An Overview of Robust Bayesian Analysis*

by

James Berger
Purdue University

Technical Report #93-53C

Department of Statistics
Purdue University

October 1993

*Research supported by the National Science Foundation, Grants DMS–8923071 and DMS 93-03556.

△

AN OVERVIEW OF ROBUST BAYESIAN ANALYSIS

by

James Berger
Purdue University

ABSTRACT

Robust Bayesian analysis is the study of the sensitivity of Bayesian answers to uncertain inputs. This paper seeks to provide an overview of the subject, one that is accessible to statisticians outside the field. Recent developments in the area are also reviewed, though with very uneven emphasis. The topics to be covered are as follows:

1. INTRODUCTION

1.1 Motivation

Robust Bayesian analysis is the study of the sensitivity of Bayesian answers to uncertain inputs. These uncertain inputs are typically the model, prior distribution, or utility function, or some combination thereof. Informal or adhoc sensitivity studies have long been a part of applied Bayesian analysis (cf. Box, 1980), but recent years have seen an explosion of interest and literature on the subject. There are several reasons for this interest:

Foundational Motivation: There is a common perception that foundational arguments lead to subjective Bayesian analysis as the only coherent method of behavior. Non-Bayesians often recognize this, but feel that the subjective Bayesian approach is too difficult to implement, and hence they ignore the foundational arguments. Both sides are partly right. Subjective Bayesian analysis is, indeed, the only coherent mode of behavior, *but only if it is assumed that one can make arbitrarily fine discriminations in judgment about unknowns and utilities.* In reality, it is very difficult to discriminate between, say, 0.10 and 0.15 as the subjective probability, $P(E)$, to assign to an event E, much less to discriminate between 0.10 and 0.100001. Yet standard Bayesian axiomatics assumes that the latter can (and will) be done. Non-Bayesians intuitively reject the possibility of this, and hence reject subjective Bayesian theory.

It is less well known that *realistic* foundational systems exist, based on axiomatics of behavior which acknowledge that arbitrarily fine discrimination is impossible. For instance, such systems allow the possibility that $P(E)$ can only be assigned the range of values from 0.08 to 0.13; reasons for such limitations range from possible psychological limitations to constraints on time for elicitation. The conclusion of these foundational systems is that a type of *robust Bayesian analysis* is the coherent mode of behavior. Roughly, coherent behavior corresponds to having *classes* of models, priors, and utilities, which yield a range of possible Bayesian answers (corresponding to the answers obtained through combination of all model-prior-utility triples from the classes). If this range of answers is too large, the question of interest may not, of course, be settled, but that is only realistic: if the inputs are too uncertain, one cannot expect certain outputs. Indeed, if one were to perform ordinary subjective Bayesian analysis without checking for robustness, one could be seriously misled as to the accuracy of the conclusion.

Extensive developments of such foundational systems can be found in Walley (1991), Rios Insua (1990, 1992) and Rios Insua and Martín (1994); see also Rios and Girón (1980) and Kouznetsov (1991). I. J. Good (cf., Good, 1983a) was the first to extensively discuss these issues. Other earlier references can be found in Berger (1984, 1985) and in Walley (1991); this latter work is particularly to be recommended for its deep and scholarly study of the foundations of imprecision and robustness. Recent developments in some of the interesting theoretical aspects of the foundations can be found in Wasserman and Kadane (1990, 1992b) and Wasserman and Seidenfeld (1994).

Practical Bayesian Motivation: Above, we alluded to the difficulty of subjective elicitation. It is so difficult that, in practice, it is rarely done. Instead, noninformative priors or other approximations (e.g., BIC in model selection) are typically used. The chief difficulties in elicitation are (i) knowing the degree of accuracy in elicitation that is necessary; (ii) knowing what to elicit. Robust Bayesian analysis can provide the tools to answer both questions.

As an example of (i), one might be able to quickly determine that $0.05 \leq P(E) \leq 0.15$, but then wonder if more accurate specification is needed. Robust Bayesian methods can operate with such partial specifications, allowing computation of the corresponding range of Bayesian answers. If this range of answers is small enough to provide an answer to the question of interest, then further elicitation is unnecessary. If, however, the range is too large to provide a clear answer, then one must attempt finer elicitation (or obtain more data or otherwise strengthen the information base).

Knowing what to elicit is even more crucial, especially in higher dimensional problems where it is completely infeasible to elicit everything that is possibly relevant. Suppose, for instance, that one believes in a 10-dimensional normal model, but that the mean vector and covariance matrix are unknown. Then there are 65 unknown parameters, and accurate elicitation of a 65-dimensional distribution is impossible (unless one is willing to introduce structure that effectively greatly reduces the number of parameters). But many of these parameters may be accurately determined by the data, or the question of interest may not depend on accurately knowing many of the parameters. In fact, there may only be a few crucial quantities that need to be elicited. Robust Bayesian techniques can help to identify these quantities.

Acceptance of Bayesian Analysis: Rightly or wrongly, the majority of the statistical world resists use of Bayesian methods. The most often vocalized reason is fear of using a subjective prior, because of a number of perceived dangers. While we do not view this fear as being particularly reasonable (assumptions made in other parts of the analysis are usually much more influential and questionable), we recognize its existence. Robust Bayesian methods, which can operate with a wide class of prior distributions (reflecting either the elicitor's uncertainty in the chosen prior or a range of prior opinions of different individuals), seems to be an effective way to eliminate this fear.

Non-Bayesian Motivation: Many classical procedures work well in practice, but some standard procedures are simply illogical. Robust Bayesian analysis can be used to determine which procedures are clearly bad. Consider, for instance, the following example:

Example 1. A series of clinical trials is performed, with trial i testing drug D_i versus a placebo. Each clinical trial is to be analyzed separately, but all can be modelled as standard normal tests of $H_0\colon \theta_i = 0$ versus $H_1\colon \theta_i \neq 0$, where θ_i is the mean effect of D_i minus the mean effect of the placebo. Suppose we know, from past experience, that about 1/2 of the drugs that are tested will end up being ineffective; i.e., will have $\theta_i = 0$. (This assumption is not essential; it merely provides a mental reference for the ensuing understanding.)

We will focus on the meaning of P-values that arise in this sequence of tests. Table 1 presents the first twelve such P-values. Consider, first, those tests for which the P-value is approximately 0.05; D_2 and D_8 are examples. A crucial question is: among the drugs for which the P-value of the test is approximately 0.05, what fraction are actually ineffective (i.e., correspond to true H_0)? Likewise, consider those D_i for which the P-value is approximately 0.01 (D_5 and D_{10} are examples) and ask: what fraction are actually ineffective?

Table 1. P-values resulting from the first twelve clinical trials, testing $H_0\colon D_i$ has no effect vs. $H_1\colon D_i$ has an effect.

DRUG	D_1	D_2	D_3	D_4	D_5	D_6
P-Value	0.41	**0.04**	0.32	0.94	**0.01**	0.28
DRUG	D_7	D_8	D_9	D_{10}	D_{11}	D_{12}
P-Value	0.11	**0.05**	0.65	**0.009**	0.09	0.66

The answers to these questions are, of course, indeterminate. They depend on the actual sequence of $\{\theta_i\}$ that arises. However, using robust Bayesian techniques one can find lower bounds on the answers that are valid for *any* sequence $\{\theta_i\}$. These can be computed as in Berger and Sellke (1987, Section 4.3), and are 0.24 for the first question and 0.07 for the second.

This is quite startling, since most statistical users would believe that, when the P-value is 0.05, H_0 is very likely to be wrong and, when the P-value is 0.01, H_0 is almost certain to be wrong. The actual truth is very different. And since 0.24 and 0.07 are lower bounds that are actually difficult to attain, the fractions of true H_0 encountered in practice would typically be much larger (on the order of 50% and 15%, respectively). Thus we have a situation where the standard classical method, or at least its standard interpretation, is highly misleading. □

There is also a more subtle potential use of robust Bayesian analysis within frequentist statistics, arising from the fact that "optimal" frequentist procedures are virtually always Bayes (or generalized Bayes) procedures. Note that this, by itself, is not a compelling reason for a frequentist to adopt the Bayesian viewpoint, because the prior distribution that is used to develop the frequentist procedure can be considered merely to be a mathematical artifact, with no inherent meaning. (Using a prior to develop the procedure but ignoring its Bayesian implications may appear to be rather myopic, but it is not illogical to do so from the frequentist perspective.)

When the statistical problem becomes even moderately difficult, however, in the sense that the frequentist accuracy or performance measure is not constant over the unknown parameters, it can become very difficult for the frequentist to recommend a particular procedure. A very appealing possibility is to then use the Bayesian perspective to choose the prior, and to consider the resulting Bayes procedure from the frequentist perspective. If the Bayesian procedure is a robust Bayesian procedure, there are numerous indications that it will have excellent frequentist properties. See Berger (1984, 1985), DasGupta and Studden (1988a, 1989), Berger and Robert (1990), Robert (1992), Mukhopadhyay and DasGupta (1993) and DasGupta and Mukhopadhyay (1994), for such arguments in general; here we content ourselves with an interesting example, from Berger, Brown, and Wolpert (1993).

Example 2. Suppose $X_1, X_2, \ldots$ are i.i.d. $\mathcal{N}(\theta, 1)$ and that it is desired to test $H_0\colon \theta = -1$

versus $H_1\colon \theta = 1$. If the hypotheses have equal prior probability, the Bayesian inference, after stopping experimentation at sample size N, will be to (i) compute the posterior probability of H_0, which can be seen to be (defining $\overline{x}_N = \sum_{i=1}^{N} x_i/N$)

$$\begin{aligned} P(H_0|x_1,\ldots,x_N) &= 1/[1+\exp\{2N\overline{x}_N\}] \\ &= 1 - P(H_1|x_1,\ldots,x_N); \end{aligned}$$

(ii) choose the hypothesis with larger posterior probability (assuming the utility structure is symmetric); and (iii) report the posterior probability of the rejected hypothesis as the error probability.

There would seem to be no problem here for a frequentist: simply choose the most powerful Neyman-Pearson test with, say, equal error probabilities. But the situation is not so clear. First, this could have been a sequential experiment (e.g., the SPRT) with N being the stopping time, and stopping rules can have a dramatic effect on classical testing. Second, even if N is fixed, the most powerful test has strange properties. For instance, if $N = 4$, the frequentist error probabilities corresponding to the test "reject if $\overline{x}_4 \geq 0$ and accept otherwise" would be 0.025, and this would be the reported error for either $\overline{x}_4 = 0$ or $\overline{x}_4 = 1.5$; this is very strange because $\overline{x}_4 = 0$ would seem to indicate no evidence for H_0 versus H_1 (since 0 is equidistant between $\theta = -1$ and $\theta = +1$), while $\overline{x}_4 = 1.5$ would indicate overwhelming evidence for H_1 (it being 5 standard errors from H_0).

When standard frequentist procedures behave unnaturally, frequentists turn to conditional frequentist procedures (cf., Kiefer, 1977). But in this problem there are a plethora of possible conditional frequentist tests, and it is unclear how one should be chosen. Also, the interpretation of conditional tests and conditional error probabilities can be very difficult for practitioners.

Now look back at the simple Bayes test described at the beginning of the example. It is easy to use; it does not depend on the stopping rule in a sequential setting; it avoids the intuitive objections to the Neyman-Pearson test (when $\overline{x} = 0$, one reports $P(H_0|x_1,\ldots,x_4) = 0.5$ and, when $\overline{x} = 1.5$, one reports $P(H_0|x_1,\ldots,x_4) \cong 6 \times 10^{-6}$); and it has a simple interpretation. This test would be delightful for a frequentist, if only it could be given a frequentist interpretation. But it can! Indeed, in Berger, Brown, and Wolpert (1993), it is shown that this is a valid conditional frequentist test, with conditional error probabilities being given by the posterior probabilities.

Because this situation involved only the testing of simple hypotheses, the choice of the prior was not particularly relevant, and hence Bayesian robustness was not a factor. In testing of composite hypotheses, however, it appears to be necessary to utilize robust Bayesian procedures if one seeks to have sensible tests with a conditional frequentist interpretation. This work is currently under development. □

1.2 Preview

First, this is not exactly a review paper. More formal and thorough reviews can be found in Berger (1984, 1990, and, to a lesser extent, 1985) and in Wasserman (1992b). We will make a somewhat uneven effort to indicate the literature that has arisen since these review papers, but there will be only moderate discussion of this literature.

The primary goals of the paper are, instead, to provide a fairly accessible discussion of Bayesian robustness for statisticians not in the field, and to summarize our views on some of the important issues and considerations in Bayesian robustness.

Section 2 considers the idea of choosing models and priors that are inherently robust. The idea is that it is perhaps easier to build robustness into the analysis at the beginning, than to attempt to verify robustness at the end.

Section 3 briefly discusses diagnostics, influence, and sensitivity. Our review of this material is admittedly too brief; it is deserving of much more coverage.

Section 4 spends a perhaps inordinate amount of space on the issue of global robustness: finding the range of Bayesian answers as the Bayesian inputs vary. This area has experienced by far the most active development in recent years.

Uses of computing in Bayesian robustness are discussed in Section 5; perhaps of particular interest is the possibility of using Bayesian robustness to enhance interactive elicitation. Section 6 summarizes some thoughts about the future.

There is one major aspect of Bayesian robustness that is essentially ignored in the paper, namely robustness with respect to the utility or loss function. This mirrors a similar avoidance of the issue in the literature. There are, perhaps, three reasons for this avoidance. First, formal statistical decision analysis is not often done in practice (at least by statisticians), because of the extreme difficulty in eliciting utilities. (But perhaps Bayesian robustness is, for this reason, even more compelling in decision problems.) Second, modelling uncertainty in utility functions is often more awkward, and more case-specific, than modelling uncertainty in distributions. Finally, robust Bayesian analysis involving utility functions can be technically more difficult than other types of Bayesian robustness. A few references to robustness involving the utility are Kadane and Chuang (1978), Moskowitz (1992), Rios Insua (1990, 1992), Rios Insua and French (1991), Drummey (1991), Basu and DasGupta (1992), and Rios Insua and Martín (1994).

We will also ignore several other important robustness issues for reasons of space. One such is the issue of model selection and Bayesian prediction in the face of model uncertainty. For discussion and references see Draper (1992), Kass and Raftery (1992), Berger and Pericchi (1993), and Pericchi and Pérez (1994).

We also will not discuss the huge literature on gamma minimax estimation, which is the frequentist version of robust Bayesian analysis. Extensive discussion of this approach, and its relationship to the posterior robust Bayesian approach discussed here, can be found in Berger (1984, 1985), which also contain numerous references. Recent references include Ickstadt (1992), Vidakovic (1992), and Eichenauer-Herrmann and Ickstadt (1993).

Finally, there have been numerous Bayesian robustness investigations in particular problems or situations. A partial list of recent works is Kass and Greenhouse (1989), Lavine and Wasserman (1992), Berger and Chen (1993), Goldstein and Wooff (1994), and O'Hagan (1994).

1.3 Notation

The entire data set will be denoted by X, which will be assumed to arise from a density $f(x|\theta_f)$ (w.r.t. a fixed dominating measure), with θ_f denoting unknown parameters of f. A prior density for θ_f will be denoted by $\pi(\theta_f)$; we will assume that this is a density w.r.t. Lebesgue measure, for notational convenience.

Key Bayesian quantities are

$$m(x|\pi, f) = \int f(x|\theta_f)\pi(\theta_f)d\theta_f,$$

which is the marginal or predictive density of X, and

$$\pi(\theta_f|x, f) = f(x|\theta_f)\pi(\theta_f)/m(x|\pi, f)$$

which, assuming the denominator is nonzero, is the posterior density of θ_f. We explicitly retain f in the notation to allow for discussion of robustness w.r.t. f. For analyses in which

f is fixed, we will simply drop f from the notation. Finally, we define $\psi(\pi, f)$ (suppressing x) to be the posterior (or other) quantity of interest. Typically,

$$\psi(\pi, f) = \int h(\theta_f)\pi(\theta_f|x)d\theta_f = \frac{\int h(\theta_f)f(x|\theta_f)\pi(\theta_f)d\theta_f}{\int f(x|\theta_f)\pi(\theta_f)d\theta_f}.$$

For instance, $h(\theta_f) = \theta_f$ yields the posterior mean and $h(\theta_f) = 1_C(\theta_f)$ (the indicator function on the set C) yields the posterior probability of C. Other types of $\psi(\pi, f)$ are, however, possible: for instance, posterior quantiles or $m(x|\pi, f)$ itself.

2. DEVELOPMENT OF INHERENTLY ROBUST PROCEDURES

2.1 Introduction

Choices of the functional form of the statistical model or prior distribution are frequently quite arbitrary.

Example 3. Suppose $X_1, \ldots, X_n$ are felt to be i.i.d. observations from the measurement error model $X_i = \mu + \varepsilon_i$, where the measurement errors, ε_i, have a symmetric, unimodal distribution with unknown variance σ^2. Very little is known about σ^2, but the unknown μ is felt, apriori, to be $0 \pm \sqrt{2.19}$; we will interpret this to mean that 0 and $\sqrt{2.19}$ are the prior mean and prior standard error, respectively.

The "standard" analysis here would be to choose $f(x_i|\mu, \sigma)$ to be $\mathcal{N}(\mu, \sigma^2)$, and to choose $\pi(\mu, \sigma) = \frac{1}{\sigma} \cdot \pi_1(\mu)$, where $\pi_1(\mu)$ is $\mathcal{N}(0, 2.19)$. (The unknown σ is here given the usual noninformative prior. Sometimes $\pi_1(\mu|\sigma) = \mathcal{N}(0, (2.19)\sigma^2)$ is used in place of $\pi_1(\mu)$.) □

While various arguments can be given for such standard choices, the fact remains that they are often quite arbitrary. Furthermore, standard choices such as these often result in models from the exponential family and conjugate priors, both of which are known to be nonrobust in various ways: models in the exponential family are very sensitive to outliers in the data, and conjugate priors can have a pronounced effect on the answers even if the data is in conflict with the specified prior information. (This last is not always bad, but most users prefer to "trust the data" in such situations.) Further discussion and other references can be found in Berger (1984, 1985).

2.2 Use of Flat-tailed Distributions

Considerable evidence has accumulated that use of distributions with flat tails tends to be much more robust than use of standard choices, such as those discussed in Section 2.1. See Dawid (1973), Box and Tiao (1973), Berger (1984, 1985), O'Hagan (1988, 1990), Angers and Berger (1991), Fan and Berger (1992), Geweke (1992), and Lucas (1992).

Example 3 (continued). Suppose, instead, that $f(x_i|\mu,\sigma)$ is chosen to be a t-distribution with, say, 4 degrees of freedom. One might actually want to introduce the degrees of freedom, α, as an unknown parameter (see Chib, Osiewalski, and Steel, 1991, for a recent study), but that is more a model elaboration than a model robustification. Also, $\pi_1(\mu)$ could be chosen to be Cauchy(0, 1) (which matches the quartiles of a $\mathcal{N}(0, 2.19)$).

This analysis would be robust in two respects. First, if there are outliers in the data, they will automatically be filtered out of the analysis. Second, if the prior information about μ turns out to be very inaccurate (due, say, to the all-too-common problem that elicitors typically choose prior variances that are much smaller than their real uncertainties), then it is automatically discounted in the analysis. Neither of these robust behaviors occurs with the standard analysis. □

The price to be paid for utilization of inherently robust procedures is computational; closed form calculation is no longer possible. Today, however, computational schemes exist for performing robust Bayes computations routinely. For instance, any situation involving normal models and normal priors that is to be analyzed with Gibbs sampling can, instead, be done with t-distributions (cf, Verdinelli and Wasserman, 1991; Geweke, 1992; and Datta and Lahiri, 1992).

Example 3 (continued). Saying that $X_i \sim \mathcal{T}_4(\mu, \sigma^2)$ is equivalent to saying that, given τ_i, $X_i \sim \mathcal{N}(\mu, \sigma^2/\tau_i)$, where $\tau_i \sim$ Gamma $(2, \frac{1}{2})$. Likewise, saying that $\mu \sim \mathcal{C}(0,1)$ is equivalent to saying that, given τ_0, $\mu \sim \mathcal{N}(0, 1/\tau_0)$, where $\tau_0 \sim$ Gamma $(\frac{1}{2}, \frac{1}{2})$. By introducing the τ_i as random unknowns, it is possible to write the conditional posterior distributions of each unknown, given the others, as simple normal, gamma, or inverse gamma distributions, allowing for straightforward Gibbs sampling. □

While the above example indicates that, in principle, robustification is always possible for normal models, the computational cost may still be severe. For instance, the original two unknowns, (μ, σ), above are replaced by the unknowns $(\mu, \sigma, \tau_0, \tau_1, \ldots, \tau_n)$. When n

is large, the Gibbs sampling simulation can be very expensive.

Introducing such robustifications in hierarchical Bayes scenarios is often much more cost effective. For instance, replacing the standard hierarchical Bayes model, $X_i \sim \mathcal{N}(\theta_i, \sigma^2)$ and $\theta_i \sim \mathcal{N}(\mu, A)$, for $i = 1, \ldots, p$, by the model $X_i \sim \mathcal{T}_4(\theta_i, \sigma^2)$ and $\theta_i \sim \mathcal{C}(\mu, A)$, and introducing τ_i to convert the latter model to a normal and inverse gamma model, would only increase the number of parameters from $p+3$ to $3p+3$. A factor of 3 in Gibbs sampling is not severe.

Note that analytic methods for doing computations in certain of these hierarchical models exist. See Spiegelhalter (1985), Fan and Berger (1990, 1992), Angers and Berger (1991), Angers (1992), and Angers, MacGibbon, and Wang (1992).

A somewhat more modest type of robust prior has long existed in multivariate problems. Suppose $\boldsymbol{X} = (X_1, \ldots, X_p)^t \sim \mathcal{N}(\boldsymbol{\theta}, \Sigma)$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^t$ is unknown and Σ is given. The conjugate prior for $\boldsymbol{\theta}$ is a $\mathcal{N}(\boldsymbol{\mu}, A)$ prior, for which the posterior mean is $\boldsymbol{\delta}^\pi(\boldsymbol{x}) = \boldsymbol{x} - \Sigma(\Sigma + A)^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$. A variety of arguments suggest that it is more robust to use "shrinkage" versions of $\boldsymbol{\delta}^\pi$; among the many approaches to developing such are minimax theory, ridge regression, empirical Bayes analysis, and BLUP theory. But the best robust alternatives to $\boldsymbol{\delta}^\pi$ are, arguably, the robust Bayes alternatives, in which the $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ prior is replaced by a $\mathcal{T}_\alpha(\boldsymbol{\mu}, \Sigma)$ prior (for, say, $\alpha = 4$) or something similar. Extensive discussion of one such alternative, that is particularly easy to work with, can be found in Berger (1985, Section 4.7.10), which also has many references. See, also, Zellner (1976) and Berger and Robert (1990).

The reasons this latter type of robustness is more limited than the earlier type are: (i) model robustness is not involved; (ii) one achieves robustness to prior misspecification only in the overall sense that if the prior and data clash, the entire prior is discounted. The earlier discussed use of independent t-distributions would allow discounting of only part of the prior.

2.3 Use of Noninformative and Partially Informative Priors

That noninformative priors often yield automatically robust answers was recognized as early as Laplace (1812). Indeed, his development of the Central Limit Theorem was essentially a demonstration that, for large sample sizes, the posterior distribution of an unknown model parameter θ is essentially the same asymptotic normal distribution for any nonzero prior density. (See Ghosh, Ghosal, and Samanta, 1994, for recent developments and references.) For this and various intuitive reasons, Laplace felt that simply using $\pi(\theta) = 1$, as the prior, would give quite robust answers.

Another sense in which use of $\pi(\theta) = 1$ is robust was formalized by L. J. Savage as the theory of *precise measurement*. See Edwards, Lindman, and Savage (1963), Moreno and Pericchi (1993b), Mukhopadhyay and DasGupta (1993), and Pericchi and Pérez (1994).

Modern noninformative prior theory takes this one step further. Noninformative priors are specifically constructed so as to have minimal influence, in some sense, on the answer. The sense in which this engenders robustness is rather weak: it seems to ensure that the Bayesian analysis, for small or moderate samples, is not affected by unintended properties of the prior. For instance, in Example 3 we saw that standard conjugate choices of the tail of the prior (or likelihood) could have a dramatic unintended effect on the posterior. In multivariate situations, the potential for such unintended effects is particularly large, since few features of the prior will actually be subjectively elicited and there is a substantial possibility that mistakes can "accumulate" across the dimensions.

The two most extensively developed noninformative prior theories of this type are the *reference prior* theory (cf., Bernardo, 1979; Berger and Bernardo, 1992; and Bernardo and Smith, 1994), and the *maximum entropy* approach (cf., Jaynes, 1983, and Fougere, 1990). Other approaches are discussed in the excellent review paper Kass and Wasserman (1993).

Partially informative priors are also of considerable interest from the robustness perspective. These priors are of two types. The first type is for use in problems where there are, say, "parameters of interest" and "nuisance parameters." The parameters of interest are basically given subjectively elicited prior distributions, perhaps with associated robustness investigations being performed, while the nuisance parameters are given noninformative priors. The idea here is that elicitation of priors for nuisance parameters is likely to be difficult and a less valuable use of available elicitation time, and that attempting formal robustness studies with respect to the nuisance parameters is likely to be ineffective. For examples and further discussion of this general notion, see Liseo (1993) and Berger and Mortera (1994).

The second type of commonly used partially informative prior is a constrained maximum entropy prior. The idea here is that one specifies certain features of the prior (or model) and then chooses that prior (or model) which maximizes entropy subject to the specified constraints. The hope is that the resulting prior (or model) will have the specified features, but be robust (in the noninformative prior sense) with respect to unspecified features. For further discussion see Jaynes (1983) and Fougere (1990). Somewhat different approaches are considered in Casella and Wells (1991) and Bernardo and Smith (1994).

2.4 Nonparametric and Infinite Parametric Bayes Procedures

Bayesian nonparametrics can be considered to be an approach to automatic robustness with respect to model choice. A large nonparametric class of models is entertained and given a prior distribution so that, hopefully, the data will cause the analysis to automatically adapt to the true model.

The majority of the work on Bayesian nonparametrics has involved use of the Dirichlet process prior on the space of all probability distributions. Recent references include Brunner and Lo (1989), Lo and Weng (1989), Gasparini (1990), Ferguson, Phadia, and Tiwari (1992), Tamura (1992), and Doss (1994).

Dirichlet process priors have a number of potentially unappealing features, such as the fact that they give probability one to the set of discrete probability measures. Hence there has been considerable effort expended to develop priors that are supported on continuous densities, such as Gaussian process priors. An example of such a prior, for the space of continuous densities, $f(t)$, on $[0, T]$, is to let

$$f(t) = \exp\{X(t)\} / \int_0^T \exp\{X(t)\} dt,$$

where $X(t)$ is the sample path of a Gaussian process. This and other such priors are studied in Leonard (1978), Lenk (1988), Angers and Delampady (1992), and Zidek and Weerahandi (1992). Computations with such priors are more difficult than with Dirichlet process priors, but the recent new Bayesian computational tools should enhance the utilization of these alternative nonparametric priors.

In regards to Gaussian process priors, the Bayesian interpretation of smoothing splines should also be mentioned. Smoothing splines can be developed as Bayesian function estimates for certain Gaussian process priors on derivates of functions. This interpretation has been important in deriving accuracy estimates for smoothing splines (utilizing the associated posterior covariance function). See Kohn and Ansley (1988), Wahba (1990) and Gu and Wahba (1993). There is considerable promise in further exploiting this relationship for higher dimensional smoothing splines, especially if structural assumptions on the function are made (e.g., $f(x_1, \ldots, x_p) = \sum_{i=1}^{p} f_i(x_i)$). Finally, other very promising nonparametric Bayes approaches are being developed, such as Lavine (1992b) and West (1992).

While Bayesian nonparametrics strives to produce inherently robust procedures, there have been a number of recent developments which suggest that caution must be exercised. For instance, a "minimal" robustness condition, that one would hope would be satisfied by any Bayes procedure, is consistency: as the sample size grows to infinity, the Bayes estimates of quantities of interest should converge to the true values. It has been discovered, however, that this need not be the case in Bayesian nonparametrics; see Diaconis and Freedman (1986), Ghosh (1993), and Berliner and MacEachern (1993). The following infinite parametric example is a very simple illustration of the phenomenon.

Example 4. J.K Ghosh (personal communication, 1992) has studied an interesting variant of the Neyman-Scott problem. Suppose we observe (all independently) $X_{ij} \sim \mathcal{N}(\mu_i, \sigma^2), i = 1, \ldots, p$ and $j = 1, 2$. It is desired to estimate σ^2. A simple consistent estimator, as $p \to \infty$, is $\hat{\sigma}^2 = \sum_{i=1}^{p} (x_{i1} - x_{i2})^2/(2p)$.

Now suppose a Bayesian were to proceed by choosing independent proper priors for all parameters $\{\sigma^2, \mu_1, \mu_2, \ldots, \mu_p\}$. Then, for "almost all" sequences $\{\mu_1, \mu_2, \ldots\}$, the Bayes estimator of σ^2 seems to be inconsistent. ("Almost all" here is in a topological sense, not probabilistic; the Bayes estimator is consistent for almost all sequences $\{\mu_1, \mu_2, \ldots\}$ in probability under the prior, but the set of such sequences becomes vanishingly small. Conditions on the priors and sequences are needed for the proof of inconsistency, but the result is probably true generally.) $\square$

Determining the extent to which such possible inconsistencies are a practical concern for Bayesians will be an important task for the future. At the very least, these concerns should significantly influence the types of priors chosen for these problems (cf., Ghosh, 1994, in regards to the above example).

3. DIAGNOSTICS, INFLUENCE, AND SENSITIVITY

3.1 Diagnostics

An important aspect of robustness is developing methods of detecting when a robustness problem exists and suggesting where the difficulty might lie. Examples include the detection of outliers and the detection of a lack of model fit.

Virtually all Bayesian diagnostic techniques are based on some type of utilization of $m(x|\pi, f)$. Interestingly, some suggested utilizations are non-Bayesian in character. For instance, Box (1980) suggests determining the adequacy of an assumed model, f_0, by choosing a noninformative prior, π_0, and then conducting a classical significance test with the null distribution being $m(x|\pi_0, f_0)$. The formal Bayesian approach would be to, instead, embed f_0 in a larger class of models $\mathcal{F}$, choose a prior distribution on $\mathcal{F}$, and infer the adequacy of f_0 relative to other distributions in $\mathcal{F}$ (through, say, Bayes factors or predictive measures). While we prefer the formal Bayesian approach if feasible, the purpose of diagnostics is often to provide an initial indication that something is wrong, and so suggest that the more formal Bayesian approach be undertaken. Evidence obtained from such initial pseudo-Bayesian diagnostics should not be trusted too far, however, and should be confirmed by the formal Bayesian approach before being considered conclusive. For further discussion of this issue, with examples, see Berger (1985, section 4.7.2.).

We do not have space to review the huge literature on Bayesian diagnostics. A few recent references are Smith (1983), Pettit (1988, 1992), Guttman and Peña (1988), Poirier (1988), Kass, Tierney and Kadane (1989), West and Harrison (1989), Carlin and Polson (1991), Verdinelli and Wasserman (1991), Geisser (1992), Kass and Slate (1992), Peña and Tiao (1992), Weiss (1992, 1993), Peña and Guttman (1993), and Meng (1994). Note that global robust Bayesian methods (see Section 4) have begun to themselves be applied to diagnostics; see Bayarri and Berger (1993b, 1994) for an application to outlier detection.

3.2 Influence and Sensitivity

Whereas diagnostics is oriented towards detecting that a problem exists with an analysis, influence and sensitivity seeks to determine which features of the model, prior, or utility, or which data, have a large effect on the answer. There are many parametric analyses of this type, including Guttman and Peña (1988, 1993), Kass, Tierney, and Kadane (1989), McCulloch (1989), Meczarski and Zieliński (1991), Geisser (1992), Lavine (1992d), and Basu and Jammalamadaka (1993).

A recent interesting approach to investigating sensitivity to the prior, in a nonparametric fashion, is to consider functional derivatives of the Bayes operator $\psi(\pi, f)$ with respect to π. (One could, similarly, take derivatives w.r.t. f, but this is usually more involved.) These derivatives, evaluated at a base prior π_0 and in "direction" g, indicate how sensitive $\psi(\pi, f)$ is to local changes in π_0. Besides indicating local sensitivity, these derivatives can be used to construct quite accurate global robustness bounds. The rapidly growing literature on functional derivatives in Bayesian robustness includes Diaconis and Freedman (1986), Cuevas and Sanz (1988), Srinivasan and Truszczynska (1990, 1993), Ruggeri and Wasserman (1991, 1993), Boratyńska and Zielińska (1991), Fortini and Ruggeri (1992, 1994), Sivaganesan (1993c), Basu, Jammalamadaka and Liu (1993a, 1993b), Gustafson and Wasserman (1993), Delampady and Dey (1994), and Salinetti (1994).

4. GLOBAL ROBUSTNESS

4.1 Introduction

In Bayesian robustness it is frequently assumed that $f \in \mathcal{F}$ and that $\pi(\theta_f) \in \Gamma_f$, where $\mathcal{F}$ and Γ_f are classes of densities. (Frequently, Γ_f will be enlarged to include distributions that do not have densities with respect to Lebesgue measure; we will abuse notation when this is needed.) If $\psi(\pi, f)$ is the posterior functional of interest (e.g., the posterior mean), global robustness is concerned with computation of

$$\underline{\psi} = \inf_{f \in \mathcal{F}} \inf_{\pi \in \Gamma_f} \psi(\pi, f), \quad \overline{\psi} = \sup_{f \in \mathcal{F}} \sup_{\pi \in \Gamma_f} \psi(\pi, f). \tag{4.1}$$

One then reports $(\underline{\psi}, \overline{\psi})$ as the range of possible answers. If this range is small enough for the conclusion to be clear, the conclusion is declared to be robust. If not, further elicitation, data collection, or analysis is necessary.

4.2 Parametric Classes

Historically, global robustness has been investigated using parametric classes of likelihoods and priors.

Example 5. In the situation of Example 3, instead of considering the $\mathcal{N}(\mu, \sigma^2)$ density for the i.i.d. $X_1, \ldots, X_n$, one could consider the class

$$\mathcal{F} = \{\mathcal{T}_\alpha(\mu, K_\alpha \sigma^2) \text{ densities for the } X_i,\ \alpha \geq 1\}, \tag{4.2}$$

where $\sqrt{K_\alpha} = (0.674)/q_\alpha$, with q_α being the third quartile of the $\mathcal{T}_\alpha(0,1)$ distribution. The reason for introducing K_α is that the $\mathcal{T}_\alpha(\mu, K_\alpha \sigma^2)$ distribution will then have the same quartiles as the $\mathcal{N}(\mu, \sigma^2)$ distribution, so that μ and σ^2 will have comparable meanings across all distributions. If the restriction $\alpha > 2$ were employed, one could instead choose K_α so that all distributions have the same mean and variance, but we generally prefer scaling by quartiles.

Suppose μ and σ^2 are thought to be independent apriori, with μ having unimodal density with quartiles $-1, 0, 1$ and nothing being known about σ^2. Then the prior, $\pi(\mu, \sigma^2)$, might be assigned the class

$$\Gamma = \{\pi(\mu, \sigma^2) = \pi_1(\mu)\pi_2(\sigma^2)\colon\ \pi_1 \text{ is } \mathcal{T}_\nu(0, q_\nu^{-2}), \nu \geq 1,\ \text{and } \pi_2(\sigma^2) = (\sigma^2)^a, -2 \leq a \leq 0.\}$$

The $\mathcal{T}_\nu(0, q_\nu^{-2})$ distributions are a fairly wide class of unimodal distributions with quartiles $-1, 0, 1$, and might appropriately represent the specified information about μ. Since nothing is specified about σ^2, it would be typical to use a range of noninformative priors as the relevant class (but see Pericchi and Walley, 1991; and Walley, 1991, for other suggestions). Note that, because of the scaling of the $f \in \mathcal{F}$ to preserve the meaning of μ and σ^2, it is not necessary to write (μ_f, σ_f^2) and define Γ_f depending on f.

For any functional $\psi(\pi, f)$ of interest, one can now compute $(\underline{\psi}, \overline{\psi})$ by minimizing and maximizing $\psi(\pi, f)$ with respect to (α, ν, a). $\square$

There are two main reasons that parametric robustness is attractive. The first is that computations are relatively straightforward. For instance, in Example 5, the maximizations are only three-dimensional. Of course, computation of the $\psi(\pi, f)$ will involve two-dimensional integration (over μ and σ^2), so the computation is not trivial (see, also, Section 5.1).

The second attractive feature of parametric classes is that they can allow for convenient communication of robust Bayesian conclusions. An example is given in Section 4.10.2 of Berger (1985).

The main disadvantage of parametric classes is that they may fail to capture realistic possible deviations from the base model or prior. Thus, in Example 5, we have robustified against normality, in the sense of allowing flatter tails for the distributions, but no allowance for, say, possible skewness has been made. Ideally, one will construct $\mathcal{F}$ and/or the Γ_f to reflect all deviations that are deemed to be possible, but it is unfortunately all-too-common to fail to anticipate the actual deviations that arise.

Recent references utilizing parametric classes of priors include Leamer (1982), Polasek (1985), Good and Crook (1987), Polasek and Pötzelberger (1988, 1994), DasGupta and Studden (1988a, 1988b, 1989, 1991), Drummey (1991), Pötzelberger and Polasek (1991), Coolen (1993), and Dette and Studden (1994).

4.3 Nonparametric Classes of Priors

The majority of recent papers on Bayesian robustness deal with a fixed likelihood and nonparametric classes of priors. This is an important problem, for several reasons. First, there are many situations in which priors are less well known than the model. Second, the major objection of non-Bayesians to Bayesian analysis is uncertainty in the prior, so eliminating this concern can make Bayesian methods considerably more appealing. Third, serious inadequacies in certain classical methods can be revealed by Bayesian prior robustness (see Sections 1.2 and 4.3.3). Finally, conclusions must frequently be reached by a group of people with differing prior opinions, and robust Bayesian analysis, with Γ equal to the class of prior opinions, can then have a variety of uses.

That said, the main reason researchers have concentrated on global prior robustness is probably its mathematical elegance. There is nothing wrong with this, of course, as long as we remember that global prior robustness is only one piece of the robustness puzzle.

In the remainder of this subsection, f will be considered fixed, so we write just θ for the unknown parameters, Γ for the class of priors being considered, and $\psi(\pi)$ (instead of $\psi(\pi, f)$) as the criterion functional.

4.3.1 Factors Involved in Choosing a Class

Several discussions and reviews concerning choice of good classes of priors already exist, including Berger (1990), Sivaganesan (1990), Lavine (1991), Pericchi and Walley (1991), Walley (1991), Moreno and Pericchi (1992a), and Wasserman (1992b). The following issues should be kept in mind in choosing a class:

(i) The class should be as easy to elicit and interpret as possible. Recall that a prime reason for considering Bayesian robustness is the difficulty of eliciting a prior; making the class, Γ, difficult to elicit would thus be self-defeating.

(ii) The class should be as easy to handle computationally as is possible. The usual computational technique is to identify "extreme points" of Γ (relative to $\psi(\pi)$) and perform maximizations over these extreme points. Typically, the extreme points will be in a low-dimensional subset, Γ^*, of Γ, so the maximizations are over a low-dimensional set. The dimension of Γ^* will depend on several factors, but primarily on the dimension of θ and the number of elicited features of the prior. Hence, rather paradoxically, the more features one elicits, the harder the robust Bayesian computation is likely to become. Part of the computability issue is also having a class, Γ, which is compatible with model and/or utility robustness.

(iii) The size of Γ should be appropriate, in the sense of being a reasonable reflection of prior uncertainty. If Γ is too small, one might fear being erroneously led to a conclusion of robustness. If Γ is too large, in the sense of containing many prior distributions that are clearly unreasonable, then one might conclude that robustness is lacking when, in fact, a reasonable Γ would imply robustness. For detecting this latter problem, it is useful to determine the $\pi \in \Gamma$ at which $\underline{\psi}$ or $\overline{\psi}$ is attained, and judge if such a π is reasonable. If not, one should try to refine Γ to eliminate such π.

(iv) Γ should be extendable to higher dimensions and adaptable in terms of allowing incorporation of constraints (e.g., shape constraints, independence, etc.). The point here is that eventual methodological implementations will need to be based on at most a few "standard" classes (for elicitational, computational, and interpretational reasons), and so these classes need to be flexible enough to handle a very wide range of problems.

The following simple example illustrates several of the above ideas.

Example 6. Suppose prior beliefs about a real parameter θ are symmetric about 0, with the third quartile, q_3, being between 1 and 2. Consider

$$\Gamma_1 = \{\mathcal{N}(0, \tau^2) \text{ priors, } 2.19 < \tau^2 < 8.76\}, \quad \Gamma_2 = \{\text{all symmetric priors with } 1 < q_3 < 2\}.$$

Both classes are easy to elicit (i.e., easy to specify from the given information; the range of τ^2 in Γ_1 yields q_3 between 1 and 2). Also, both are easy to handle computationally; indeed, maximization over Γ_2 will often only involve maximization over the "extreme points"

$$\Gamma_2^* = \{\text{distributions giving probability } \tfrac{1}{2} \text{ each to } \pm q_3\text{: } 1 < q_3 < 2\}.$$

Although Γ_1 can be appropriate for some situations, it will often be considered "too small" because of its specified prior shape and because it has only sharp-tailed distributions.

In contrast, Γ_2 will typically be a "too big" reflection of the prior information, in the sense that it contains prior distributions which, upon reflection, are probably unreasonable.

Very sensible classes can be formed by taking "too large" classes, such as Γ_2, and adding shape constraints. For instance, if it is also believed that the prior density is unimodal, then one obtains

$$\Gamma_3 = \{\text{unimodal, symmetric densities with } 1 < q_3 < 2\}.$$

Such classes are often very sensible, in that they are large enough to include all reasonable priors compatible with prior information, but small enough that unreasonable priors are excluded. □

4.3.2 Common Classes

We briefly review the common classes of priors that are used. For extensive discussion, comparisons, and examples, see the references listed under each class.

Classes of Given Shape or Smoothness: An example of a class based on shape is $\Gamma = \{\text{all symmetric, unimodal priors}\}$. Such classes have interesting uses in hypothesis testing (see Section 4.3.3). Usually, however, shape is used as an additional constraint in one of the other classes (cf, Example 6), so as to eliminate unreasonable priors from the class. Note that general shape features are often relatively easy to elicit, even in higher dimensions.

Smoothness constraints typically limit the rate of change of the prior density. (Note that requiring only continuity adds nothing, because arbitrary distributions can typically be approximated, arbitrarily well, by continuous densities.) Although one could define a class of priors based only on smoothness, it is typically used, instead, as a supplemental constraint for other classes (cf., Bose, 1990, 1994).

Moment Class: This is defined as the set of all priors with a specified collection of moments. Analysis using probabilistic moment theory is typically straightforward. See Sivaganesan and Berger (1989, 1993), Goutis (1991), Betró, Meczarski and Ruggeri (1994), and Sivaganesan (1992).

Moments are quite difficult to elicit. For this reason, moment conditions are also typically used merely as additional constraints in other classes, in the hope that misspecification of moments will then have a reduced effect.

Contamination Class: This is defined by

$$\Gamma = \{\pi = (1-\varepsilon)\pi_0 + \varepsilon q, \quad q \in \mathcal{Q}\}, \tag{4.3}$$

where π_0 is a base prior (for instance, the prior elicited in a standard Bayesian analysis), ε is the perceived possible error in π_0, and $\mathcal{Q}$ is the allowed class of contaminations. In terms of the four criteria of Section 4.3.1, this class is easy to elicit; computation is relatively easy for many reasonable choices of Q; and the class can easily incorporate additional constraints and be used in higher dimensions. The class can be "too big" if $\mathcal{Q}$ is "too big" and ε is appreciable. In one dimension this is rarely a problem, but it can be a severe problem in higher dimensions. References include Berger and Berliner (1986), Sivaganesan (1988, 1989, 1993a), Sivaganesan and Berger (1989), Moreno and Pericchi (1990, 1991), Dey and Birmiwal (1991), Gelfand and Dey (1991), Boratyńska (1991), Lavine (1991b), Moreno and Cano (1991), and Bose (1994).

Density Ratio (or Density Band) Class: This is defined by

$$\Gamma = \{\text{generalized } \pi\colon\ L(\theta) \le \pi(\theta) \le U(\theta)\}. \tag{4.4}$$

(A "generalized" prior is one which does not integrate to 1; typically the posterior will, nevertheless, be proper.) Often this class is the simplest to handle computationally, and is reasonable in higher dimensions. Its main disadvantage is that it is very hard to elicit; choosing L and U appropriately can be quite difficult.

A useful modification of this class is the *Density Bounded Class*, which is as in (4.4), but with the additional constraint that π must be proper. The class then becomes much easier to elicit and interpret, but can be more challenging computationally.

References working with these classes include DeRobertis (1978), DeRobertis and Hartigan (1981), Hartigan (1983), Lavine (1991a, 1991b, 1992c), Ruggeri and Wasserman (1991), Wasserman (1991, 1992a, 1992b, 1992c), Moreno and Pericchi (1992b), and Sivaganesan (1994).

Quantile Class: This is defined by

$$\Gamma = \{\pi\colon\ \alpha_i \le Pr(\theta \in \Theta_i) \le \beta_i, \quad i = 1, \ldots, m\},$$

where the Θ_i are specified subsets of Θ. (Usually $\{\Theta_i;\ i = 1, \ldots, m\}$ is a partition of Θ.) This class is probably the most natural of all from the viewpoint of elicitation, and is computationally manageable. It tends to be "too big" in higher dimensions, however, unless additional shape constraints are added. References to this class include Cano, Hernandez, and Moreno (1985), Berger and O'Hagan (1988), O'Hagan and Berger (1988), Moreno and Cano (1989), Moreno and Pericchi (1990), Ruggeri (1990, 1991, 1992), and Sivaganesan (1991).

Mixture Classes: These are of the form

$$\Gamma = \{\pi(\theta) = \int \pi(\theta|\alpha) dG(\alpha),\ G \in \mathcal{G}\}. \tag{4.5}$$

Most other classes are actually themselves mixture classes.

Example 7. Suppose $\boldsymbol{\theta} \in R^p$, and the prior distribution is known to depend only on $|\boldsymbol{\theta}|$. The class, Γ_S, of all such priors is typically too big, in the sense of containing many unreasonable distributions. Often, however, unimodality is also believed to hold, leading to Γ_{US}, the class of unimodal spherically symmetric priors. Interestingly, this class can be written as

$$\Gamma_{US} = \{\pi(\boldsymbol{\theta}) = \int_0^\infty 1_{(0,\alpha)}(|\boldsymbol{\theta}|) V_\alpha^{-1} dG(\alpha),\quad G \text{ any c.d.f. on } [0, \infty)\}, \tag{4.6}$$

where V_α is the volume of the ball in R^p of radius α. This is a much smaller class than Γ_S, and would be reasonable for most purposes, but it may be possible to refine the class even further. In particular, if prior beliefs are felt to be "bell-shaped," a class such as

$$\Gamma_{NS} = \{\pi(\boldsymbol{\theta}) = \int_0^\infty (2\pi\alpha)^{-p/2} e^{-|\boldsymbol{\theta}|^2/(2\alpha)} dG(\alpha),\quad G \text{ any c.d.f. on } [0, \infty)\}$$

could be employed. This is easily seen to be a subset of Γ_{US} that contains only bell-shaped distributions (though admittedly not all bell-shaped distributions). Recall that we earlier encountered such priors in Section 2.2, as being "inherently robust" for certain G. □

Example 8. An archeological artifact is θ years old, θ unknown. It could have been produced by any one of 3 civilizations that occupied the given site. For civilization i, a $\mathcal{N}(\mu_i, A_i)$ distribution (to be denoted π_i) is thought to describe the likelihood of artifact production at any given time. (All μ_i and A_i are assumed to be specified.)

Several experts are asked to classify the object, based on its style. They do not agree completely, but conclude that all their opinions are contained in

$$\mathcal{G} = \{\boldsymbol{g} = (g_1, g_2, g_3):\ 0.1 \le g_1 \le 0.2,\ 0.6 \le g_2 \le 0.7\},$$

where $g_i = Pr$(the artifact is from civilization i), and $g_1 + g_2 + g_3 = 1$. Then $\pi(\theta)$, the overall prior distribution for θ, is in

$$\Gamma = \left\{ \pi(\theta) = \sum_{i=1}^{3} g_i \pi_i(\theta):\ \boldsymbol{g} \in \mathcal{G} \right\},$$

which is of the form (4.5) with $\alpha = i$ and G being discrete. $\square$

Mixture classes will play a very prominent role in Bayesian robustness because of several key properties:

(i) Mixture classes are often computationally simple. In Example 7, for instance, maximization over Γ_{US} or Γ_{NS} will typically reduce to maximization over $\pi(\boldsymbol{\theta}|\alpha) = 1_{(0,\alpha)}(|\boldsymbol{\theta}|)V_\alpha^{-1}$ or $\pi(\boldsymbol{\theta}|\alpha) = (2\pi\alpha)^{-p/2}\exp\{-|\boldsymbol{\theta}|^2/(2\alpha)\}$, respectively, both of which are simple one-dimensional maximizations (over α).

(ii) Mixture classes can flexibly represent prior information about structure or shape, as in Example 7, or information arising from several sources, as in Example 8.

(iii) Mixture classes are often not "too big," in the sense of containing unreasonable distributions. This is particularly crucial for multivariate $\boldsymbol{\theta}$, where the range of Bayesian answers as π varies over Γ will typically be huge, unless Γ is somehow constrained so as not to contain unreasonable distributions. Operating with mixture classes seems to be the only effective way of avoiding the problem (other than using parametric classes, of course).

As a final comment about mixture classes, note that they can also arise as refinements in the elicitation process. In Example 8, for instance, suppose $X \sim \mathcal{N}(\theta, \sigma^2)$ is observed (say, X is a radiocarbon dating of the artifact). One might first consider just $\Gamma = \{\pi_1, \pi_2, \pi_3\}$, and compute the Bayesian answer (e.g., posterior mean of θ) for each prior in Γ. If the range of answers is small enough, there would be no need to look further. If, however, there are substantial differences between the answers, then one might go to the next "level" of elicitation, obtaining $\boldsymbol{g}$. In this example, uncertainty thus ends up residing in the higher level elicitation (see also Good, 1983b; and Pericchi and Nazaret, 1988). Note that there could easily be virtually complete robustness with respect to $g \in \mathcal{G}$, even if there is no robustness with respect to the π_i in Γ. For other examples of mixture classes, see Bose (1990, 1993), Cano (1993), Moreno and Pericchi (1993a), and Liseo, Petrella, and Salinetti (1993). West (1992) discusses their uses in modelling.

Marginal and Independence Classes: When $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ is multi-dimensional, elicitation of $\pi(\boldsymbol{\theta})$ is particularly difficult. One could hope that elicitation of, say, the marginal densities, $\pi_i(\theta_i)$, could be effective, in the sense that robustness over

$$\Gamma = \{\pi(\boldsymbol{\theta}) \text{ having the specified marginals}\} \tag{4.7}$$

would often obtain. Alas, this is not the case, as was dramatically shown by Lavine, Wasserman, and Wolpert (1991); the class in (4.7) is so large that the range of resulting Bayesian answers is typically enormous. See, also, Moreno and Cano (1992) for related results.

Of course, if the θ_i were, apriori, judged to be independent, then one would simply have the single prior $\pi(\boldsymbol{\theta}) = \prod_{i=1}^{p} \pi_i(\theta_i)$. It is sometimes possible to make the judgment of independence, and it is then natural to consider its effect on other classes. This is studied for contamination classes and density ratio classes in Berger and Moreno (1994), where it is shown that the assumption of independence of coordinates does have a dramatic effect on robustness; the range of Bayesian answers can decrease dramatically. Independence is admittedly a strong assumption, but one typically must make strong assumptions in multi-dimensional problems to obtain a moderate range of Bayesian answers.

Other Classes: An interesting alternative to the Density Ratio class for one-dimensional θ is the *Distribution Band* class of all priors whose c.d.f. lies between two nondecreasing functions. This is studied in Basu (1992a, 1992b) and Basu and DasGupta (1992).

Neighborhood classes can be defined by choosing a "distance measure," $d(\pi_1, \pi_2)$, between priors (it need not be a true distance function), and defining $\Gamma = \{\pi\colon\ d(\pi, \pi_0) \leq \varepsilon\}$, where π_0 is again a "base" prior. Related classes can be developed using "concentration functions"; see Regazzini (1992) and Fortini and Ruggeri (1990, 1992, 1993, 1994).

Belief Function classes use belief functions (a type of generalization of probability) to generate the class of priors. See Wasserman (1990) for an example.

Classes based on *Choquet Capacities* are defined and studied in Wasserman and Kadane (1990, 1992a) and Wasserman (1992b). Two-alternating capacities have particularly attractive theoretical and computational properties.

Pericchi and Walley (1991) and Walley (1991) (see also Sansó and Pericchi, 1992) propose *Near Ignorance* classes of priors to provide a robust Bayesian alternative to noninformative priors. Their approach provides an interesting contrast to typical constructions of Γ, which seek to construct Γ so as to contain all the "nice," believable priors. Instead, Pericchi and Walley argue that one can construct a "nice" class from a collection of "not nice" or not compatible priors, and that there can be positive advantages in doing so. In part, this notion arises from the axiomatic development in Walley (1991), which effectively shows that rationality corresponds to operating with some Γ, but does not require or imply that the priors in Γ correspond to actual subjective beliefs. Although counterintuitive to standard Bayesian thinking, this approach should not be casually dismissed. Its counterintuitive nature, however, poses real difficulties for elicitation of Γ.

Conclusions: No single class of priors is likely to dominate robust Bayesian analysis. Our personal favorites are contamination, quantile, and mixture classes, with shape and structural restrictions as appropriate. We prefer the contamination and quantile classes because they are easiest to elicit and interpret. They can be considerably more difficult than, say, the density ratio class in terms of computation, but computations will eventually just be hidden within software. The important issue will be how user-friendly is the software in terms of choice of Γ, so the most easily elicitable classes are to be preferred.

The argument for mixture classes is somewhat different, although they too are often natural from an elicitation viewpoint. The argument is simply the necessity, in multidimensional problems, of doing something fairly drastic to reduce the size of Γ in order to avoid excessively large ranges of Bayesian answers. It is important to be clear here: in very low dimensional problems one can often verify Bayesian robustness, even when the prior inputs are very weak. In high-dimensional problems this is typically impossible, and one must accept the need to make rather strong and "dangerous" assumptions if an answer is to be obtained. The point, of course, is to only make those strong assumptions which seem plausible. The next section contains an example illustrating some of these notions.

4.3.3 Application to Hypothesis Testing and Ockham's Razor

Some of the most interesting applications of robust Bayesian analysis have been to hypothesis testing, and related model selection ideas. We review these here, in part as an illustration of points made in the preceding section.

Suppose $\boldsymbol{X} = (X_1, \ldots, X_p) \sim \mathcal{N}_p(\boldsymbol{\theta}, \sigma^2 \boldsymbol{I})$, σ^2 known, and that model M_1 specifies $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, while model M_2 has $\boldsymbol{\theta}$ unrestricted. Under M_2, consider the following two classes of priors for $\boldsymbol{\theta}$:

$$\Gamma_A = \{\text{all prior distributions}\},$$

$$\Gamma_{\boldsymbol{\mu}} = \{\text{all prior densities of the form } \pi(\boldsymbol{\theta}) = h(|\boldsymbol{\theta} - \boldsymbol{\mu}|),\ h \text{ nonincreasing}\}.$$

Here $\boldsymbol{\mu}$ is fixed, corresponding to a prior "most likely" value of $\boldsymbol{\theta}$ under M_2 or, perhaps, to the "center of symmetry" of π under M_2. Often, $\boldsymbol{\mu}$ will equal $\boldsymbol{\theta}_0$, but other values are possible.

The Bayes factor of M_1 to M_2, corresponding to a prior density $\pi(\boldsymbol{\theta})$ under M_2, is

$$B(\pi) = f(\boldsymbol{x}|\boldsymbol{\theta}_0) \Big/ \int f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}.$$

Define $\underline{B}_A$ and $\underline{B}_{\boldsymbol{\mu}}$ as the lower bounds on $B(\pi)$ as π ranges over Γ_A and $\Gamma_{\boldsymbol{\mu}}$, respectively.

For the case $\boldsymbol{\mu} = \boldsymbol{\theta}_0$, which arises naturally in testing H_0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus H_1: $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ with prior opinions under H_1 being symmetric about $\boldsymbol{\theta}_0$, Table 2 gives values of $\underline{B}_A$ and $\underline{B}_{\boldsymbol{\theta}_0}$; instead of presenting the values as a function of $\boldsymbol{x}$, we state them as a function of the P-value associated with $\boldsymbol{x}$. (See Delampady, 1989, for computation of $\underline{B}_{\boldsymbol{\theta}_0}$.)

Table 2. Lower Bounds, $\underline{B}_A$ and $\underline{B}_{\boldsymbol{\theta}_0}$, corresponding to various P-values for testing H_0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus H_1: $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.

		dimension p								
P-value		1	2	3	4	5	10	15	20	40
0.05	$\underline{B}_A$	.146	.050	.020	.009	.004	10^{-4}	10^{-6}	10^{-7}	10^{-12}
	$\underline{B}_{\boldsymbol{\theta}_0}$	.409	.348	.326	.314	.307	.293	.288	.284	.279
0.01	$\underline{B}_A$	.036	.010	.003	.001	.0005	10^{-5}	10^{-7}	10^{-8}	10^{-14}
	$\underline{B}_{\boldsymbol{\theta}_0}$	.123	.098	.090	.085	.082	.078	.075	.074	.073
0.001	$\underline{B}_A$	.004	.001	.0003	.0001	10^{-5}	10^{-7}	10^{-8}	10^{-10}	10^{-16}
	$\underline{B}_{\boldsymbol{\theta}_0}$	.018	.014	.012	.011	.010	.009	.009	.009	.009

This table reveals the, by now familiar, discrepancy between P-values and Bayes factors. In one dimension for instance, when the P-value is 0.05 the lower bound on the Bayes factor over all symmetric (about $\boldsymbol{\theta}_0$) unimodal priors is 0.409, and the lower bound over *all* priors is 0.146. This proves that a P-value of 0.05, in this situation, is at best quite weak evidence against H_0.

A secondary point is the demonstration that classes of priors which are "too big" fail to give useful bounds in high dimensions. Thus $\underline{B}_A$ becomes uselessly small as the dimension increases. In contrast, $\underline{B}_{\boldsymbol{\theta}_0}$ is very stable as the dimension increases. And note that, as in (4.6), $\Gamma_{\boldsymbol{\theta}_0}$ can be written as a one-dimensional mixture class; this is an example of why we view mixture classes as promising in high dimensions.

For $p = 1$ and general μ, a very accurate approximation to $\underline{B}_\mu$ is

$$\underline{B}_\mu = 2\varphi(d_0)[d_1 + \sqrt{2\log(d_1 + 1.2)}], \tag{4.8}$$

where $d_0 = |x - \theta_0|/\sigma$, $d_1 = |x - \mu|/\sigma$, and φ is the standard normal density. This is argued, in Berger and Jefferys (1992) and Jefferys and Berger (1992), to be a "Bayesian Ockham's razor" for comparing models M_1 and M_2. For instance, those papers discussed the situation of comparing, in about the year 1920, M_1: *Einstein's general relativity* versus M_2: *Newcomb's gravity theory*, based on data from unexplained perturbations in the orbit of Mercury. The situation fits our framework with $\theta_0 = 42.9$ (the perturbation predicted by M_1); $\mu = 0$ (Newcomb's theory made no prediction about the size or sign of the perturbation, so centering prior opinion at zero is natural); and $x = 41.6$ (the observed perturbation, with a standard error of $\sigma = 2$). Computation yields $\underline{B}_\mu = \underline{B}_0 = 15.04$; since this is a lower bound over $\Gamma_\mu = \Gamma_0$, we can conclude that the evidence favors M_1 by *at least* 15 to 1. This relates to Ockham's razor because M_1 was the "simple" model, in the sense of having no free parameter (it specified $\theta_0 = 42.9$), while M_2 allowed θ to float freely. "Ockham's razor" argues that one should prefer a simple model that adequately explains the data to a complex model that does so, which is precisely what $\underline{B}_\mu$ established quantitatively.

As a final point, it must be remembered that the $\underline{B}$ above are lower bounds on the Bayes factor, and can be much lower than actual reasonable Bayes factors (cf, Bayarri and Berger, 1994). If the lower bound, itself, answers the question of interest, then all is well. If not, substantial refinement of Γ is needed. Note that, in contrast to estimation problems, there are not (in general) "robust" noninformative priors for testing or model selection problems. See Kass and Raftery (1992) and Berger and Pericchi (1993) for discussion and default methods of proceeding.

Robust Bayesian analysis of testing problems can be found in Edwards, Lindman, and Savage (1963), Berger and Sellke (1987), Berger and Delampady (1987), Casella and Berger

(1987), Delampady (1989a, b), Moreno and Cano (1989), DasGupta and Delamapady (1990), Delampady and Berger (1990), Berger and Mortera (1991, 1994), Berger (1992), and Berger and Jefferys (1992).

4.4 Nonparametric Classes of Likelihoods

Dealing with likelihoods via the global robustness approach varies from trivial to nearly impossible. One approach is to take the observed likelihood function, $\ell_0(\theta) = f_0(x|\theta)$, for a hypothesized model f_0 and given data, and embed it in a larger class of likelihoods, such as $\mathcal{F}_\varepsilon = \{\ell(\theta) = (1-\varepsilon)\ell_0(\theta) + \varepsilon q(\theta),\ q \in \mathcal{Q}\}$. Since $\ell(\theta)$ and $\pi(\theta)$ operate interchangeably in Bayesian computations, this approach to likelihood robustness is equivalent to the global prior robustness approach (with the contamination class).

The difficulty with this approach is that such classes of likelihoods do not reflect typical types of uncertainty in f. For instance, if $X_1, X_2, \ldots, X_n$ are i.i.d. $g(x_i|\theta)$, uncertainty would typically reside in g, reflected by, say,

$$\mathcal{F}_1^g = \{g = (1-\varepsilon)g_0(x_i|\theta) + \varepsilon q(x_i|\theta),\ q \in \mathcal{Q}\} \quad \text{or}$$

$$\mathcal{F}_2^g = \{\text{densities } g\colon\ g_1(x_i|\theta) \leq g(x_i|\theta) \leq g_2(x_i|\theta)\}.$$

(Even these may not be completely natural, but they suffice for making the point.) The resulting classes of likelihoods are

$$\mathcal{F}_j = \left\{\ell\colon\ \ell(\theta) = \prod_{i=1}^{n} g(x_i|\theta),\ g \in \mathcal{F}_j^g\right\}, \quad j = 1, 2, \tag{4.9}$$

and these are very complex and difficult to work with. For instance, the relevant subclasses of extreme points are typically at least n-dimensional, which can become prohibitively expensive computationally for large n.

A second difficulty with classes of likelihoods, as in (4.9), is that they can be too large, unless the $\mathcal{F}_j^g$ are very small. One approach that does seem to give useful answers is that of Lavine (1991a, 1991b, and 1994).

For special or restricted problems, robustness analysis can be much easier. Robustness among certain generalized elliptical distributions is studied in Osiewalski and Steel (1993a, b, c), and Fernández, Osiewalski, and Steel (1993). The following example is from another special situation, studied in Bayarri and Berger (1993a).

Example 9 (*Weighted Distributions*): Assume that the random variable $X \in \mathbb{R}^1$ is distributed over some population of interest according to $f(x|\theta)$, $\theta \in (r, s)$, a (possibly infinite) interval in $\mathbb{R}^1$, but that, when $X = x$, the probability of recording x (or the probability that x is *selected* to enter the sample) is $w(x)$. Then the true density of an actual observation is

$$f_w(x|\theta) = \frac{w(x)f(x|\theta)}{\nu_w(\theta)}, \tag{4.10}$$

where $\nu_w(\theta) = E_\theta[w(X)]$. Selection models occur often in practice (Rao, 1985; Bayarri and DeGroot, 1992).

Often the specification of $w(\cdot)$ is highly subjective. It is thus of considerable interest to study the robustness of the analysis to choice of w. The problem becomes particularly important in the multi-observational setting, because the effect of the weight function can then be extremely dramatic. Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. from the density (4.10), so that the likelihood function for θ is

$$L_w(\theta) \propto \ell(\theta)[\nu_w(\theta)]^{-n}, \tag{4.11}$$

where $\ell(\theta) \propto \prod_{i=1}^{n} f(x_i|\theta)$ would be the likelihood function for the unweighted base density. If $\pi(\theta)$ is the prior density for θ, the posterior density is then

$$\pi(\theta|x, w) = \frac{\ell(\theta)[\nu_w(\theta)]^{-n}\pi(\theta)}{\int \ell(\theta)[\nu_w(\theta)]^{-n}\pi(\theta)d(\theta)}, \tag{4.12}$$

assuming π is such that the denominator is finite. Expression (4.12) suggests that, at least for large n, the weight function w can have a much more significant effect on $\pi(\theta|x, w)$ than might the prior π. Hence we will treat $\pi(\theta)$ as given here; for instance, it might be chosen to be a noninformative prior for the base model $f(x_i|\theta)$.

In Bayarri and Berger (1993), this problem is studied for the class of weight functions

$$\mathcal{W} = \{\text{nondecreasing } w\text{: } w_1(x) \le w(x) \le w_2(x)\}, \tag{4.13}$$

where w_1 and w_2 are specified nondecreasing functions representing the extremes of beliefs concerning w. Posterior functionals

$$\psi(w) = \int \xi(\theta)\pi(\theta|x, w)d\theta \tag{4.14}$$

are studied for a variety of shapes of the target $\xi(\theta)$. When $\xi(\theta)$ is monotonic (e.g., $\xi(\theta) = \theta$ or $\xi(\theta) = 1_{(c,\infty)}(\theta)$), the extreme points in $\mathcal{W}$ at which $\overline{\psi} = \sup_w \psi(w)$ and $\underline{\psi} = \inf_w \psi(w)$

are attained were shown to have one of the following two forms:

$$w(x) = \begin{cases} w_1(x) & \text{if } r < x \le a \\ w_2(x) & \text{if } a < x < s \end{cases}, \tag{4.15}$$

$$w(x) = \begin{cases} w_2(x) & \text{if } r \le x < h_2(c) \\ c & \text{if } h_2(c) < x < h_1(c), \\ w_1(x) & \text{if } h_1(c) < x < s \end{cases} \tag{4.16}$$

where $h_1(c) = \inf\{x\colon\ w_1(x) \le c\}$ and $h_2(c) = \sup\{x\colon\ w_2(x) \ge c\}$. The condition needed for this result is primarily that $f(x|\theta)$ have monotone likelihood ratio.

As a specific example, suppose $f(x_i|\theta) = \theta \exp\{-\theta x_i\}$ for $i = 1, \ldots, n$, where $x_i > 0$ and $\theta > 0$. Any x_i that is less than a value T_1 is, however, not observed. Any x_i that is greater than T_2 is observed. For $T_1 \le x_i \le T_2$, the probability of its being observed is not known, but the probability is known to be nondecreasing. This specifies the class of weight functions in (4.13), with $w_1(x) = 1_{(T_2,\infty)}(x)$ and $w_2(x) = 1_{(T_1,\infty)}(x)$.

Suppose $\xi(\theta) = \theta$ is of interest, so that $(\underline{\psi}, \overline{\psi})$ is the range of the posterior mean as w ranges over $\mathcal{W}$. Then one can explicitly minimize and maximize (4.14) over w of the form (4.15) and (4.16), obtaining $\underline{\psi} = 1/(\overline{x} - T_1)$ and $\overline{\psi} = 1/(\overline{x} - T_2)$. Whether or not robustness is achieved is thus easy to determine. Note that it depends on the size of $\overline{x}$ as well as the closeness of T_1 and T_2. □

4.5 Limitations of Global Robustness

Global robustness ignores a very important quantity, namely $m(x|\pi, f)$, which can be considered to be the "likelihood" of π and/or f. A full Bayesian analysis automatically takes this into account.

Example 10. Suppose $X_i \sim \mathcal{N}(\theta_i, 1),\ i = 1, \ldots, p$. The class of prior distributions under consideration for $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ is

$$\Gamma = \{\pi(\boldsymbol{\theta})\colon \text{ the } \theta_i \text{ are i.i.d. } \mathcal{N}(\mu, 1),\quad -8 < \mu < 12\}.$$

Suppose we are interested in the posterior mean for θ_1. This is given by $\hat{\theta}_1 = (x_1 + \mu)/2$. Thus the range of posterior means, as π varies over Γ, is $(\frac{1}{2}x_1 - 4, \frac{1}{2}x_1 + 6)$.

Calculation shows that, here,

$$m(\boldsymbol{x}|\pi) = \tfrac{1}{(4\pi)^{p/2}} \exp\left\{-\tfrac{1}{4}[p(\overline{x} - \mu)^2 + \sum_{i=1}^{p}(x_i - \overline{x})^2]\right\}.$$

Thus values of μ close to $\overline{x}$ are far more "likely" than values far from $\overline{x}$. For instance, if $p = 8$ and $\overline{x} = 3$, this likelihood is a normal likelihood with mean 3 and variance 1/4, so that only the μ between 2 and 4 have appreciable likelihood. Note that, if a full Bayesian analysis were done with, say, μ being given the noninformative prior $\pi(\mu) = 1$, then the posterior mean for θ_1 would be

$$\begin{aligned}
\hat{\theta}_1^* &= \int \tfrac{1}{2}(x_1 + \mu)\pi(\mu|\overline{x})d\mu \\
&= \int \tfrac{1}{2}(x_1 + \mu)\frac{1}{\sqrt{2\pi(2/p)}} \exp\{-\tfrac{p}{4}(\overline{x} - \mu)^2\}d\mu \\
&= \tfrac{1}{2}(x_1 + \overline{x}),
\end{aligned}$$

which has effectively "weighted" the $(x_1 + \mu)/2$ by the likelihood of μ. □

The message here is that a global robustness analysis might erroneously indicate a lack of robustness, erroneous in the sense that, were $m(x|\pi, f)$ taken into account, robustness might obtain. There are two possibilities for formally investigating if this is so. The first is to go to a "higher level" Bayesian robustness investigation, as in Cano (1993) and Moreno and Pericchi (1993a).

Example 10 (continued). It is determined that $(\frac{1}{2}x_1 - 4, \frac{1}{2}x_1 + 6)$ is too large an interval to reach a conclusion. Prior information about μ, the presumed common mean of the θ_i, is thus considered. A "best guess" for μ is 2, but there is considerable uncertainty in this guess. It is decided that the standard error of this guess is at least 2, but that finer elicitation would be difficult. This information can reasonably be modeled by the class of priors (for μ)

$$\Gamma^* = \{\mathcal{N}(2, \tau^2) \text{ densities}, \tau^2 \geq 4\}.$$

For given τ^2, an easy computation yields that the posterior mean for θ_1 is

$$\hat{\theta}_1^{**} = \tfrac{1}{2}(x_1 + \overline{x}) + \tfrac{1}{(p\tau^2+2)}(2 - \overline{x}).$$

The range of possible posterior means as $\pi(\mu)$ varies over Γ^* (i.e., for $\tau^2 \geq 4$) is thus (if, say, $2 - \overline{x} \leq 0$)

$$(\tfrac{1}{2}(x_1 + \overline{x}) + \tfrac{1}{(4p+2)}(2 - \overline{x}),\ \tfrac{1}{2}(x_1 + \overline{x})).$$

For the case $p = 8$ and $\overline{x} = 3$, this range is

$$(\tfrac{1}{2}(x_1 + 3) - \tfrac{1}{34},\ \tfrac{1}{2}(x_1 + 3)), \tag{4.17}$$

which would typically be considered to be a highly robust conclusion. $\square$

The second possibility for utilizing $m(x|\pi)$ (similar ideas apply if dealing with $m(x|\pi, f)$) is to replace Γ by

$$\Gamma^* = \{\pi \in \Gamma\colon\ m(x|\pi) \geq K\}; \tag{4.18}$$

here K could be chosen by likelihood or noninformative prior Bayesian methods. (See Sivaganesan and Berger, 1993, for development of this approach.)

Example 10 (continued). For the case $p = 8$ and $\overline{x} = 3$, we observed that $m(x|\pi)$ is essentially a $\mathcal{N}(3, \frac{1}{4})$ likelihood for μ. Likelihood or Bayesian noninformative prior methods would suggest that (2, 4) is a "95% confidence or credible set" for μ, so we might replace Γ by

$$\Gamma^* = \{\pi \in \Gamma\colon\ 2 \leq \mu \leq 4\}$$

(which can easily be seen to be of the form (4.18)). The range of the posterior mean for θ_1, as π ranges over Γ^*, is $(\frac{1}{2}x_1 + 1, \frac{1}{2}x_1 + 2)$, which might well be small enough to claim that the conclusion is reasonably robust. $\square$

This second method of incorporating $m(x|\pi)$ (or $m(x|\pi, f)$) is appealing because it seems to avoid the need to put "priors on priors", etc. It also is related to empirical Bayes techniques; indeed, empirical Bayes analysis can be thought of as simply replacing Γ by the prior $\pi^* \in \Gamma$ for which $m(x|\pi)$ is maximized (clearly the degenerate limit of (4.18)). Unfortunately, this second method can give the wrong answer (as can empirical Bayes analysis). A rather startling example of this is given in Bayarri and Berger (1994). Hence we cannot definitively recommend this second method.

4.6 Optimal Robust Procedures

Global Bayesian robustness lends itself naturally to defining notions of optimality. Here is an example, from Sivaganesan, Berliner, and Berger (1993).

Example 11. We observe $X \sim \text{Cauchy}(\theta, 1)$. Elicitation yields $-0.3, 0.0, 0.3$ as the prior quartiles for θ. The usual "inherently robust" prior density for θ would be the Cauchy(0, 0.3) density; call this π_0. Even though one expects considerable inherent robustness in this situation, it is decided to formally consider global robustness with respect to the contamination class of priors Γ, in (4.3), with $\varepsilon = 0.01$.

Suppose now that a credible set, C, for θ, is desired and that it is (conservatively) decided to require that the posterior probability of C satisfy

$$Pr(\theta \in C|x, \pi) \geq 0.90 \quad \text{for all } \pi \in \Gamma. \tag{4.19}$$

Under this condition, one can be assured that C is a 90% credible set.

A natural notion of optimality, here, is to define C^* as optimal if C^* has minimal size (e.g., Lebesgue measure) among all C satisfying (4.19). In Sivaganesan, Berliner, and Berger (1993), it is shown how to find such optimal C^* for quite general problems of this type. For the specific case considered here, and when $x = 6$ is observed, the optimal C^* is $C^* = (-1.22, 2.70) \cup (3.56, 8.43)$; note that this is the union of two intervals, one where the likelihood is large and one where the prior is large. □

Many other notions of optimality w.r.t. Γ in global robustness are discussed in Berger (1985), Wasserman (1989), Li and Saxena (1990), DasGupta (1991), Meczarski (1991), Basu (1992c), De la Horra and Fernandez (1993, 1994), and Sivaganesan (1993b).

Optimal global robustness is potentially useful. For instance, if C^* in Example 11 is deemed to be a small enough set for practical purposes then, in light of (4.19), one can be quite satisfied. It can even be possible to design the experiment so as to achieve this with high predictive probability (cf, Mukhopadhyay and DasGupta, 1993; and DasGupta and Mukhopadhyay, 1994).

There is a serious danger with some optimality notions, however: the optimal procedure can be terrible from a "real" Bayesian perspective. This is because, as discussed in Section 4.5, it can be important to take $m(x|\pi, f)$ into account. (See, also, Berger, 1985; DasGupta and Studden, 1988; Sivaganesan and Berger, 1993; Zen and DasGupta, 1993; and Bayarri and Berger, 1994.)

Example 10 (continued). The initial global robustness analysis yielded $(\frac{1}{2}x_1 - 4, \frac{1}{2}x_1 + 6)$ as the range of posterior means for θ_1. Many notions of optimality would suggest that the midpoint of this interval, $\frac{1}{2}x_1 + 1$, is optimal. However, this corresponds to the value $\mu = 2$, which has very low likelihood, $m(\boldsymbol{x}|\mu)$; indeed, we saw that $\mu = 2$ is at the edge of the "95% noninformative prior credible set" for μ. Furthermore, $\frac{1}{2}x_1 + 1$ is well outside the interval of possibilities in (4.17) that was obtained by a "higher level" robustness analysis. □

5. COMPUTING

5.1 Computational Issues

In discussing creation of classes of likelihoods or priors, it was observed that computational considerations are crucial. Here we briefly review several generally useful computational techniques.

Linearization: It is easy to see that, under mild conditions, $\overline{\psi}$ (see (4.1)) is the solution to

$$0 = \sup_{f \in \mathcal{F}} \sup_{\pi \in \Gamma_f} \int [h(\theta_f) - \overline{\psi}] f(x|\theta_f)\pi(\theta_f) d\theta_f. \tag{5.1}$$

The point here is that maximization over $\psi(\pi, f)$ is a non-linear operation, but it can be converted, via (5.1), to a linear maximization together with a root-finding operation. This can be a useful simplification. (However, if one can theoretically determine the relevant functional extreme points of the class, (5.1) is unnecessary.) Development and discussion of this algorithm can be found in DeRobertis and Hartigan (1981), Lavine (1991b), Lavine, Wasserman, and Wolpert (1993), Wasserman (1992b), and Wasserman and Kadane (1992a). The latter two papers discuss computation of $\overline{\psi}$, via (5.1), in the important case when it is necessary to utilize Monte-Carlo techniques for computation of the integral.

Reweighting: When computing Bayesian integrals via Monte-Carlo techniques, there are opportunities for relatively easy robustness investigations. To take the simplest case, suppose we approximate $\psi(\pi)$ by

$$\begin{aligned} \psi(\pi) &= \int h(\theta)\pi(\theta|x)d\theta \\ &\cong \frac{\sum_{i=1}^{N} h(\theta^{(i)}) f(x|\theta^{(i)})\omega_\pi(\theta^{(i)})}{\sum_{i=1}^{N} f(x|\theta^{(i)})\omega_\pi(\theta^{(i)})}, \end{aligned} \tag{5.2}$$

where $\omega_\pi(\theta^{(i)}) = \pi(\theta^{(i)})/g(\theta^{(i)})$, and $\theta^{(1)}, \ldots, \theta^{(N)}$ is an i.i.d. sample from the "importance function" g. (See Berger, 1985, for background.) Then switching from one prior to another simply requires recomputing the "weights" $\omega_\pi(\theta^{(i)})$, a relatively simple operation. Indeed, a scheme such as this is virtually necessary for efficient maximization of $\psi(\theta)$, since (5.2) provides a well-defined function to maximize over π. (The alternative, of, say, trying to maximize over π with $\psi(\pi)$ being computed anew by numerical integration at each step, is very unstable.) For formal discussion as to when this scheme for maximization is convergent, see Salinetti (1994).

Reweighting schemes are, unfortunately, not useful if too wide a range of π is being considered. This is because the approximation in (5.2) need not be accurate if the weights can be extremely large. If, however, g can be chosen so that $\omega_\pi(\theta) \leq K$ (moderate) for all π under consideration, then (5.2) can be extremely effective.

Reweighting schemes are also possible for more complicated Markov Chain simulation procedures. See Stephens and Smith (1992) for discussion.

5.2 Interactive Robustness

In the Introduction, the possibility of using Bayesian robustness to guide the elicitation process was mentioned. Developing computer-interactive methods of doing this is particularly appealing. Ultimately, one could hope to have a robust Bayesian computer package that processed any given partial information, provided the implied range of Bayesian answers (see Moskowitz, 1992, for a description of such a system for discrete problems) and suggested what additional elicitations would be most desirable, if needed to increase robustness. Here is a simple example we are in the process of developing.

Example 12. Suppose that elicited information, at stage m of the interactive elicitation process for a real-valued parameter θ, will be a set of quantiles $q_1 < q_2 < \ldots < q_m$, with elicited $p_i = Pr(\theta \in (q_i, q_{i+1}]), i = 1, \ldots, m$. (Allowing for uncertainty in the p_i would be an easy modification.) Also, suppose that the prior distribution for θ is felt to be unimodal. Then, at stage m, one has effectively specified the class of priors

$$\Gamma_m(\boldsymbol{q}, \boldsymbol{p}) = \{\text{all unimodal distributions with the given quantiles}\}. \tag{5.3}$$

Suppose $\psi(\pi)$ is the posterior functional of interest (the likelihood is being considered fixed), and that the degree of robustness is reasonably measured by

$$\overline{\psi}_m - \underline{\psi}_m = \sup_{\pi \in \Gamma_m} \psi(\pi) - \inf_{\pi \in \Gamma_m} \psi(\pi).$$

The problem of computing $\overline{\psi}_m$ and $\underline{\psi}_m$ is discussed in Berger and O'Hagan (1988), and O'Hagan and Berger (1988).

Suppose $\overline{\psi}_m - \underline{\psi}_m$ is deemed to be too large, and that additional refinement of the prior is needed. Since we are eliciting in terms of quantiles, this means that a new quantile, q^*, must be chosen, with the associated p^* (for the new interval created) being elicited. Which q^* should be chosen? It is quite natural to make the choice so that the ensuing $\overline{\psi}_{m+1} - \underline{\psi}_{m+1}$ is likely to be smallest; this would make q^* maximally efficient in terms of Bayesian robustness.

A reasonable scheme for implementing this idea is to consider each possible candidate location, b, for q^*. If $q_i < b < q_{i+1}$, one could assume that the "least informative" elicitation will be done, resulting in

$$p(b) = Pr(\theta \in (q_i, b]) = \frac{(b - q_i)}{(q_{i+1} - q_i)} \cdot p_i.$$

(This just assumes that the prior probability p_i, assigned to $(q_i, q_{i+1}]$, is distributed uniformly over the interval.) Special adjustments have to be made for $b < q_1$ or $b > q_m$. Assuming b and $p(b)$ specify the new quantile, one would have the new class $\Gamma_{m+1}(\{\boldsymbol{q} \cup b\}, \{\boldsymbol{p} \cup p(b)\})$ as in (5.3), and could compute the corresponding range $(\overline{\psi}_{m+1}(b) - \underline{\psi}_{m+1}(b))$. Minimizing over b would yield the quantile that, if elicited, would be most likely to result in a substantial gain in robustness. (Of course, once $q^* = b$ is chosen, the actual p^* would be elicited; $p(b)$ would not be used.) $\square$

Schemes for interactive elicitation could also be developed based on notions of "most sensitive direction in prior space," as discussed in Section 3.2. The difficulty with such an approach is that the optimal direction in which to focus elicitation efforts may not correspond to quantities that are easy to elicit. Hence we prefer to consider the types of allowed elicitations (e.g., quantiles) as being specified in advance, at which point it is probably easier to consider $\overline{\psi} - \underline{\psi}$ directly, rather than look at local sensitivity.

6. FUTURE DIRECTIONS

Many of the theoretical and methodological directions in which Bayesian robustness is developing were discussed in the paper. Rather than attempting to summarize that discussion, it is useful to focus here on the *types* of statistical problems in which Bayesian robustness can be most usefully applied.

Statistical problems fall into several different categories. The most difficult are problems in which it is a challenge to perform any Bayesian analysis whatsoever. For such problems it will inherently be the case that formal Bayesian robustness cannot be investigated; at best, the informal "try a few models and priors" will be done.

The next category consists of those problems in which subjective Bayesian analysis is feasible, but objective (noninformative) Bayesian methods are also available (and perhaps classical methods that are very similar to the objective Bayesian methods). For such problems, subjective Bayesian analysis is typically performed only when the subjective information is quite influential, relative to the information in the data. Until subjective Bayesian methods become more widely used in these problems, the scope for utilization of formal Bayesian robustness methods will be limited. Of course, as mentioned in the Introduction, the capability to routinely augment a subjective analysis with robustness determinations might increase the willingness to use subjective methods. Probably the most immediate contribution of Bayesian robustness for this class of problems is the possibility of using relatively sophisticated automatically robust methods, such as those discussed in Section 2, as an alternative to the standard methodology.

The third category of statistical problems consists of problems for which objective Bayesian methods do not exist (i.e., the answers typically depend significantly on prior opinions) or are to be avoided (see below). One example that we have discussed is precise hypothesis testing, where the prior on the parameter space of the alternative hypothesis always has a significant effect. Other examples are mentioned below.

For problems in this category, it can be argued that robust Bayesian analysis is *required*; since the answer depends strongly on prior opinions, it is important to show that any conclusions are valid over the range of sensible prior opinions. Problems in this category also typically lack sensible classical answers, so that many non-Bayesian are more willing to consider Bayesian approaches to these problems. It is thus this third category of problems that promises to provide the most immediate applications of robust Bayesian theory. A brief, partial listing of these problems follows.

Similarly to precise hypothesis testing, in *Model Selection* from among models of differing complexity, the prior distributions always have a significant effect. The challenge here, for the robust Bayesian approach, is to choose classes of priors that are appropriately "tied together" for the differing models (see Berger and Pericchi, 1993, for discussion of what this means). Simply having unrelated classes of priors is likely to result in uselessly wide ranges of answers (see Berger and Mortera, 1994, for an illustration in a simple setting).

Extrapolation beyond the range of the data inherently involves subjective opinion, and is very non-robust. Hence it is a natural problem in which to consider Bayesian robustness. See Berger and Chen (1993) for an example.

A common aspect of *Meta-Analysis* is the need to relate the various studies or experiments that are to be combined; the protocol, populations studied, and experimental conditions will often vary from study to study, requiring adjustments if the studies are to be combined. In Bayesian analysis, these adjustment factors, which are typically highly subjective, are built in through the prior distributions (cf, DuMouchel and Harris, 1983; Morris and Normand, 1992; and Wolpert and Warren Hicks, 1992). Since the adjustments are typically highly uncertain, robust Bayesian analysis is natural. (See Berger and Mortera, 1991, for study of one such situation.)

We have already illustrated robust Bayesian analysis for *Selection Models* or *Weighted Distributions*. Because the selection or weighting mechanism can have an enormous effect and is often uncertain, there is clear motivation for studying Bayesian robustness in these problems.

Clinical Trials provide a natural domain for various types of Bayesian robustness investigations. The reason is that attention is increasingly being paid to conducting clinical trials in a fashion that is as ethical as possible towards the patients in the trial. There are two aspects of this that are particularly relevant to Bayesian robustness. First, the trials may assign patients to treatments in a partially non-random way that involves medical opinion. Second, prior opinion may be used to allow the trial to stop earlier, not only because of the effect of the additional information, but also because Bayesian sequential trials will naturally stop earlier (since repeated looks at the data are not penalized). In both cases, there are typically a variety of prior opinions that must be taken into account, and so some type of Bayesian robustness investigation is needed. For discussion and examples, see Berry, Wolf, and Sack (1992), Carlin and Louis (1993), Carlin, Chaloner, Louis, and Rhame (1993), Sedransk (1993), and Kadane (1994).

Group Decision Making is a related domain in which there naturally exist a variety of prior opinions. Group decision making often begins by seeing if there is a possible action that is simultaneously optimal for all members of the group. This would involve a type of robust Bayesian computation. Only if there is not an optimal answer, in this sense, would more involved group decision making techniques be utilized. (Note, however, that it is not necessarily correct to behave in this way; indeed, it is easy to construct examples where every member of the group initially thinks that a certain action is optimal but that, after sharing information, a different action is seen to be optimal. Ideally, therefore, complete information sharing should be done before applying the robust Bayesian methods.) See Genest and Zidek (1986) and Van Eeden and Zidek (1994) for discussion and references.

REFERENCES

Angers, J. F. (1992). Use of the student-t prior for the estimation of normal means: a computational approach. In *Bayesian Statistics 4*, J. M. Bernardo, et.al. (Eds.). Oxford University Press, London.

Angers, J. F. and Berger, J. O. (1991). Robust hierarchical Bayes estimation of exchangeable means. *Canadian J. Statist.*, **19**, 39–56.

Angers, J. F. and Delampady, M. (1992). Hierarchical Bayesian curve fitting and smoothing. *Canadian J. Statist.*, **20**, 35–49.

Angers, J. F., MacGibbon, B., and Wang, S. (1992). A robust Bayesian approach to the estimation of intra-block exchangeable normal means with applications. D.M.S. No. 92-7, Département de Mathématiques et de Statistique, Université de Montréal.

Basu, S. (1992a). Variations of posterior expectations for symmetric unimodal priors in a distribution band. Technical Report #214, Department of Statistics and Applied Probability, University of California, Santa Barbara.

Basu, S. (1992b). Ranges of posterior probability: symmetry, unimodality and the likelihood. Technical Report #214, Department of Statistics and Applied Probability, University of California, Santa Barbara.

Basu, S. (1992c). A new look at Bayesian point null hypothesis testing: HPD sets, volume minimizing sets, and robust Bayes. Technical Report, Department of Statistics and Applied Probability, University of California, Santa Barbara.

Basu, S. and DasGupta, A. (1992). Bayesian analysis with distribution bands: the role of the loss function. Technical Report #208, Department of Statistics and Applied Probability, University of California, Santa Barbara.

Basu, S. and Jammalamadaka, S. R. (1993). Local posterior robustness with parametric priors: maximum and average sensitivity. Technical report #239, Department of Statistics and Applied Probability, University of California, Santa Barbara.

Basu, S., Jammalamadaka, S. R., and Liu, Wei (1993a). Local posterior robustness: total derivatives and functional derivatives. Technical Report #239, Department of Statistics and Applied Probability, University of California, Santa Barbara.

Basu, S. Jammalamadaka, S. R., and Liu, W. (1993b). Qualitative robustness and stability of the posterior distributions and posterior quantities. Technical Report #238, Department of Statistics and Applied Probability, University of California, Santa Barbara.

Bayarri, M. J. and Berger, J. (1993a). Robust Bayesian analysis of selection models. Technical Report #93-6, Department of Statistics, Purdue University, W. Lafayette.

Bayarri, M. J. and Berger, J. (1993b). Robust Bayesian bounds for outlier detection. In *Proceedings of the* 4^{th} *International Meeting of Statistics in the Basque Country–IMSIBAC4*, M. L. Puri and J. P. Vilaplana (Eds.). North-Holland, Amsterdam.

Bayarri, M. J. and Berger, J. (1994). Applications and limitations of robust Bayesian bounds and type II MLE. In *Statistical Decision Theory and Related Topics V*, S. S. Gupta and J. Berger (Eds.). Springer–Verlag, New York.

Bayarri, M. J. and DeGroot, M. (1992). A BAD view of weighted distributions and selection models. In *Bayesian Statistics IV*, J. M. Bernardo, et.al. (Eds.). Oxford University Press, London.

Berger, J. (1984). The robust Bayesian viewpoint (with discussion). In *Robustness of Bayesian Analysis*, J. Kadane, (Ed.), North-Holland, Amsterdam.

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

Berger, J. O. (1990). Robust Bayesian analysis: sensitivity to the prior. *J. Statist. Plann. Inference*, **25**, 303–328.

Berger, J. (1992). A comparison of minimal Bayesian test of precise hypotheses. *Rassegna di Metodi Statistici ed Applicazioni*, **7**, Pitagora Editrice, Bologna.

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

Berger, J. and Bernardo, J. M. (1992). On the development of the reference prior method.

Bayesian Statistics 4, J. M. Bernardo, et. al. (Eds.). Oxford University Press, London.

Berger, J. O., Brown, L., and Wolpert, R. (1993). A unified conditional frequentist and Bayesian test for fixed and sequential hypothesis testing. Technical Report #93-14C, Department of Statistics, Purdue University, W. Lafayette.

Berger, J. O., and Chen, M. H. (1993). Determining retirement patterns: prediction for a multinomial distribution with constrained parameter space. *The Statistician* **42**, 427–443.

Berger, J. O. and Delampady, M. (1987). Testing precise hypotheses (with discussion). *Statist. Sci.*, **2**, 317-352.

Berger, J. O., and Jefferys, W. (1992). The application of robust Bayesian analysis to hypothesis testing and Occam's razor. *J. Ital. Statist. Society*, **1**, 17-32.

Berger, J. O. and Moreno, E. (1994). Bayesian robustness in bidimensional models: prior independence. To appear in *J. Statist. Planning and Inference*.

Berger, J. O. and Mortera, J. (1991). Interpreting the stars in precise hypothesis testing. *Int. Statist. Review.*, **59**, 337–353.

Berger, J. O. and Mortera, J. (1994). Robust Bayesian hypothesis testing in the presence of nuisance parameters. To appear in *J. Statist. Planning and Inference*.

Berger, J. O. and O'Hagan, A. (1988). Ranges of posterior probabilities for unimodal priors with specified quantiles. In *Bayesian Statistics 3*, J. M. Bernardo et al. (Eds.), Oxford University Press, Oxford.

Berger, J. O. and Pericchi, L. R. (1993). The intrinsic Bayes factor for model selection and prediction. Technical Report #93-43C, Department of Statistics, Purdue University, W. Lafayette.

Berger, J. O. and Robert, C. (1990). Subjective hierarchical Bayes estimation of a multivariate normal mean: on the frequentist interface. *Ann. Statist.*, **18**, 617–651.

Berger, J. O. and Sellke, T. (1987). Testing a point null hypothesis: the irreconcilability of significance levels and evidence (with discussion). *J. Amer. Statist. Assoc.*, **82**, 112–122.

Berliner, L. M. and MacEachern, S. N. (1993). Examples of inconsistent Bayes procedures based on observations on dynamical systems. *Statistics and Probab. Letters*, **17**, 355–360.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference (with discussion). *J. Roy. Statist. Soc.*, **41**, 113–147.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory.* Wiley, New York.

Berry, D., Wolff, M. C., and Sack, D. (1992). Public health decision making: a sequential vaccine trial. In *Bayesian Statistics 4*, J. Bernardo, J. Berger, A. Dawid, and A. F. M. Smith (Eds.). Oxford Univeristy Press, London.

Betrò, B., Meczarski, M., and Ruggeri, F. (1994). Robust Bayesian analysis under generalized moment conditions. To appear in *J. Statist. Planning and Inference.*

Boratyńska, A. (1991). On Bayesian robustness with the ε–contamination class of priors. Technical Report, Institute of Mathematics, University of Warsaw, Warsaw.

Boratyńska, A. and Zieliński, R. (1991). Infinitesimal Bayes robustness in the Kolmogorov and the Lévy metrics. Institute of Mathematics, University of Warsaw, Warsaw.

Bose, S. (1990). Bayesian robustness with shape-constrained priors and mixtures of priors. Ph.D. Thesis, Purdue University.

Bose, S. (1993). Bayesian robustness with mixture classes of priors. Technical Report #93-1, Department of Statistics, George Washington University, Washington D.C.

Bose, S. (1994). Bayesian robustness with more than one class of contaminations. To appear in *J. Statist. Planning and Inference.*

Box, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness (with discussion). *J. Roy. Statist. Soc. Ser. A*, **143**, 383–430.

Box, G. and Tiao, G. (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Reading.

Brunner, L. and Lo, A. (1989). Bayes methods for a symmetric unimodal density and its mode. *Ann. Statist.*, **17**, 1550–1566.

Cano, J. A. (1993). Robustness of the posterior mean in normal hierarchical models. *Commun. Statist.–Theory Meth.* **22(7)**, 1999–2014.

Cano, J. A., Hernández, A. and Moreno, E. (1985). Posterior measures under partial prior information. *Statistica*, **2**, 219–230.

Carlin, B. P., Chaloner, K. M., Louis, T. A., and Rhame, F. S. (1993). Elicitation, monitoring, and analysis for an AIDS clinical trial. Technical Report, Division of Biostatistics, University of Minnesota.

Carlin, B. P. and Louis, T. A. (1993). Identifying prior distributions that produce specific decisions, with application to monitoring clinical trials. To appear in *Bayesian Inference in Econometrics and Statistics*.

Carlin, B. P., and Polson, N. G. (1991). An expected utility approach to influence diagnostics. *J. Amer. Statist. Assoc.*, **86**, 1013–1021.

Casella, G. and Berger, R. (1987). Reconciling Bayesian and frequentist evidence in the one-sided testing problem (with discussion). *J. Amer. Statist. Assoc.*, **82**, 106–111.

Casella, G. and Wells, M. (1991). Noninformative priors for robust Bayesian inference. Technical Report, Cornell University, Ithaca.

Chib, S., Osiewalski, J. and Steel, M. F. J. (1991). Posterior inference on the degrees of freedom parameter in multivariate-t regression models. *Economics Letters*, **37**, 391–397.

Coolen, F. P. A. (1993). Imprecise conjugate prior densities for the one-parameter exponential model. *Statistics and Probability Letters*, **16**, 337–342.

Cuevas, A. and Sanz, P. (1988). On differentiability properties of Bayes Operators. In *Bayesian Statistics 3*, J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith (eds.), 569–577, Oxford University Press.

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

DasGupta, A. and Delampady, M. (1990). Bayesian testing with symmetric and unimodal priors. Technical Report #90-47, Department of Statistics, Purdue University, W. Lafayette.

DasGupta, A. and Mukhopadhyay, S. (1994). Uniform and subuniform posterior robustness: the sample size problem. To appear in *J. Statist. Planning and Inference*.

DasGupta, A. and Studden, W. J. (1988a). Frequentist behavior of robust Bayes procedures: new applications of the Lehmann-Wald minimax theory to a novel geometric game. Technical Report #88-36C, Department of Statistics, Purdue University, W. Lafayette.

DasGupta, A. and Studden, W. J. (1988b). Robust Bayesian analysis in normal linear models with many parameters. Technical Report #88-14, Department of Statistics, Purdue University, W. Lafayette.

DasGupta, A. and Studden, W. J. (1989). Frequentist behavior of robust Bayes estimates of normal means. *Statist. and Decisions*, **7**, 333–361.

DasGupta, A. and Studden, W. J. (1991). Robust Bayesian experimental designs in normal linear models. *Ann. Statist.*, **19**, 1244–1256.

Datta, G. and Lahiri, P. (1992). Robust hierarchical Bayes estimation of small area characteristics in presence of covariates. Technical Report #92-28, Department of Statistics, University of Georgia, Athens.

Dawid, A. P. (1973). Posterior expectations for large observations. *Biometrika*, **60**, 664–667.

De la Horra, J. and Fernandez, C. (1993). Bayesian analysis under ε-contaminated priors: a trade-off between robustness and precision. *J. Statist. Planning and Inference*, **38**, 13–30.

De la Horra, J. and Fernández, C. (1994). Bayesian robustness of credible regions in the presence of nuisance parameters. To appear in *Communications in Statistics*, **23**.

Delampady, M. (1989a). Lower bounds on Bayes factors for interval null hypotheses. *J. Amer. Statist. Assoc.*, **84**, 120–124.

Delampady, M. (1989b). Lower bounds on Bayes factors for invariant testing situations. *J. Multivariate Anal.*, **28**, 227–246.

Delampady, M. and Berger, J. (1990). Lower bounds on Bayes factors for Multinomial and chi-squared tests of fit. *Ann. Statist.*, **18**, 1295–1316.

Delampady, M. and Dey, D. (1994). Bayesian robustness for multiparameter problems. To appear in *J. Statist. Planning and Inference*.

DeRobertis, L. (1978). The use of partial prior knowledge in Bayesian inference. Ph.D. Thesis, Yale University, New Haven, CT.

DeRobertis, L. and Hartigan, J. A. (1981). Bayesian inference using intervals of measures. *Ann. Statist.*, **1**, 235–244.

Dette, H. and Studden, W. J. (1994). A geometric solution of the Bayesian E-optimal design problem. In *Statistical Decision Theory and Related Topics V*, S. S. Gupta and J. Berger (Eds.). Springer-Verlag, New York.

Dey, D. and Birmiwal (1991). Robust Bayesian analysis using entropy and divergence measures. Technical Report, Department of Statistics, University of Connecticut, Storrs.

Diaconis, P. and Freedman, D. (1986). On the consistency of Bayes estimates. *Ann. Statist.*, **14**, 1–67.

Doss, H. (1994). Bayesian estimation for censored data: an experiment in sensitivity analysis. In *Statistical Decision Theory and Related Topics V*, S. S. Gupta and J. Berger (Eds.). Springer-Verlag, New York.

Draper, D. (1992). Assessment and propagation of model uncertainty. Technical Report, Department of Mathematics, University of California, Las Angeles.

Drummey, K. W. (1991). Robust Bayesian estimation in the normal, gamma, and binomial probability models: a computational approach. Ph.D. Thesis, University of Maryland Baltimore County.

DuMouchel, W. and Harris, J. (1983). Bayes methods for combining the results of cancer studies in human and other species (with discussion). *J. Amer. Statist. Assoc.*, **78**, 293–315.

Edwards, W., Lindman, H. and Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Rev.*, **70**, 193–242.

Eichenauer-Herrmann, J. and Ickstadt, K. (1993). A saddle point characterization for classes of priors with shape-restricted densities. *Statistics and Decisions*, **11**, 175–179.

Fan, T. H. and Berger, J. O. (1990). Exact convolution of t–distributions, with application to Bayesian inference for a normal mean with t prior distributions. *J. Statist. Computation and Simulation*, **36**, 209–228.

Fan, T. H. and Berger, J. O. (1992). Behavior of the posterior distribution and inferences for a normal mean with t prior distributions. *Statistics and Decisions*, **10**, 99-120.

Ferguson, T.S., Phadia, E.G., and Tiwari, R.C. (1992). Bayesian nonparametric inference. In *Current Issues in Statistical Inference: Essays in Honor of D. Basu*, M. Ghosh, and P. K. Pathak (Eds.). Institute of Mathematical Statistics, Hayward CA.

Fernández, C., Osiewalski, J., and Steel, M.F.J. (1993). Marginal equivalence in ν-spherical models. Technical Report, Universidad Autónoma, Madrid.

Fortini, S. and Ruggeri,F. (1990). Concentration function in a robust Bayesian framework. *Quaderno IAMI 90.6*, CNR-IAMI, Milano.

Fortini, S. and Ruggeri, F. (1992). On defining neighborhoods of measures through the concentration function. To appear in *Sankhyā A*.

Fortini, S. and Ruggeri, F. (1993). Concentration function and coefficient of divergence for signed measures. To appear in *J. Ital. Statist. Soc.*

Fortini, S. and Ruggeri, F. (1994). Concentration functions and Bayesian robustness. To appear in *J. Statist. Planning and Inference.*

Fougere, P. (Ed.) (1990). *Maximum Entropy and Bayesian Methods.* Klower Academic Publ., Dordrecht.

Gasparini, M. (1990). Nonparametric Bayes estimation of a distribution function with truncated data. Technical Report #182, Statistics Department, University of Michigan, Ann Arbor.

Geisser, S. (1992). Bayesian perturbation diagnostics and robustness. In *Bayesian Analysis in Statistics and Econometrics*, P. K. Goel and N. S. Iyengar (Eds.). Springer-Verlag, New York.

Gelfand, A. and Dey, D. (1991). On Bayesian robustness of contaminated classes of priors. *Satistics and Decisions*, **9**, 63–80.

Genest, C., and Zidek, J. (1986). Combining probability distributions: A critique and an annotated bibliography. *Statist. Science*, **1**, 114–135.

Geweke, J. (1992). Priors for macroeconomic time series and their application. To appear in the *Proceedings of the Conference on Bayes Methods and Unit Roots*, a special issue of *Econometric Theory.*

Ghosh, J. K., Ghosal, S., and Samanta, T. (1994). Stability and convergence of the posterior in non-regular problems. In *Statistical Decision Theory and Related Topics V*, S. S. Gupta and J. Berger (Eds.). Springer-Verlag, New York.

Ghosh, M. (1993). Inconsistent maximum likelihood estimators for the Rasch model. Technical Report, Department of Statistics, University of Florida, Gainesville.

Ghosh, M. (1994). On some Bayesian solutions of the Neyman-Scott problem. In *Statistical Decision Theory and Related Topics V*, S. S. Gupta and J. Berger (Eds.). Springer-Verlag, New York.

Goldstein, M. and Wooff, D.A. (1994). Robustness measures for Bayes linear analyses. To appear in *J. Statist. Planning and Inference.*

Good, I. J. (1983a). *Good Thinking: The Foundations of Proability and Its Applications.* University of Minnesota Press, Minneapolis, MN.

Good, I. J. (1983b). The robustness of a hierarchical model for multinomials and contingency tables. In *Scientific Inference, Data Analysis, and Robustness,* G. E. P. Box et al. (Eds.), Academic Press, New York.

Good, I. J. and Crook, J. F. (1987). The robustness and sensitivity of the mixed-Dirichlet Bayesian test for 'independence' in contingency tables. *Ann. Statist.*, **15**, 694–711.

Goutis, C. (1991). Ranges of posterior measures for some classes of priors with specified moments. Research Report #70, Department of Statistical Science, University College London.

Gu, C. and Wahba, G. (1993). Smoothing spline ANOVA with component-wise Bayesian confidence intervals. *J. Comput. and Graphical Stat.*, **2**, 97–117.

Gustafson, P. and Wasserman, L. (1993). Local sensitivity diagnostics for Bayesian inference. Technical Report #574, Department of Statistics, Carnegie-Mellon University, Pittsburgh.

Guttman, I. and Peña, D. (1988). Outliers and influence: evaluation by posteriors of parameters in the linear model. In *Bayesian Statistics 3*, J.M. Bernardo, et.al. (Eds.). Oxford University Press, London.

Guttman, I. and Peña, D. (1993). A Bayesian look at diagnostics in the univariate linear model. *Statistica Sinica*, **3**, 367–390.

Hartigan, J.A. (1983). *Bayes Theory.* Springer-Verlag, New York.

Ickstadt, K. (1992). Gamma-minimax estimators with respect to unimodal priors. In *Operations Research '91*, P. Gritzmann, et.al. (Eds.). Physica-Verlag, Heidelberg.

Jaynes, E. T. (1983). *Papers on Probability, Statistics and Statistical Physics*, R. Rosenkrantz (Ed.), Reidel, Dordrecht.

Jefferys, W. and Berger, J. (1992). Ockham's razor and Bayesian analysis. *American Scientist*, **80**, 64–72.

Kadane, J. (1994). An application of robust Bayesian analysis to a medical experiment. To appear in *J. Statist. Planning and Inference.*

Kadane, J. B. and Chuang, D. T. (1978). Stable decision problems. *Ann. Statist.*, **6**, 1095–1110.

Kass, R. and Greenhouse, J. (1989). Investigating therapies of potentially great benefit: A Bayesian perspective. Comments on "Investigating therapies of potentially great benefit: ECMO," by J. H. Ware. *Statistical Science 4*, 310–317.

Kass, R. E. and Raftery, A. (1992). Bayes factors and model uncertainty. Technical Report #571, Department of Statistics, Carnegie-Mellon University, Pittsburgh.

Kass, R. E., and Slate, E. H. (1992). Reparametrization and diagnostics of posterior non-normality. In *Bayesian Statistics 4*, J. Bernardo, J. Berger, A. Dawid, and A. F. M. Smith (Eds.). Oxford University Press, London.

Kass, R. E., Tierney, L. and Kadane, J. B. (1980). Approximate methods for assessing influence and sensitivity in Bayesian analysis. *Biometrika*, **76**, 663–674.

Kass, R. E. and Wasserman, L. (1993). Formal rules of selecting prior distributions: a review and annotated bibliography. Technical Report, Department of Statistics, Carnegie-Mellon University, Pittsburgh.

Kiefer, J. (1977). Conditional confidence statements and confidence estimators (theory and methods). *J. Amer. Statist. Assoc.*, **72**, 789–827.

Kohn, R., and Ansley, C. F. (1988). The equivalence between Bayesian smoothness priors and optimal smoothing for function estimation. In *Bayesian Analysis of Time Series and Dynamic Models*, J. C. Spall (Ed.). Marcel Dekker, New York.

Kouznetsov, V. P. (1991). *Interval Statistical Models.* Moscow, Radio and Communication; Moscow.

Laplace, P. S. (1812). *Theorie Analytique des Probabilites.* Courcier, Paris.

Lavine, M. (1989). The boon of dimensionality: how to be a sensitive multidimensional Bayesian. Discussion Paper 89-14, ISDS, Duke University, Durham.

Lavine, M. (1991a). Sensitivity in Bayesian statistics: the prior and the likelihood. *J. Amer. Statist. Assoc.*, **86**, 396–399.

Lavine, M. (1991b). An approach to robust Bayesian analysis with multidimensional spaces. *J. Amer. Statist. Assoc.*, **86**, 400–403.

Lavine, M. (1992a). Sensitivity in Bayesian statistics: the prior and the likelihood. *J. Amer. Statist. Assoc.*, **86**, 396–399.

Lavine, M. (1992b). Some aspects of Polya tree distributions for statistical modelling. *Ann. Statist.*, **20**, 1222–1235.

Lavine, M. (1992c). A note on bounding Monte Carlo variances. *Commun. Statist.–Theory Meth.*, **21 (10)**, 2855–2860.

Lavine, M. (1992d). Local predictive influence in Bayesian linear models with conjugate priors. *Commun. Statist.-Simula*, **21 (1)**, 269–283.

Lavine, M. (1994). An approach to evaluating sensitivity in Bayesian regression analysis. To appear in *J. Statist. Planning and Inference*.

Lavine, M. and Wasserman, L. (1992). Can we estimate N? Technical Report #546, Department of Statistics, Carnegie-Mellon University.

Lavine, M., Wasserman, L. and Wolpert, R. (1991). Bayesian inference with specified prior marginals. *J. Amer. Statist. Assoc.*, **86**, 964–971.

Lavine, M., Wasserman, L., and Wolpert, R. (1993). Linearization of Bayesian robustness problems. *J. Statist. Planning and Inference*, **37**, 307–316.

Leamer, E. E. (1982). Sets of posterior means with bounded variance prior. *Econometrica*, **50**, 725–736.

Lenk, P. J. (1988). The logistic normal distribution for Bayesian, nonparametric, predictive densities. *J. Amer. Statist. Assoc.*, **83**, 509–516.

Leonard T. (1978). Density estimation, stochastic processes, and prior information. *J. Roy. Statist. Soc. B*, **40**, 113–146.

Li, Y. and Saxena, K. M. L. (1990). Optimal robust Bayesian estimation. Technical Report #4, Department of Mathematics and Statistics, University of Nebraska, Lincoln.

Liseo, B. (1993). Elimination of nuisance parameters with reference priors. *Biometrika*, **80**, 295–304.

Liseo, B., Petrella, L., and Salinetti, G. (1993). Block unimodality for multivariate Bayesian robustness. To appear in *J. Ital. Statist. Soc.*

Lo, A.Y. and Weng, C. S. (1989). On a class of Bayesian nonparametric estimates: II. Hazard rate estimates. *Ann. Inst. Statist. Math.*, **41**, 227–245.

Lucas, T. W. (1992). When is conflict normal? Technical Report, the RAND Corporation, Santa Monica.

Meczarski, M. (1991). Stable Bayesian estimation in the Poisson model: a nonlinear problem. *Quaderno IAMI 91.14*, CNR-IAMI, Milano.

Meczarski, M. and Zieliński, R. (1991). Stability of the Bayesian estimator of the Poisson mean under the inexactly specified gamma prior. *Statist. Prob. Letters*, **12**, 329–333.

Meng, X. L. (1994). Posterior predictive P-values. To appear in *Ann. Statist.*

McCulloch, R. E. (1989). Local model influence. *J. Amer. Statist. Assoc.*, **84**, 473–478.

Moreno, E. and Cano, J. A. (1989). Testing a point null hypothesis: Asymptotic robust Bayesian analysis with respect to the priors given on a subsigma field. *Int. Statist. Rev.*, **57**, 221–232.

Moreno, E. and Cano, J. A. (1991). Robust Bayesian analysis for ε-contaminations partially known. *J. Royal Statist. Soc., Series B*, **53**, 143–155.

Moreno, E. and Cano, J. A. (1992). Classes of bidimensional priors specified on a collection of sets: Bayesian robustness. Technical Report, Universidad de Granada.

Moreno, E. and Pericchi, L. R. (1990). Sensitivity of the Bayesian analysis to the priors: structural contaminations with specified quantiles of parametric families. *Actas III Cong. Latinoamericano Probab. Estad. Mat.*, 143–158.

Moreno, E. and Pericchi, L. R. (1991). Robust Bayesian analysis for ε-contaminations with shape and quantile constraints. *Proc. Fifth Inter. Symp. on Applied Stochastic Models and Data Analysis*. World Scientific Publ., 454–470.

Moreno, E. and Pericchi, L. R. (1992a). Subjetivismo sin dogmatismo: análisis Bayesiano robusto (with discussion). *Estadistica Española*, **34 (129)**, 5–60.

Moreno, E. and Pericchi, L. R. (1992b). Bands of probability measures: A robust Bayesian analysis. In *Bayesian Statistics 4*, J. M. Bernardo, et.al. (Eds.). Oxford University Press, London.

Moreno, E. and Pericchi, L. R. (1993a). Bayesian robustness for hierarchical ε-contamination models. *J. Statist. Planning and Inference*, **37**, 159–168.

Moreno, E. and Pericchi, L. R. (1993b). Precise measurement theory: robust Bayesian analysis. Technical Report, Department of Statistics, Universidad de Granada.

Morris, C. N. and Normand, S. L. (1992). Hierarchical models for combining information and for meta-analyses. In *Bayesian Statistics 4*, J. Bernardo, J. Berger, A. Dawid, and A. F. M. Smith (Eds.). Oxford University Press, London.

Moskowitz, H. (1992). Multiple-criteria robust interactive decision analysis for optimizing public policies. *European J. of Operations Research*, **56**, 219–236.

Mukhopadhyay, S. and DasGupta, A. (1993). Uniform approximation of Bayes solutions and posteriors: frequentistly valid Bayes inference. Technical Report #93-12C, Department of Statistics, Purdue University, W. Lafayette.

O'Hagan, A. (1988). Modelling with heavy tails. In *Bayesian Statistics 3*, J. M. Bernardo, et.al (Eds.). Oxford University Press, London.

O'Hagan, A. (1990). Outliers and credence for location parameter inference. *J. Amer. Statist. Assoc.*, **85**, 172–176.

O'Hagan, A. (1994). Robust modelling for asset management. To appear in *J. Statist. Planning and Inference*.

O'Hagan, A. and Berger, J. O. (1988). Ranges of posterior probabilities for quasi-unimodal priors with specified quantiles. *J. Amer. Statist. Assoc.*, **83**, 503–508.

Osiewalski, J. and Steel, M. F. J. (1993a). Robust Bayesian inference in ℓ_q-spherical models. To appear in *Biometrika*.

Osiewalski, J. and Steel, M. F. J. (1993b). Robust Bayesian inference in elliptical regression models. *Journal of Econometrics*, **57**, 345–363.

Osiewlaski, J. and Steel, M. F. J. (1993c). Bayesian marginal equivalence of elliptical regression models. To appear in *Journal of Econometrics*.

Peña, D. and Guttman, I. (1993). Comparing probabilistic methods for outlier detection in linear models. *Biometrika*, **80**, 603–610.

Peña, D. and Tiao, G. C. (1992). Bayesian outlier functions for linear models. In *Bayesian Statistics 4*, J. Bernardo, et.al. (Eds.). Oxford University Press, London.

Pericchi, L. R. and Nazaret, W. (1988). On being imprecise at the higher levels of a hierarchical linear model. In *Bayesian Statistics 3*, J. M. Bernardo, et. al. (Eds.). Oxford University Press, London.

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

Pericchi, L. R. and Walley P. (1991). Robust Bayesian credible intervals and prior ignorance. *Internat. Statist. Rev.*, **58**, 1–23.

Pettit, L. I. (1988). Bayes methods for outliers in exponential samples. In *Bayesian Statistics 4*, J. Bernardo, et.al (Eds.). Oxford University Press, London.

Pettit, L. (1992). Bayes factors for outlier models using the device of imaginary observations. *J. Amer. Statist. Assoc.*, **87**, 541–545.

Poirier, D. J. (1988). Bayesian diagnostic testing in the general linear normal regression model. In *Bayesian Statistics 3*, J. M. Bernardo, et.al. (Eds.). Oxford University

Press, London.

Polasek, W. (1985). Sensitivity analysis for general and hierarchical linear regression models. In *Bayesian Inference and Decision Techniques with Applications*, P. K. Goel and A. Zellner, (Eds.), North-Holland, Amsterdam.

Polasek, W. and Pötzelberger, K. (1988). Robust Bayesian analysis in hierarchical models. In *Bayesian Statistics 3*, J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (Eds.). Oxford University Press.

Polasek, W. and Pötzelberger, K. (1994). Robust Bayesian methods in simple ANOVA models. To appear in *J. Statist. Planning and Inference*.

Pötzelberger, K. and Polasek, W. (1991). Robust HPD-regions in Bayesian regression models. *Econometrica*, **59**, 1581–1590.

Rao, C. R. (1985). Weighted distributions arising out of methods of ascertainment: what population does a sample represent? In *A Celebration of Statistics: The ISI Centenary Volume*, A. G. Atkinson and S. Fienberg (Eds.). Springer-Verlag, New York.

Regazzini, E. (1992). Concentration comparisons between probability measures. *Sankhyā*, **B54**, 129–149.

Rios Insua, D. (1990). *Sensitivity Analysis in Multiobjective Decision Making*. Springer-Verlag, New York.

Rios Insua, D. (1992). Foundations for a robust theory of decision making: the simple case. *Test*, **1**, 69–78.

Rios Insua, D. and French, S. (1991). A framework for sensitivity analysis in discrete multiobjective decision making. *EJOR*, **54**, 176–190.

Rios Insua, D. and Martín, J. (1994). Robustness issues under precise beliefs and preferences. To appear in *J. Statist. Planning and Inference*.

Rios, S. and Girón, F. J. (1980). Quasi-Bayesian behavior: a more realistic approach to decision making? In *Bayesian Statistics*, J. M. Bernardo et al., (Eds.), University Press, Valencia.

Robert, C. (1992). *L'Analyse Statistique Bayesienne*. Economica, Paris.

Ruggeri, F. (1990). Posterior ranges of functions of parameters under priors with specified quantiles. *Commun. Statist. Theory Methods*, **19**, 127–144.

Ruggeri, F. (1991). Robust Bayesian analysis given a lower bound on the probability of a set. *Commun. Statist.-Theory and Methods.*, **20**, 1881–1891.

Ruggeri, F. (1992). Bounds on the prior probability of a set and robust Bayesian analysis. *Theory of Probability and Its Applications*, **37 (2)**.

Ruggeri, F. and Wasserman, L. (1991). Density based classes of priors: Infinitesimal properties and approximations. Technical Report #528, Department of Statistics, Carnegie-Mellon University.

Ruggeri, F. and Wasserman, L. (1993). Infinitesimal sensitivity of posterior distributions. *Canad. J. Statist.*, **21**, 195–203.

Salinetti, G. (1994). Stability of Bayesian decisions. To appear in *J. Statist. Planning and Inference*.

Sansó, B. and Pericchi, L. R. (1992). Near ignorance classes of log-concave priors for the location model. *Test*, **1**, 39–46.

Sedransk, N. (1993). Admissibility of treatment. In *Clinical Trials: Bayesian Methods and Ethics*, J. Kadane (Ed.). Wiley and Sons, New York.

Sivaganesan, S. (1988). Range of posterior measures for priors with arbitrary contaminations. *Comm. Statst.*, **17**, 1591–1612.

Sivaganesan, S. (1989). Sensitivity of posterior mean to unimodality preserving contaminations. *Statist. and Decisions*, **7**, 77–93.

Sivaganesan, S. (1990). Sensitivity of some standard Bayesian estimates to prior uncertainty – a comparison. *J. Statist. Planning and Inference*, **27**, 85–103.

Sivaganesan, S. (1991). Sensitivity of some posterior summaries when the prior is unimodal with specified quantiles. *Canad. J. Stat.*, **19**, 57–65.

Sivaganesan, S. (1992). An evaluation of robustness in binomial empirical Bayes testing. In *Bayesian Statistics 4*, J. Bernardo, et. al. (Eds.). Oxford University Press, London.

Sivaganesan, S. (1993a). Range of the posterior probability of an interval for priors with unimodality preserving contaminations. *Ann. Inst. of Statist. Math.*

Sivaganesan, S. (1993b). Optimal robust sets for a density bounded class. Technical Report, Department of Mathematical Statistics, University of Cincinnati.

Sivaganesan, S. (1993c). Robust Bayesian diagnostics. *J. Statist. Plann. Inf.*, **35**, 171–188.

Sivaganesan, S. (1994). Bounds on posterior expectations for density bounded classes with constant bandwidth (with discussion). To appear in *J. Statist. Planning and Inference.*

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

Sivaganesan, S. and Berger, J. (1993). Robust Bayesian analysis of the binomial empirical Bayes problem. *Canadian J. Statist.*, **21**, 107–119.

Sivaganesan, S., Berliner, L. M., and Berger, J. (1993). Optimal robust credible sets for contaminated priors. To appear in *Statist. and Probab. Letters*.

Smith, A. F. M. (1983). Bayesian approaches to outliers and robustness. In *Specifying Statistical Models*, J. P. Florens et.al. (Eds.). Springer-Verlag, New York.

Spiegelhalter, D. J. (1985). Exact Bayesian inference on the parameters of a Cauchy distribution with vague prior information. In *Bayesian Statitics 2*, J. M. Bernardo, et.al. (Eds.), North-Holland, Amsterdam.

Srinivasan, C. and Trusczcynska, H. (1990). Approximation to the range of a ratio linear posterior quantity based on Frechet derivative. Technical Report #289, Department of Statistics, University of Kentucky.

Srinivasan, C. and Trusczcynska, H. (1993). Ranges of non-linear posterior quantities. To appear in the *Ann. Statist.*

Stephens, D. A. and Smith, A. F. M. (1992). Sampling-resampling techniques for the computation of posterior densities in normal means problems. *Test*, **1**, 1–18.

Tamura, H. (1992). Robust Bayesian auditing. Technical Report, School of Business Administration, University of Washington.

VanEeden, C. and Zidek, J. V. (1994). Group Bayes estimation of the exponential mean: a retrospective view of the Wald theory. In *Statistical Decision Theory and Related Topics V*, S. S. Gupta and J. Berger (Eds.). Springer-Verlag, New York.

Verdinelli, I. and Wasserman, L. (1991). Bayesian analysis of outlier problems using the Gibbs sampler. *Statistics and Computing*, **1**, 105–117.

Vidakovic, B. (1992). A study of properties of computationally simple rules in estimation problems. Ph.D. Thesis, Purdue University, W. Lafayette.

Wahba, G. (1990). *Spline Models for Observational Data.* SIAM, Philadelphia.

Walley, P. (1991). *Statistical Reasoning With Imprecise Probabilities.* Chapman and Hall, London.

Wasserman, L. (1989). A robust Bayesian interpretation of likelihood regions. *Ann.*

Statist., **17**, 1387–1393.

Wasserman, L. (1990). Prior envelopes based on belief functions. *Ann. Statist.*, **18**, 454–464.

Wasserman, L. (1992a). The conflict between improper priors and robustness. Technical Report #559, Department of Statistics, Carnegie-Mellon University, Pittsburgh.

Wasserman, L. (1992b). Recent methodological advances in robust Bayesian inference. In *Bayesian Statistics 4*, J. M. Bernardo, et.al. (Eds.). Oxford University Press, Oxford.

Wasserman, L. (1992c). Invariance properties of density ratio priors. *Ann. Statist.*, **20**, 2177–2182.

Wasserman, L. and Kadane, J. (1990). Bayes' theorem for Choquet capacities. *Ann. Statist.*, **18**, 1328–1339.

Wasserman, L. and Kadane, J. B. (1992a). Computing bounds on expectation. *J. Amer. Statist. Assoc.*, **87**, 516–522.

Wasserman, L. and Kadane, J. (1992b). Symmetric upper probabilities. *Ann. Statist*, **20**, 1720–1736.

Wasserman, L. and Seidenfeld, T. (1994). The dilation phenomenon in robust Bayesian inference. To appear in *J. Statist. Planning and Inference.*

Weiss, R. E. (1992). Influence diagnostics with the Gibbs sampler. In *Computing Science and Statistics: Proceedings of the* 24^{th} *Symposium on the Interface*, H.J. Newton (Ed.), American Statistical Association, Alexandria.

Weiss, R. (1993). Identification of outlying perturbations. Technical Report #594, School of Statistics, University of Minnesota, Minneapolis.

West, M. (1987). On scale mixtures of normal distributions. *Biometrika*, **74**, 646–648.

West, M. (1992). Modeling with mixtures. In *Bayesian Statistics 4*, J. M. Bernardo, et.al. (Eds.). Oxford University Press, London.

West, M. and Harrison, J. (1989). *Bayesian Forecasting and Dynamic Models.* Springer-Verlag, New York.

Wolpert, R. L. and Warren Hicks, W. J. (1992). Bayesian hierarchical logistic models for combining field and laboratory survival data. In *Bayesian Statistics 4*, J. M. Bernardo, J. Berger, A. Dawid, and A. F. M. Smith (Eds.). Oxford University Press, London.

Zellner, A. (1976). Bayesian and non-Bayesian analysis of regression models with multivariate student-t error terms. *J. Amer. Statist. Assoc.*, **71**, 400–405.

Zen, M. M. and DasGupta, A. (1993). Estimating a binomial parameter: is robust Bayes real Bayes? *Statistics and Decisions*, **11**, 37–60.

Zidek, J. V. and Weerahandi, S. (1992). Bayesian predictive inference for samples from smooth processes. In *Bayesian Statistics 4*, J. M. Bernardo, J. Berger, A. Dawid, and A. F. M. Smith (Eds.). Oxford University Press, London.