OPTIMAL ROBUST CREDIBLE SETS FOR CONTAMINATED PRIORS

by

Siva Savaganesan and L. Mark Berliner
University of Cincinnati — Ohio State University

James Berger
Purdue University

Technical Report # 92-46C

Department of Statistics
Purdue University

October, 1992

Optimal Robust Credible Sets for Contaminated Priors

by

Siva Sivaganesan
University of Cincinnati

L. Mark Berliner
Ohio State University

James Berger
Purdue University

Abstract

In robust Bayesian analysis, it is of interest to find the optimal robust credible set, viz: the smallest set with posterior probability at least, say γ, with respect to each prior in the class. Here, we derive the optimal robust credible set for the ε–contamination class of priors with arbitrary contaminations.

Key words : optimal robust sets, arbitrary contaminations, Bayesian robustness

1 Introduction

When eliciting prior information about an unknown parameter $\theta \in \Theta$ in terms of a single prior distribution, one is usually left with some uncertainty about the elicited prior. In most of the recent developments in robust Bayesian analysis, this uncertainty is (formally) accommodated by the use of a class Γ consisting of all prior distributions that are a priori deemed plausible. When a $100\gamma\%$ credible set for θ is desired, it is of interest in robust Bayesian analysis to seek a set $C \subset \Theta$ which has a posterior probability at least γ with respect to every prior $\pi \in \Gamma$. In fact, the set with the smallest size, in terms of the Lebesgue measure, among all such sets might seem particularly desirable. This set is known as the $100\gamma\%$ optimal robust credible set. Formally, let $\rho_\pi(C)$ denote the posterior probability of C w.r.t the prior π, and define

$$\underline{\rho}(C) = inf_{\pi\in\Gamma}\rho_\pi(C).$$

Then, when θ is continuous and certain regularity conditions such as the likelihood being continuous hold, C_0 is a $100\gamma\%$ optimal robust credible set if the following two conditions hold:

1. $\underline{\rho}(C_0) = \gamma$
2. $\mu(C_0) \leq \mu(C)$ whenever $\underline{\rho}(C) \geq \gamma$,

where $\mu(C)$ is the Lebesgue measure of C. In this paper, we consider the situation where θ is a continuous real parameter, and consider the ε-contamination class of priors with arbitrary contaminations, given by

$$\Gamma = \{\pi = (1-\varepsilon)\pi_0 + \varepsilon q : q \text{ any prob. dist. }\}.$$

Here, π_0 is usually a (computationally) convenient choice for the prior, and ε is the amount of uncertainty about π_0.

The goal of this paper is to find the optimal robust credible set C_0 for the class Γ above. In Section 2, we give the general form of this set. In Section 3, we assume that the likelihood, and the posterior with respect to π_0 are both unimodal and give further results that are useful in actually determining the optimal robust credible set. These assumptions, which are made only to reduce the complexity of the proofs, can actually be relaxed while keeping the results intact. We discuss this at the end of Section 3, and conclude the paper with some examples.

The notion of an optimal robust credible set was first introduced in an unpublished work by Berger and Berliner(1983). Articles where this or a similar notion is used to derive credible sets are DasGupta(1991), where the Lebesgue measure and diameter minimizing sets are considered for density ratio classes, and Sivaganesan(1992) where the optimal robust credible set is derived for the density bounded class with constant bandwidth. In a closely related paper, Wasserman(1989) considered what he termed the maximally robust credible set, namely the set which minimizes the range (or the sensitivity) of $\rho_\pi(C)$ over Γ, among those sets C for which $\rho_{\pi_0}(C) = \gamma$. It was shown there that such sets are in fact given by regions where the likelihood is highest.

There is a rapidly growing literature in the more general area of robust Bayesian analysis. For reviews and discussion of the various issues and approaches, see Berger(1984, 1985, 1990, 1992) and Wasserman(1992).

2 Form of the Optimal Robust Credible Set

In this section, we show that the optimal robust credible set, C_0, can be expressed in a simple form. But, before we state the result, we introduce some notation and review some results that will be useful later. Although our focus in this paper is on a real parameter θ, the result and the proof in this section are also valid when θ is a vector.

Throughout the paper, we let $l(\theta)$ denote the likelihood and $g(\theta)$ denote the posterior density (w.r.t Lebesgue measure) corresponding to the prior π_0. We will also assume that both $l(\theta)$ and $g(\theta)$ are continuous everywhere in the interior of the parameter space, and non-constant in any open interval. Thus, we have (see Huber(1973) and Berger and Berliner(1986))

$$\underline{\rho}(C) = \frac{\int_C g(\theta)d\theta}{1 + r \sup_{\theta \notin C} l(\theta)},$$

where $r = \varepsilon/(1-\varepsilon)m_0$ and $m_0 = \int l(\theta)\pi_0(\theta)d\theta$.

Now, we define, for given $s > 0$ and $t > 0$, two sets A_s and B_t based, respectively, on the likelihood $l(\theta)$ and the posterior $g(\theta)$ by

$$A_s = \{\theta : l(\theta) > s\}$$

and

$$B_t = \{\theta : g(\theta) > t\}.$$

These two sets, as will be seen later, play a fundamental role in the determination of C_0. For use here and in the next section, define, for each set $C \subset \Theta$,

$$p(C) = \int_C g(\theta)d\theta - r\gamma \sup_{\theta \notin C} l(\theta). \tag{1}$$

Thus, note that

$$\underline{\rho}(C) > \text{ or } = \gamma \iff p(C) > \text{ or } = \gamma. \tag{2}$$

Theorem 2.1 *C_0 is of the form*

$$C_0 = A_\eta \cup B_\psi$$

for some $\eta > 0$ and $\psi > 0$.

Proof: Let $\eta = \sup_{\theta \notin C_0} l(\theta)$. Then, clearly $A_\eta \subseteq C_0$. Now, since $\underline{\rho}(C_0) = \gamma$, we have using (1) and (2),

$$\int_{C_0} g(\theta)d\theta = \gamma + \gamma r\eta. \tag{3}$$

Now, letting $B_0 = C_0 \setminus A_\eta$, we have $B_0 \cap A_\eta = \phi$ and $C_0 = A_\eta \cup B_0$. Thus, using (3), we get $B = B_0$ satisfies

$$\int_B g(\theta)d\theta = \gamma + \gamma r\eta - \int_{A_\eta} g(\theta)d\theta. \tag{4}$$

Note that there exists a unique ψ such that $B = B_\psi \backslash A_\eta$ satisfies (4). Let $B_1 = B_\psi \backslash A_\eta$. Then, it is easy to see that $\mu(B_1) \leq \mu(B)$ for all sets B which satisfy $B \cap A_\eta = \phi$ and (4), with strict inequality holding whenever $B \neq B_1$. (We say two sets are equal (or unequal) whenever their symmetric difference has zero (or non-zero) Lebesgue measure.) Now, we show that $B_0 = B_1$, which will complete the proof. To prove this, suppose otherwise. Then, $\mu(B_1) < \mu(B_0)$. Letting $C = A_\eta \cup B_\psi$, we have $\mu(C) < \mu(C_0)$, and $\sup_{\theta \notin C} l(\theta) = \eta' \leq \eta$. Hence, using the fact that $B = B_1$ satisfies (4), we get

$$\begin{aligned}\underline{\rho}(C) &= \frac{\int_C g(\theta)d\theta}{1 + r\eta'} \\ &= \frac{\gamma + \gamma r\eta}{1 + r\eta'} \geq \gamma,\end{aligned}$$

which is a contradiction since $\mu(C) < \mu(C_0)$. This concludes the proof. □

3 Further Results

The results in the previous section alone may actually be used, as we now describe, to find C_0. In fact, for each s, $(0 \leq s \leq \sup_\theta l(\theta))$, one can find $t = t(s)$ so that $C = A_s \cup B_t$ satisfies $\underline{\rho}(C) = \gamma$. The size of C, $\mu(C)$, can be found for each s, and hence be minimized (w.r.t s) to find C_0. Although this program is conceptually straightforward, its implementation would usually be difficult because of the complexity of the function $\psi = \psi(s)$, and the need to consider two separate cases for $\mu(C)$, determined by whether C is connected or not.

In this section, we give certain results, under some additional conditions on $l(\theta)$ and $g(\theta)$, which will substantially facilitate the calculation of C_0. Assume that both $l(\theta)$ and $g(\theta)$ are strictly unimodal with respective modes $\hat{\theta}$ and $\tilde{\theta}$, and that $l(\theta)$ is differentiable everywhere in the interior of the parameter space. Also, let $C_0 = A_\eta \cup B_\psi$ be the $100\gamma\%$ optimal robust credible set, and assume that A_η and B_ψ are non-empty proper subsets of C_0. Here, as before, $\eta = \sup_{\theta \notin C_0} l(\theta)$. In the following theorem, we give a necessary condition satisfied by the end points of C_0. More specifically, we give an expression for ψ in terms of η.

Theorem 3.1

(i) If $A_\eta \cap B_\psi \neq \phi$, then

$$\psi = g(a) + r\gamma|l'(a)|, \tag{5}$$

where a is that end point of A_η which is also an end point of C_0.

(ii) If $A_\eta \cap B_\psi = \phi$, then

$$\psi = \alpha(a_2)g(a_1) + \alpha(a_1)g(a_2) + r\gamma\alpha(a_2)|l'(a_1)|, \tag{6}$$

where a_1 and a_2 are the end points of A_η, and

$$\alpha(a_1) = 1 - \alpha(a_2) = \frac{|l'(a_1)|}{|l'(a_1)| + |l'(a_2)|}.$$

Proof: Let $A_\eta = (a_1, a_2)$ and $B_\psi = (b_1, b_2)$. Without loss of generality, assume that $\tilde{\theta} < \hat{\theta}$, and therefore $b_1 < a_2$. Then, using (2),

$$p(C_0) = \int_{C_0} g(\theta)d\theta - r\gamma l(a_2) = \gamma. \tag{7}$$

Proof of (i) Here A_η and B_ψ overlap, i.e., $a_1 < b_2$ and $a = a_2$. We now want to prove that $\leq$ holds in (5). Suppose that

$$g(b_1) = \psi > g(a_2) + r\gamma|l'(a_2)|. \tag{8}$$

For a given $\delta > 0$ small, define intervals $N_1 = (a_2 - \delta, a_2)$ and $N_2 = (b_1 - \delta, b_1)$. We now claim that there exists $\delta > 0$ such that

$$\int_{N_2} g(\theta)d\theta - \int_{N_1} g(\theta)d\theta > r\gamma\left[l(a_2 - \delta) - l(a_2)\right]. \tag{9}$$

Suppose not. Then, the converse of the above holds for all $\delta > 0$. This gives,

$$\frac{1}{\delta}\int_{N_2} g(\theta)d\theta - \frac{1}{\delta}\int_{N_1} g(\theta)d\theta \leq r\gamma\frac{l(a_2 - \delta) - l(a_2)}{\delta} \text{ for all } \delta > 0.$$

Thus, letting $\delta \to 0$, we obtain

$$\psi - g(a_2) \leq r\gamma|l'(a_2)|,$$

which contradicts (8). Hence, there exists $\delta > 0$ satisfying (9). If we let $C_1 = (C_0 \setminus N_1) \cup N_2$, then $\mu(C_1) = \mu(C_0)$. Using (2), (7) and the unimodality of $l(\cdot)$, we have

$$\begin{aligned} p(C_1) &= \int_{C_1} g(\theta)d\theta - r\gamma l(a_2 - \delta) \\ &= p(C_0) + \{\int_{N_2} g(\theta)d\theta - \int_{N_1} g(\theta)d\theta - r\gamma\,[l(a_2-\delta) - l(a_2)]\} \\ &> \gamma. \end{aligned}$$

The last inequality follows from (7) and (9). Thus, C_1 is such that $\underline{\rho}(C_1) > \gamma$ and $\mu(C_1) = \mu(C_0)$. This is a contradiction (since one could then find a set C with smaller Lebesgue measure than C_0 such that $\underline{\rho}(C) = \gamma$).Thus (8) cannot be true, and hence

$$\psi \leq g(a_2) + r\gamma|l'(a_2)|. \tag{10}$$

Now, we show that strict inequality cannot hold in the above. Let $N_2 = (b_1, b_1+\delta)$ and $N_1 = (a_2, a_2+\delta)$, and let $C_1 = (C_0 \setminus N_2) \cup N_1$. Since $A_\eta \subset C_0$, there exists $\delta_1 > 0$ such that $\sup_{\theta \notin C_1} l(\theta) = l(a_2 + \delta)$ for all $0 < \delta < \delta_1$. Suppose that strict inequality holds in (10). Using similar arguments as before, we can show that, given any $\delta_1 > 0$, there exists $0 < \delta < \delta_1$ such that

$$\int_{N_1} g(\theta)d\theta - \int_{N_2} g(\theta)d\theta > r\gamma(l(a_2+\delta) - l(a_2)),$$

which then implies $\underline{\rho}(C_1) > \gamma$. This is again a contradiction since $\mu(C_1) = \mu(C_0)$. Thus strict inequality cannot hold in (10) proving the desired result.

Proof of (ii) The basic approach of the proof is the same as before, although the details are somewhat different. Suppose that strict inequality > holds in (6), i.e.,

$$\psi > \alpha(a_2)g(a_1) + \alpha(a_1)g(a_2) + r\gamma\alpha(a_2)|l'(a_1)|. \tag{11}$$

Let $B_\psi = (b_1, b_2)$, and define $N_1 = (b_1 - \delta, b_1) \cup (b_2, b_2 + \delta)$ and $N_2 = (a_1, a_1 + \delta_1) \cup (a_2 - \delta_2, a_2)$. Here, $\delta > 0$ is small enough, and δ_1 and δ_2 are both positive and satisfy the conditions

$$\delta_1 + \delta_2 = 2\delta \quad \text{and} \quad l(a_1 + \delta_1) = l(a_2 - \delta_2). \tag{12}$$

Now, we claim that given $\delta' > 0$, there exists $0 < \delta < \delta'$ such that

$$\int_{N_1} g(\theta)d\theta - \int_{N_2} g(\theta)d\theta > l(a_1 + \delta_1) - l(a_1). \tag{13}$$

Else, there would exist $\delta' > 0$ such that $\leq$ holds in (13) for all $0 < \delta < \delta'$. Replacing $>$ in (13) by $\leq$, dividing by δ, and letting $\delta \to 0$ gives

$$2g(b_1) - \left(g(a_1)\frac{d\delta_1}{d\delta} + g(a_2)\frac{d\delta_2}{d\delta}\right) \leq l'(a_1)\frac{d\delta_1}{d\delta}. \tag{14}$$

Now, it can be shown from (12) that

$$\frac{d\delta_1}{d\delta} = 2 - \frac{d\delta_2}{d\delta} = 2\alpha(a_1).$$

This, along with (14), gives

$$\psi \leq \alpha(a_2)g(a_1) + \alpha(a_1)g(a_2) + r\gamma\alpha(a_2)|l'(a_1)|, \tag{15}$$

which contradicts (11), thus proving the claim (13). As before, letting $C_1 = (C_0 \setminus N_2) \cup N_1$, we have $\mu(C_1) = \mu(C_0)$. Moreover, it can be verified, using (13), that $\underline{\rho}(C_1) > \gamma$, again leading to a contradiction. Thus, $\leq$ must hold in (6). Using similar arguments, one can also show that the converse also must hold true, concluding the proof. □

The above theorem gives different expressions for ψ depending on whether $A_\eta \cap B_\psi$ is empty or not. The following corollary gives a necessary and sufficient condition, in terms of η, for $A_\eta \cap B_\psi$ to be empty. For use here, let $D(\eta) = |g(a_1) - g(a_2)|$, and a_2 be that end point of A_η farthest from B_ψ.

Corollary 1 *$A_\eta \cap B_\psi = \phi$ if and only if $D(\eta) < r\gamma|l'(a_2)|$.*

Proof: The necessary part can be easily verified using (6). To prove the sufficiency part, suppose that $A_\eta \cap B_\psi \neq \phi$. Then, by (5), we must have

$$\psi = g(a_2) + r\gamma|l'(a_2)|.$$

Now, let $B_\psi = (b_1, b_2)$ and assume $\hat{\theta} > \tilde{\theta}$. Then, by the assumption that A_η and B_ψ are proper subsets of C_0, we must have $a_1 \in B_\psi$, and hence $\psi = g(b_2) \leq g(a_1)$. This implies $D \geq r\gamma|l'(a_2)|$, a contradiction, proving the result. □

Summary of the results and their implications

The optimal robust credible region C_0 satisfies the condition $p(C_0) = \gamma$, and is of the form $C_0 = A_\eta \cup B_\psi$, where $\eta = \sup_{\theta \notin C_0} l(\theta)$. There are (typically) two possibilities (except perhaps in the uninteresting case of very small γ). Either, $C_0 = A_\eta$ for suitable η, say η_0, or $C_0 \supset A_\eta$. In the latter case, if $D(\eta) < r\gamma|l'(a_2)|$ then $A_\eta \cap B_\psi = \phi$, and ψ is given by (6). Otherwise, $A_\eta \cap B_\psi \neq \phi$, in which case ψ is given by (5). Without loss of generality, assume $\tilde{\theta} < \hat{\theta}$. Then, for a given s, $(0 < s < l(\hat{\theta}))$, we can find $t = t(s)$ either by using the following version of equation (5),

$$t = g(a) + r\gamma|l'(a)|$$

with a being the right end point of A_s, or by the similar version of equation (6), as the case may be, determined by whether $D(s) \geq$ or $< r\gamma l'(a)$. Thus, for given s, we can construct a set $C^s = A_s \cup B_t$. Now, the optimal robust credible set C_0 can be found by solving the equation (in s)

$$p(C^s) = \gamma. \tag{16}$$

Typically, we found $p(C^s)$ to be increasing with decreasing s, leading either to a unique solution to (16) and hence to C_0, or to the value η_0 for which $p(A_{\eta_0}) = \gamma$, in which case $C_0 = A_{\eta_0}$.

<u>Remarks</u>

1. If $l(\theta)$ is, in addition, symmetric, (5) simplifies to the following:

$$\psi = \frac{1}{2}\{g(a_1) + g(a_2) + r\gamma|l'(a_2)|\}.$$

2. When Θ is a half-open interval such as $(0, \infty)$, for sufficiently small values of η's, one end point of A_η may coincide with (the finite) end point of Θ. For such η values, the corresponding ψ is given by

$$\psi = g(a) + r\gamma|l'(a)|,$$

where a is the other end point of A_η. This can be shown using similar arguments, while observing that we need only consider the case where the sets A_η and B_ψ are disjoint.

3. The results in this section remain valid when π_0, rather than g, is assumed unimodal. With sufficient book-keeping, the proofs in this section can easily be adapted to show this. A crucial modification would be to consider two separate cases determined by whether the intersection of the boundary of A_η, and the

set B_ψ is a singleton or a two-point set. The key here is that that all modes of g would lie between the modes of the prior $\pi_0(\cdot)$ and the likelihood $l(\cdot)$, and hence we would still know, in the case where the intersection is a singleton, which of the end points of A_η intersects with B_ψ.

Examples

1. As an illustration, consider the normal mean problem, where $X \sim N(\theta, \sigma^2)$ and $\pi_0 \equiv N(\mu, \tau^2)$. Now let $\tau^2 = 1$, $\mu = 0$, $\sigma^2 = 1/5$ and $\varepsilon = 0.05$. In each example that follows, we will also let $\gamma = 0.9$, and use C_0 to denote the 90% optimal robust credible set. When $X = 0.3$, the calculation as described above gives $C_0 = A_\eta = (-0.44, 1.04)$. When $X = 1.0$, we get $C_0 = (0.13, 1.66)$. This corresponds to $A_\eta = (0.34, 1.66)$ and $B_\psi = (0.13, 1.54)$. When $X = 1.5$, $C_0 = (0.56, 2.34)$, which is again in the form $A_\eta \cup B_\psi = (0.66, 2.34) \cup (0.56, 2.34)$. As the value of X gets further away from $\mu = 0$, C_0 coincides with A_η. For example, when $X = 3$, $C_0 = A_\eta = (1.71, 4.29)$. The other case in which $A_\eta \neq C_0$ and $A_\eta \cap B_\psi = \phi$ does not usually occur. As can be seen from Corollary 1, this case may (only) occur when ε and γ are large, and τ^2 is small compared to σ^2 (yielding a π_0-posterior mean adequately far from the value of X). Note that these conditions would rarely occur in practice. To give an example of this case, let $\tau^2 = 1/40$, $\sigma^2 = 1$ and $\varepsilon = 0.4$. For this case, when $X = 1.5$, calculation yields $C_0 = A_\eta \cup B_\psi = (-.20, .28) \cup (.32, 2.68)$. The 'pattern' we observe is that for $|x - \mu|$ small, $C_0 = A_\eta$; for $|x - \mu|$ larger $C_0 = A_\eta \cup B_\psi$ where $A_\eta \neq C_0$; then as $|x - \mu|$ gets even larger, returning to the initial form $C_0 = A_\eta$.

2. Assume that the lifetime of an electronic component has an exponential distribution with mean θ (in units of hours), and that the prior information about θ is elicited by the distribution π_0 = Inverse Gamma(9, .01). Suppose that we are interested in a 90% credible set for θ when the uncertainty in π_0 is expressed by the class Γ with $\varepsilon = 0.1$. Suppose also that a sample of 5 components tested has a total lifetime of 65. Then, using the method of calculations described above, the optimal robust credible set for θ is $C_0 = A_\eta \cup B_\psi = (7.67, 24.65) \cup (6.35, 21.57) = (6.35, 24.65)$.

3. Suppose $X|\theta \sim$ Cauchy$(\theta, 1)$, $\pi_0 \equiv$ Cauchy$(0, .3)$ and $\varepsilon = 0.01$. When $X = 6$, the 90% optimal robust credible set for θ is $C_0 = (-1.22, 2.70) \cup (3.56, 8.43)$.

References

[1] Berger, J. (1984). The robust Bayesian viewpoint. *Robustness in Bayesian Statistics* (J. Kadane ed.), Amsterdam, North-Holland, 63-124, (with discussion).

[2] Berger, J. (1985). *Statistical Decision Theory and Bayesian analysis.* Springer-Verlag, New York.

[3] Berger, J. (1990) Robust Bayesian analysis: sensitivity to the prior. *Jour. Stat. Planning and Inf.*, 25:303–328.

[4] Berger, J. (1992) An overview of Bayesian robustness. Technical Report, Purdue University.

[5] Berger, J. and Berliner, L. M. (1983). Robust Bayes and empirical Bayes analysis with ε-contaminated priors. Technical Report 83-35, Purdue University.

[6] Berger, J. and Berliner, L. M. (1986). Robust Bayes and empirical Bayes analysis with ε-contaminated priors. *Ann. Statist.*, 14, 461-486.

[7] DasGupta, A. (1991). Diameter and volume minimizing confidence sets in Bayes and classical problems. *Ann. Statist.* 19, 1225-1243.

[8] de la Horra, J. and Fernandez, C. (1992). Bayesian analysis under $\varepsilon-$contaminated priors: a trade-off between robustness and precision. Tecnical Report, Universidad Autonoma, Madrid.

[9] Huber, P. J. (1973). The use of Choquet capacities in statistics. *Proc. 39th Session ISI*, 45, 181-188.

[10] Sivaganesan, S. (1992). Robust confidence sets for density bounded class with constant bandwidth. Manuscript under preparation.

[11] Wasserman, L. A. (1989). A robust Bayesian interpretation of likelihood regions. *Ann. Statist.*, 17, 1387-1393.

[12] Wasserman, L. A. (1992). Recent methodological advances in robust Bayesian inference. *Bayesian Statistics 4* (J.M. Bernardo, J.O. Berger, A. P. Davwid and A. F. M. Smith, eds.), Oxford University Press, 483-502, (with discussion).