}(\theta)U_1(\theta)\sum_{\mathcal{I}}\ U_{\alpha_i}(x_1-\theta)U_{\alpha_j}(x_2-\theta)\ d\theta}{\int U_1(\theta)\sum_{\mathcal{I}}\ U_{\alpha_i}(x_1-\theta)U_{\alpha_j}(x_2-\theta)\ d\theta},$$

where $\mathcal{I} = \{(1,2),(2,1)\}$ and $(x_1, x_2) = (0.25, 1.0)$. It can be seen that this supremum is 2/3, which gives us an upper bound for the original supremum. Now we shall show that a lower bound of the supremum is in fact 2/3, which completes the solution. Let α_S, α_L be two positive real numbers and let $\alpha_S < 0.25$. The model that we assign for X_1 would be $U_{(-Y,Y)}$ where Y takes values α_S, α_L with equal probability. With this assignment the posterior probability is given by,

$$\frac{1+4\alpha_L}{4+6\alpha_L},$$

after some simplifications for large values of α_L. It follows that, as we move α_L to infinity, the posterior probability would approach 2/3 from below. □

Next we shall consider a more traditional statistical problem. Here, unlike the previous example, we do not arrive at the exact infimum but are quite close.

Example 2: Suppose that X_1, X_2, X_3, X_4 are four i.i.d observations from a symmetric unimodal location model. The statistician's guess for the likelihood is the normal likelihood. He models uncertainty in the likelihood by the following nonparametric class of models, which we denote by $\mathcal{M}$. By nonparametric, we mean that there is no obvious finite dimensional parametrization of the models in the class. The class $\mathcal{M}$ is given by

$$\{(P_\theta)_{\{\theta \in \mathbf{R}\}} : P_\theta = 0.9N_\theta + 0.1Q_\theta, (Q_\theta)_{\{\theta \in \mathbf{R}\}} \in \mathcal{M}_{su}\},$$

where the class of models $\mathcal{M}_{su}$ is described below:

$$\mathcal{M}_{su} = \{(Q_\theta)_{\{\theta \in \mathbf{R}\}} : Q_\theta = Q(. - \theta), Q \in \mathcal{Q}\}$$

where $\mathcal{Q}$ is the class of all symmetric unimodal probability measures on the real line. In words, the uncertainty in the likelihood is modeled by an ϵ-contamination class with the base model being normal with unknown location and unit variance and the contaminants allowed to belong to the class of all symmetric unimodal location models. The statistician assigns, as the noninformative prior λ, the Lebesgue measure on the real line. He is interested in a confidence interval for the unknown location and would like to know the infimum of the posterior confidence for the 95% posterior confidence interval derived in the usual manner by assuming that the model is exactly the base normal model. The mathematical and computational problem for evaluating the supremum, though mathematically similar to that of the infimum, is of minimal interest.

Note that, by the Khintchine representation, we know that every symmetric unimodal distribution can be represented as a mixture of symmetric uniforms. So we do have the required Choquet representation.

The desired lower bound on the infimum was computed using the above described method. We have tabulated the results using this method under the column titled 'Exchangeable'. We have also evaluated the infimum of the posterior probability over the class of independent (but not necessarily identical) symmetric unimodal contaminants in order to see the improvement obtained by our method. These results are tabulated under the column titled 'Independent'.

$\mathbf{x} = (x_1, x_2, x_3, x_4)$	Independent	Exchangeable
(-1.0,-0.5,0.5,1.0)	0.886	0.916
(-2.0,-1.0,1.0,2.0)	0.742	0.757
(-1.0,0.0,0.1,3.0)	0.670	0.718

Table 1: Infimum Posterior Probability of the 95% Confidence Interval.

The numbers in the second column were seen to be very close to the infimum that would be obtained using a mixing distribution with mass at three points. □

The following is an example of a problem in Bayesian robustness with respect to the prior where the method of the paper is also useful.

Example 3: This example is from Berger & Moreno (1991). Let $\mathbf{X} = (X_1, X_2)$ be a $\mathcal{N}_2((\theta_1, \theta_2), I)$ random variable, and suppose the base prior density π_0 of the form

$$\pi_0(\theta_1, \theta_2) = \pi_{01}(\theta_1)\pi_{02}(\theta_2)$$

is elicited, where π_{0i} are the $\mathcal{N}(0, 2)$ densities. We are interested in studying the robustness of the posterior probability of $H_0 : \theta_1 < \theta_2$ to departures from the base prior. In Berger & Moreno (1991), they study the sensitivity of the above probability in the class Γ_{IC} which is defined as

$$\Gamma_{IC} = \{\pi : \pi(\theta_1, \theta_2) = [(1-\epsilon_1)\pi_{01}(\theta_1) + \epsilon_1 q_1(\theta_1)][(1-\epsilon_2)\pi_{02}(\theta_2) + \epsilon_2 q_2(\theta_2)], q_1, q_2 \in \mathcal{Q}\},$$

where $\mathcal{Q}$ denotes the set of all probability measures. That is they assume prior independence to hold. They also have tabulated the results when you assume virtually nothing about the contaminants, i.e. when you model uncertainty by Γ_C, which is given by

$$\Gamma_C = \{\pi(\theta_1, \theta_2) : \pi(\theta_1, \theta_2) = (1-\epsilon)\pi_0(\theta_1, \theta_2) + \epsilon q(\theta_1, \theta_2), q \in \mathcal{Q}\}$$

Here, we find the ranges of the posterior probability of the above set when we model the uncertainty by Γ_E, where

$$\Gamma_E = \{\pi(\theta_1, \theta_2) : \pi(\theta_1, \theta_2) = (1-\epsilon)\pi_0(\theta_1, \theta_2) + \epsilon q(\theta_1) q(\theta_2), q \in \mathcal{Q}\},$$

or by Γ_{II}, where

$$\Gamma_{II} = \{\pi(\theta_1, \theta_2) : \pi(\theta_1, \theta_2) = [(1-\epsilon_1)\pi_{01}(\theta_1) + \epsilon_1 q(\theta_1)][(1-\epsilon_2)\pi_{02}(\theta_2) + \epsilon_2 q(\theta_2)], q \in \mathcal{Q}\}.$$

As is clear, these classes assume the identical nature of the prior opinion. It is clear that Γ_{II} is a subset of Γ_{IC} and that Γ_E is a subset of Γ_C. It is also easy to see that Γ_C and Γ_{IC} have only π_0 in common. We assign the value 0.106 to both ϵ_1 and ϵ_2 and the value 0.2 to ϵ. For the reason behind the odd value of ϵ_i, see Berger & Moreno (1991).

These new classes are meaningful when apriori it is felt that the distributions of the two parameters are the same. By imposing this condition, one also brings in a sort of objectivity to the bounds.

From table 2 it is clear that assumption of identical marginals does shrink the bounds dramatically when the data is not too supportive of H_0.

The notations used in the table below are; $\underline{P}_C = \inf_{\pi\in\Gamma_C} P^\pi(H_{\pi_0}|\mathbf{x})$, $\overline{P}_C = \sup_{\pi\in\Gamma_C} P^\pi(H_{\pi_0}|\mathbf{x})$, $\underline{P}_E = \inf_{\pi\in\Gamma_E} P^\pi(H_{\pi_0}|\mathbf{x})$, $\overline{P}_E = \sup_{\pi\in\Gamma_E} P^\pi(H_{\pi_0}|\mathbf{x})$, $\underline{P}_{IC} = \inf_{\pi\in\Gamma_{IC}} P^\pi(H_{\pi_0}|\mathbf{x})$, $\overline{P}_{IC} = \sup_{\pi\in\Gamma_{IC}} P^\pi(H_{\pi_0}|\mathbf{x})$, $\underline{P}_{II} = \inf_{\pi\in\Gamma_{II}} P^\pi(H_{\pi_0}|\mathbf{x})$, and $\overline{P}_{II} = \sup_{\pi\in\Gamma_{II}} P^\pi(H_{\pi_0}|\mathbf{x})$.

$\mathbf{x} = (x_1, x_2)$	$P^{\pi_0}(H_0\|x)$	$(\underline{P}_C, \overline{P}_C)$	$(\underline{P}_E, \overline{P}_E)$	$(\underline{P}_{IC}, \overline{P}_{IC})$	$(\underline{P}_{II}, \overline{P}_{II})$
(-0.1,0.1)	0.546	(.313,.740)	(.527,.552)	(.408,.680)	(.537,.557)
(-0.2,0.2)	0.591	(.342,.768)	(.553,.602)	(.508,.672)	(0.573,0.613)
(-0.5,0.5)	0.718	(.439,.845)	(.633,.740)	(.646,.786)	(.679,.763)
(-0.8,0.8)	0.822	(.552,.908)	(.716,.845)	(.766,.873)	(.772,.870)
(-1.0,1.0)	0.876	(.632,.924)	(.773,.888)	(.792,.934)	(0.826,0.919)
(-1.5,1.5)	0.958	(.821,.973)	(.894,.963)	(.913,.981)	(.925,.979)

Table 2: Ranges of the posterior probability of H_0

Acknowledgement

I would like to thank my advisor, Prof. James O. Berger, for suggesting to think about the likelihood robustness problem and continued encouragement and support without which this paper would not have seen it's present form. Thanks to Prof. Burgess Davis for some very useful discussion.

References

Aldous, D. J. (1985). Exchangeability and Related Topics. *Ecole dEte de Probabilités de Saint-Flour XIII-1983*, Lecture Notes in Mathematics 1117, Springer Verlag.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, Springer Verlag, New York, second edition.

Berger, J. O. and Moreno E. (1992). BAYESIAN ROBUSTNESS IN BIDIMENSIONAL MODELS: PRIOR INDEPENDENCE. Technical Report No. 92-52C, Purdue University.

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Reading, Mass., Addison-Wesley Pub. Co.

Diaconis, P. and Freedman, D. (1980). Finite Exchangeable Sequences. *Ann. Probab.* **8**, 745-764.

Feller, W. (1965). *An Introduction To Probability Theory And Its Applications Vol. II*, Wiley Eastern Limited, New Delhi.

Freedman, D. (1977). A Remark on the Difference Between Sampling With and Without Replacement. *J. Amer. Statist. Assoc.* **72**, 681.

Lavine, M. (1992). Sensitivity in Bayesian statistics: the prior and the likelihood. *J. Amer. Statist. Assoc.* **86**, 1222-1235.

Pericchi, L. R. and Perez, M. E. (1994). Posterior robustness with more than one sampling model. To appear in *J. Statist. Plann. Inference*

DeRobertis, L. (1978). The Use Of Partial Prior Knowledge In Bayesian Inference. Doctoral dissertation submitted to Yale University.

Sivaganesan, S. and Berger, J. (1989). Ranges of posterior measures for priors with unimodal contaminations. *Ann. Statist.* **17**, 868-889.

Department of Statistics
Purdue University
West Lafayette, Indiana 47907