BAYESIAN ROBUSTNESS IN BIDIMENSIONAL MODELS: 
PRIOR INDEPENDENCE 

by 
James Berger and Elías Moreno 
Purdue University and Universidad de Granada 

Technical Report #92-52C 

Department of Statistics 
Purdue University 

November 1992 
Revised - February 1993
BAYESIAN ROBUSTNESS IN BIDIMENSIONAL MODELS:
PRIOR INDEPENDENCE*

by

James Berger and Elías Moreno
Purdue University Universidad de Granada

Abstract

When $\theta$ is a multidimensional parameter, the issue of prior dependence or independence of coordinates is a serious concern. This is especially true in robust Bayesian analysis; Lavine, Wasserman, and Wolpert (1991) show that allowing a wide range of prior dependencies among coordinates can result in near vacuous conclusions. It is sometimes possible, however, to confidently make the judgement that the coordinates of $\theta$ are independent apriori and, when this can be done, robust Bayesian conclusions improve dramatically. In this paper, it is shown how to incorporate the independence assumption into robust Bayesian analysis involving $\epsilon$-contamination and density band classes of priors. Attention is restricted to the case $\theta = (\theta_1, \theta_2)$ for clarity, although the ideas generalize.

* Research supported by the National Science Foundation, Grant DMS-8923071.
1. Introduction

A common approach to prior uncertainty in robust Bayesian analysis is to choose a class, \( \Gamma \), of prior distributions, and compute the range of Bayesian answers as the prior, \( \pi \), ranges over \( \Gamma \). Two common classes considered are the \( \epsilon \)-contamination class

\[
\Gamma_C = \{ \pi : \pi = (1 - \epsilon)\pi_0 + \epsilon q, \quad q \in Q \},
\]

(1)

where \( \pi_0 \) is the base elicited prior and \( Q \) is the class of allowed contaminations, and the density band class

\[
\Gamma_B = \{ \pi : L(\theta) \leq \pi(\theta) \leq U(\theta) \},
\]

(2)

where \( L \) and \( U \) are specified measurable functions on the parameter space \( \Theta \). For motivation, discussion, and examples of analyses with these classes see DeRobertis and Hartigan (1981), Berger and Berliner (1986), Sivaganesan and Berger (1989), Bose (1990), Berger (1990), Moreno and Cano (1991), Moreno and Pericchi (1992), Wasserman and Kadane (1992), and Wasserman (1992).

For a multidimensional parameter \( \theta = (\theta_1, \ldots, \theta_n) \), the classes (1) and (2) can be quite large; so large that the range of Bayesian answers can become uselessly wide. One of the main difficulties with multidimensional Bayesian robustness is that of prior dependence among coordinates. If too much dependence is allowed, the range of Bayesian answers can become extremely wide. See Lavine, Wasserman, and Wolpert (1991) for illustration. This problem may well be unavoidable in certain applications, but in others the elicitor may be willing to make the prior judgement of independence of the \( \theta_i \). This is, admittedly, a strong assumption, but it is a judgement that is accessible to intuition. And making some type of strong restriction is typically necessary in multidimensional Bayesian robustness. Thus we encourage elicitors to carefully consider whether or not they would be willing to judge the \( \theta_i \) to be apriori independent in a given application.

For simplicity of exposition, we restrict ourselves here to the two-dimensional case, \( \theta = (\theta_1, \theta_2) \), where \( \theta_1 \in \Theta_1 \) and \( \theta_2 \in \Theta_2 \). Let \( \Theta = \Theta_1 \times \Theta_2 \). The likelihood function of the data \( x \) will be denoted by \( f(x|\theta_1, \theta_2) \), and will be assumed to be appropriately measurable in \( \theta \).

We will use the notation

\[
E^\pi(\varphi f) = \int_\Theta \varphi(\theta_1, \theta_2)f(x|\theta_1, \theta_2)\pi(d\theta_1, d\theta_2),
\]

\[
E^\pi(\varphi|x) = \int_\Theta \varphi(\theta_1, \theta_2)f(x|\theta_1, \theta_2)\pi(d\theta_1, d\theta_2)/E^\pi(f).
\]

For simplicity, we henceforth write \( \pi(\theta_1, \theta_2) \) as a density (w.r.t. Lebesgue measure); this is a slight abuse of notation, since \( \pi \) will be seen to possibly have discrete components, but no confusion should result.
Directly imposing the constraint of independence in (1) results in a very complicated class to analyze. We thus consider, instead, the contamination class

$$\Gamma_{IC} = \{ \pi(\theta_1, \theta_2) : \pi(\theta_1, \theta_2) = [(1 - \varepsilon_1)\pi_{01}(\theta_1) + \varepsilon_1 q_1(\theta_1)][(1 - \varepsilon_2)\pi_{02}(\theta_2) + \varepsilon_2 q_2(\theta_2)],$$

$$q_1 \in Q_1, q_2 \in Q_2 \};$$

(3)

this not only imposes independence on \( \theta_1 \) and \( \theta_2 \), but also permits the introduction of different degrees of confidence in the base priors \( \pi_{01}(\theta_1) \) and \( \pi_{02}(\theta_2) \) through choice of the constants \( \varepsilon_1, \varepsilon_2, 0 < \varepsilon_1 < 1, 0 < \varepsilon_2 < 1 \). The contaminating classes can be chosen in accordance with features of \( \pi_{01}(\theta_1) \) and \( \pi_{02}(\theta_2) \) that it is desired to keep fixed. This class is studied in Section 2, when the \( Q_i \) are quantile classes of distributions or unimodal quantile classes.

For the class (2), directly imposing independence again results in excessive complications. Thus we consider the related density band class

$$\Gamma_{IB} = \{ \pi(\theta_1, \theta_2) : \pi(\theta_1, \theta_2) = \pi_1(\theta_1)\pi_2(\theta_2), \ L_i(\theta) \leq \pi_i(\theta) \leq U_i(\theta), \ i = 1, 2 \}. \quad (4)$$

Robustness results for this class are presented in Section 3 for certain quantities of interest. Note that analysis with this class is typically considerably more difficult than analysis with \( \Gamma_{IC} \). Examples that are considered include the well-studied ECMO trial.

2. Posterior Ranges for the \( \varepsilon \)-Contamination Class with Independence

2.1. Contaminating Classes Defined by Quantile Constraints

Consider the class of priors \( \Gamma_{IC} \) in (3) for contaminating classes \( Q_1, Q_2 \) given by

$$Q_1 = \{ q_1(\theta_1) : \int_{A_i} q_1(d\theta_1) = \alpha_i, \ i \geq 1, \ \sum \alpha_i = 1, \ 0 < \alpha_i < 1 \},$$

$$Q_2 = \{ q_2(\theta_2) : \int_{B_j} q_2(d\theta_2) = \beta_j, \ j \geq 1, \ \sum \beta_j = 1, \ 0 < \beta_j < 1 \}. \quad (5)$$

where \( \{A_i, \ i \geq 1\}, \{B_j, \ j \geq 1\} \) are specified measurable partitions of the spaces \( \Theta_1, \Theta_2 \), respectively, and the \( \alpha_i \) and \( \beta_j \) are specified constants. Let \( \varphi(\theta_1, \theta_2) \) be an arbitrary measurable function. Ranges of the posterior expectation of \( \varphi(\theta_1, \theta_2) \) as \( \pi \) ranges over \( \Gamma_{IC} \) are given in the following theorem.
Theorem 1.

\[ \inf_{\pi \in \Gamma_{IC}} E^\pi(\varphi(\theta_1, \theta_2)|x) = \inf \frac{(1 - \epsilon_1)(1 - \epsilon_2)E^{\pi_01\pi_02}(f|\varphi) + \sum_{i,j} K(a_i, b_j)\alpha_i\beta_j}{(1 - \epsilon_1)(1 - \epsilon_2)E^{\pi_01\pi_02}(f) + \sum_{i,j} H(a_i, b_j)\alpha_i\beta_j}, \]

where the infimum is over \( \{a_i \in A_i, \ i \geq 1, \ and \ b_j \in B_j, \ j \geq 1\} \),

\[ K(\theta_1, \theta_2) = (1 - \epsilon_1)\epsilon_2E^{\pi_01}(f|\varphi) + (1 - \epsilon_2)\epsilon_1E^{\pi_02}(f|\varphi) + \epsilon_1\epsilon_2f(x|\theta_1, \theta_2)\varphi(\theta_1, \theta_2), \]

\[ H(\theta_1, \theta_2) = (1 - \epsilon_1)\epsilon_2E^{\pi_01}(f) + (1 - \epsilon_2)\epsilon_1E^{\pi_02}(f) + \epsilon_1\epsilon_2f(x|\theta_1, \theta_2). \]

(ii) The supremum of \( E^\pi(\varphi(\theta_1, \theta_2)|x) \), as \( \pi \) varies over \( \Gamma_{IC} \), is obtained by interchanging inf with sup in (i).

Proof. Employing the linearization technique (cf. Lavine, Wasserman, and Wolpert, 1992), define

\[ \lambda = \inf_{\pi \in \Gamma_{IC}} E^\pi[\varphi(\theta_1, \theta_2)|x], \]

and note that the \( \pi \in \Gamma_{IC} \) (or sequence \( \{\pi_i\} \) of priors in \( \Gamma_{IC} \)) that minimize \( E^\pi[\varphi(\theta_1, \theta_2|x)] \) also minimizes

\[ \psi(q_1, q_2) = \int [\varphi(\theta_1, \theta_2) - \lambda]f(x|\theta_1, \theta_2)\pi(\theta_1, \theta_2)d\theta_1d\theta_2 \]

\[ = \int [\varphi(\theta_1, \theta_2) - \lambda]f(x|\theta_1, \theta_2)[(1 - \epsilon_1)\pi_01(\theta_1) + \epsilon_1q_1(\theta_1)] \]

\[ \cdot[(1 - \epsilon_2)\pi_02(\theta_2) + \epsilon_2q_2(\theta_2)]d\theta_1d\theta_2. \]

Write

\[ q_1(\theta_1) = \Sigma_i\alpha_iq_{i1}(\theta_1), \quad q_2(\theta_2) = \Sigma_j\beta_jq_{2j}(\theta_2), \]

where the \( q_{i1} \) and \( q_{2j} \) are (unrestricted) distributions on \( A_i \) and \( B_j \), respectively, and note that \( \psi(q_1, q_2) \) is linear in each of the \( q_{i1} \) and \( q_{2j} \). Hence the infimum of \( \psi(q_1, q_2) \) is achieved when the \( q_{i1} \) and \( q_{2j} \) are point masses, proving (i). The proof of (ii) is similar. \( \square \)

Example 1. Let \( X = (X_1, X_2) \) be a \( N_2((\theta_1, \theta_2), I) \) random variable, and suppose the base prior \( \pi_0(\theta_1, \theta_2) = \pi_01(\theta_1)\pi_02(\theta_2) \) is elicited, where the \( \pi_0i \) are \( N(0, 2) \) densities. We are interested in the robustness of the posterior probability of \( H_0 : \theta_1 < \theta_2 \) to departures from this base prior, and are willing to assume prior independence of \( \theta_1 \) and \( \theta_2 \). Thus we consider \( \Gamma_{IC} \) in (3), and choose the conservative \( \Omega_1 = \{ \text{all distributions on } \Theta_1 \} \), \( \Omega_2 = \{ \text{all distributions on } \Theta_2 \} \), and \( \epsilon_1 = \epsilon_2 = 0.106 \) (so that \( (1 - \epsilon_1)(1 - \epsilon_2) = 0.8 \); see below).

To see the effect of the independence constraint, we also consider \( \Gamma_C \) in (1) with \( \pi_0(\theta_1, \theta_2) \) as above, \( \epsilon = 0.2, \) and \( Q = \{ \text{all distributions} \} \). Although, as mentioned earlier, \( \Gamma_C \) is not exactly \( \Gamma_{IC} \) with the independence constraint removed, it is approximately so. Indeed, the prior precision, of \( H_0 \), defined as
\[ \Delta_{\Gamma} P^\pi(H_0) \equiv \sup_{\pi \in \Gamma} P^\pi(H_0) - \inf_{\pi \in \Gamma} P^\pi(H_0), \]

is the same for both classes, since

\[ \Delta_{\Gamma_C}(H_0) = \left[ \frac{1}{2}(1 - \varepsilon) + \varepsilon \right] - \frac{1}{2}(1 - \varepsilon) = 0.2, \]
\[ \Delta_{\Gamma IC}(H_0) = \left[ \frac{1}{2}(1 - \varepsilon_1)(1 - \varepsilon_2) + (1 - \varepsilon_1)\varepsilon_2 + \varepsilon_2(1 - \varepsilon_2) + \varepsilon_1\varepsilon_2 \right] - \frac{1}{2}(1 - \varepsilon_1)(1 - \varepsilon_2) = 0.2. \]

For various \( x \), the posterior ranges of \( H_0 \) as the prior ranges over \( \Gamma IC \) and \( \Gamma C \) are displayed in the third and fourth columns of Table 1, respectively. The corresponding values of the posterior probability of \( H_0 \), for the base prior \( \pi_0 \), are given in the second column. The notation used in this table is \( P_C(H_0|x) = \inf_{\pi \in \Gamma C} P^\pi(H_0|x), \quad P IC(H_0|x) = \sup_{\pi \in \Gamma IC} P^\pi(H_0|x) \).

Table 1. Ranges of the Posterior Probability of \( H_0 \)

| \( x = (x_1, x_2) \) | \( P^\pi_0(H_0|x) \) | \( (P_C(H_0|x), P IC(H_0|x)) \) |
|----------------|----------------|------------------|
| (0.2, 0.2)    | 0.591          | (0.342, 0.768), (0.508, 0.672) |
| (0.5, 0.5)    | 0.718          | (0.439, 0.845), (0.646, 0.786) |
| (0.8, 0.8)    | 0.822          | (0.552, 0.908), (0.766, 0.873) |
| (1.1, 1.1)    | 0.898          | (0.673, 0.952), (0.860, 0.930) |
| (1.4, 1.4)    | 0.947          | (0.787, 0.978), (0.924, 0.967) |
| (1.7, 1.7)    | 0.975          | (0.879, 0.992), (0.963, 0.986) |

Table 1 reveals the dramatic impact of the independence assumption, even though the \( Q \)'s are otherwise unconstrained. The effect of the assumption is to reduce the range of the posterior probability of \( H_0 \) by a factor of from 3 to 5.

2.2. Unimodal Contaminating Classes

The extreme priors for the posterior expectation \( E^\pi(\phi(\theta_1, \theta_2)|x) \), as \( \pi \) ranges over the \( \Gamma IC \) considered in Section 2.1, is given by the mixture

\[ [(1 - \varepsilon_1)\pi_{01}(\theta_1) + \varepsilon_1 q_1(\theta_1)][(1 - \varepsilon_2)\pi_{02}(\theta_2) + \varepsilon_2 q_2(\theta_2)], \]

where \( q_1(\theta_1) \) and \( q_2(\theta_2) \) are discrete distributions. If \( \pi_0(\theta_1, \theta_2) \) is a continuous distribution, allowing such unusual priors is often inappropriate. A popular method for eliminating such priors from the class is to also impose the restriction that \( \pi \) must be unimodal, assuming of course, that \( \pi_0 \) is unimodal. Adding the independence constraint, together with unimodality, turns out to be quite simple. Indeed, the infimum of \( E^\pi(\phi(\theta_1, \theta_2)|x) \), as \( \pi \) varies over \( \Gamma IC \) in (3) with

\[ Q_1 = \{ q_1(\theta_1) : q_1 \text{ is unimodal with the same mode as that of } \pi_{01} \}, \]
\[ Q_2 = \{ q_2(\theta_2) : q_2 \text{ is unimodal with the same mode as that of } \pi_{02} \}, \]
is given (assuming, without loss of generality, that the modes are zero) by

\[ \inf_{\pi \in \Gamma_{TC}} E^\pi(\varphi(\theta_1, \theta_2)|x) = \inf_{a_1, a_2} \frac{(1 - \varepsilon_1)(1 - \varepsilon_2)E^{\pi_0}(f\varphi) + K(a_1, a_2)}{(1 - \varepsilon_1)(1 - \varepsilon_2)E^{\pi_0}(f)} + H(a_1, a_2) \]

where the \( a_i \) are allowed to be positive or negative and

\[ K(a_1, a_2) = (1 - \varepsilon_1)\varepsilon_2 \frac{1}{a_2} \int_0^{a_2} E^{\pi_0}(f\varphi)d\theta_2 + (1 - \varepsilon_2)\varepsilon_1 \frac{1}{a_1} \int_0^{a_1} E^{\pi_0}(f\varphi)d\theta_1 \]

\[ + \varepsilon_1\varepsilon_2 \frac{1}{a_1} \int_0^{a_1} \left\{ \frac{1}{a_2} \int_0^{a_2} (f\varphi)d\theta_1 \right\} d\theta_2, \]

\[ H(a_1, a_2) = (1 - \varepsilon_1)\varepsilon_2 \frac{1}{a_2} \int_0^{a_2} E^{\pi_0}(f)d\theta_2 + (1 - \varepsilon_2)\varepsilon_1 \frac{1}{a_1} \int_0^{a_1} E^{\pi_0}(f)d\theta_1 \]

\[ + \varepsilon_1\varepsilon_2 \frac{1}{a_1} \int_0^{a_1} \left\{ \frac{1}{a_2} \int_0^{a_2} (f)d\theta_2 \right\} d\theta_1, \]

This follows directly from the Khinchine representation of a unimodal distribution as a mixture of uniforms, and Theorem 1. For the supremum of \( E^\pi(\varphi(\theta_1, \theta_2)|x) \), simply replace "inf" by "sup" above.

3. Posterior Ranges for the Density Band Class with Independence

The density band class \( \Gamma_B \), in (2), is popular because it allows considerable freedom in the tail behavior of the prior (providing the upper and lower functions \( L(\theta) \) and \( U(\theta) \) have substantially different tails) and because it is typically the easiest class to handle computationally. In this section we explore situations in which the associated independence class, \( \Gamma_{IB} \), is also computationally tractable.

Suppose the quantity of interest is \( \varphi(\theta_1) \) (e.g., \( \varphi(\theta_1) = \theta_1 \) or \( \varphi(\theta_1) = 1_A(\theta_1) \)), so that \( \theta_2 \) is effectively a nuisance parameter. The following theorem reduces determination of the range of \( E^\pi(\varphi(\theta_1)|x) \) to a two dimensional optimization problem.

**Theorem 2.**

\[ \sup_{\pi \in \Gamma_{IB}} E^\pi(\varphi(\theta_1)|x) = \sup_{k_1, k_2} \frac{\int m(x|\theta_2, \pi_1)E^{\pi_1}(\varphi(\theta_1|x_2, \theta_2)\pi_2(\theta_2)d\theta_2}{\int m(x|\theta_2, \pi_1)\pi_2(\theta_2)d\theta_2} \]

where

\[ \pi_1(\theta_1) = U_1(\theta_1)1_{\varphi(\theta) \geq k_1}(\theta_1) + L_1(\theta_1)1_{\varphi(\theta) < k_1}(\theta_1), \]

\[ E^{\pi_1}(\varphi(\theta_1)|x, \theta_2) = \{m(x|\theta_2, \pi_1)\}^{-1}\int \varphi(\theta_1)f(x|\theta_1, \theta_2)\pi_1(\theta_1)d\theta_1, \]

and

\[ \pi_2(\theta_2) = U_2(\theta_2)1_{A(k_1, k_2, x)}(\theta_2) + L_2(\theta_2)1_{A^c(k_1, k_2, x)}(\theta_2), \]

\[ A(k_1, k_2, x) \text{ being the set defined as } \{\theta_2 : E^{\pi_1}(\varphi(\theta_1)|x, \theta_2) \geq k_2\}. \]
(ii) To compute \( \inf_{\pi \in \Gamma_{IB}} E^\pi(\varphi(\theta_1)|x) \), simply replace "sup" by "inf" in part (i) and interchange the \( U_i \) and the \( L_i \) in the expressions for \( \pi_i(\theta_i) \).

**Proof.** Since the proofs of part (i) and part (ii) are virtually identical, we only prove part (i). Let \( \tilde{\pi}(\theta_1, \theta_2) = \tilde{\pi}_1(\theta_1) \cdot \tilde{\pi}_2(\theta_2) \) be a prior density in \( \Gamma_{IB} \) attaining the supremum of \( E^\pi(\varphi(\theta_1)|x) \). Defining \( \Gamma_2 = \{ \pi_2(\theta_2) : L_2(\theta_2) \leq \pi_2(\theta_2) \leq U_2(\theta_2) \} \), it is clear that

\[
\sup_{\pi \in \Gamma_{IB}} E^\pi(\varphi(\theta_1)|x) = \sup_{\pi_2 \in \Gamma_2} \frac{\int \varphi(\theta_1)f(x|\theta_1, \theta_2)\tilde{\pi}_1(\theta_1)\pi_2(\theta_2)d\theta_1d\theta_2}{\int f(x|\theta_1, \theta_2)\tilde{\pi}_1(\theta_1)\pi_2(\theta_2)d\theta_1d\theta_2} \]

\[
= \sup_{\pi_2 \in \Gamma_2} \frac{\int \tilde{\varphi}(\theta_2)m(x|\theta_2, \tilde{\pi}_1)\pi_2(\theta_2)d\theta_2}{\int m(x|\theta_2, \tilde{\pi}_1)\pi_2(\theta_2)d\theta_2},
\]

where \( m(x|\theta_2, \tilde{\pi}_1) = \int f(x|\theta_1, \theta_2)\tilde{\pi}_1(\theta_1)d\theta_1 \) and \( \tilde{\varphi}(\theta_2) = E^{\tilde{\pi}_1}(\varphi(\theta_1)|x, \theta_2) \). Treating \( m \) as the density of \( x \), DeRobertis and Hartigan (1981) can be applied to show that \( \pi_2(\theta_2) \) is as in (7), with \( A(k_1, k_2, x) \) computed using \( \tilde{\pi}_1 \). A similar argument shows that \( \pi_1(\theta_1) \) should be of the form (6), completing the proof. \( \square \)

**Corollary 1.** For any measurable set \( A \subset \Theta_1 \), \( \sup_{\pi \in \Gamma_{IB}} P^\pi(A|x) \) can be found from Theorem 2 by setting \( \varphi(\theta_1) = 1_A(\theta_1) \),

\[
\pi_1(\theta_1) = U_1(\theta_1)1_A(\theta_1) + L_1(\theta_1)1_{A^c}(\theta_1),
\]

and eliminating \( k_1 \) from all expressions. To find \( \inf_{\pi \in \Gamma_{IB}} P^\pi(A|x) \), simply interchange the \( U_i \) and \( L_i \) in all expressions. (Note that the problem reduces to a one-dimensional optimization, over \( k_2 \).)

**Proof.** Because \( \varphi(\theta_1) \) is itself an indicator function, the priors in (6) either have form (8) or are \( \pi_1(\theta_1) = U_1(\theta_1) \) or \( \pi_1(\theta_1) = L_1(\theta_1) \). These last two possibilities are easily shown to be inferior to the optimal prior of form (8). \( \square \)

**Example 2.** A clinical trial: ECMO

Nine patients were treated with ECMO (extra corporeal membrane oxygenation) of whom all nine survived. Ten patients were given standard therapy of whom six survived. Let \( p_1 \) be the probability of success under standard therapy and let \( p_2 \) be the probability of success under ECMO. It is desired to compare the two treatments.

This example has been deeply analyzed by Ware (1989), Kass and Greenhouse (1989) and Lavine, Wasserman, and Wolpert (1991) for a variety of priors. Here we shall carry out a robust Bayesian analysis for the density band classes of priors, with and without the independence constraint.

Let \( \eta_i = \log\left(p_i/(1 - p_i)\right), i = 1, 2 \), and consider the parameters \( \delta = \eta_2 - \eta_1 \) and \( \eta_1 \). With this reparametrization, the quantity of interest is the posterior probability that \( \delta > 0 \), and \( \eta_1 \) is a nuisance parameter.

Priors considered by Kass and Greenhouse (1989) for \( \eta_1 \) included Cauchy and Normal priors with location parameter equal to -1.7 or 0 and scale parameter equal to 0.769,
0.419, or 0.838; for δ the priors were Cauchy and Normal with location at 0 and different scale values. To globally capture the prior uncertainty underlying their analysis, we shall consider density band classes defined by the following procedure: if a collection \{π_i\} of priors is being considered for a parameter \(\xi\), let \(L(\xi)\) be the prior with sharpest tails and, if \(\pi^*\) denotes the prior with the thickest tails, let \(U(\xi)\) be defined by

\[
U(\xi) = \pi^*(\xi) \max_i \left\{ \sup_{\xi} \frac{L(\xi)}{\pi_i(\xi)} \right\}.
\]

It is easy to see that the ensuing density band class includes all of the \{π_i\}. (Note that multiplicative constants are irrelevant.)

Choosing two, rather extreme, priors considered by Kass and Greenhouse for \(\eta_1\), say \(N(-1.7, 769^2)\) and \(C(0, 419^2)\), we have that the above lower and upper densities for \(\eta_1\), are \(L_1(\eta_1) = N(-1.7, 769^2)\) and \(U_1(\eta_1) = (15.79)C(0, 419^2)\). Using the same idea, the lower and upper densities for \(\delta\) are \(L_2(\delta) = N(0, \delta^2)\) and \(U_2(\delta) = (2.29)C(0, 1.099^2)\). The actual prior classes we consider are

\[
\Gamma_B = \{π(\eta_1, \delta) : L_1(\eta_1)L_2(\delta) ≤ π(\eta_1, \delta) ≤ U_1(\eta_1)U_2(\delta)\}
\]

and

\[
\Gamma_{IB} = \{π(\eta_1, \delta) : \pi(\eta_1, \delta) = π_1(\eta_1)π_2(\delta), \quad L_1(\eta_1) ≤ π_1(\eta_1) ≤ U_1(\eta_1),
L_2(\delta) ≤ π_2(\delta) ≤ U_2(\delta)\}.
\]

Note that Kass and Greenhouse suggested that the independence assumption might be reasonable here.

The infima of the posterior probabilities of \(H_0 : \delta > 0\) for these classes are

\[
\inf_{\pi \in \Gamma_B} P^\pi(\delta > 0| \text{ Data } ) = 0.02159, \quad \inf_{\pi \in \Gamma_{IB}} P^\pi(\delta > 0| \text{ Data } ) = 0.6738.
\]

The suprema are both close to one. Again, the independence constraint seems to sharply reduce the posterior range. Note that the range without independence is uselessly large, in line with the conclusion obtained by Lavine, Wasserman, and Wolpert (1991) in their analysis without independence. The lower bound of 0.6738 under the independence assumption certainly suggests that the ECMO treatment is superior, but is not conclusive. Of course, we were still using a perhaps excessively large class of priors.

**Example 3.** Let \(X\) be a normal random variable with unknown mean \(\theta\) and unknown precision \(r\). The parameter of interest is \(\theta\) with \(r\) being a nuisance parameter.

For \(\theta\), a \(N(0, 1)\) prior was elicited, but the elicitor was known to typically give overly precise elicitations. Also, thick tails for \(π(\theta)\) were quite possible. These considerations led to choice of a flat upper bound, \(U_1(\theta) = (2π)^{-\frac{1}{2}}\), while only a modestly smaller lower bound, \(L_1(\theta) = (.9)N(0, 1)\), was deemed necessary. For \(r\), a Gamma (2,1) prior was elicited, but as above could well be too concentrated, so that the upper bound \(U_2(\gamma) = 1\) was employed. A lower bound, \(L_2(\gamma)\), of 2/3 times the Gamma (2,1) density was chosen,
there being somewhat more uncertainty for \( \pi_2(r) \) than for \( \pi_1(\theta) \). Finally, independence of \( \theta \) and \( r \) is assumed. Thus the elicited class of priors is

\[
\Gamma_{IB} = \{ \pi(\theta, r) : \pi(\theta, r) = \pi_1(\theta)\pi_2(r), \quad 2 \frac{r}{3} \exp(-r) \leq \pi_1(\theta) \leq 1, \\
(0.9)(2\pi)^{-\frac{1}{2}} \exp(-\theta^2/2) \leq \pi_2(r) \leq (2\pi)^{-\frac{1}{2}} \}. \]

Suppose the data yields \( \bar{x} = 0, s^2 = 2, n = 5 \). Then the sup and inf of the posterior mean for \( \theta \) is given by \( \sup_{\pi \in \Gamma_{IB}} E^\pi(\theta|\bar{x} = 0, s^2 = 2, n = 5) = 0.1244 \) and \( \inf_{\pi \in \Gamma_{IB}} E^\pi(\theta|\bar{x} = 0, s^2 = 2, n = 5) = -0.1872 \). If the prior independence of \( \theta \) and \( r \) is dropped and \( \Gamma_B \) is formed from the lower bound \( L_1(\theta)L_2(r) \) and upper bound \( U_1(\theta)U_2(r) \), then \( \sup_{\pi \in \Gamma_B} E^\pi(\theta|\bar{x} = 0, s^2 = 2, n = 5) = 0.533 \) and \( \inf_{\pi \in \Gamma_B} E^\pi(\theta|\bar{x} = 0, s^2 = 2, n = 5) = -0.533 \).

### 4.2 Comparing Two Parameters and Generalizations

Another situation in which a robust Bayesian analysis for \( \Gamma_{IB} \) can be carried out is the following. Suppose that interest is focused on a quantity \( h(\theta_1, \theta_2) \) satisfying:

**Condition 1.** \( h(\theta_1, \theta_2) \) is nonincreasing in \( \theta_1 \) and nondecreasing in \( \theta_2 \).

Note that any function \( g(\theta_1 - \theta_2) \), where \( g \) is a nonincreasing real function, satisfies Condition 1. An important example is \( h(\theta_1, \theta_2) = 1_{\{(\theta_1, \theta_2) : \theta_1 \leq \theta_2\}}(\theta_1, \theta_2) \), which arises in testing \( H_0 : \theta_1 \leq \theta_2 \). We also will need:

**Condition 2.** \( f(x|\theta_1, \theta_2) = f_1(x|\theta_1)f_2(x|\theta_2) \).

Then, for \( \pi(\theta_1, \theta_2) = \pi_1(\theta_1)\pi_2(\theta_2) \in \Gamma_{IB} \), it is clear that, under Condition 2,

\[
\pi(\theta_1, \theta_2|x) = \pi_1(\theta_1|x)\pi_2(\theta_2|x),
\]

while, under Condition 1 and for \( i \neq j \),

\[
\psi_i(\theta_i|x_j) = E^{\pi_i(\theta|\theta_j)}(h(\theta_1, \theta_2)|x)
\]

is nonincreasing for \( i = 1 \) and nondecreasing for \( i = 2 \). The following theorem shows that, under these conditions, obtaining the posterior range reduces to a two-dimensional optimization problem.

**Theorem 3.** Under Conditions 1 and 2,

(i) \( \sup_{\pi \in \Gamma_{IB}} E^\pi(h(\theta_1, \theta_2)|x) = E^{\pi^*(\theta_1, \theta_2)}(h(\theta_1, \theta_2)|x) \),

where \( \pi^*(\theta_1, \theta_2) = \pi_1^*(\theta_1)\pi_2^*(\theta_2) \) and \( \pi_1^* \) and \( \pi_2^* \) are of the form, for some \( k_1 \) and \( k_2 \),

\[
\pi_1^*(\theta_1) = U_1(\theta_1)1_{(\theta_1 \leq k_1)}(\theta_1) + L_1(\theta_1)1_{(\theta_1 > k_1)}(\theta_1), \quad (9)
\]
\[ \pi^*_2(\theta_2) = L_2(\theta_2)1(\theta_2 \leq k_2)(\theta_2) + U_2(\theta_2)1(\theta_2 > k_2)(\theta_2); \]  

(ii) \[ \inf_{\pi \in \Gamma_{IB}} E^\pi(h(\theta_1, \theta_2)|x) = E^{\pi^*}(h(\theta_1, \theta_2)|x), \]

where \( \pi^* \) is as in part 1, but with the \( U_i \) and \( L_i \) being interchanged.  

Proof. To prove part (i), let \( \bar{\pi} = \pi_1 \bar{\pi}_2 \) be a prior in \( \Gamma_{IB} \) that achieves the supremum (this is easily shown to exist), and observe that

\[ \sup_{\pi \in \Gamma_{IB}} E^\pi(h(\theta_1, \theta_2)|x) = \sup_{\pi_1 \in \Gamma_1} E^{\pi_1}(\psi_1(\theta_1|\bar{\pi}_2)|x), \]

where \( \Gamma_1 = \{\pi_1(\theta_1) : L_1(\theta_1) \leq \pi_1(\theta_1) \leq U_1(\theta_1)\} \). Since \( \psi_1(\theta_1|\bar{\pi}_2) \) is nonincreasing, DeRobertis and Hartigan (1981) can be applied to show that the right hand side of (11) is maximized at some \( \pi_1 \) of the form (9). An identical argument shows that the maximizing \( \pi_2 \) is of the form (10), completing the proof. The proof for part (ii) is identical. \( \square \)

Example 4. Suppose \( X = (X_1, X_2) \) is \( N_2((\theta_1, \theta_2), I) \), and that it is desired to test \( H_0 : \theta_1 \leq \theta_2 \). It is known that \( \theta_1 \) and \( \theta_2 \) are apriori independent, and \( N(0,1) \) prior distributions are thought to be reasonable for the \( \theta_i \). There is, however, considerable uncertainty concerning the \( N(0,1) \) assessment, and it is determined that this uncertainty is (conservatively) captured by the class \( \Gamma_{IB} \) with, for \( i = 1, 2 \),

\[(0.9)(2\pi)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \theta_i^2\right\} \leq \pi_i(\theta_i) \leq (2\pi)^{-\frac{1}{2}}. \]

For several observations \( x \), the range of the posterior probability of \( H_0 \), as \( \pi \) varies over \( \Gamma_{IB} \), is displayed in Table 2. The ranges are rather moderate in size. If the independence constraint is dropped, (i.e., an analogous class of the form \( \Gamma_B \) in (2) is considered), it can be shown that the infimum of the posterior probability of \( H_0 \) is close to zero, and the supremum is close to one. Thus, once again, we see that the adoption of prior independence has a dramatic effect.

| \( x = (x_1, x_2) \) | (inf \( \pi \in \Gamma_{IB} \) \( P^\pi(H_0|x) \), sup \( \pi \in \Gamma_{IB} \) \( P^\pi(H_0|x) \)) |
|-----------------------|--------------------------------------------------|
| (-0.2, 0.2)           | (0.43, 0.60)                                     |
| (-0.5, 0.5)           | (0.57, 0.73)                                     |
| (-0.8, 0.8)           | (0.70, 0.84)                                     |
| (-1.1, 1.1)           | (0.80, 0.91)                                     |
| (-1.7, 1.7)           | (0.94, 0.98)                                     |

Table 2. Range of Posterior Probabilities of \( H_0 \)
References


