The Annals of Statistics
1989, Vol. 17, No. 3, 1387–1393

A ROBUST BAYESIAN INTERPRETATION OF LIKELIHOOD REGIONS[1]

By Larry Alan Wasserman

Carnegie Mellon University

Likelihood regions are shown to be robust in the sense that their posterior probability content is relatively insensitive to contaminations of the prior. This provides a Bayesian interpretation of regions that are commonly used by frequentists to construct confidence intervals and whose use are also advocated by the pure likelihood approach.

1. Introduction. As inferential summaries, subsets of relatively high likelihood, or likelihood regions, are pervasive in classical and likelihood based inference. In Bayesian inference, highest posterior density credible regions are usually preferred to likelihood regions. In this article, we show that there is a robust Bayesian interpretation of likelihood regions.

We are concerned here with robustness with respect to the prior distribution. To study this type of rubustness, the prior distribution is replaced with a class of priors. This leads to a class of posterior distributions arrived at by applying Bayes' theorem to each prior distribution. Bounds may then be placed on the posterior probability content of a given set, over this class of distributions. An analysis carried out in this way is useful since elicitation of a prior is a difficult task. See Good (1950), Berger (1984, 1985) and references contained therein for details on Bayesian robustness. Also, see Smith (1961), Williams (1976) and Walley (1981, 1982) for a formal justification for using classes of probability measures. One could also consider robustness with respect to the model—an issue that is dealt with, for example, in Box and Tiao (1973) and Huber (1973).

The result proved in this article is that, of all sets with a given posterior probability content for a specified prior P, the likelihood region is robust in the sense that its probability content is least sensitive to changes in the prior with respect to the class of ε-contaminated priors. Likelihood regions were considered from a Bayesian perspective by Box and Tiao (1965) who pointed out that they may be regarded as an equivariant summary of a posterior distribution. They were also studied by Piccinato (1984) who showed that a generalized version of likelihood regions have the property of always having posterior probability greater than or equal to their prior probability.

In Section 2 we review robust Bayesian inference, paying special attention to the theory of ε-contaminated priors as discussed by Huber (1973) and Berger and

Received March 1987; revised October 1988.

[1]The original version of this article was written at the University of Toronto. Research supported by a studentship from the Medical Research Council of Canada and by a postdoctoral fellowship from the Natural Sciences and Engineering Research Council of Canada.

AMS 1980 *subject classifications*. Primary 62A15; secondary 62F15.

Key words and phrases. Contaminated priors, credible regions, likelihood regions, robust Bayesian inference.

Berliner (1986). In Section 3 the robustness of likelihood regions is demonstrated. Section 4 presents some numerical examples. Section 5 contains discussion and conclusions.

2. Robust Bayesian inference. Consider a model $M = \{f(x|\theta);\ \theta \in \Theta\}$, where each $f(x|\theta)$ is a probability density with respect to Lebesgue measure and Θ is a subset of $\mathscr{R}^k$. We shall assume that $f(x|\theta)$ is a bounded function of θ for each x. Let $\mathscr{B}(\Theta)$ be the Borel subsets of Θ and let P be a prior probability measure on $\mathscr{B}(\Theta)$ with density $\pi(\theta)$. The posterior distribution Q on $\mathscr{B}(\Theta)$, given $X = x$, has density function q defined by $q(\theta) \propto f(x|\theta)\pi(\theta)$. To carry out robust Bayesian inference, we replace the prior distribution, P, with a class $\Pi(P)$ of probability measures containing P. Bayes' theorem is applied to each member of $\Pi(P)$ which leads to a class $\mathscr{Q}(P)$ of posterior distributions.

The class of priors we are considering is defined by

$$\Pi_\varepsilon(P) = \left\{R \in \mathscr{P};\ R = (1-\varepsilon)P + \varepsilon\tilde{P},\ \tilde{P} \in \mathscr{P}\right\},$$

where $\varepsilon \in [0,1]$ and $\mathscr{P}$ is the set of all probability measures on $\mathscr{B}(\Theta)$. The number ε may be regarded as a measure of our uncertainty about the prior P. This is the class of ε-contaminated priors discussed in Huber (1973) and Berger and Berliner (1986). Berger and Berliner (1986) consider more restricted classes of priors by letting $\tilde{P}$ range over a set that is smaller than $\mathscr{P}$. We refer the reader to their article for a detailed analysis of ε-contaminated priors. There are other sets of priors that may be used to carry out a robust analysis, but we shall focus only on the class $\Pi_\varepsilon(P)$.

We denote the set of posteriors resulting from the class $\Pi_\varepsilon(P)$ by $\mathscr{Q}_\varepsilon(P)$. Henceforth, we shall suppress the dependence on P. For any measurable set $A \subset \Theta$ define

$$Q^*(A) = \sup_{Q \in \mathscr{Q}_\varepsilon} Q(A) \quad \text{and} \quad Q_*(A) = \inf_{Q \in \mathscr{Q}_\varepsilon} Q(A).$$

The functions Q^* and Q_* are called the upper and lower posterior probabilities, respectively. Let $\delta(A) = Q^*(A) - Q_*(A)$. We regard $\delta(A)$ as a measure of how robust our analysis is in the sense that it measures the stability of the probability content of A over the class of distributions. If $\delta(A)$ is small, then our posterior probability statements are not greatly affected by the uncertainty about the prior. Although we shall focus on this aspect of robustness, we point out that one could define robustness in many other ways. The following theorem is due to Huber (1973).

THEOREM 2.1.

$$Q^*(A) = \frac{Q(A) + s(A)}{1 + s(A)} \quad \textit{and} \quad Q_*(A) = \frac{Q(A)}{1 + s(A^c)},$$

where $s(\phi) \equiv 0$,

$$s(A) = \left(\frac{\varepsilon}{1-\varepsilon}\right)\frac{l_A}{m(x)} \quad \textit{for } A \neq \varnothing,$$

$l_A = \sup_{\theta \in A} f(x|\theta)$ *and* $m(x) = \int_\Theta f(x|\theta)\pi(\theta)\,d\theta$.

Suppose that an analysis is carried out with prior P, leading to a posterior measure Q. It is common practice to summarize Q by way of a credible region—a subset A with a specified posterior probability content. See Box and Tiao (1973) and Berger (1985). This is the Bayesian analog of the frequentist confidence interval. Box and Tiao (1965, 1973) considered two methods for constructing credible regions. The first is called the *highest posterior density region*, or *HPD region*, denoted by H_γ. This is the region with minimal Lebesgue measure of all regions with posterior probability content at least γ.

The second is the likelihood region E_γ defined in the following way. For each $c \in [0,1]$, let

$$R_c = \{\theta;\ r(\theta) \geq c\} \quad \text{and} \quad c(\gamma) = \sup\{c;\ Q(R_c) \geq \gamma\},$$

where

$$r(\theta) = f(x|\theta) / \sup_{\theta \in \Theta} f(x|\theta)$$

is the relative likelihood function. The γ-level likelihood region E_γ is defined by $E_\gamma = R_{c(\gamma)}$. Box and Tiao (1965, 1973) expressed doubts about using this region. They referred to it as an "artificial construction" [Box and Tiao (1973), page 124]. The next section demonstrates that these regions are robust in the sense of minimizing $\delta(A)$ over all sets with a specified probability content for the prior P. This gives a robust Bayesian interpretation of these regions. Also, note that E_γ is equivariant under one-to-one, monotone transformations of the parameter space. This may be a virtue in problems when there is no particular reason to construct a credible region in one particular parametrization. The region H_γ is not equivariant. Berger (1985), page 144, gives a demonstration of this lack of equivariance.

The likelihood regions $\{R_c;\ c \in [0,1]\}$ play an important role in frequentist theory. By choosing the constant c appropriately, R_c can be used as an approximate confidence interval. See Cox and Hinkley (1974). Also, the pure likelihood approach to statistics advocates the use of R_c as a region of parameter values highly supported by the data [Edwards (1972)].

3. Robust credible regions. Suppose that a robust Bayesian analysis has been carried out with a class of priors $\Pi(P)$. The problem of choosing a robust credible region A, of smallest Lebesgue measure subject to $Q_*(A)$ being greater than a fixed number, is considered in Berger and Berliner (1983).

Here, we show that the likelihood region E_γ is maximally robust over the class of ε-contaminated priors, with respect to the robustness measure δ introduced in Section 2. Let Q be the posterior measure based on P. Formally, we call

W_γ a *maximally δ-robust credible region* if $Q(W_\gamma) = \gamma$ and if, for any other measurable set A with $Q(A) = \gamma$, we have $\delta(W_\gamma) \leq \delta(A)$. Thus, W_γ is maximally robust if it minimizes the range of posterior probabilities over $\Pi_\varepsilon(P)$. Note that the class of sets being compared to W_γ all have posterior probability content γ under Q. Since P, and hence Q, are not considered completely accurate, we must be cautious in interpreting robustness in this way.

In what follows, we assume that γ is some fixed, predetermined probability level not equal to zero or one. We extend the function r to $\mathscr{B}(\Theta)$ by defining $r(A) = \sup_{\theta \in A} r(\theta)$. We shall also assume the following three regularity conditions:

(i) For each $\gamma \in (0,1)$, $Q(E_\gamma) = \gamma$,
(ii) for each c, $Q(r^{-1}(\{c\})) = 0$ and
(iii) there exists $\hat{\theta} \in \Theta$ such that $r(\hat{\theta}) = 1$.

LEMMA 3.1. *For any measurable set A,*

$$\delta(A) = \frac{Q(A)k(r(A^c) - r(A))}{(1 + kr(A))(1 + kr(A^c))} + \frac{kr(A)}{1 + kr(A)},$$

where k is a positive constant.

PROOF. Note that

$$s(A) = \left(\frac{\varepsilon}{1-\varepsilon}\right)\frac{l_A}{m(x)} = \left(\frac{\varepsilon}{1-\varepsilon}\right)\frac{l_\Theta}{m(x)}\frac{l_A}{l_\Theta} = kr(A),$$

where

$$k = \left(\frac{\varepsilon}{1-\varepsilon}\right)\frac{l_\Theta}{m(x)}.$$

The result follows by applying Theorem 2.1 and calculating $Q^*(A) - Q_*(A)$. □

We can now state the main result.

THEOREM 3.1. *If $\gamma > 0.5$, then the maximally δ-robust region is the likelihood region E_γ.*

PROOF. Define

$$R^+ = \{R \in \mathscr{B}(\Theta);\ Q(R) = \gamma \text{ and } \hat{\theta} \in R\}$$

and

$$R^- = \{R \in \mathscr{B}(\Theta);\ Q(R) = \gamma \text{ and } \hat{\theta} \notin R\}.$$

First we show that E_γ is maximally δ-robust in R^+.

For any $A \in R^+$, $r(A) = 1$ since $\hat{\theta} \in A$, hence

$$\delta(A) = \frac{\gamma k(r(A^c) - 1)}{(1+k)(1+kr(A^c))} + \frac{k}{1+k}.$$

Since

$$\frac{d\delta(A)}{dr(A^c)} = \frac{\gamma k}{(1+kr(A^c))^2} > 0,$$

$\delta(A)$ is an increasing function of $r(A^c)$ and hence is minimized by making $r(A^c)$ as small as possible. Set $\tilde{A} = \{\theta;\ r(\theta) > r(A^c)\}$ and $\dot{A} = \{\theta;\ r(\theta) \geq r(A^c)\}$. Now, $\tilde{A} \subset A$ so that $Q(\tilde{A}) \leq \gamma$. By (ii), $Q(\dot{A}) = Q(\tilde{A}) \leq \gamma$ and then by (i), $\dot{A} \subset E_\gamma$ so that $r(\dot{A}^c) \geq r(E_\gamma^c)$. But $r(\dot{A}^c) \leq r(A^c)$ which implies that $r(A^c) \geq r(E_\gamma^c)$. Thus, δ is minimized in R^+ by E_γ.

Next we claim that $E_{1-\gamma}^c$ is maximally δ-robust in R^-. To see this, note that $\delta(E_{1-\gamma}^c) = \delta(E_{1-\gamma}) \leq \delta(A^c) = \delta(A)$ for each $A \in R^-$.

Since R^+ and R^- form a partition of the set of all γ-level credible regions, we need only compare the maximally robust regions from each class. In other words, we need only compare E_γ and $E_{1-\gamma}^c$. By hypothesis, $\gamma > 0.5$. It follows easily that $\delta(E_{1-\gamma}) > \delta(E_\gamma)$ which completes the proof since $\delta(E_{1-\gamma}^c) = \delta(E_{1-\gamma})$. □

4. Examples. To demonstrate the δ-robustness of the likelihood regions, we examine two examples presented in Berger and Berliner (1986). For the sake of comparison, we will also compute HPD regions. The difference in performance of the regions should not be regarded as a criticism of HPD regions since these regions were designed to minimize Lebesgue measure, a goal that is clearly incompatible with δ-robustness. For further comparison, we shall also find the HPD region of length equal to that of E_γ.

Suppose X is normally distributed with unknown mean θ and known variance σ^2. Suppose our prior for θ is also normal, with mean μ and variance τ^2. It follows that the posterior for θ, given $X = x$, is normal with mean d and variance V^2, where

$$d = \frac{\mu/\tau^2 + x/\sigma^2}{1/\tau^2 + 1/\sigma^2} \quad \text{and} \quad V^2 = \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}.$$

Note that in the normal case, the HPD regions are identical to the equal tailed credible regions.

EXAMPLE 1. Suppose $\sigma^2 = 1$, $\tau^2 = 2$, $\mu = 0$, $\varepsilon = 0.2$ and $x = 0.5$. The 95% HPD region H_γ turns out to be the interval $(-1.27, 1.93)$ and the likelihood region E_γ is $(-1.13, 2.13)$. The HPD region A with length equal to the length of E_γ is $(-1.30, 1.96)$. For this region $Q(A) = 0.954$. The upper and lower posterior probabilities, the value of δ and the length of each interval are displayed in the first part of Table 1. E_γ is slightly more robust than the other regions. We remark that $Q^*(H_\gamma) = Q^*(E_\gamma)$ since Q^* is a function of the probability content

TABLE 1
Comparison of credible regions

	$x = 0.5$				$x = 4.0$			
	Q_*	Q^*	δ	Length	Q_*	Q^*	δ	Length
H_γ	0.817	0.965	0.148	3.20	0.136	0.993	0.857	3.20
E_γ	0.848	0.965	0.117	3.26	0.811	0.993	0.182	5.36
A	0.826	0.968	0.142	3.26	0.283	1.000	0.717	5.36

under Q and the maximum of the likelihood obtained over the set, both of which are identical in this example.

EXAMPLE 2. Assume the same situation as in Example 1 except now take $x = 4$. Then $H_\gamma = (1.07, 4.27)$, $E_\gamma = (1.32, 6.68)$ and the region A, as defined in Example 1, is equal to $(-0.02, 5.34)$. We find that $Q(A) = 0.999$. The results are summarized in the second part of Table 1. Here, the robustness of E_γ is substantial. Clearly, having a more extreme observation makes the analysis more sensitive to the prior.

5. Discussion. Theorem 3.1 provides a robust Bayesian interpretation of likelihood regions. Nonetheless, many questions still need to be addressed. Following are some remarks and open questions.

1. Our definition of robustness is rather specific and suffers certain limitations. As pointed out in Section 3, the prior P plays a significant role in the definition of the robust region. In particular, E_γ is maximally robust only among those regions whose posterior probability content is γ, where the posterior probability is calculated using the prior P. This is reasonable only if P is felt to be a good initial choice as a prior. If one feels that P should not hold such a special role, then our definition of a robust region is less convincing. This strong dependence on P is troubling and deserves further investigation.
2. We have restricted ourselves to ε-contaminated priors and we have allowed all distributions as possible contaminations. It would be interesting to extend the results to a more restrictive set of contaminations as in Berger and Berliner (1986).
3. One might feel that attention should be focused on maximizing $Q_*(A)$ rather than minimizing $\delta(A)$. But an argument like that in Theorem 3.1 shows that the region that maximizes $Q_*(A)$ over all subsets with $Q(A) = \gamma$ is again the likelihood region E_γ. However, the same problems discussed in the first comment apply here as well.

Acknowledgments. I would like to thank Jim Berger, Rob Tibshirani and Isabella Verdinelli for many helpful comments and suggestions. I would also like to thank an Associate Editor and two referees for a thorough review of this article that led to considerable improvements in the presentation.

REFERENCES

BERGER, J. (1984). The robust Bayesian viewpoint (with discussion). In *Robustness in Bayesian Statistics* (J. Kadane, ed.). North-Holland, Amsterdam.

BERGER, J. (1985). *Statistical Decision Theory*, 2nd ed. Springer, New York.

BERGER, J. and BERLINER, M. (1983). Robust Bayes and empirical Bayes analysis with ε-contaminated priors. Technical Report 83-85, Dept. Statistics, Purdue Univ.

BERGER, J. and BERLINER, M. (1986). Robust Bayes and empirical Bayes analysis with ε-contaminated priors. *Ann. Statist.* **14** 461–486.

BOX, G. E. P. and TIAO, G. C. (1965). Multiparameter problems from a Bayesian point of view. *Ann. Math. Statist.* **36** 1468–1482.

BOX, G. E. P. and TIAO, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, Mass.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

EDWARDS, A. W. F. (1972). *Likelihood*. Cambridge Univ. Press, Cambridge.

GOOD, I. J. (1950). *Probability and the Weighing of Evidence*. Griffin, London.

HUBER, P. J. (1973). The use of Choquet capacities in statistics. *Bull. Internat. Statist. Inst.* **45** 181–191.

PICCINATO, L. (1984). A Bayesian property of the likelihood sets. *Statistica* **44** 197–204.

SMITH, C. A. B. (1961). Consistency in statistical inference and decision (with discussion). *J. Roy. Statist. Soc. Ser. B* **23** 1–37.

WALLEY, P. (1981). Coherent lower and upper probabilities. Statistics Research Report, Univ. Warwick.

WALLEY, P. (1982). The elicitation and aggregation of beliefs. Statistics Research Report, Univ. Warwick.

WILLIAMS, P. M. (1976). Indeterminate probabilities. In *Formal Methods in the Methodology of Empirical Sciences* (M. Przełecki, K. Szaniawski, R. Wójcicki and G. Malinowski, eds.) 229–246. Reidel, Dordrecht.

DEPARTMENT OF STATISTICS
CARNEGIE MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213