Applying Non-parametric Robust Bayesian Analysis to Non-Opinionated Judicial Neutrality

Joseph B. Kadane* **Elias Moreno †** **Maria Eglee Perez‡** **Luis Raul Pericchi§**

Abstract

This paper explores the usefulness of robust Bayesian analysis in the context of an applied problem, finding priors to model judicial neutrality in an age discrimination case. We seek large classes of prior distributions without trivial bounds on the posterior probability of a key set, that is, without bounds that are independent of the data. Such an exploration shows qualitatively where the prior elicitation matters most, and quantitatively how sensitive the conclusions are to specified prior changes. The novel non-parametric classes proposed and studied here represent judicial netrality and are reasonably wide, so that when a clear conclusion emerges from the data at hand, this is arguably *very reliable.*

Keywords. Discrimination, elicitation, law, linearization, moment problem

1 Introduction

Robust Bayesian analysis (RBA), as championed by Berger and others [1] examines the maximum and minimum over a set of prior distributions, of a quantity of interest, like the posterior expectation of a function in the parameter space, for instance, the posterior probability of some set C. This paper explores what can be learned from such an analysis in the context of a specific example, the modeling of judicial neutrality in employment discrimination lawsuits.

What might be learned from a robust Bayesian analysis? We think there are at least two senses in which it can shed light on a standard Bayesian analysis that used a single, or a very few, prior distributions. The first is qualitative. If the class of prior distributions is unrestricted, for example if it contains both priors putting probability zero and probability one on the set C, the posterior probability will also be bounded by zero and one, independent of the data. This phenomenon, bounds attained whatever be the data, is referred in this paper as a set of priors leading to trivial bounds. At the other extreme, a class of prior distributions consisting of a single prior distribution is indistinguishable from an ordinary Bayesian analysis. However, if one can find a large class of prior distributions leading to non-trivial bounds, such an analysis can be informative about what aspects of the prior are particularly important to the determination of C's posterior probability. This qualitative information, in turn, can help direct attention in elicitation to the aspects of the prior that matter most for the application.

The second kind of information that might be gleaned from an RBA is quantitative, namely how much the probability of C changes as the class expands in a qualitatively sensitive direction. This in turn can lead to a judgment of whether a Bayesian analysis is sufficiently robust to changes in details of the prior specification to be relied upon in a specific application with specific data. Robust Bayesian analyses, with reasonable wide classes of priors ("reasonable" meaning that dogmatic priors are excluded), are particularly well suited to judicial weighting of evidence.

While there has been a rich theoretical development in RBA (which this paper uses to guide the calculations), applications have been fewer. One reason is that in settings involving several parameters, the extrema over large classes of prior distributions tend to be difficult to find (analytically or computationally), and the bounds tend to be extreme.

[9] discuss the famous ECMO trial ([16]) from a robust Bayesian viewpoint. They report having calculated with 84 different priors, spanning what they claim is

* L.J. Savage Professor of Statistics and Social Sciences at Carnegie Mellon University.

† Professor of Statistics at the University of Granada, Spain.

‡ Associate Professor of Statistics and Computational Mathematics, Simon Bolivar University, Caracas.

§ Professor of Statistics and Computational Mathematics, Simon Bolivar University, Caracas.

"a satisfactory range of appropriate skeptical opinions," and report reasonable robustness in stopping the trial.

[6] uses an ϵ-contamination model in which the class of priors considered is

$$\{(1-\epsilon)\pi_o + \epsilon q : q \epsilon \mathcal{Q}\}$$

where π_o is the tentatively believed prior, and $\mathcal{Q}$ is all possible prior distributions. As $\epsilon \downarrow 0$, the posterior probability of a set C approaches that under π_o; as $\epsilon \uparrow 1$, the bounds on a set C become the trivial bounds 0 and 1. They find that a cancer trial was appropriately stopped, but that the ECMO decision is more equivocal.

Finally,[7] considers a testing framework in which they want to discern which of four models is favored by the data. They choose a main prior, and consider four variants of it to show robustness.

A general and more radical approach that focus on upper and lower probabilities, is presented in [15].

This paper re-analyses the age-discrimination data of [8]. In that paper, four firing waves were studied using two-by-two doubly constrained contingency tables. Since the likelihood in this case has a single parameter with a natural interpretation, it is reasonable to hope that a broad class of priors might lead to reasonable bounds. This application has the difficulty, however, that the decision maker (judge or juror) is unavailable for prior elicitation. For this reason, it might be argued that RBA is unavoidable here. Hence we seek the prior of an idealized "judicially neutral" person.

Section 2 reviews the application and previous efforts at a robust analysis of it, which are still too narrow. Section 3 proposes some new classes of priors, and examines the conclusions that might be drawn from them. Section 4 concludes the paper with a discussion of the implications of our findings for the application, and for robust Bayesian analysis in general. The computational methods we used are described in the appendix.

2 Age Discrimination

[8] analyzed the data in Table 1 using a 2×2 contingency table. He took both the marginal distribution by age (over and under 40) and the marginal distribution by employment status (fired or retained) as fixed, since they are legally irrelevant. Thus both the age structure of the work force and the number fired are neutral happenstances, and the examination is based on who (by age) was chosen to be fired or retained.

The likelihood for a single firing wave is

$$f(n_{11} \mid L) = \frac{\binom{n}{n_{11}, n_{1+}-n_{11}, n_{+1}-n_{11}, n-n_{1+}-n_{+1}+n_{11}} e^{n_{11}L}}{\sum_{j=v}^{w} \binom{n}{j, n_{1+}-j, n_{+1}-j, n-n_{1+}-n_{+1}+j} e^{jL}}, \qquad (1)$$

where $v = \max(0, n_{1+}+n_{+1}-n)$, $w = \min(n_{1+}, n_{+1})$. The parameter L has the interpretation of being the log-odds-ratio, i.e. $L = \log(p_{11}p_{22}/p_{12}p_{21})$, the log-odds of being fired if an employee is over forty compared to one under forty. Therefore $L = 0$ means that the firing policy is age-neutral, $L > 0$ means that the firing policy disadvantages people older than 40, and conversely, $L < 0$ means that younger workers are disadvantaged. Since $L \leq 0$ is legal, while $L > 0$ is not, there are only two decisions available to the decision maker. Thus we concentrate on $P[L > 0]$. Notice, in passing, that one of the advantages of the Bayesian approach, preserved under the RBA, is its probability coherence. Thus the conclusions are invariant under smooth transformations of the parameter L, although this is a very natural parametrization of the likelihood.

To be useful in court, a prior for this problem needs to represent not the real prior opinion of an expert witness, but rather an opinion that is judicially neutral, i.e. that does not bias the analysis in either direction. This consideration led Kadane to propose symmetry for L around zero and unimodality as reasonable features for such a neutral prior to have.

For convenience, Kadane chose the normal family with zero mean, and calculated posteriors using standard deviations of 1, 2, 4, 8, and ∞. This is an example of parametric robust inference, in that the families of priors permitted are indexed by a parameter, here the standard deviation. Kadane put greatest stress on large standard deviations, infinity for waves II, III, and IV, and eight for wave I (in wave I, the zero in the table leads to an improper posterior under an unbounded uniform prior).

[13] noticed that the likelihood (1) belongs to the Exponential Family which can be written, in obvious notation as:

$$f(n_{11} \mid L) \propto \exp[n_{11}L - M(L)]g(\mathbf{n}),$$

so that a natural conjugate prior exists of the form,

$$\pi(L \mid n_0, x_0) \propto \exp[n_0 x_0 L - n_0 M(L)].$$

[13] explore this conjugate family, which has hiperparameters x_0 and n_0. They suggest using $n_0 = 1$

				Cell		
Notation:		1,1	1,2		2,1	2,2
Numbers		n_{11}	n_{12}		n_{21}	n_{22}
Probabilities		p_{11}	p_{12}		p_{21}	p_{22}
	I	18	0		129	102
Wave	II	26	10		105	83
	III	13	14		92	66
	IV	13	2		81	52
Category		Fired, 40+	Fired, 40-		Retained, 40+	Retained, 40-
			(from [8])			

Table 1: Ages of those fired and retained in four firing waves: Notation and Data

Prior			Wave		
	I	II	III	IV	Combined
$N(0,1)$	1.000	.960	.183	.965	.999
$N(0,2^2)$	1.000	.967	.169	.981	.999
$N(0,4^2)$	1.000	.969	.165	.985	.999
$N(0,8^2)$	1.000	.970	.164	.986	.999
$N(0,\infty)$	NA	.970	.164	.987	.999
C	.997	.906	.248	.930	.984
C_E	.997	.904	.246	.925	. 984
J	1.000	.968	.165	.980	.998
J_E	1.000	.969	.170	.984	.999
		(from [8] and [13])			

Table 2: Posterior Probability of $L > 0$ (older than 40 disadvantaged) as a function of the prior.

and $x_0 = n_{1+}n_{+1}/n$, denoted C here. This choice is nearly, but not exactly, symmetric around zero. Consequently they suggest the even part of the conjugate prior, which is symmetric by construction (C_E). Additionally, they report the consequences of the Jeffreys prior, and its even part (denoted here J and J_E respectively) shown in Perez (1994) to be proper. These results of Kadane and Perez and Pericchi are summarized in Table 2, which shows a broad, general consistency in the results.

None of these reflect a class of priors, so, even together, they do not address fully the issue of the robustness of the inference. To address this, Perez and Pericchi study two kinds of classes of prior.

The first, following [5], and [4], considers the class of all priors $\pi(L)$ satisfying

$$\frac{\pi(L)}{\xi(L)} \leq \lambda \quad \forall L, \tag{2}$$

where $\xi(L)$ is a fixed operational prior and $\lambda \leq 1$. Using as operational prior C, J, and N(0,1), they compute lower bounds for $P(L > 0)$ for several λ's.

The second class studied by Perez and Pericchi uses an idea introduced by [2]. They consider the class of unnormalized prior distributions $\pi(L)$ satisfying $\ell(L) \leq \pi(L) \leq u(L)$ for almost all L, where $\ell(L)$ and $u(L)$ are specified and need not integrate to unity. Using the choices $\ell(L) = N(0,\sigma^2)$, where σ^2 is chosen so that $\ell(0) = C(0)$ and $u(L) = \max_L (C_E(L)) * \max_L \left(\frac{\ell(L)}{C_E(L)}\right)$, they again derive lower bounds for the posterior probability that L is positive. These results are given in Table 3. Again, they indicate a certain qualitative robustness. But these classes do not articulate well with the idea that the class of priors should represent judicial neutrality in some sense.

3 Robust Neutrality

The most natural class, from the viewpoint of judicial neutrality, is the class of unimodal priors symmetric around 0. However, as the analysis below shows, this class turns out to be too broad.

According to the Khinchine representation theorem (see [3] p. 10, every prior in this class can be represented as an arbitrary mixture of uniform densities symmetric around zero. [8] remarks that a prior that puts all its weight uniformly on $[-\epsilon, \epsilon]$ will put posterior probability 1/2 on the set $L > 0$, (as $\epsilon \to 0$) regardless of what the data are. Denote this prior (really a limit of priors) as $U(0^+)$. Another important prior is $U(\infty-)$, found by taking a uniform prior on $[-m, m]$ as $m \to \infty$. Consider now the prior mixture of two uniforms:

$$\pi_p(L) = \begin{cases} p & U(0+) \\ 1-p & U(\infty-) \end{cases} \tag{3}$$

for some $p \,\epsilon\, (0,1]$. Also suppose there are no zeros in the data table (thus excluding Table I), which corresponds to $v < n_{11} < w$. Then as $\mid L \mid \to \infty$, it is easy

Operational Prior	λ	Wave I	II	III	IV	Combined
	1.5	1.000	.952	.116	.971	.997
	2	1.000	.937	.090	.961	.996
J	3	1.000	.909	.062	.944	.993
	4	1.000	.882	.047	.926	.991
	5	1.000	.857	.038	.909	.989
	1.5	1.000	.865	.128	.899	.976
	2	1.000	.828	.099	.896	.969
C	3	1.000	.763	.069	.816	.953
	4	1.000	.707	.052	.769	.939
	5	1.000	.658	.042	.939	.925
	1.5	1.000	.942	.132	.949	.998
	2	1.000	.924	.103	.933	.997
N(0,1)	3	1.000	.891	.071	.903	.996
	4	1.000	.859	.054	.875	.995
	5	1.000	.830	.044	.848	.993
DeRobertis/ Hartigan Class		.996	.893	.268	.916	.982

(From [13])

Table 3: Lower bound for the Posterior Probability that $L > 0$ as a function of the class of prior

to see that $f(n_{11} \mid L) \to 0$ exponentially fast. Now consider the posterior that results from the prior (3).

$$P\{L \leq 0 \mid n_{11}\} = \frac{(1/2)pf(n_{11} \mid L=0)}{pf(n_{11} \mid L=0) + (1-p)A}$$
$$\text{where } A = \lim_{m\to\infty} \int_{-m}^{m} \frac{1}{2m} f(n_{11} \mid L)dL \qquad (4)$$

But since $f(n_{11} \mid L) \to 0$ exponentially fast, the limit in the denominator of (4) is zero, so

$$P\{L \leq 0 \mid n_{11}\} = 1/2 \qquad (5)$$

for all members of the class (3), irrespective of the data n_{11}. We were surprised by this property of the class (3).

This fact already suggests at least one qualitative result: a subclass of all unimodal priors symmetric around zero can lead to a non-trivial bound only if it avoids putting too much probability close to zero, and avoids allowing too much probability to be put on extremely high and low values of L. This explains the failure of a number of our early attempts to restrict the class of unimodal, symmetric priors: (i) by fixing the height of π at 0,[1](ii) by bounding the variance of π from below, and (iii) by fixing a quantile of π. In each case, a member of (3) can be found to satisfy the constraint, thus showing that the class, even restricted, is uninterestingly broad, in the sense that 1/2 would be one of the possible values of $P\{L > 0\}$, regardless of the data. This demonstration that class (3) leads to trivial bounds cannot be taken as evidence that only class (3) leads to trivial bounds.

The priors $U(0+)$ and $U(\infty-)$ are not satisfactory representations of judicial neutrality. The former says essentially that the neutral arbitrator is sure, before hearing any evidence, that if there were age discrimination at all, its magnitude is so small as to have a negligible effect. The prior $U(\infty-)$ is also unreasonable, as it puts all its predictive weight on v and w. Thus a neutral arbitrator holding this prior is sure, before seeing the data, that all the firings will be of employees under forty or all will be of employees over forty. Not only is this not generally the case, but it is not a good model for a neutral arbitrator either.

[1]To see this, suppose that $\pi(0) \leq h$ for some fixed h. Consider a prior that puts $\pi(x) = h$ for $-\epsilon \leq x \leq \epsilon$, and puts $1 - h\epsilon$ probability on $U(\infty-)$. For each $\epsilon > 0$, such a prior has $P\{L \leq 0 \mid n_{11}\}$ nearly 1/2 independent of the data n_{11}. Hence as ϵ decreases, this continues to be the case.

Predictive distributions are a particularly useful way to think about the consequences of a prior distribution because it refers to what a neutral arbitrator might expect about the number of employees over forty fired, after learning the age structure of the work force and the number of employees to be fired. Berger (1994) interestingly points out that the predictive distribution is in fact the likelihood of the prior (for a fixed likelihood), and a limitation of some RBA studies is that robustness might be missing due to priors which have a very low (posterior) likelihood. In other words lack of robustness might be caused by priors which are ruled out by the data. Berger's insight seems to be consistent with the following principle in our situation: Neutrality might be considered in terms of not being too surprised at any way the data might come out. More formally, suppose that the prior is $\pi(L)$, and the likelihood is $f_i(L)$ where $i = n_{11}$ is the datum. Let

$$g_\pi(i) = \int_{-\infty}^{\infty} f_i(L)\pi(L)dL. \tag{6}$$

Then the neutral class A can be defined as

$$A = \{\pi(L) : \pi(L) \text{is unimodal and symmetric around } 0, \text{and } g_\pi(i) \geq \epsilon \text{ for all } i, v \leq i \leq w\} \tag{7}$$

The parameter ϵ of this class is then the minimum prior predictive probability of the possible data. The idea of this class is that it constrains the neutral arbitrator to have at least probability $\epsilon > 0$ on each possible data point. In other words, only priors which have a non-negligable likelihood, for all possible data, are allowed in our *neutrality class*. Every $\epsilon > 0$ prevents $U(\infty-)$ as a possible prior. To prevent $U(0+)$ from being a possible prior, it would be necessary to have

$$\epsilon > \min_i f(0 \mid i).$$

On the other hand, it is necessary to have $\epsilon < 1/(w - v + 1)$ in order for the class to be non-empty.

For each wave and for the combined waves as given in Table 1, there is a maximum value of epsilon, above which the class A is empty. For waves I to IV, and the combined waves, these maxima are .032, .016, .023, .041, and 8.57 10^{-32}, respectively. Thus any useful choice for epsilon must not exceed these numbers for the associated data set. It seems reasonable to allow epsilon to be one tenth or one hundredth of this maximum value, and to see the extent to which the posterior probability of positive L varies as a result.

Another base that might be used is the height that the predictive distribution would have if it were uniform, namely $1/(w-v+1)$ and then taking 1/10 or 1/100 of this base. For waves I to IV, and the combined waves, these heights are .053, .027, .036, .063, and 3.2×10^{-6}, respectively. For the individual waves, these numbers are larger than but the same order of magnitude as the maxima reported above. However, this idea is infeasible for the combined table, as it leads to an empty class.

Epsilon	Inf.	Sup
4 10^{-8}	.50	1.00
5 10^{-8}	.69	
10^{-7}	.86	
10^{-6}	.98	
10^{-5}	.997	

Wave I

Epsilon	Inf.	Sup
10^{-16}	.50	.97
5 10^{-16}	.67	
10^{-15}	.72	
10^{-14}	.83	
10^{-12}	.92	
10^{-8}	.97	

Wave II

Epsilon	Inf.	Sup
10^{-11}	.17	.50
5 10^{-11}		.45
10^{-10}		.32
10^{-8}		.20
10^{-6}		.17

Wave III

Epsilon	Inf.	Sup
10^{-8}	.50	.99
10^{-7}	.64	
10^{-6}	.86	
10^{-5}	.94	
.011	.98	

Wave IV

Epsilon	Inf.	Sup
10^{-42}	.50	.999
10^{-41}	.68	
10^{-40}	.81	
10^{-39}	.89	
10^{-37}	.96	
10^{-35}	.98	

Combined Waves

Table 4: Upper and lower bounds on the posterior probability of age discrimination for Class A as a function of ϵ.

Table 4 records the results for the data in Table 1. (The appendix to this paper shows how the computations were done). The computations in Table 4 show that the upper and lower bounds come together as the class of priors narrows by increasing epsilon, as expected. It also shows that each computation is unaffected by epsilons much smaller than one or two orders of magnitude below the maximum or the height the predictive distribution would have if it were uniform. Thus the calculations look quite insensitive to selection in class A with ϵ's chosen as suggested above.

s	Inf.	Sup
10^5	.99	1.00
10^6	.92	
$3\ 10^6$	.77	
$4.5\ 10^6$	.61	
$5\ 10^6$	.5	

Wave I

s	Inf.	Sup
10^7	.97	.97
10^{10}	.95	
10^{12}	.90	
10^{14}	.74	
10^{15}	.54	

Wave II

s	Inf.	Sup
10^6	.17	.18
10^8		.24
10^9		.31
10^{10}		.41
$3\ 10^{10}$		.50

Wave III

s	Inf.	Sup
10^3	.97	.99
10^5	.89	
10^6	.75	
$2\ 10^6$	.66	
$3.2\ 10^6$	.54	

Wave IV

s	Inf.	Sup
10^{24}	.998	.999
10^{30}	.992	
10^{34}	.957	
10^{36}	.886	
10^{37}	.810	
10^{38}	.680	
$3\ 10^{38}$	.569	
$4\ 10^{38}$	.500	

Combined Waves

Table 5: Upper and lower bounds on the posterior probability of age discrimination for Class B as a function of s.

It may be, however, that predictive probabilities themselves are hard to think about, because they depend so much on the margins taken to be fixed. This consideration led to the construction of a second class of prior distributions, class B. The idea here is to constrain the ratio of predictive probabilities, i.e. so that

$$\frac{\max_i g_\pi(i)}{\min_i g_\pi(i)} \leq s \tag{8}$$

for some $s > 1$ (among π's unimodal and symmetric around zero). Here there is a minimum value of s below which the class again goes empty. Respectively, those values are 2.94, 4.52, 3.40, 3.57 and 77.1. Here again it makes sense to allow variations of one or two orders of magnitude (factors of 10 to 100). Again, the appendix shows how the calculations were done. Table 5 records the results.

Again, variations of one or two orders of magnitude on the minimum s do not affect the results. Thus these calculations confirm the results of [8] and [13] that the calculations are robust. The classes considered are indeed very wide, allowing a pletora of different behaviour, but at the same time obeying the natural requirement of being neutral and non-dogmatic, in order to be fair in the final judgement, whatever the outcome of the data.

4 Conclusions

The Robust Bayesian analyses performed here lead to two kinds of conclusions. Qualitatively, they show that there are two sensitive areas of prior elicitation, close to $L = 0$, and L very large in absolute value. Neither of these comports well with the idea of judicial neutrality: not favoring either litigant and being open to being influenced by the data. Quantitatively, the analyses confirm the impression gained from the earlier studies, that for this particular application and data sets, the results are satisfactorily robust.

What have we learned from RBA to judicial weighting of the evidence? That a RBA is the natural implementation of the Bayesian approach to convey that a conclusion is very reliable.

On the other hand, what have we learned from this representative case study to RBA in general? That a RBA should take into account the likelihood of the priors in the class. This is encapsulated in the principle that classes that claim to model judicial neutrality, should yield non-negligible likelihoods of *any* possible outcome. It is our hope that this study will motivate similar principles in other realms of knowledge, and make closer to current judicial practice sensible Robust Bayesian analyses.

A Finding the Extrema for Classes A and B

The data may consist of a single table, or of several tables. In either case, the sample space is discrete. For a single table, it consists of integers i such that $v \leq i \leq w$. For several tables indexed by t, it consists of a vector of integers of length T, $\boldsymbol{i} = (i_1, \ldots, i_T)$ such that $v_t \leq i_t \leq w_t$ for all t, $1 \leq t \leq T$. Let N denote this sample space.

The likelihood function may then be written as $f_{\boldsymbol{i}}(L)$. The class A is then defined in equation (7), with the modification above in the case of more than one table. Using Khinchine's representation, the set of priors in A can be expressed as the set of distribution functions F satisfying

$$\pi(L)=\int_L^\infty \frac{1}{a}dF(a), \tag{9}$$

where F is a distribution function in the set $\mathcal{F} = \{F(\cdot) : \int_0^\infty dF(a) = \frac{1}{2}$,

$$\int_0^\infty f^*_{\boldsymbol{i}}(a)dF(a) \geq \epsilon > 0 \text{ for all } \boldsymbol{i}\ \epsilon\ N\} \tag{10}$$

where

$$f^*_{\boldsymbol{i}}(a)=\frac{1}{a}\int_0^a (f_{\boldsymbol{i}}(-L)+f_{\boldsymbol{i}}(L))dL\ ,\ a>0$$

and $f^*_i(0)$ is defined by continuity.

The quantity of interest is the probability of $\{L \geq 0\}$ posterior to observing $\boldsymbol{i}$ when the prior is $\pi(L)\epsilon A$. This can be rewritten as

$$P^\pi[L\geq 0 \mid \boldsymbol{i}] = \left[1+\frac{\int_{L\geq 0} f_{\boldsymbol{i}}(-L)\pi(L)dL}{\int_{L\geq 0} f_{\boldsymbol{i}}(L)\pi(L)dL}\right]^{-1} \tag{11}$$

or, equivalently as

$$P^F[L\geq 0 \mid \boldsymbol{i}] = \left[1+\frac{\int_0^\infty f^1_{\boldsymbol{i}}(a)dF(a)}{\int_0^\infty f^2_{\boldsymbol{i}}(a)dF(a)}\right]^{-1} \tag{12}$$

where

$$\begin{array}{rcl} f^1_{\boldsymbol{i}}(a) & = & \frac{1}{a}\int_0^a f_{\boldsymbol{i}}(-L)dL, \\ f^2_{\boldsymbol{i}}(a) & = & \frac{1}{a}\int_0^a f_{\boldsymbol{i}}(L)dL, \\ \text{and} & & F(\cdot)\ \epsilon\ \mathcal{F}. \end{array}$$

Then the supremum of the posterior probability that $L \geq 0$ can be written as

$$\begin{aligned} \sup_{\pi \epsilon A} p^\pi(L\geq 0 \mid \boldsymbol{i}) &= \sup_{F\epsilon\mathcal{F}} p^F(L\geq 0\mid \boldsymbol{i}) \\ &= \left[1+\inf_{F\epsilon\mathcal{F}} \frac{\int_0^\infty f^1_{\boldsymbol{i}}(a)dF(a)}{\int_0^\infty f^2_{\boldsymbol{i}}(a)dF(a)}\right]^{-1}. \end{aligned} \tag{13}$$

By the linearization algorithm ([11], the infimum in (13) is the unique solution in λ of the equation

$$\inf_{F\epsilon\mathcal{F}} \int_0^\infty \left[f^1_{\boldsymbol{i}}(a) - \lambda f^2_{\boldsymbol{i}}(a)\right] dF(a) = 0. \tag{14}$$

Once λ_0 has been found,

$$\sup_{\pi\epsilon A} P^\pi(L\geq 0 \mid \boldsymbol{i}) = (1+\lambda_0)^{-1}. \tag{15}$$

Using [10] (see also [14] and [12], (14) can be rewritten

$$0 = \sup_{\substack{d_{\boldsymbol{i}}\geq 0 \\ \boldsymbol{i}\epsilon N}} \left\{\epsilon \sum_{\boldsymbol{i}'\epsilon N} d_{\boldsymbol{i}'} + g\right. \tag{16}$$

where

$$g = \left. \inf_{F\epsilon\mathcal{F}_0} \int_0^\infty \left[f^1_{\boldsymbol{i}}(a) - \lambda f^2_{\boldsymbol{i}}(a) - \sum_{\boldsymbol{i}'\epsilon N} d_{\boldsymbol{i}'} f^*_{\boldsymbol{i}'}(a)\right] dF(a)\right\}$$

and where $\mathcal{F}_0$ is the class $\{F(\cdot) : \int_0^\infty dF(a) = 1/2\}$.

While (16) may look formidable, and hence not a simplification, it has important consequences. First, the internal infimum occurs at an F that puts all its probability at a single point a. This means that the extremum will occur at a single uniform distribution for L. Thus

$$\begin{aligned} &\inf_{F\epsilon\mathcal{F}_0} \int_0^\infty \left[f^1_{\boldsymbol{i}}(a) - \lambda f^2_{\boldsymbol{i}}(a) - \sum_{\boldsymbol{i}'\epsilon N} d_{\boldsymbol{i}'} f^*_{\boldsymbol{i}'}(a)\right] dF(a) \\ &\quad = \inf_{a\geq 0} \frac{1}{2}\left[f^1_{\boldsymbol{i}}(a) - \lambda f^2_{\boldsymbol{i}}(a) - \sum_{\boldsymbol{i}'\epsilon N} d_{\boldsymbol{i}'} f^*_{\boldsymbol{i}'}(a)\right], \end{aligned} \tag{17}$$

which permits reduction of (16) to

$$0 = \sup_{\substack{d_{\boldsymbol{i}}\leq 0 \\ \boldsymbol{i}\epsilon N}} \inf_{a\geq 0} \left\{f^1_{\boldsymbol{i}}(a) - \lambda f^2_{\boldsymbol{i}}(a) - \sum_{\boldsymbol{i}'\epsilon N} (f^*_{\boldsymbol{i}'}(a) - 2\epsilon) d_{\boldsymbol{i}'}\right\}. \tag{18}$$

From (18), if for any i, $f_i^*(a) < 2\epsilon$, allowing the corresponding d_i to go to infinity results in a sup of infinity, so (18) cannot be satisfied. Hence the supremum is attained when $d_i = 0$ for all i. Thus (18) further simplifies to finding a value λ_0 of λ such that

$$\inf_{a\geq 0}\left\{(f_{\boldsymbol{i}}^1(a) - \lambda f_{\boldsymbol{i}}^2(a))I_{\{a\geq 0: f_{\boldsymbol{i}}^*(a)\geq 2\epsilon, \forall \boldsymbol{i}\epsilon N}(a)\right\} = 0. \tag{19}$$

Now there are two cases to be considered separately. Since the supremum of $P^\pi(L \geq 0 \mid i)$ corresponds to a small value of λ, find the value of a for which $f_{\boldsymbol{i}}^1(a)/f_{\boldsymbol{i}}^2(a)$ is a minimum. If that value of a satisfies the constraint $\min_{\boldsymbol{i}} f_{\boldsymbol{i}}^*(a) > 2\epsilon$, then the supremum has been found. If not, then the constraint is binding. In this case, because $f_{\boldsymbol{i}}^1(a) - \lambda f_{\boldsymbol{i}}^2(a)$ is continuous in a, the infimum in (19) occurs when $f_{\boldsymbol{i}}^*(a) = 2\epsilon$ for some $\boldsymbol{i}\epsilon N$.

Thus the search for a solution of (19) in the second case can be found at the points a at which

$$\min_{\boldsymbol{i}\epsilon N} f_{\boldsymbol{i}}^*(a) = 2\epsilon, \tag{20}$$

and then

$$\lambda(a) = f_i^1(a)/f_i^2(a). \tag{21}$$

If there are several points a satisfying (20), the smallest $\lambda(a)$ in the set corresponds to the infimum in (13). This can be accomplished by a one-dimensional search over possible values a.

To find $\inf_{\pi\epsilon A} p^\pi(L \geq 0 \mid i)$, simply reverse the roles of inf and sup in (13). This can be done by reversing the roles of $f_i^1(a)$ and $f_i^{(2)}(a)$ in each of the subsequent formulas.

Finding the extrema for the class B is quite similar to finding them for class A. Here the constraints (10) are replaced by the constraints (8), which are equivalent to

$$sg_\pi(\boldsymbol{i}') \geq g_\pi(\boldsymbol{i}'') \text{ for all } \boldsymbol{i}', \boldsymbol{i}''\epsilon N. \tag{22}$$

Hence equation (10) is replaced by

$$\int_0^\infty f_{\boldsymbol{i}',\boldsymbol{i}''}^*(a)dF(a) \geq 0 \text{ for all } \boldsymbol{i}', \boldsymbol{i}''\epsilon N, \text{ where} \tag{23}$$

$$f_{\boldsymbol{i}',\boldsymbol{i}''}^*(a) = \frac{1}{a}\int_0^a s(f_{\boldsymbol{i}'}(L) + f_{\boldsymbol{i}'}(-L)) - (f_{\boldsymbol{i}''}(L) + f_{\boldsymbol{i}''}(-L))dL,$$

and $f_{\boldsymbol{i}',\boldsymbol{i}''}^*(0)$ is defined by continuity.

The same linearization can be applied, leading to the following analog of (16):

$$0 = \sup_{\substack{a_{\boldsymbol{i}',\boldsymbol{i}''}\geq 0 \\ \boldsymbol{i}',\boldsymbol{i}''\epsilon N}} \inf_{F\epsilon\mathcal{F}_0} \int_0^\infty h(a)dF(a) \tag{24}$$

where

$$h(a) = \left[f_{\boldsymbol{i}',\boldsymbol{i}''}^1(a) - \lambda f_{\boldsymbol{i}',\boldsymbol{i}''}^2(a) - \sum_{\boldsymbol{i}',\boldsymbol{i}''\epsilon N} d_{\boldsymbol{i}',\boldsymbol{i}''} f_{\boldsymbol{i}',\boldsymbol{i}'',s}^*(a)\right]$$

By exactly the same arguments, this results, in the case in which the constraints are binding, in finding the set of a's such that

$$\min_{\boldsymbol{i}',\boldsymbol{i}''} f_{\boldsymbol{i}',\boldsymbol{i}'',s}(a) = 0$$

and then choosing the smallest $\lambda_0(a)$ from the resulting set. Again, the infimum over the class B is found by reversing the roles of f^1 and f^2.

Acknowledgements

JBK's research was supported in part by U.S. National Science Foundation Grant DMS-9303557, EM's by Spanish Ministry of Education and Science Grant DIGYCIT, PB93-1154, and MEP and LRP's by GID-CONICIT.

References

[1] J. Berger. An overview of bayesian robustness. *TEST*, 3:1–125, 1994.

[2] L. DeRobertis and J. Hartigan. Bayesian inference using intervals of measures. *Annals of Statistics*, 9:235–244, 1981.

[3] S. Dharmadhikari and K. Joag-Dev. *Unimodality, Convexity and Applications.* Academic Press, Boston, 1988.

[4] J. Dickey. Approximate posterior distributions. *Journal of the American Statistical Association*, 71:680–9, 1976.

[5] W. Edwards, H. Lindman, and L.J. Savage. Bayesian statistical inference for psychological research. *Psychic Review*, (70):193–242, 1963.

[6] J. Greenhouse and L. Wasserman. Robust bayesian methods for monitoring clinical trials. *Statistics in Medicine*, 14:1379–1391, 1995.

[7] J. Greenhouse and L. Wasserman. A practical, robust method for bayesian model selection: A case study in the analysis of clinical trials. In Berger et al., editor, *Bayesian Robustness: Proceedings of the Workshop on Bayesian Robustness*, IMS Lecture Notes-Monograph, pages 41–62, 1996. (with discussion).

[8] J.B. Kadane. A statistical analysis of adverse impact of employer decisions. *Journal of the American Statistical Association*, (85):925–933, 1990.

[9] R. Kass and J. Greenhouse. Comment: A bayesian perspective. *Statistical Science*, (4):310–317, 1989.

[10] J. Kemperman. Geometry of the moment problem. In *Proceedings of Symposia in Applied Mathematics*, number 37, pages 16–53, 1987.

[11] M. Lavine, L. Wasserman, and R. Wolpert. Linearization of bayesian robustness problems. *Journal of Statistical Planning and Inference*, (37):307–316, 1993.

[12] B. Liseo, E. Moreno, and G. Salinetti. Bayesian robustness of the class with given marginals: An approach based on moment theory. In Berger et al., editor, *Bayesian Robustness: Proceedings of the Workshop on Bayesian Robustness*, volume 29 of *IMS Lecture Notes-Monograph*, pages 101–118, 1996. (with discussion).

[13] M. E. Perez and R. Pericchi. A case study on the bayesian analysis of 2×2 tables with all margins fixed. *Revista Brasileira de Probabilidatde e Estatistica*, (1):27–37, 1994.

[14] G. Salinetti. Discussion to Berger (1994). *TEST*, (3):1–125, 1994.

[15] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[16] J. Ware. Investigating therapies of potentially great benefit: Ecmo. *Statistical Science*, (4):298–340, 1989.