ROBUST BAYESIAN ANALYSIS: SENSITIVITY TO THE PRIOR *

by

James O. Berger
Purdue University

Technical Report #87-10

Department of Statistics
Purdue University

July 1987

* Research supported by the National Science Foundation, Grant DMS-8401996

*

1. INTRODUCTION

It is a great pleasure to write a review paper on robust Bayesian analysis for the occasion of I. J. Good's 70th birthday conference. Good was the first Bayesian to clearly make uncertainty in the prior distribution an integral part of his statistical philosophy. Indeed Good's descriptions of the robust Bayesian viewpoint, in terms of both philosophy and practice, have never been improved upon. For this reason, Section 2 is devoted to a discussion of Good's view of the subject.

A general review of all the approaches to robust Bayesian analysis was given in Berger (1984), which also included an extensive bibliography. Thus this paper will concentrate on reviewing recent activity in the subject, of which there has been an explosion in recent years. The bulk of this recent work has been concerned with:

(i) Modelling uncertainty in the prior by specifying a class, Γ, of possible prior distributions; and

(ii) Determining the range of the posterior quantity of interest as the prior ranges over Γ.

This can be thought of as an implementation of the "black box" model for Bayesian robustness which was introduced by Good (and is discussed in Section 2.2). Section 3 presents a review of this material, comparing the strengths and weaknesses of the various methodologies that have been proposed, and illustrating the methodologies with numerical examples.

Section 4 discusses one of the most immediate and important applications of robust Bayesian methodology, namely the calculation of lower bounds on Bayes factors in hypothesis testing.

The notation that will be used is as follows. The unknown parameter of interest will be denoted by θ, assumed to lie in the parameter space Θ, and the experimental evidence about θ will be given by the observed likelihood function $\ell(\theta)$. We will not be considering

frequentist measures of robustness, and hence the dependence of $\ell(\theta)$ on the data will be suppressed.

For a prior distribution π, the posterior distribution of θ is then given (under mild conditions) by

$$\pi^*(d\theta) = \pi(d\theta)\ell(\theta)/m,$$

where $m = \int \ell(\theta)\pi(d\theta)$. In most of our examples, Θ will be Euclidean and π will be assumed to have a density w.r.t. Lebesgue measure. For simplicity in such cases we will let the prior density be denoted by $\pi(\theta)$.

Instead of supposing the specification of a single prior π_0, suppose we know only that $\pi \epsilon \Gamma$, a class of distributions on Θ. This class could arise in at least two ways:

(i) Γ could be used to represent uncertainty in the prior elicitation process;

(ii) Γ could consist of the differing prior distributions of a group of individuals who must arrive at a joint decision.

In either case, there will be some posterior quantity $\rho(\pi)$ of interest (e.g. the posterior mean, posterior variance, posterior probability of a credible region or hypothesis, or posterior expected loss) and we will seek

$$\underline{\rho}_\Gamma = \inf_{\pi \epsilon \Gamma} \rho(\pi), \quad \overline{\rho}_\Gamma = \sup_{\pi \epsilon \Gamma} \rho(\pi).$$

The hope, of course, is that the range $(\underline{\rho}, \overline{\rho})$ (we will often suppress Γ) is small enough that the indeterminacy in the prior is deemed to be essentially irrelevant, allowing a claim of robustness with respect to the prior.

Our discussion will focus on the case (i) scenario above, though many of the methodologies presented would also apply to the group problem. The following example is typical, and will be among those analyzed later.

Example 1. Suppose it is specified that π has median 0, quartiles $\pm$ 0.675, and is symmetric and unimodal. These are among the types of judgements that an elicitor can be expected to make. Formally we thus have

$$\Gamma = \{\text{symmetric, unimodal } \pi \text{ with quartiles} - 0.675, 0.0, \text{ and } 0.675\}.$$

Note that specification of a specific functional form for π is likely to be very difficult; overall shape features, such as above, are all that might typically be possible. (Exact specification of the quartiles might similarly be criticized, but this is easily corrected by allowing the quartiles to vary within intervals.)

Results will be presented only for real parameters θ. Many of the results are valid in higher dimensions, but notation and comparisons are facilitated by considering only the real case. Also, we will not explicitly consider uncertainty in the likelihood function $\ell(\theta)$. At an abstract level this is not a limitation, in that θ could be understood to represent *all* unknown aspects of the model. In practice, however, techniques for dealing with uncertain models (in a Bayesian framework) will usually differ substantially from methods of dealing with uncertain priors. Techniques which can be adapted to deal with both will be mentioned as we proceed.

Another possible source of uncertainty arises in decision problems, namely uncertainty in the loss or utility function. This issue will not be addressed here. See Kadane and Chuang (1978) and Berger (1984) for discussion and references.

2. GOOD'S CONTRIBUTION TO BAYESIAN ROBUSTNESS

Quite simply, Good was the originator of the modern robust Bayesian viewpoint. We briefly review this viewpoint here, calling it the "Doogian" philosophy for historical reasons. Section 2.1 reviews the principles upon which Doogianism is based. Section 2.2 discusses the Doogian "black box" implementation of these principles. Section 2.3 briefly reviews other contributions of Good to robustness. Section 2.4 provides some historical context.

2.1 Doogian Robust Bayesian Principles

In Good (1973) are listed twenty seven Priggish Principles of the Doogian philosophy. If all statisticians were to read and understand these principles, our profession would be immeasurably improved. Priggish Principles 3, 4, 5, 6, 9, 21, and 24 directly bear on the issue of Bayesian robustness. For convenience of exposition we will distill these principles into three general principles.

Principle 1. "*The principle of rationality is the principle to maximize expected utility*" (Priggish Principle 4).

Loosely stated, this urges one to think like a Bayesian decision-theorist. But Good does not necessarily mean this to be a dogmatic requirement to write down a prior and utility function and optimize. Instead, this is more a proscription of behavior not consistent with rationality.

> " The freedom to be irrational, is a freedom that some of us can do without." (Good, 1982b).

Principle 1 is to be a filter eliminating our illogical excesses.

Principle 2. *Subjective judgements are imprecise, and can typically be quantified only as intervals of possible probabilities, utilities, etc.*

> " For it would only be a joke if you were to say that the probability of rain tomorrow ... is 0.3057876289." (Good, 1979).

Good observes that the intervals could result either from the imprecision of a single individual, or from the necessity of reaching a conclusion in the face of differing judgements of a group of people;

> " though the widths of the intervals ... might be much larger for groups than for individuals" (Good, 1975).

The emphasis that Good places on the necessity of acknowledging imprecision in judgement is evidenced by imprecision being mentioned in Priggish Principles 3, 5, 9, 21, and 24.

Principle 3 (Type II Rationality). *When practical considerations of time and cost are taken into account, many compromises with the other principles may be necessary.*

In particular it may, quite frequently, be necessary to ignore Principle 2:

> " It is not obvious whether it is ever reasonable to judge that a probability is precisely equal to a definite number such as $\frac{1}{2}$. But it may often be judged that such an equality is a sufficiently good approximation for some particular purpose." (Good, 1950).

Even non-Bayesian methods might be used because of Type II rationality:

> " It often justifies adhoc and non-Bayesian procedures such as confidence methods ..." (Priggish Principle 6).

A sample of Good's writings on these issues is Good (1950, 1952, 1956b, 1957b, 1959, 1961, 1962(a,b,c), 1965, 1969, 1973, 1974b, 1975, 1976, 1979, 1980, 1982(a,b), 1983(a,b)).

2.2 Good's Black Box Model

Principles 1 and 2 do not actually determine a unique statistical mode of operation. Good, however, suggests an approach to implement these axioms, which he calls the *black box model.* The idea is to imagine a black box, which essentially contains the rules of probability, Bayes theorem, the algorithm of maximizing expected utility, etc. In other words, this box contains all the usual Bayesian methods associated with precise judgements.

This box is to be thought of as a processor, which takes, as input, the intervals resulting from actual imprecise judgements and produces, as output, other intervals reflecting the possible range of derived quantities of interest. For instance, the input might be an "interval" of prior distributions and an "interval" of likelihoods for a parameter, with the output being the interval of posterior means that are possible. This output interval will be formed by taking all possible prior-likelihood pairs, processing each pair by the black box, and reporting all possible results.

The black box model does *not* directly provide guidance as to what to do with the output intervals. Implicit in the model is that, if the output intervals are small enough, then the conclusion or decision will be clear. And if the output intervals are too large, the only solution is to attempt refinement of the input intervals.

All of this may seem natural and obvious, but the fact is that many investigators have tried something very different and much more complicated, creating a completely new black box designed to directly combine the interval inputs in some fashion, e.g. developing a probability-type calculus for "upper and lower" probabilities. This has provided interesting challenges and led to entirely new paradigms for dealing with uncertainty, but a basic question is: Why go to all the trouble? Good's black box model provides a straightforward way of simultaneously satisfying Principles 1 and 2, and no example has yet been given which clearly demonstrates the need for a different approach.

> " The main merit that I claim for the Doogian philosophy is that it codifies and exemplifies an adequately complete and simple theory of rationality, complete in the sense that it is I believe not subject to the criticisms that are usually directed at other forms of Bayesianism, and simple in the sense that it attains realism with a minimum of machinery." (Good, 1976).

As further support for the black box model, consider the situation "in reverse." Suppose the output intervals are too large for conclusiveness, and that no further reduction of the input intervals is deemed to be possible. (This could certainly happen when the input is from a group of individuals, but even a single individual might be unable to decide between the probability of rain being 0.305 and 0.306, and it could conceivably matter.) Then there are legitimate differences or uncertainties in opinion which lead to different conclusions, and it seems wisest to just conclude that there is no answer; more evidence is needed to resolve the ambiguity. Any "alternative" black box which claims to do more, would simply be masking legitimate uncertainty by "sweeping it under the carpet" (Good, 1976).

A further advantage of Good's black box is that it can be automated. For instance, an interactive computer package can be imagined which elicits intervals of probabilities (or whatever) from users, and processes these into the output intervals. Or, alternatively, single prior inputs could automatically be embedded in intervals of priors, with the corresponding output intervals being calculated. More discussion of this will be presented as we proceed, but automation is surely desirable to remove the subject from the domain of the expert to that of the practitioner.

The writings of Good which explicitly discuss the black box model include Good (1959, 1961, 1962a, 1975, 1976, 1982(a,b)). The model is implicit in many other writings, however. For instance, he writes concerning probability axioms:

" ... the product axiom ... is $P(E \text{ and } F|H) = P(E|H) \cdot P(F|E \text{ and } H)$ and its meaning is that if it is assumed that any two of the probabilities belong to certain intervals, then the third probability can be inferred to belong to some interval using the equation." (Good, 1981).

2.3 Good's Other Works on Robustness

Much work on robustness has been directed towards clarifying situations in which robustness is, and is not, present. Identification of robustness eliminates the necessity of trying to implement Good's black box, and is a prime example of Type II rationality.

Good frequently discusses the robustness of Bayesian procedures he develops. For instance, he observed very early that hierarchical priors have the property of being robust at the higher levels:

" It might be objected that the higher the type the woollier (fuzzier) the probabilities. It will be found, however, that the higher the type, the less the woolliness matters ..." (Good, 1952).

Among the many works of Good discussing robustness of hierarchical models are Good (1952, 1956a, 1965, 1979, 1983b), Good and Crook (1974, 1987), and Crook and Good (1980). Good (1983b) and Good and Crook (1987) are specifically devoted to the discussion of robustness for hierarchical models in multinomial problems.

Another major body of work by Good concerns robustness in testing. The Bayes/non-Bayes compromise of using a Bayes factor as a test statistic and looking at its tail areas often has considerable robustness with respect to the prior. His writings on this subject are mentioned in Section 4.

Of course, as important (and perhaps more important) than identification of robuustness, is identification of situations in which robustness is lacking. Interesting examples of nonrobustness can be found in Good (1967, 1983b).

2.4 Historical Context

Keynes (1921) considered interval valued probabilities, but dealt only with "logical" probabilities. Koopman (1940a,b) discussed a rather involved set of axioms for interval valued probabilities. Good (1950, 1962a) developed quite simple axioms, similar to those given extensive justification in Smith (1961). A sample of subsequent axiomatic developments includes Fine (1973), Levi (1980), Rios and Girón (1980), Wolfenson and Fine (1982), and Walley (1987).

There have been relatively few works directed towards implementation of Good's black box approach. The recent papers in this direction are explicitly discussed in the following sections. Earlier related papers include Isaacs (1963), Fishburn (1965), Dempster (1975,1976), Suppes (1975), Rubin (1977), West (1979), and Hill (1980a,b). Again, an extensive set of references concerning robust Bayesian analysis can be found in Berger (1984).

3. RANGES OF BAYESIAN QUANTITIES

3.1 Quantities of Interest

There are three categories of Bayesian quantities that are typically of interest. These categories are determined by the type of dependence on the prior π; this dependence strongly affects the ease of analysis.

I. Linear Functionals

The easiest quantities to handle, from a robust Bayesian perspective, are linear functionals of π, such as

$$\rho(\pi) = \int_{\Theta} h(\theta)\pi(d\theta), \tag{3.1}$$

where h is a given function. Perhaps the most important linear functional is

$$m(\pi) = \int_{\Theta} \ell(\theta)\pi(d\theta), \tag{3.2}$$

i.e., the marginal density of the data which yields the likelihood function ℓ. Maximization (over π) of this quantity is often done to select a prior for use in empirical Bayes and other problems; Good calls this Type II maximum likelihood (Good, 1965).

Another important linear functional is frequentist Bayes risk. If a procedure $\delta(x)$ (here x denotes the data) is to be used in conjunction with some criterion $L(\theta, \delta)$ (which could be a loss, an indicator function, etc.), the frequentist risk is $R(\theta, \delta) = E_\theta L(\theta, \delta(X))$ (the expectation being over the random outcome X), and the Bayes risk is the linear functional

$$\rho(\pi) = \int_{\Theta} R(\theta, \delta)\pi(d\theta).$$

Such measures are basic to problems such as the design of experiments.

II. Ratio-Linear Posterior Quantities

Many posterior quantities of interest can be written as

$$\rho(\pi) = \int_{\Theta} h(\theta)\ell(\theta)\pi(d\theta)/m(\pi), \tag{3.3}$$

for some function h. The choice $h(\theta) = \theta$ yields the posterior mean. The choice $h(\theta) = I_C(\theta)$ (the indicator function on the set C) yields the posterior probability of C. The choice $h(\theta) = L(\theta, a)$, where a is an action and L is the corresponding loss when θ obtains, yields the posterior expected loss of a. We will call such posterior quantities *ratio-linear*, because they can be written as a ratio of linear functionals of π.

III. Ratio Non-Linear Posterior Quantities

Some posterior quantities are of the form

$$\rho(\pi) = \int_{\Theta} h(\theta, \varphi(\pi))\ell(\theta)\pi(d\theta)/m(\pi). \tag{3.4}$$

The most common example arises from

$$h(\theta, \varphi(\pi)) = (\theta - \mu(\pi))^2,$$

where $\mu(\pi)$ is the posterior mean; $\rho(\pi)$ is then the posterior variance. More generally, if $L(\theta, a)$ is a loss function and $a(\pi)$ is the Bayes action (i.e., that which minimizes the posterior expected loss), then setting $h(\theta, \varphi(\pi)) = L(\theta, a(\pi))$ results in the posterior Bayes risk.

It is sometimes of interest to convert non-linear quantities into conditional linear quantities, when determining ranges. The idea is to replace the constraint $\pi \epsilon \Gamma$ by the constraint

$$\pi \epsilon \Gamma_{\varphi_o} = \{\pi \epsilon \Gamma : \varphi(\pi) = \varphi_0\}, \tag{3.5}$$

Figure 1. Range of Posterior Variance for given $\mu(\pi)$ when x = 4.0, as $\mu(\pi)$ varies from 1.96 to 3.34: Normal Example.

φ_o being a specific value; $\rho(\pi)$ is ratio-linear for $\pi \epsilon \Gamma_{\rho_0}$. This can be useful as a technical device; for each ρ_0 one finds the range of $\rho(\pi)$ over $\pi \epsilon \Gamma_{\rho_0}$, and then maximizes and minimizes over ρ_0.

This can also be useful from a methodological perspective. For instance, it can be argued that the range of the posterior variance is not of inherent interest; variances are only of interest in association with means, so one should determine the range of posterior means corresponding to $\pi \epsilon \Gamma$ and find, for each such posterior mean, the possible range of posterior variances. Figure 1 presents the results of such an analysis for an example to be discussed in Section 3.4.2. The posterior mean, $\mu(\pi)$, here ranges between 1.96 and 3.34, and the upper and lower curves exhibit the possible range of variances for each mean. Thus when $\mu(\pi) = 2.4$, the posterior variance can vary between 0.55 and 1.19.

3.2 Classes of Priors

In selecting a class, Γ, of priors to model prior uncertainty, there are four possibly competing goals. These are:

(i) Calculation of $\underline{\rho}$ and $\overline{\rho}$ should be as easy as possible;

(ii) Γ should contain as many "reasonable" priors as possible to ensure robustness;

(iii) Γ should not contain unreasonable priors, or robustness may be erroneously judged to be absent;

(iv) Γ should correspond to easily elicitable prior information.

Formal approaches to robustness have generally involved compromises among these goals. The most common Γ that have been considered are discussed below, and used to illustrate these concepts.

I. Conjugate Classes

Let π_λ be a natural conjugate prior (see Berger, 1985, for definition) corresponding to the likelihood $\ell(\theta)$, and define

$$\Gamma_C = \{\pi_\lambda : \lambda \epsilon \Lambda\}, \tag{3.6}$$

where Λ allows for a range of possible values for the hyperparameter λ.

Example 1 (continued). Suppose $\ell(\theta)$ is a normal likelihood. The natural conjugate prior which matches the specified quartiles is the $N(0,1)$ prior. Allowing for a degree of uncertainty in the prior specification, one might consider

$$\Gamma_C = \{N(\mu, \tau^2) \text{ distributions: } -0.2 \leq \mu \leq 0.2 \text{ and } 0.7 \leq \tau^2 \leq 1.3\}. \tag{3.7}$$

The great advantage of conjugate classes is that posterior quantities can often be calculated in closed form for natural conjugate priors; minimizing and maximizing over $\lambda \epsilon \Lambda$ is then straightforward. Recent references include Leamer (1978, 1982) and Polasek (1985), to which we refer the reader for discussion and examples.

When the Bayesian analysis is robust with respect to the exact functional form of π, use of conjugate classes is desirable because of the ensuing calculational simplicity. Unfortunately, conjugate classes are very small and omit many reasonable priors. Thus, in Example 1 (continued) one could have matched the specified prior information equally well with a *Cauchy* (0,.675) prior. It is easy to construct situations where one is robust with respect to Γ_C, yet not robust with respect to such "similar" priors (cf. Berger (1985)). Thus, in general, we prefer classes Γ which are large enough to include all reasonable priors, or at least priors with a variety of functional forms.

II. Classes With Approximately Specified Moments

Consider

$$\Gamma_M = \{\pi : \alpha_i \leq E^\pi[\theta^i] \leq \beta_i, i = 1, \ldots, k\}.$$

Most common is to consider specification of the first two moments (cf. Stone (1963), Hartigan (1969) and Goldstein (1980)). Such classes are typically employed in conjunction with a restriction on the set of procedures being considered. For instance, if one considers only linear estimators of θ, then behavior (e.g. posterior variance) is frequently determined by only the first two moments of π, and classes such as Γ_M are easy to work with.

The main problem with Γ_M is similar to that with Γ_C; both place strong restrictions on the tails of π, typically prohibiting reasonable priors such as the Cauchy prior. Inded, when Γ_M is used in conjunction with, say, linear estimation, there is typically a 1-1 relationship between robustness under Γ_C and robustness under Γ_M. Thus we will not consider Γ_M further. (Note that we are not impuning the use of Γ_M in all situations; in many physical situations, in particular, knowledge of moments of the system - e.g. temperature - may be available.)

III. Neighborhood Classes

Suppose, following the usual Bayesian paradigm, that π_0 is a single elicited prior. Prior uncertainty can be modelled by considering

$$\Gamma_N = \{\pi \text{ which are in a neighborhood of } \pi_0\}.$$

A variety of neighborhoods could be considered, based (say) on various notions of distance between priors, but the most extensively studied neighborhood is the ε-contamination neighborhood. This yields the *ε-contamination class* of priors, namely

$$\Gamma_\varepsilon = \{\pi = (1-\varepsilon)\pi_0 + \varepsilon q : q \epsilon \mathcal{Q}\}. \tag{3.8}$$

Here ε reflects the amount of uncertainty in π_0, and $\mathcal{Q}$ determines the allowed contaminations which are mixed with π_0.

The choice of $\mathcal{Q}$ can have a great effect, and is often chosen with an eye towards the four goals mentioned at the beginning of this subsection. Possibilities that have been

considered include (where θ_0 is the mode of π_0)

$$\mathcal{Q}_{\mathcal{A}} = \{\text{all distributions } q\}, \tag{3.9}$$

$$\mathcal{Q}_{U*} = \{\text{all distributions } q : \pi = (1-\varepsilon)\pi_0 + \varepsilon q \text{ is unimodal}\}, \tag{3.10}$$

$$\mathcal{Q}_{U} = \{\text{all unimodal distributions } q, \text{ with mode } \theta_0\}, \tag{3.11}$$

$$\mathcal{Q}_{SU} = \{\text{all symmetric unimodal distributions } q, \text{ with mode } \theta_0\}. \tag{3.12}$$

The last three classes are designed for situations where π_0 is unimodal, and in which one is confident in the assessment of overall unimodality of the prior. This type of overall shape specification is frequently feasible.

Working with $\mathcal{Q}_{\mathcal{A}}$ is typically easiest, but Γ_ε then contains many unreasonable distributions. ($\mathcal{Q}_{\mathcal{A}}$ allows, for instance, point masses at very extreme values of θ.) Thus $(\underline{\rho}, \overline{\rho})$ will often be an excessively large interval.

When unimodality of the prior is believed, Γ_ε with $\mathcal{Q}_{U*}$ is very sensible; it will include virtually all reasonable priors (those which are unimodal and close to π_0), and no unreasonable priors. Unfortunately, computation with this class is quite difficult.

Use of $\mathcal{Q}_U$ or $\mathcal{Q}_{SU}$ results in still smaller Γ_ε, raising the possibility that certain reasonable priors will have been omitted. Note, however, that Γ_ε will still contain a very wide range of priors; in particular, much thicker tails than that of π_0 are allowed. The gain in utilizing $\mathcal{Q}_U$ or $\mathcal{Q}_{SU}$ is that calculation is comparatively simple.

Section 3.4.4 contains further discussion of the choice of $\mathcal{Q}$. References discussing ε-contamination classes, from our Bayesian perspective, include Huber (1973), Berger and Berliner (1984, 1986), Berger (1985), Sivaganesan (1986), and Sivaganesan and Berger (1987). Berger and Berliner (1986) also discuss a number of more involved ε-contamination classes which are suitable for hierarchical models.

IV. Sub-Sigma Field Classes

Prior elicitation typically involves the actual specification of only a finite number of features of the prior. Often, this can be viewed as actually defining the prior on a sub-sigma field of sets. Such ideas were theoretically explored by Kudō (1967), Fine (1973), Manski (1981), Lambert and Duncan (1981) and Cano, Hernández, and Moreno (1985), among others.

The actual examples of developed methodology in this direction have involved specification of quantiles. Thus suppose Θ is an interval (a_0, a_m) (where a_0 and a_m could be -∞ and ∞ respectively), which is partitioned into the intervals $I_i = (a_{i-1}, a_i), i = 1, \ldots, m$, where $a_0 < a_1 < \ldots < a_m$. The prior probability assigned to I_i will be denoted by p_i. Suppose that elicitation yields the bounds

$$\ell_i \leq p_i \leq u_i, \quad i = 1, \ldots, m.$$

This defines the *quantile class*

$$\Gamma_Q = \{\pi : \ell_i \leq \int_{I_i} \pi(d\theta) \leq u_i, \quad i = 1, \ldots, m\}. \tag{3.13}$$

The great appeal of this class is that specification of probabilities is the most natural elicitation mechanism.

DeRobertis (1978) considered Γ_Q with $u_i = \ell_i + \ell_0$ (here all $\ell_i > 0$ and $\sum_{i=0}^{m} \ell_i = 1$), while Berliner and Goel (1986) considered $\ell_i = u_i$. Versions of Γ_Q for finite Θ were considered in Fishburn (1965).

The class Γ_Q unfortunately contains unreasonable distributions, e.g. discrete distributions assigning masses p_i to individual points in I_i. This again can lead to overly wide ranges, $(\underline{\rho}, \overline{\rho})$, of posterior measures. One attractive solution is to also impose shape constraints on π. One example, considered in Berger and O'Hagan (1987), is

$$\Gamma_{QU} = \{\pi \epsilon \Gamma_Q : \pi \text{ is unimodal}\}. \tag{3.14}$$

O'Hagan and Berger (1987) consider a related, but somewhat easier to analyze, class.

V. Density Ratio Classes

DeRobertis and Hartigan (1981) introduced a very interesting class of priors which can be considered to be a type of limit of the quantile class of DeRobertis (1978). Assuming (without loss of generality, as it turns out) the existence of a dominating measure ν for all priors in the class, the class can be given in terms of generalized prior densities (w.r.t. ν) as

$$\Gamma_{DR} = \{\pi : L(\theta) \leq \alpha\, \pi(\theta) \leq U(\theta) \text{ for some } \alpha > 0\}. \tag{3.15}$$

Here L and U are specified nonnegative functions, and π is not required to have mass one, or even to be proper. (Normalization of π is not important for ratio linear or ratio nonlinear posterior quantities.) An alternative definition of this class is

$$\Gamma_{DR} = \{\pi : \frac{L(\theta)}{U(\theta')} \leq \frac{\pi(\theta)}{\pi(\theta')} \leq \frac{U(\theta)}{L(\theta')} \text{ for all } \theta, \theta'\},$$

from which it is clear that Γ_{DR} specifies ranges for the ratios of the prior density between any two points.

Example 2. Suppose $L(\theta) = 1$ and $U(\theta) = K$. Then Γ_{DR} consists of all prior densities for which the density ratio between any two points lies between K^{-1} and K. This class is a reasonable representation of prior vagueness, from a robust Bayesian perspective.

Example 3. Let π_0 be a single elicited prior density arising from the usual Bayesian paradigm. To express uncertainty in π_0, it is natural to consider $L(\theta) = \pi_0(\theta), U(\theta) = (1+\varepsilon)\pi_0(\theta)$. This can be shown to define a type of neighborhood of π_0 (see DeRobertis (1978)). Unfortunately, all priors in Γ_{DR} will have essentially the same tail behavior as π_0; this neighborhood is thus perhaps not very satisfactory as a representation of prior uncertainty. Instead, $U(\theta)$ should perhaps be chosen to be $\pi_0(\theta) + \varepsilon g(\theta)$, where $g(\theta)$ is some larger tailed density.

Determining a sensible Γ_{DR}, in a given problem, is clearly not always easy. Thinking in terms of comparative density ratios is certainly possible, however; indeed, with practice it can become a valuable elicitation tool. The big advantage of density ratio classes will be seen to be calculational; ranges of posterior quantities are comparatively easy to determine.

Density ratio classes are related to implicit classes studied in the "precise measurement" Bayesian robustness literature. Typical theorems in that literature place bounds on the allowed variation in the prior density, and then show that for very concentrated likelihoods (usually arising from large sample sizes) the variation in the Bayes answer over the class of priors is negligible. A few references to such work are Edwards, Lindman, and Savage (1963), DeGroot (1970), Dickey (1976), and Davis (1979).

3.3 Methods for Calculation of $\underline{\rho}$ and $\overline{\rho}$

Methods of minimization and maximization of $\rho(\pi)$ over $\pi \epsilon \Gamma$ are often specific to the criterion ρ and the class Γ being considered. Frequently, the idea is to identify a low dimensional subclass of Γ in which the overall minimizing or maximizing prior must lie; the optimization need then be carried out numerically only over this low dimensional class.

Although any given problem or formulation may require a specialized argument, there are a few broadly applicable methods of optimization. Several will be discussed here, organized according to the criterion classification scheme in Section 3.1.

I. Linear Functionals

Maximization and minimization of linear functionals of π is typically quite easy. Often Γ is convex, with identifiable extreme points in Γ_0. Then

$$\sup_{\pi \epsilon \Gamma} \rho(\pi) = \sup_{\pi \epsilon \Gamma} \int h(\theta)\pi(d\theta) = \sup_{\pi \epsilon \Gamma_0} \rho(\pi), \text{ and } \inf_{\pi \epsilon \Gamma} \rho(\pi) = \inf_{\pi \epsilon \Gamma_0} \rho(\pi).$$

Furthermore, Γ_0 is often a low dimensional set, so that the final optimization can often be easily done numerically.

Example 4. Consider Γ_ε in (3.8) with $\mathcal{Q}_{SU}$ as in (3.12). Suppose the criterion of interest is the marginal density $m(\pi)$ defined in (3.2). Clearly

$$\begin{aligned} m(\pi) &= \int_\Theta \ell(\theta)[(1-\varepsilon)\pi_0(\theta) + \varepsilon q(\theta)]d\theta \\ &= (1-\varepsilon)m(\pi_0) + \varepsilon m(q). \end{aligned} \tag{3.16}$$

Note next that any unimodal symmetric (about θ_0) distribution q can be represented as a mixture of symmetric uniforms. Thus the extreme points of $\mathcal{Q}_{SU}$ are simply the *Uniform* $(\theta_0 - z, \theta_0 + z)$ densities, and

$$\begin{aligned} \Gamma_0 &= \{\text{extreme points of } \Gamma\} \\ &= \{(1-\varepsilon)\pi_0 + \varepsilon \mathcal{U}(\theta_0 - z, \theta_0 + z), z > 0\}. \end{aligned}$$

Thus

$$\sup_{\pi \epsilon \Gamma} m(\pi) = (1-\varepsilon)m(\pi_0) + \sup_{z>0} \int_{\theta_0 - z}^{\theta_0 + z} \frac{1}{2z} \ell(\theta) d\theta. \tag{3.17}$$

II. Ratio-Linear Posterior Quantities

Consider a quantity of the form

$$\rho(\pi) = \frac{\int f(\theta)\pi(d\theta)}{\int g(\theta)\pi(d\theta)}, \tag{3.18}$$

where $g(\cdot) > 0$. Then

$$\sup_\pi \rho(\pi) = \sup_\theta \frac{f(\theta)}{g(\theta)}, \quad \inf_\pi \rho(\pi) = \inf_\theta \frac{f(\theta)}{g(\theta)}. \tag{3.19}$$

This is a standard result (c.f. Sivaganesan and Berger, 1987), yet is extremely powerful for finding ranges of many posterior measures.

III. Ratio-Nonlinear Posterior Measures

In Section 3.1, it was observed that interesting ratio-nonlinear measures can often be handled by working with constrained classes Γ_{φ_0}. Specifically, suppose one has a class Γ specified by

$$\Gamma = \{\pi : \int b_i(\theta)\pi(d\theta) = 0, \quad i = 1, \ldots, k\}. \tag{3.20}$$

Example 5. Suppose μ_0 is a possible value of the posterior mean, and that it is desired to find the range of possible posterior variances among all priors with posterior mean μ_0, i.e., for which

$$\frac{\int \theta \ell(\theta)\pi(d\theta)}{m(\pi)} = \mu_0,$$

where, as usual, $m(\pi) = \int \ell(\theta)\pi(d\theta)$. To reduce this to a class of the form (3.20), we further constrain π so that $m(\pi)$ is fixed at a value m_0. Thus we define

$$\Gamma_{m_0} = \{\pi : \int [\ell(\theta) - m_0]\pi(d\theta) = 0, \int [\theta\ell(\theta) - m_0\mu_0]\pi(d\theta) = 0\}.$$

If we can find

$$\overline{\rho}_{m_0} = \sup_{\pi \in \Gamma_{m_0}} \rho(\pi), \quad \text{and} \quad \underline{\rho}_{m_0} = \inf_{\pi \in \Gamma_{m_0}} \rho(\pi)$$

for each m_0, then simple optimization over m_0 will yield the overall $\underline{\rho}$ and $\overline{\rho}$.

The constraints in (3.20) are quite flexible, allowing not only moment type constraints on π, but also probabilistic constraints. For instance, setting $b_i(\theta) = I_C(\theta) - p$, one has the constraint that π gives probability p to the set C.

Maximization or minimization of a linear functional, over a class Γ as in (3.20), is called a generalized moment problem (cf. Kemperman (1968)). Specifically, if one wants to maximize

$$\rho(\pi) = \int h(\theta)\pi(\theta)d\theta$$

over Γ in (3.20), there will typically exist a maximizing π_0 satisfying

$$\textit{support } (\pi_0) = \{\theta : h(\theta) = a_0 + \sum_{i=1}^{k} a_i b_i(\theta)\}, \tag{3.21}$$

where $\{a_0, a_1, \ldots, a_k\}$ is some set of constants satisfying

$$h(\theta) \leq a_0 + \sum_{i=1}^{k} a_i b_i(\theta) \quad \text{for all } \theta. \tag{3.22}$$

Such a result is typically utilized by showing that *support* (π_0) contains, say, at most n points, so that maximization of $\rho(\pi)$ over Γ can be done by maximizing over all n point

distributions in Γ. For precise theorems and arguments along these lines in robust Bayesian settings, see Sivaganesan and Berger (1987).

Analysis is also possible when the constraints in (3.20) are replaced by the constraints $\int b_i(\theta)\pi(d\theta) \geq 0$. See DeRobertis (1978) for discussion.

Analysis for Unimodal Γ

When dealing with classes of unimodal priors, the problem can usually be recast so as to apply the above methods of optimization. This is done by representing the unimodal prior as a mixture of uniforms, and applying the methods to the mixing distribution. Example 4 and Section 3.4.2 illustrate the idea.

3.4 Results for ε-Contamination Classes

3.4.1 When $\mathcal{Q}$ = All distributions

For any $\pi \epsilon \Gamma_\varepsilon$, one can write a ratio-linear posterior quantity as

$$\begin{aligned}\rho(\pi) &= E^{\pi *}[h(\theta)] \\ &= \frac{(1-\varepsilon)\int h(\theta)\ell(\theta)\pi_0(d\theta) + \varepsilon \int h(\theta)\ell(\theta)q(d\theta)}{(1-\varepsilon)\int \ell(\theta)\pi_0(d\theta) + \varepsilon \int \ell(\theta)q(d\theta)} \\ &= \frac{\int f(\theta)q(d\theta)}{\int g(\theta)q(d\theta)},\end{aligned}$$

where

$$\begin{aligned}f(\theta) &= (1-\varepsilon)\int h(\xi)\ell(\xi)\pi_0(d\xi) + \varepsilon h(\theta)\ell(\theta), \\ g(\theta) &= (1-\varepsilon)\int \ell(\xi)\pi_0(d\xi) + \varepsilon \ell(\theta).\end{aligned}$$

From (3.18) and (3.19) it follows that $\overline{\rho}$ and $\underline{\rho}$ can be determined by simply maximizing and minimizing, respectively, the ratio $f(\theta)/g(\theta)$ over θ, a simple numerical problem.

Huber (1973) developed this result for $h(\theta) = 1_C(\theta)$ (so that $\rho(\pi)$ is then the posterior probability of C), while Sivaganesan (1986) considered other posterior quantities such as

the mean and variance. Further simplification of the numerical problem to solution of iterative equations, and even closed form solution, is often possible.

Sivaganesan (1986) also considered the problem of finding the range of the posterior variance over Γ_ε for each fixed possible value of the posterior mean. The approach outlined in part III of Section 3.3 was used.

3.4.2 When $\mathcal{Q} = \mathcal{Q}_{SU}$

The opposite extreme to $\mathcal{Q}_A$ that has been considered in $\mathcal{Q}_{SU}$, the class of all symmetric (about θ_0) and unimodal densities. This class may be appropriate when one is quite confident that the prior is symmetric about θ_0 and unimodal (and, of course, chooses π_0 to be so). This can be considered to be a "minimal" robustness check, mainly investigating sensitivity to the exact functional form and tail of the prior. (In contrast, $\mathcal{Q}_A$ yields a "maximal" robustness check.)

In Sivaganesan and Berger (1987) this problem is solved, using the idea of representing q as a mixture of uniforms, namely

$$q(\theta) = \int_0^\infty \frac{1}{2z} I_{(\theta_0 - z, \theta_0 + z)}(\theta) G(dz),$$

where G is an arbitrary distribution on $(0, \infty)$. Any ratio-linear posterior quantity can then be written as

$$\begin{aligned} \rho(\pi) &= E^{\pi *}[h(\theta)] \\ &= \frac{(1-\varepsilon)\int h(\theta)\ell(\theta)\pi_0(d\theta) + \varepsilon \int_0^\infty H_1(z)G(dz)}{(1-\varepsilon)\int \ell(\theta)\pi_0(d\theta) + \varepsilon \int_0^\infty H_2(z)G(dz)}, \end{aligned}$$

where

$$H_1(z) = \frac{1}{2z}\int_{\theta_0 - z}^{\theta_0 + z} h(\theta)\ell(\theta)d\theta,$$

$$H_2(z) = \frac{1}{2z}\int_{\theta_0 - z}^{\theta_0 + z} \ell(\theta)d\theta.$$

Defining

$$f(z) = (1-\varepsilon)\int h(\xi)\ell(\xi)\pi_0(d\xi) + \varepsilon H_1(z),$$

$$q(z) = (1-\varepsilon)\int \ell(\xi)\pi_0(d\xi) + \varepsilon H_2(z),$$

it is clear that

$$\rho(\pi) = \int f(z)G(dz) / \int g(z)G(dz),$$

and (3.18) and (3.19) can again be applied to yield $\overline{\rho}$ and $\underline{\rho}$ as the maximum and minimum over z of $f(z)/g(z)$.

To determine the range of the posterior variance for a fixed posterior mean, Sivaganesan and Berger (1987) used the generalized moment approach outlined in part III of Section 3.3. Under certain conditions on $\ell(\theta)$, they were able to show that the extremes would occur among all two point mixing distributions G yielding the given posterior mean. The numerical problem thus was only two-dimensional.

Example 6. Suppose that $X \sim N(\theta, 1), \pi_0$ is $N(0,1), \varepsilon = 0.1$, and Γ_ε is used with $\mathcal{Q}_{SU}$. When $x = 4$ is observed, the range of the posterior mean is 1.96 to 3.34 and, for each possible value of the mean, the associated range of variances was given in Figure 1 (Section 3.1). This figure is from Sivaganesan and Berger (1987).

3.4.3 Other $\mathcal{Q}$

The analysis for $\mathcal{Q}_U$ is similar to that for $\mathcal{Q}_{SU}$, and can also be found in Sivagenesan and Berger (1987). The analysis for $\mathcal{Q}_{U^*}$ is much more difficult, since $\mathcal{Q}_{U^*}$ is only defined implicitly by the constraint that the overall prior be unimodal. Indeed it is somewhat surprising that the class is amenable to analysis.

The determination of $\underline{\rho}$ and $\overline{\rho}$ for $\mathcal{Q}_{U^*}$ has only been carried out when ρ is the marginal m (done in Berger and Berliner (1986)) or the posterior mean (done in Sivaganesan (1987)). The details are too complicated to present here, though the basic shape of the optimizing

$\pi \epsilon \Gamma$ can be described: it will be continuous, uniform within some interval, and equal to $(1-\varepsilon)\pi_0(\theta)$ outside the interval.

3.4.4 Considerations in Choosing $\mathcal{Q}$

As indicated previously, choice of Γ (here $\mathcal{Q}$) involves several trade-offs. Use of $\mathcal{Q}_{\mathcal{A}}$ typically leads to easy calculations, and assures that no reasonable priors are missed; thus if robustness is found to be present when $\mathcal{Q}_{\mathcal{A}}$ is used, one can unarguably claim robustness.

The negative side of $\mathcal{Q}_{\mathcal{A}}$ is that it contains many distributions which are undoubtedly unreasonable apriori, and this can lead to intervals $(\underline{\rho}, \overline{\rho})$ which are much wider than necessary. As an indication of this, it is interesting to observe (cf. Sivaganesan (1986) or Berger and Berliner (1986)) that asymptotics are "wrong" when $\mathcal{Q}_{\mathcal{A}}$ is used. For instance, the range of the posterior mean goes to zero at a rate $0(\sqrt{\ell n(n)/n})$, instead of the rate $0(1/\sqrt{n})$ that one would expect.

Restricting $\mathcal{Q}$ in some way is thus very desirable. Any of $\mathcal{Q}_{U*}, \mathcal{Q}_U$, and $\mathcal{Q}_{SU}$ are reasonable, and can be shown to yield sensible asymptotics. To indicate the comparative sizes of the intervals $(\underline{\rho}, \overline{\rho})$ they yield, consider the following example.

Example 6 (continued). Suppose that $X \sim N(\theta, 1), \pi_0$ is $N(0,1)$, and $\varepsilon = 0.1$. For each of the four $\mathcal{Q}$ considered and various values of x, the range $(\underline{\rho}, \overline{\rho})$ of the posterior mean for θ is given in Table 1. Note that, as discussed above, the bounds for $\mathcal{Q}_{SU}$ are substantially smaller than those for $\mathcal{Q}_{\mathcal{A}}$. Note also that larger x correspond to a "clash" between the data and π_0, and hence result in greater uncertainty.

Table 1. Ranges of the Posterior Mean for Γ_ε

	x					
	0	1.0	2.0	3.0	4.0	6.0
$\mathcal{Q}_{\mathcal{A}}$	(-0.10, 0.10)	(0.45, 0.69)	(0.93, 1.45)	(1.42, 2.67)	(1.85, 4.48)	(2.61, 8.48)
$\mathcal{Q}_{U*}$	(-0.08, 0.08)	(0.46, 0.66)	(0.95, 1.42)	(1.44, 2.44)	(1.87, 3.83)	(2.67, 5.93)
$\mathcal{Q}_U$	(-0.06, 0.06)	(0.47, 0.60)	(0.96, 1.20)	(1.45, 2.09)	(1.92, 3.50)	(2.78, 5.90)
$\mathcal{Q}_{SU}$	(-0.03, 0.03)	(0.48, 0.53)	(0.97, 1.12)	(1.46, 1.89)	(1.96, 3.34)	(2.87, 5.87)

From a calculational perspective, analyses with $\mathcal{Q}_U$ and $\mathcal{Q}_{SU}$ are both quite easy; $\mathcal{Q}_{U*}$ is substantially more difficult to deal with. All things considered, we recommend working with the smallest $\mathcal{Q}$ that can be used in good conscience, $\mathcal{Q}_{SU}$ if symmetry and unimodality are plausible. For situations in which symmetry is inappropriate, additional constraints should perhaps be applied to $\mathcal{Q}_U$ to prevent all the mass of q from concentrating on one side of θ_0.

Another issue, in the use of Γ_ε, is the choice of π_0. Γ_ε is actually somewhat limited, in that priors in Γ_ε have tails which are at least as large as $(1-\varepsilon)\pi_0$. Thus, if one chooses π_0 to be a Cauchy or t distribution, no prior in Γ_ε will have exponentially decreasing tails. Γ_ε will, however, contain all priors with larger tails than π_0, which is generally more important.

The choice of π_0 can have a substantial effect, especially when ℓ and π_0 clash (i.e., are concentrated in different regions). This is because the tails of π_0 can then drastically affect the answer, with the results for exponential tails and polynomial tails often being markedly different. Thus if one chooses π_0 to be, say, normal and π_0 clashes with ℓ, then the range of posterior means over Γ_ε will typically be very large. If, on the other hand, one chooses π_0 to be Cauchy, only Cauchy-like tails will occur in Γ_ε and the range of posterior means will be much smaller.

Which π_0 to choose is, of course, a subjective decision. If one feels essentially certain that sharp tails are inappropriate, then choosing a flat-tailed π_0 will certainly aid in verifying robustness. The subjective decision as to whether or not sharp tails are appropriate can perhaps be answered by determining one's likely reaction to a hypothetical clash between ℓ and π_0; if one would react to such a clash by essentially ignoring π_0 and trusting ℓ, then considering only fatter-tailed priors is reasonable.

3.5 Results for Quantile Classes

DeRobertis (1978) considered Γ_Q in (3.13) with $u_i = \ell_i + \ell_0$, and determined ranges of posterior probabilities of sets and posterior means. The priors at which minimums and maximums obtain are, not surprisingly, discrete measures. Various computational simplifications were developed in DeRobertis (1978).

This work did not cover the special case in which $\ell_i = u_i = p_i$. This case is of substantial interest in that even unsophisticated practitioners might be able to specify a few prior quantiles; more elaborate interval specification is a higher order activity. The solution to this problem, when the range of the posterior probability of a set is of interest, was given in Berliner and Goel (1987). Again the optimizing priors are discrete, and here the optimizing prior can be given explicitly.

As discussed earlier, a class which contains unreasonable priors (such as discrete priors) will result in overly wide ranges $(\underline{\rho}, \overline{\rho})$. The above quantile classes clearly suffer from this potential problem. Indeed, in DeRobertis (1978), it is demonstrated that the asymptotics for this class are again "wrong".

Thus a further refinement of Γ_Q, such as the addition of the unimodality condition leading to Γ_{QU}, is desirable. This class (together with further specializations and generalizations) is considered in Berger and O'Hagan (1987) and O'Hagan and Berger (1987). The posterior criterion considered is posterior probability of a set. For unimodal likelihoods, it is shown that $\underline{\rho}$ and $\overline{\rho}$ are attained at priors which are step functions having at most one step per interval I_i. Finding the optimizing priors can, indeed, be reduced to a simple convex programming problem, for which numerical solution is straightforward.

Example 7. An engineer specifies his prior probabilities of six intervals, I_i, for the mean life, θ, of a certain type of machine. These are given in Table 2. The engineer also states that the overall shape of his prior is unimodal. Thus Γ_{QU} in (3.14) is the implied class of

priors with $\ell_i = u_i =$ prior probability of I_i.

Data about θ is obtained in the form of two independent exponential observations of actual lifetimes, resulting in the likelihood function $\ell(\theta) = \theta^{-2}exp(-4500/\theta)$. The last column in Table 2 gives the ranges of the posterior probabilities of the I_i over π in Γ_{QU}. For comparison purposes, the corresponding ranges for Γ_Q (the class without the unimodality contraint) are given. (These results are from Berger and O'Hagan (1987) and Berliner and Goel (1987), respectively.) Note that the unimodality constraint can significantly reduce the ranges.

Table 2. Prior Probabilities and Posterior Ranges: Exponential Example

I_i	Prior Probability	Posterior Range: Γ_Q	Posterior Range: Γ_{QU}
[0,1000)	0.01	(0, 0.006)	(0.001, 0.004)
[1000,2000)	0.04	(0.019, 0.057)	(0.037, 0.049)
[2000,3000)	0.20	(0.214, 0.291)	(0.225, 0.260)
[3000,4000)	0.50	(0.476, 0.613)	(0.517, 0.584)
[4000,5000)	0.15	(0.106, 0.164)	(0.121, 0.147)
[5000,∞)	0.10	(0, 0.083)	(0, 0.071)

3.6 Results for Density Ratio Classes

When $\rho(\pi) = E^{\pi *}[h(\theta)]$, DeRobertis and Hartigan (1981) establish that $\underline{\rho}$ and $\overline{\rho}$ for Γ_{DR} are the unique solutions, respectively, of

$$\int_{\{\theta:h(\theta)<\underline{\rho}\}} (h(\theta) - \underline{\rho})\ell(\theta)U(\theta)\nu(d\theta) + \int_{\{\theta:h(\theta)>\underline{\rho}\}} (h(\theta) - \underline{\rho})\ell(\theta)L(\theta)\nu(d\theta) = 0, \qquad (3.23)$$

$$\int_{\{\theta:h(\theta)<\overline{\rho}\}} (h(\theta) - \overline{\rho})\ell(\theta)L(\theta)\nu(d\theta) + \int_{\{\theta:h(\theta)>\overline{\rho}\}} (h(\theta) - \overline{\rho})\ell(\theta)U(\theta)\nu(d\theta) = 0. \qquad (3.24)$$

These can sometimes be solved in closed form and are, in any case, quite easy to solve numerically.

Example 8. (DeRobertis and Hartigan, 1981): Suppose $\ell(\theta)$ arises from $\overline{X} \sim N(\theta, \sigma^2/n)$, σ^2 known. Let $L(\theta) = 1$ and $U(\theta) = K$, as discussed in Example 2. Set $h(\theta) = \theta$, so that

$\rho(\pi)$ is the posterior mean. Then the interval $(\underline{\rho}, \overline{\rho})$ can be shown to be

$$(\overline{x} - \frac{\sigma}{\sqrt{n}}\gamma(K), \overline{x} + \frac{\sigma}{\sqrt{n}}\gamma(K)), \tag{3.25}$$

where $\gamma(K)$ is a function tabulated in Table 3 for various K.

Table 3. Half Width of $(\underline{\rho}, \overline{\rho})$ for the Posterior Mean

K	1	1.25	1.50	2	2.5	3	4	6	8	10
$\gamma(K)$	0	.089	.162	.276	.364	.436	.549	.707	.817	.901

Thus, if one feels sure that the prior density ratio between 2 points is always between $\frac{1}{2}$ and 2, then the posterior mean is guaranteed to be within 0.276 standard errors of $\overline{x}$.

Example 9. Suppose $h(\theta) = I_C(\theta)$, so that $\rho(\pi)$ is the posterior probability of the set C. Then the solutions to (3.23) and (3.24) are (letting $\overline{C}$ denote the complement of C)

$$\underline{\rho} = \left[1 + \frac{\int_{\overline{C}} U(\theta)\ell(\theta)\nu(d\theta)}{\int_C L(\theta)\ell(\theta)\nu(d\theta)}\right]^{-1},$$

$$\overline{\rho} = \left[1 + \frac{\int_{\overline{C}} L(\theta)\ell(\theta)\nu(d\theta)}{\int_C L(\theta)\ell(\theta)\nu(d\theta)}\right]^{-1}.$$

DeRobertis and Hartigan (1981) were also able to partially solve the difficult problem of finding $\underline{\rho}$ and $\overline{\rho}$ for ratio-nonlinear quantities. Indeed for decision problems, in which (3.4) is of interest with $h(\theta, \varphi(\pi)) = L(\theta, a(\pi))$, they determine $\underline{\rho}$, and an upper bound on $\overline{\rho}$ which is frequently sharp.

It is also shown in DeRobertis and Hartigan (1981) that the asymptotics for Γ_{DR} is "right". This in indicated by the range in (3.25) for the posterior mean; as $n \to \infty$ the width decreases by the "right" order of $(1/\sqrt{n})$.

3.7 Comparing Classes of Priors

In this section we summarize and augment the arguments for and against the various classes of priors. Although often easy to work with, the conjugate and moment classes are typically too narrow to provide reliable indicators of robustness; hence we will concentrate on the other classes.

The ε-contamination class, with $\mathcal{Q}$ chosen in a constrained fashion, is usually quite easy to work with, gives sensible robustness ranges, and has a simple interpretation.. Furthermore, it can be used easily in "automatic" robustness checks. Thus, if a Bayesian specifies a single prior π_0, one could have a computer program which printed out $(\underline{\rho}, \overline{\rho})$ for "default" ε ($\varepsilon = 0.1$ or $\varepsilon = 0.2$ are reasonable) and $\mathcal{Q}$ equal to, say, $\mathcal{Q}_U$ or $\mathcal{Q}_{SU}$.

On the negative side, $\mathcal{Q}_A$ allows in too many silly priors, $\mathcal{Q}_U$ and $\mathcal{Q}_{SU}$ might be deemed to be too small for some applications, and the appealing $\mathcal{Q}_{U*}$ is very hard to work with. Also, the choice of π_0, in particular its tail behavior, can have a pronounced effect.

The chief advantage of the quantile classes, Γ_Q and Γ_{QU}, is that they relate naturally to the most basic of elicitation mechanisms, specification of probabilities of sets. While Γ_Q again contains many silly priors, Γ_{QU} does not and will tend to give reasonably tight posterior ranges. The disadvantage of Γ_{QU} is that it is the most difficult class to work with; indeed the determination of $(\underline{\rho}, \overline{\rho})$ has been done only for $\rho =$ posterior probability of a set, and only when $\ell(\theta)$ is itself unimodal. In O'Hagan and Berger (1987), however, a related "quasi-unimodal" class is considered which gives very similar answers to Γ_{QU} and is much easier to analyze. Indeed the analysis with this class could easily be extended to handle other posterior quantities.

The density ratio class Γ_{DR} is the nicest to work with mathematically. The range of virtually any posterior quantity of interest can be found using the results of DeRobertis and Hartigan (1981).

There are two disadvantages with Γ_{DR}. The first is that it is more difficult to elicit than the other classes. "Natural" methods of elicitation are not necessarily successful, as Example 3 demonstrated. However, we feel that effort directed at the elicitation problem has a good chance of success, and is well worthwhile given the calculational simplicity of the class.

The second disadvantage of Γ_{DR} is that one cannot control the overall shape of priors in the class. For instance, one cannot impose a constraint such as unimodality on Γ_{DR}, and preserve the calculational simplicity. While not always reasonable, such shape constraints arise frequently and are desirable in terms of their effect on $(\underline{\rho}, \overline{\rho})$. Of course, one can choose both $L(\theta)$ and $U(\theta)$ in Γ_{DR} to follow the desired shape; the extent of the allowed deviation from this shape might then be minimal for many applications.

In conclusion, neither the ϵ-contamination, quantile, or density ratio classes can be said to be clearly superior to the others. They all correspond to legitimate and accessible ways of modelling prior uncertainty, and all should be available for use and studied.

4. LOWER BOUNDS ON BAYES FACTORS IN TESTING SCENARIOS

Testing precise hypotheses has been extensively studied from the robust Bayesian perspective because of the large discrepancy observed between P-values and Bayes factors (or posterior probabilities). Suppose it is desired to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. The Bayes factor against H_0 is

$$B = \ell(\theta_0) / \int_{\{\theta \neq \theta_0\}} \ell(\theta)\pi_1(d\theta),$$

where π_1 is the conditional prior distribution given H_1 is true.

Of interest, when π_1 is known to belong to a class Γ, is

$$\underline{B} = \inf_{\pi_1 \in \Gamma} B = \frac{\ell(\theta_0)}{\sup_{\pi_1 \in \Gamma} \int_{\{\theta \neq \theta_0\}} \ell(\theta)\pi_1(d\theta)}.$$

Typically $\pi_1(d\theta)$ has a continuous density which, if defined at $\theta = \theta_0$ by continuity, gives no mass to θ_0. Then the denominator above is simply

$$\overline{m} = \sup_{\pi_1 \in \Gamma} m(\pi_1),$$

calculation of which was discussed previously.

The startling conflict with P-values can be illustrated by taking Γ to be a large class, such as

$$\Gamma_{SU} = \{\pi_1 : \pi_1 \text{ is unimodal and symmetric about } \theta_0\}.$$

This can be thought of as the ε-contamination class discussed earlier with $\varepsilon = 1$. The analysis in Example 4 then applies (with $\varepsilon = 1$) yielding

$$\underline{B} = \frac{\ell(\theta_0)}{\overline{m}} = \frac{\ell(\theta_0)}{\sup_z \left[\frac{1}{2z}\int_{\theta_0 - z}^{\theta_0 + z} \ell(\theta)d\theta\right]}.$$

Example 10. Suppose we will be observing $\overline{X} \sim N(\theta, \sigma^2/n), \sigma^2$ known. The usual P-value of observed data $\overline{x}$ against $H_0 : \theta = \theta_0$ is

$$\alpha = P_{\theta_0}(|Z| \geq \sqrt{n}|\overline{x} - \theta_0|/\sigma),$$

where Z is $N(0,1)$. For this situation, Berger and Sellke (1987) provide an explicit formula for $\underline{B}$. (It can be given as a function of $\sqrt{n}|\bar{x} - \theta_0|/\sigma$.) Values of $\underline{B}$ for various α are reproduced in Table 4.

Table 4. Values of $\underline{B}$ and Corresponding P-values

P-value (α)	0.10	0.05	0.01	0.001
$\underline{B}$	1/1.56	1/2.45	1/8.17	1/54.56

The surprisingly large discrepancy between $\underline{B}$ and P-values is clear. When the P-value is 0.05 for instance, which is typically viewed as significant evidence against H_0, $\underline{B}$ is 1/2.45. Thus there is *no* symmetric, unimodal prior which, when combined with the data, would indicate more than $2\frac{1}{2}$ times as much support for H_1 as for H_0. A P-value of 0.05 would appear to be at best very weak evidence against H_0.

A variety of results in this direction have been obtained. Edwards, Lindman, and Savage (1963) were the first to formally pursue such problems, and considered conjugate classes of priors and Γ = {all priors}. (Very surprisingly, $\underline{B}$ is typically bigger than the P-value even if π_1 is allowed to be any prior.) Berger and Sellke (1987) gave a variety of one-dimensional results for symmetric and unimodal-symmetric classes of π_1. Delampady (1986b) considered invariant testing problems, where Γ was (essentially) the class of invariant priors. Delampady and Berger (1987) considered multinomial testing problems and chi-squared tests of fit, with Γ being a broad class of priors symmetric in a reparameterization. Berger and Delampady (1987) considered, for the binomial testing problem, not only the above classes but also classes such as all unimodal priors with median (or mode) equal to θ_0. In all cases large discrepancies between P-values and $\underline{B}$ were observed.

One-sided testing problems were considered in Casella and Berger (1987); here $\underline{B}$ was often found to equal the P-value. Delampady (1986a) considered the general case of an interval null hypothesis, when Γ is the class of all unimodal, symmetric about θ_0 priors. A continuum of behavior between the point null and one-sided cases was observed, as

would be expected. A related analysis was carried out in Berger and Delampady (1987), establishing conditions under which testing an interval null could be closely approximated by testing a point null. Dickey (1977) considered the relationship between bounds on P-values and error probabilities.

Of course, all of this was qualitatively well understood by Good. In Good (1957) it was explicitly observed that Bayes factors against a point null tend to be between $\frac{10}{3}\alpha$ and 30α, the larger factors corresponding to smaller P-values. Good's views on the subject, including discussion of the Bayes/Non-Bayes compromise, are reviewed in Good (1981). (See also Good, 1950, 1958, 1965, 1967, 1983a, 1984, 1985, and 1986, Good and Crook, 1974 and 1987, and Crook and Good, 1980).

BIBLIOGRAPHY

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

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

[3] Berger, J. (1986). Are P-values reasonable measures of accuracy? In *Pacific Statistical Congress*, I. Francis et al. (eds.). North-Holland, Amsterdam.

[4] Berger, J. and Berliner, L. M. (1984). Bayesian input in Stein estimation and a new minimax empirical Bayes estimator. *J. of Econometrics 25*, 87–108.

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

[6] Berger, J. and Delampady, M. (1987). Testing precise hypotheses. To appear in *Statist. Science.*

[7] Berger, J. and O'Hagan, A. (1987). Ranges of posterior probabilities for unimodal priors with specified quantiles. *Technical Report #86-59*, Department of Statistics, Purdue University.

[8] Berger, J. 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.

[9] Berliner, L. M. and Goel, P. (1986). Incorporating partial prior information: ranges of posterior probabilities. *Technical Report, No. 357*, Department of Statistics, Ohio State University.

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

[11] 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.

[12] Crook, J. F. and Good, I. J. (1980). On the application of symmetric Dirichlet distributions and their mixtures to contingency tables, Part II. *Ann. Statist. 8*, 1198–1218.

[13] Davis, W. A. (1979). Approximate Bayesian predictive distributions and model selection. *J. Amer. Statist. Assoc. 74*, 312–317.

[14] DeGroot, M. H. (1970). *Optimal Statistical Decisions.* McGraw-Hill, New York.

[15] Delampady, M. (1986a). Lower bounds on Bayes factors for interval null hypotheses. *Technical Report 86-35*, Department of Statistics, Purdue University.

[16] Delampady, M. (1986b). Lower bounds on Bayes factors for invariant testing situations. *Technical Report 86-36*, Department of Statistics, Purdue University.

[17] Delampady, M. and Berger, J. (1987). Lower bounds on Bayes factors for multinomial and chi-squared tests of fit. *Technical Report 86-37*, Department of Statistics, Purdue University.

[18] Dempster, A. P. (1975). A subjectivist look at robustness. *Bull. Int. Statist. Inst.* 46, 349–374.

[19] Dempster, A. P. (1976). Examples relevant to the robustness of applied inferences. In *Statistical Decision Theory and Related Topics II*, S. S. Gupta and D. S. Moore (eds.). Academic Press, New York.

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

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

[22] Dickey, J. M. (1976). Approximate posterior distributions. *Journal of the American Statistical Association*, 71, 680–689.

[23] Dickey, J. M. (1977). Is the tail area useful as an approximate Bayes factor. *Journal of the American Statistical Association*, 72, 138–142.

[24] Edwards, W., Lindman, H., and Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, 70, 193–242. (Reprinted in *Robustness of Bayesian Analyses*, J. Kadane (Ed.), Amsterdam: North-Holland.)

[25] Fine, T. (1973). *Theories of Probability.* Academic Press, New York.

[26] Fishburn, P. C. (1965). Analysis of decisions with incomplete knowledge of probabilities. *Operations Research 13*, 217–237.

[27] Goldstein, M. (1980). The linear Bayes regression estimator under weak prior assumptions. *Biometrica* 67, 621–628.

[28] Good, I. J. (1950). *Probability and the Weighing of Evidence.* Charles Griffin, London.

[29] Good, I. J. (1952). Rational decisions. *J. Roy. Statist. Soc. B14*, 107–114.

[30] Good, I. J. (1956a). On the estimation of small frequencies in contingency tables. *J. Roy. Statist. Soc. B 18*, 113–124.

[31] Good, I. J. (1956b). Review of M. Allais, *Foundements d'un théorie positive des choix comportant un risque et critique des postulats et axiomes de l'école Américaine. J. Roy. Statist. Soc. A 119*, 213–214.

[32] Good, I. J. (1957a). Saddle-point methods for the multinomial distribution. *Annals of Math. Statist. 28*, 861–881.

[33] Good, I. J. (1957b). The appropriate mathematical tools for describing and measuring uncertainty. In *Uncertainty and Business Decisions.* Liverpool University Press, Liverpool.

[34] Good, I. J. (1958). Significance tests in parallel and in series. *Journal of the American Statistical Association*, 53, 799—813.

[35] Good, I. J. (1959). Could a machine make probability judgements? *Computers and Automation* 8, 14–16 and 24–26.

[36] Good, I. J. (1961). Discussion of C. A. B. Smith "Consistency in statistical inference and decision." *J. Roy. Statist. Soc. B 23*, 28–29.

[37] Good, I. J. (1962a). Subjective probability as the measure of a nonmeasurable set. In *Logic, Methodology, and Philosophy of Science.* University Press, Stanford.

[38] Good, I. J. (1962b). How rational should a manager be? *Management Science 8*, 383–393.

[39] Good, I. J. (1962c). Review of H. Jeffreys *Theory of Probability* (3rd ed.). *J. Roy. Statist. Soc. A 125*, 487–489.

[40] Good, I. J. (1965). *The Estimation of Probabilities: As Essay on Modern Bayesian Methods.* M. I. T. Press, Cambridge.

[41] Good, I. J. (1967). A Bayesian significance test for the multinomial distribution. *Journal of the Royal Statistical Society B*, 29, 399–431.

[42] Good, I. J. (1969). What is the use of a distribution? In *Multivariate Analysis II*, P. R. Krishnaiah (ed.). Academic Press, New York.

[43] Good, I. J. (1973). The probabilistic explication of evidence, surprise, causality, explanation, and utility. In *Foundations of Statistical Inference*, V. P. Godambe and D. A. Sprott (eds.). Holt, Rinehart and Winston, Toronto.

[44] Good, I. J. (1974). Review of T. L. Fine *Theories of Probability: An Examination of Foundations. IEEE Trans. Inform. Theory IT-20*, 298–300.

[45] Good, I. J. (1975). Explicativity, corroboration, and the relative odds of hypotheses. *Synthese 30*, 39–73.

[46] Good, I. J. (1976). The Bayesian influence, or how to sweep subjectivism under the carpet. In *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, C. A. Hooker and W. Harper (eds.). D. Reidel, Dordrecht.

[47] Good, I. J. (1979). Some history of the hierarchical Bayesian methodology. In *Bayesian Statistics*, J. M. Bernardo et. al. (eds.). University Press, Valencia.

[48] Good, I. J. (1980). The contributions of Jeffreys to Bayesian statistics. In *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys*, Arnold Zellner (ed.). North-Holland, Amsterdam.

[49] Good, I. J. (1981). Some logic and history of hypothesis testing. In *Philosophical Foundations of Economics*, J. C. Pitt (ed.). D. Reidel, Dordrecht.

[50] Good, I. J. (1982a). The axioms of probability. *Encyclopedia of Statist. Sciences*, Vol. 1, N. L. Johnson and S. Kotz (eds.). Wiley, New York.

[51] Good, I. J. (1982b). Degrees of belief. *Encyclopedia of Statist. Sciences*, Vol. 2, N. L. Johnson and S. Kotz (eds.). Wiley, New York.

[52] Good, I. J. (1983a). *Good Thinking: The Foundations of Probability and Its Applications.* University of Minnesota Press, Minneapolis.

[53] 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.

[54] Good, I. J. (1984). Notes C140, C144, C199, C200 and C201, *Journal of Statistical Computation and Simulation*, 19.

[55] Good, I. J. (1985). Weight of evidence; A brief survey. *In Bayesian Statistics 2*, J. M. Bernardo, et. al. (Eds). North-Holland, Amsterdam.

[56] Good, I. J. (1986). A flexible Bayesian model for comparing two treatments. *J. Statist. Comput. and Simul.* 26, 301–305.

[57] Good, I. J. and Crook, J. F. (1974). The Bayes/non-Bayes compromise and the multinomial distribution. *J. Amer. Statist. Assoc. 69*, 711–720.

[58] 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**.

[59] Hartigan, J. A. (1969). Linear Bayesian methods. *J. Roy. Statist. Soc. Ser. B* 31, 446–454.

[60] Hill, B. (1980a). Robust analysis of the random model and weighted least squares regression. *In Evaluation of Econometric Models.* Academic Press, New York.

[61] Hill, B. (1980b). Invariance and robustness of the posterior distribution of characteristics of a finite population, with reference to contingency tables and the sampling of species. In *Bayesian Analysis in Econometrics and Statistics*, A. Zellner (ed.). North-Holland, Amsterdam.

[62] Huber, P. J. (1973). The use of Choquet capacities in statistics. *Bull. Int. Statist. Inst.* 45, 181–191.

[63] Isaacs, H. H. (1963). Sensitivity of decisions to probability estimation errors. *Operations Research 11*, 536–552.

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

[65] Kemperman, J. H. B. (1968). The general moment problem; a geometric approach. *Ann. Math. Statist.* 39, 93–122.

[66] Keynes, J. M. (1921). *A Treatise on Probability.* Macmillan, London.

[67] Koopman, B. O. (1940a). The axioms and algebra of intuitive probability. *Ann. of Math. 41*, 269–292.

[68] Koopman, B. O. (1940b). The bases of probability. *Bull. Amer. Math. Soc. 46*, 763–774.

[69] Kudō, H. (1967). On partial prior information and the property of parametric sufficiency. *Proc. Fifth Berkeley Symp. Statist. and Prob. 1*, University of California Press, Berkeley.

[70] Lambert, D. and Duncan, G. (1981). Bayesian learning based on partial prior information. *Technical Report*, Department of Statistics, Carnegie-Mellon University.

[71] Leamer, E. E. (1978). *Specification Searches: Ad Hoc Inference With Nonexperimental Data.* Wiley, New York.

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

[73] Levi, I. (1980). *The Enterprise of Knowledge.* MIT Press, Cambridge.

[74] Manski, C. F. (1981). Learning and decision making when subjective probabilities have subjective domains. *Ann. Statist. 9*, 59–65.

[75] O'Hagan, A. and Berger, J. (1987). Ranges of posterior probabilities for quasi-unimodal priors with specified quantiles. *Technical Report* #87-11, Department of Statistics, Purdue University.

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

[77] 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.

[78] Rubin, H. (1977). Robust Bayesian estimation. *In Statistical Decision Theory and Related Topics II*, (S. S. Gupta and D. Moore, Eds.). Academic Press, New York.

[79] Sivaganesan, S. (1986). Range of posterior measures for priors with arbitrary contaminations. *Technical Report*, Department of Statistics, Southern Methodist University, Dallas.

[80] Sivaganesan, S. (1987). Sensitivity of posterior mean to unimodality preserving contaminations. *Technical Report*, Department of Statistics, Southern Methodist University.

[81] Sivaganesan, S. and Berger, J. (1987). Ranges of posterior measures for priors with

unimodal contaminations. *Technical Report* #86-41, Department of Statistics, Purdue University, West Lafayette.

[82] Smith, C. A. B. (1961). Consistency in statistical inference and decision. *J. Roy. Statist. Soc. B 23*, 1–25.

[83] Stone, M. (1963). Robustness of nonideal decision procedures. *J. Amer. Statist. Assoc. 58*, 480–486.

[84] Suppes, P. (1975). Approximate probability and expectation of gambles. *Erkenntnis* 9, 153–161.

[85] Walley, P. (1986). *Rationality and Vagueness.* Manuscript in preparation.

[86] West, S. (1979). Upper and lower probability inferences for the logistic function. *Ann. Statist. 7*, 400–413.

[87] Wolfenson, M. and Fine, T. L. (1982). Bayes-like decision making with upper and lower probabilities. *J. Amer. Statist. Assoc.* 77, 80–88.