# ROBUST BAYESIAN ANALYSIS OF SELECTION MODELS

by

M. J. Bayarri and James Berger
Universitat de Valencia    Purdue University

Technical Report #93-6

Department of Statistics
Purdue University

August 1993

# ROBUST BAYESIAN ANALYSIS OF SELECTION MODELS

by

M. J. Bayarri and James Berger
Universitat de Valencia / Purdue University

**Abstract**

Selection models arise when the data is selected to enter the sample only if it occurs in a certain region of the sample space. When this selection occurs according to some probability distribution, the resulting model is often instead called a weighted distribution model. In either case the "original" density becomes multiplied by a "weight function" $w(x)$. Often there is considerable uncertainty concerning this weight function; for instance, it may only be known that $w$ lies between two specified weight functions. We consider robust Bayesian analysis for this situation, finding the range of posterior quantities of interest, such as the posterior mean or posterior probability of a set, as $w$ ranges over the class of weight functions. The variational analysis utilizes concepts from variation diminishing transformations.

## 1. INTRODUCTION

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

$$f_w(x|\theta) = \frac{w(x)f(x|\theta)}{\nu_w(\theta)}, \tag{1.1}$$

where $\nu_w(\theta) = E_\theta[w(X)]$. There is, actually, no reason to require $w(x)$ to be a probability; all we henceforth require is that $w$ be nonnegative and that $0 < E_\theta[w(X)] < \infty$ for all $\theta$. Then $w$ can be interpretated as a weight function that distorts (multiplies) the density $f(x|\theta)$ that observation $x$ gets selected. Selection models occur very often in practice (Patil and Rao, 1977; Rao, 1985; Bayarri and DeGroot, 1992).

---

Research supported by the National Science Foundation, Grants DMS-8923071 and DMS-9303556, and by the Universitat de Valencia, Ajuda de la Convocatoria de 1993 d'Estades en Centres d'Investigacio Nacionals y Estrangers.

Often the specification of $w(\cdot)$ is highly uncertain. It is thus of particular interest to study the robustness of the analysis to choice of $w$. We do so here using the global Bayesian robustness approach of considering a class, $\mathcal{W}$, of possible weight functions, and computing the range of posterior functionals of interest as $w$ ranges over $\mathcal{W}$.

Previous efforts in this direction for selection models have been informal and mainly confined to the study of parametric classes of weight functions, such as $w(x) = x^\alpha$ (when $X > 0$). There is rarely scientific justification for such specific parametric models; we will thus consider nonparametric classes of weight functions, such as

$$\mathcal{W}_1 = \{w\colon\ w_1(x) \le w(x) \le w_2(x)\},$$

$$\mathcal{W}_2 = \{\text{nondecreasing }\ w\colon\ w_1(x) \le w(x) \le w_2(x)\},$$

where $w_1(\cdot)$ and $w_2(\cdot)$ are nondecreasing. The upper and lower limits, $w_1$ and $w_2$, are to be chosen subjectively, representing the extremes of beliefs concerning $w$.

**Example 1.** Studies are reported in a journal only if $(i)$ the result is significant at the 0.05 level of significance (one-sided), or $(ii)$ it is significant at the 0.1 level and is deemed to be exceptionally "important" by the editors. In terms of, say, a standardized normal test statistic, $X$, we might conclude that $w \in \mathcal{W}_1$ (or $\mathcal{W}_2$) with $w_1(x) = 1_{(1.645,\ \infty)}(x)$ and $w_2(x) = 1_{(1.282,\ \infty)}(x)$, where "1" stands for the indicator function on the given set. The multiobservational version of this example can arise in meta-analysis.

The problem becomes particularly interesting in the multi-observational setting, because the effect of the weight function can then be extremely dramatic. Suppose $X_1, X_2, \ldots, X_n$ are *i.i.d.* from the density (1.1), so that the likelihood function for $\theta$ is

$$L_w(\theta) \propto l(\theta)[\nu_w(\theta)]^{-n}, \tag{1.2}$$

where $l(\theta) \propto \prod_{i=1}^{n} f(x_i|\theta)$ would be the likelihood function for the unweighted base density. If $\pi(\theta)$ is the prior density for $\theta$ (assumed to be w.r.t. Lebesgue measure) the posterior density is then

$$\pi^*(\theta) = \frac{l(\theta)[\nu_w(\theta)]^{-n}\pi(\theta)}{\int l(\theta)[\nu_w(\theta)]^{-n}\pi(\theta)d\theta}, \tag{1.3}$$

assuming $\pi$ is such that the denominator is finite. Expression (1.3) suggests that, at least for large $n$, the weight function $w$ can have a considerably greater effect on $\pi^*$ than might the prior $\pi$. Hence we will treat $\pi(\theta)$ as given here; for instance, it might be chosen to be a noninformative prior for the base model $f(x_i|\theta)$. Section 2 in the paper generalizes trivially to the scenario in which $\pi$ is also specified only up to a class $\Gamma$, but Section 3 is more difficult to generalize.

We are interested in posterior functionals of the form

$$H_\psi(\nu_w) = \int \psi(\theta)\pi^*(\theta)d\theta = \frac{\int \psi(\theta)l(\theta)[\nu_w(\theta)]^{-n}\pi(\theta)d\theta}{\int l(\theta)[\nu_w(\theta)]^{-n}\pi(\theta)d\theta}. \tag{1.4}$$

(We assume that these integrals exist for all $w$, guaranteed if they exist for $w_1$.) Typical $\psi$ of interest include $\psi(\theta) = \theta$ (yielding the posterior mean, $\mu$), $\psi(\theta) = (\theta - \mu)^2$ (yielding the posterior variance corresponding to $\mu$), and $\psi(\theta) = 1_C(\theta)$ (yielding the posterior probability of the set $C$). The sensitivity of $H_\psi(\nu_w)$ to $w$ will be determined by finding

$$\underline{H}_\psi = \inf_{w\in\mathcal{W}} H_\psi(\nu_w) \text{ and } \overline{H}_\psi = \sup_{w\in\mathcal{W}} H_\psi(\nu_w). \tag{1.5}$$

As usual in Bayesian robustness, if $(\underline{H}_\psi, \overline{H}_\psi)$ is a small enough interval then the effect of uncertainty in $w$ is minor but, if the interval is large, one cannot be assured of a robust conclusion, and must either collect more data or refine subjective opinion (about $w$ or $\pi$). For general discussion and references concerning this type of Bayesian robustness, see Berger (1990) and Wasserman (1992). Note that virtually the entire literature considers robustness w.r.t. the prior — not the likelihood, as here. (Lavine, 1991, is an exception.)

Section 2 exploits rather trivial inequalities to obtain a simple lower bound on $\underline{H}_\psi$ and upper bound on $\overline{H}_\psi$, using the technique of DeRobertis and Hartigan (1981). Unfortunately, these simple bounds are too disperse to be of much use (unless $n$ is quite small). Hence, in Section 3 we tackle the variational problem of finding $\underline{H}_\psi$ and $\overline{H}_\psi$ directly. Rather simple characterizations of the "extreme points" for these optimizations are possible when $\psi(\theta)$ is monotonic, unimodal, or bowl-shaped. The theory of variation diminishing transformations (cf., Brown, Johnstone, and MacGibbon, 1981) is used in this analysis.

## 2. EMPLOYING DEROBERTIS AND HARTIGAN BOUNDS

If $w \in \mathcal{W}_1$, then clearly

$$\nu_w(\theta) \in \Gamma_1 = \{\nu(\theta)\colon\ \nu_{w_1}(\theta) \le \nu(\theta) \le \nu_{w_2}(\theta)\}.$$

Also, if $w \in \mathcal{W}_2$ and $f(x|\theta)$ has decreasing monotone likelihood ratio in $\theta$ (i.e., $\theta_1 < \theta_2 \Rightarrow f(x|\theta_1)/f(x|\theta_2)$ is nonincreasing), then

$$\nu_w(\theta) \in \Gamma_2 = \{\text{ nondecreasing } \ \nu\colon\ \nu_{w_1}(\theta) \le \nu(\theta) \le \nu_{w_2}(\theta)\}.$$

(This follows from the MLR property; since $w \in \mathcal{W}_2$ is nondecreasing, so is $\nu_w(\theta)$.) Define, for $i = 1, 2$,

$$\underline{H}_\psi^* = \inf_{\nu \in \Gamma_i} H_\psi(\nu), \quad \overline{H}_\psi^* = \sup_{\nu \in \Gamma_i} H_\psi(\nu).$$

Since $w \in \mathcal{W}_i \Rightarrow \nu_w \in \Gamma_i$, it is clear that $\underline{H}_\psi^* \le \underline{H}_\psi$ and $\overline{H}_\psi \le \overline{H}_\psi^*$. Thus the bounds obtained by employing the $\Gamma_i$ are conservative, in that they contain the desired $\mathcal{W}_i$ bounds. The reason for considering the $\Gamma_i$ bounds is that they can be obtained from a relatively simple DeRobertis and Hartigan (1981) type of analysis.

For use in the following theorems, define

$$\Omega_a = \{\theta: \psi(\theta) \ge a\},$$

$$L(\theta) = l(\theta)[\nu_{w_2}(\theta)]^{-n}\pi(\theta), \quad U(\theta) = l(\theta)[\nu_{w_1}(\theta)]^{-n}\pi(\theta),$$

$$g_a(\theta) = \begin{cases} L(\theta) & \text{if } \theta \le a \\ U(\theta) & \text{if } \theta > a \end{cases}$$

$$h_a(\theta) = \begin{cases} U(\theta) & \text{if } \theta \le a \\ U(a) & \text{if } a \le \theta \le a^* \\ L(\theta) & \text{if } a^* \le \theta \end{cases},$$

where $a^* = \inf\{y \ge a: U(a) = L(y)\}$; note that $h_a$ is defined only for those values of $a$ for which $a^*$ is well-defined.

**Theorem 1.** *If* $\Gamma_1$ *is considered, then*

$$\underline{H}_\psi^* = \inf_a \frac{\int_{\Omega_a} \psi(\theta)L(\theta)d\theta + \int_{\Omega_a^c} \psi(\theta)U(\theta)d\theta}{\int_{\Omega_a} L(\theta)d\theta + \int_{\Omega_a^c} U(\theta)d\theta};$$

$$\overline{H}_\psi^* = \sup_a \frac{\int_{\Omega_a} \psi(\theta)U(\theta)d\theta + \int_{\Omega_a^c} \psi(\theta)L(\theta)d\theta}{\int_{\Omega_a} U(\theta)d\theta + \int_{\Omega_a^c} L(\theta)d\theta}.$$

**Proof.** This is essentially just Theorem 4.1 of DeRobertis and Hartigan (1981). □

**Theorem 2.** *If* $\psi(\theta)$ *is nondecreasing and* $\Gamma_2$ *is considered, then*

$$\underline{H}_\psi^* = \inf_a \ [\int \psi(\theta)h_a(\theta)d\theta \ / \int h_a(\theta)d\theta],$$

$$\overline{H}_\psi^* = \sup_a \ [\int \psi(\theta)g_a(\theta)d\theta \ / \int g_a(\theta)d\theta].$$

*If $\psi(\theta)$ is nonincreasing, these expressions hold with $h_a$ and $g_a$ reversed.*

**Proof.** This is essentially Theorem 2.3.1 of Bose (1990). □

An analogous result could be given for unimodal or bowl-shaped $\psi(\theta)$, but we defer such a result until Section 3 and determination of the more accurate $\underline{H}_\psi$ and $\overline{H}_\psi$.

**Example 2.** Suppose $f(x_i|\theta) = \theta^{-1}\exp\{-x_i/\theta\}$, where $x_i > 0$ and $\theta > 0$. Then $\ell(\theta) = \theta^{-n}\exp\{-n\overline{x}/\theta\}$. We will employ the usual noninformative prior, $\pi(\theta) = 1/\theta$. Consider $w_i(x) = 1_{(\tau_i,\infty)}(x)$, $\tau_2 < \tau_1$, as in Example 1. Then $\nu_{w_i}(\theta) = \exp\{-\tau_i/\theta\}$, so that $\Gamma_1$ and $\Gamma_2$ are quite simple.

Let $\psi(\theta) = \theta$, so that $H_\psi(\nu_w)$ is the posterior mean of $\theta$. We then have that $\Omega_a = [a,\infty)$, and

$$L(\theta) = \theta^{-(n-1)}e^{-n(\overline{x}-\tau_2)/\theta}, \quad U(\theta) = \theta^{-(n-1)}e^{-n(\overline{x}-\tau_1)/\theta}.$$

Theorem 1 can thus be used to numerically compute $\underline{H}^*_\psi$ and $\overline{H}^*_\psi$, the minimum and maximum of the posterior mean as $\nu_w$ ranges over $\Gamma_1$.

Similarly, for $\Gamma_2$ we can numerically compute $\underline{H}^*_\psi$ and $\overline{H}^*_\psi$ using Theorem 2. Note that the range of $a$ for which $h_a$ is defined can be shown to be $(0, a_0)$, where $a_0$ is the solution to

$$L(a_0)\exp\{n(\tau_1-\tau_2)/a_0\} = L(n[\overline{x}-\tau_2]/(n+1)).$$

Since $\psi(\theta) = \theta$ is increasing, it is easy to see that $\overline{H}^*_\psi$ is the same for $\Gamma_1$ and $\Gamma_2$, but that the lower bounds, $H^*_\psi$, differ. These bounds are all given in Figure 1, as a function of $d$, for the case $\tau_1 = 1 + d$, $\tau_2 = 1 - d$, and $n = 50$. The dashed lines are the upper and lower bounds corresponding to $\Gamma_1$, and the dotted line is the lower bound corresponding to $\Gamma_2$. Since the upper bound is unchanged for $\Gamma_2$, it is clear that imposing monotonicity on $w$ in $\Gamma_2$ provides only a slight improvement over the $\Gamma_1$ bounds.

It is of considerable interest to study the effect of the sample size, $n$. This is done in Table 1, for the case $\tau_1 = 0.8$ and $\tau_2 = 1.2$. The startling feature of the results is that the range of the posterior mean *increases* with $n$; thus larger sample sizes result in less robustness. This is a clear indication that replacing $\mathcal{W}$ by $\Gamma_1$ or $\Gamma_2$ and using the DeRobertis and Hartigan theory, is too crude; it appears to be necessary to work directly with the original $\mathcal{W}$.

For this type of situation, it will be shown in the next section that the exact bounds, $\underline{H}_\psi$ and $\overline{H}_\psi$, corresponding to $\mathcal{W}_2$, are the minimum and maximum of $H_\psi(\nu_{w_\tau})$, where

$w_\tau(x) = 1_{(\tau,\infty)}(x)$. It is straightforward to show, for such $w_\tau$, that the posterior distribution is *Inverse Gamma* $(n,\ [n(\overline{x}-\tau)]^{-1})$, so that the posterior mean is $H_\psi(\nu_{w_\tau}) = n(\overline{x}-\tau)/(n-1)$. It is then obvious that $\underline{H}_\psi = n(\overline{x}-\tau_1)/(n-1)$ and $\overline{H}_\psi = n(\overline{x}-\tau_2)/(n-1)$.

Besides being available in closed form, these exact bounds are considerably tighter than the $\underline{H}^*_\psi$ and $\overline{H}^*_\psi$. In Figure 1, the solid lines are the exact bounds. And, in Table 1, one sees that the range of the exact bounds decreases with $n$, as intuition would suggest. But note that this range decreases to the constant $(\tau_1 - \tau_2)$ so that, even for an arbitrarily large sample size, the uncertainty in the posterior mean is not completely resolved. This is the nature of selection models, and indicates why robustness studies are particularly important for their analysis.

Table 1. Ranges of the posterior mean over $\Gamma_1$, $\Gamma_2$, and $\mathcal{W}_2$ in the exponential example with sample size $n$.

| | | $n=5$ | $n=10$ | $n=50$ |
|---|---|---|---|---|
| $(\underline{H}^*_\psi,\ \overline{H}^*_\psi)$ | for $\Gamma_1$ | (0.813,1.711) | (0.610, 1.681) | (0.421, 1.976) |
| | for $\Gamma_2$ | (1.117, 1.711) | (0.864, 1.681) | (0.582, 1.976) |
| $(\underline{H}_\psi, \overline{H}_\psi)$ | for $\mathcal{W}_2$ | (1.0, 1.5) | (0.889, 1.333) | (0.816, 1.224) |
| $(\overline{H}^*_\psi - \underline{H}^*_\psi)$ | for $\Gamma_1$ | 0.898 | 1.071 | 1.555 |
| | for $\Gamma_2$ | 0.594 | 0.817 | 1.394 |
| $(\overline{H}_\psi - \underline{H}_\psi)$ | for $\mathcal{W}_2$ | 0.5 | 0.444 | 0.408 |

## 3. DETERMINING THE POSTERIOR BOUNDS

Example 2 in Section 2 demonstrated the need for exact calculation of $\underline{H}_\psi$ and $\overline{H}_\psi$ in (1.4). In this section, we indicate how this can be done when $\mathcal{W} = \mathcal{W}_2$. To keep the description of the solution manageable, we will only consider the case in which $f(x|\theta)$ is a continuous density with respect to Lebesgue measure with support $(r,s)$ ($r$ and $s$ could be infinite) and we will impose the weak condition on $\mathcal{W}_2$ that, for every $r < x < s$,

$$\lim_{\epsilon\to 0} w_1(x+\epsilon) \le \lim_{\epsilon\to 0} w_2(x-\epsilon). \tag{3.1}$$

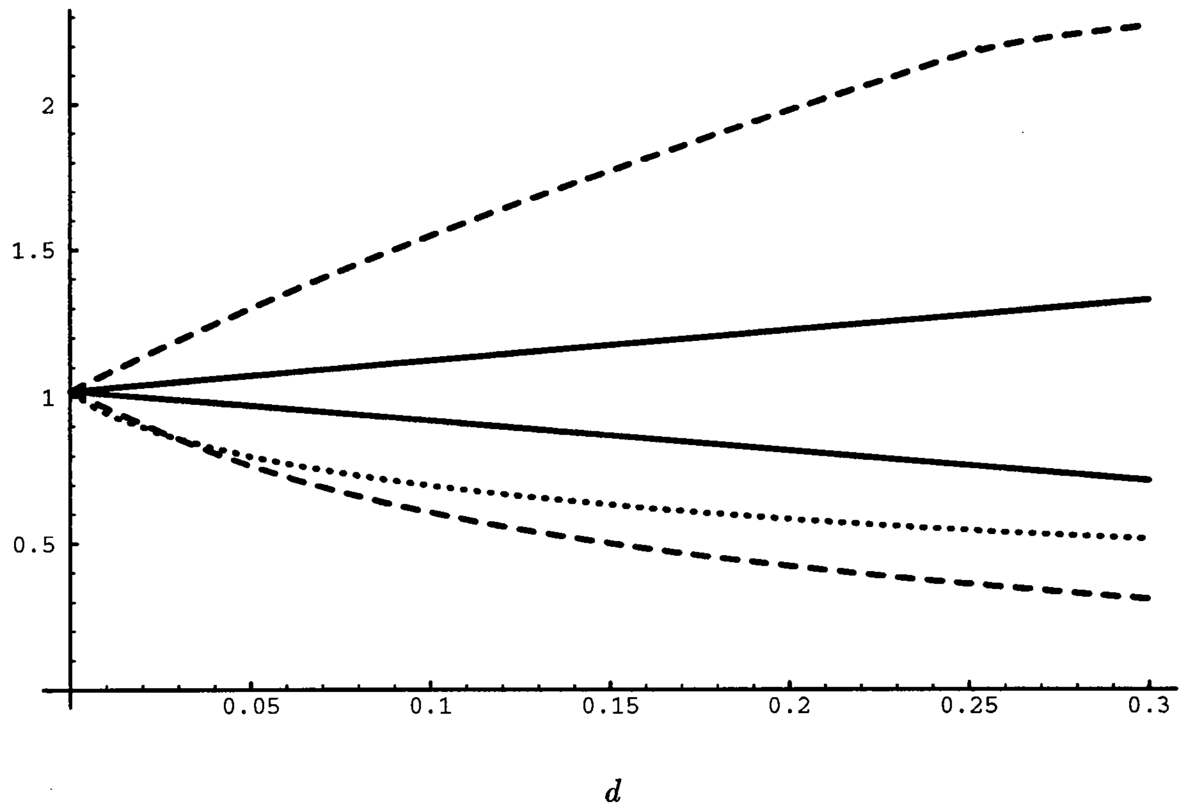


Figure 1. Ranges of the posterior mean over $\Gamma_1$ (dashed lines), $\Gamma_2$ (upper dashed and dotted lines), and $\mathcal{W}_2$ (solid lines) in the exponential example, when $\tau_1 = 1 + d$, $\tau_2 = 1 - d$, and $n = 50$.

This simply ensures that $w_1$ never jumps past where $w_2$ just was. We also assume that $w_1$ is not identically zero.

The following assumptions are needed in the optimization proof. The assumptions utilize the concept of variation diminishing transformations (cf., Brown, Johnstone, and MacGibbon, 1981). We will require that $f(x|\theta)$ be $\mathrm{SVR}_2$ or $\mathrm{SVR}_3$ (strictly variation reducing of order 2 or 3, respectively). Note that being $\mathrm{SVR}_2$ is equivalent to having strict monotone likelihood ratio (decreasing, by convention). Any distribution in the exponential family is $\mathrm{SVR}_3$ (indeed, is $SVR_\infty$); so is the noncentral $t$, noncentral $\chi^2$, noncentral $F$, and many others (see Karlin, 1968, 3.4).

**Assumption 1.** $\psi(\theta)$ and $f(x|\theta)$ satisfy either

(i) $\psi(\theta)$ is nondecreasing (to be denoted $\uparrow$) or nonincreasing ($\downarrow$), and $f(x|\theta)$ has strictly decreasing monotone likelihood ratio; or

(ii) $\psi(\theta)$ is nondecreasing for $\theta \leq \theta_0$ and nonincreasing for $\theta \geq \theta_0(\uparrow\downarrow)$, or $\psi(\theta)$ is nonincreasing for $\theta \leq \theta_0$ and nondecreasing for $\theta \geq \theta_0(\downarrow\uparrow)$, and $f(x|\theta)$ is $\mathrm{SVR}_3$.

**Assumption 2.** For all $r < x < s$

$$\int (1 + |\psi(\theta)|) l(\theta) [\nu_{w_1}(\theta)]^{-(n+1)} \pi(\theta) f(x|\theta) d\theta < \infty. \tag{3.2}$$

Note that (3.2) then also holds with $w_1$ replaced by any $w \in \mathcal{W}_2$.

It will be seen that $\overline{H}_\psi$ and $\underline{H}_\psi$ are achieved at a $w \in \mathcal{W}_2$ which has one of the four following forms. Define $h_1(c) = \inf\{x : w_1(x) \leq c\}$, and $h_2(c) = \sup\{x : w_2(x) \geq c\}$. Note that, at points of continuity of $w_i$, $h_i(c) = w_i^{-1}(c)$. Also, let $a \wedge b$ and $a \vee b$ denote the minimum and maximum, respectively, of $a$ and $b$.

**Solution Forms:**

$$\begin{aligned}
&\text{I.} \quad w(x) = \begin{cases} w_1(x) & \text{if } r < x \leq a \\ w_2(x) & \text{if } a < x < s \end{cases} \\
&\text{II.} \quad w(x) = \begin{cases} w_2(x) & \text{if } r \leq x < h_2(c) \\ c & \text{if } h_2(c) < x < h_1(c) \\ w_1(x) & \text{if } h_1(c) < x < s \end{cases} \\
&\text{III.} \quad w(x) = \begin{cases} w_1(x) & \text{if } r \leq x < a \\ w_2(x) & \text{if } a < x < a \vee h_2(c) \\ c & \text{if } a \vee h_2(c) < x < h_1(c) \\ w_1(x) & \text{if } h_1(c) < x < s \end{cases} \\
&\text{IV.} \quad w(x) = \begin{cases} w_2(x) & \text{if } r \leq x < h_2(c) \\ c & \text{if } h_2(c) < x < a \wedge h_1(c) \\ w_1(x) & \text{if } a \wedge h_1(c) < x < a \\ w_2(x) & \text{if } a < x < s. \end{cases}
\end{aligned} \tag{3.3}$$

It can be seen that I and II are both limiting cases of III and IV. This might be missed by an optimization program, however, so it is wise, when optimizing over classes III or IV, to also check classes I and II.

**Note:** When $w_1$ and $w_2$ are (nondecreasing) indicator functions, as in Examples 1 and 2, it is easy to see that the solution forms I through IV are themselves simply (nondecreasing) indicator functions.

**Theorem 3.** *$\underline{H}_\psi$ and $\overline{H}_\psi$ exist, and are attained, respectively, at some $\underline{w}$ and $\overline{w}$ in $\mathcal{W}_2$. If Assumptions 1 and 2 hold, then $\underline{w}$ and $\overline{w}$ can be chosen to be of the form indicated in* *the following table.*

Extreme Points for Optimization

| Shape of $\psi$ | $\uparrow$ | $\downarrow$ | $\uparrow\downarrow$ | $\downarrow\uparrow$ |
|---|---|---|---|---|
| Form of $\overline{w}$ | II | I | IV | III |
| Form of $\underline{w}$ | I | II | III | IV |

**Proof.** See the Appendix. □

In the following two examples, we illustrate application of Theorem 3 as well as the nature of solution forms I through IV.

**Example 3.** Consider the exponential scenario of Example 2, but now suppose that "size-biased" weights of the form $w(x) = x^\tau$ are under consideration. In particular, $\tau_1 = 0.8$ and $\tau_2 = 1.2$ are considered to be "extreme" weights, and it is decided to consider any nondecreasing weight function that lies between these extremes. The resulting class is clearly $\mathcal{W}_2$, with

$$w_1(x) = \begin{cases} x^{1.2} & \text{if } x \le 1 \\ x^{0.8} & \text{if } x > 1 \end{cases}; \quad w_2(x) = \begin{cases} x^{0.8} & \text{if } x \le 1 \\ x^{1.2} & \text{if } x > 1 \end{cases}.$$

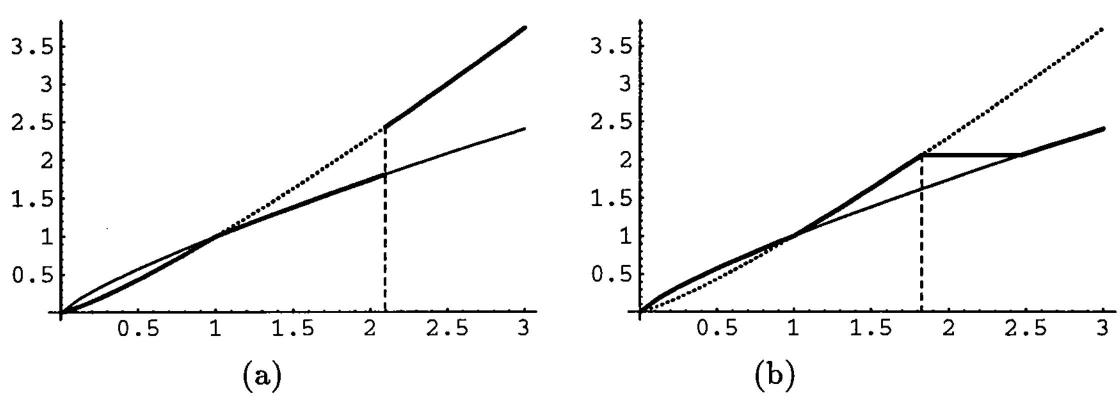


Figure 2. Graphs of $\underline{w}(x)$ and $\overline{w}(x)$ (dark lines in (a) and (b), respectively), together with $w_1(x)$ and $w_2(x)$ for Example 3.

If we are again interested in the posterior mean, so that $\psi(\theta) = \theta$ which is increasing, Theorem 3 states that $\underline{H}_\psi$ is achieved at $\underline{w}$ of form I while $\overline{H}_\psi$ is achieved at $\overline{w}$ of form II. Numerical computation for the situation $\overline{x} = 2$ and $n = 10$ shows that $\underline{w}$ is of form

I with $a = 2.097$, while $\overline{w}$ is of form II with $h_2(c) = 1.826$, $h_1(c) = (1.826)^{1.2/0.8}$, and $c = (1.826)^{1.2}$. These are graphed, as the dark lines, in Figures 2a and 2b, respectively; the lighter lines are $w_1$ and $w_2$; and the dashed lines are $a$ and $h_2(c)$, respectively. The corresponding bounds $(\underline{H}_\psi, \overline{H}_\psi)$ are (0.943, 1.189).

**Example 4.** Consider the exponential scenario of previous examples, but now suppose that "length-biased" weights of the form $w(x) = \tau x$ are considered. The "extremes" are thought to be $w_1(x) = (0.8)x$ and $w_2(x) = (1.2)x$, which we directly use to define $\mathcal{W}_2$.

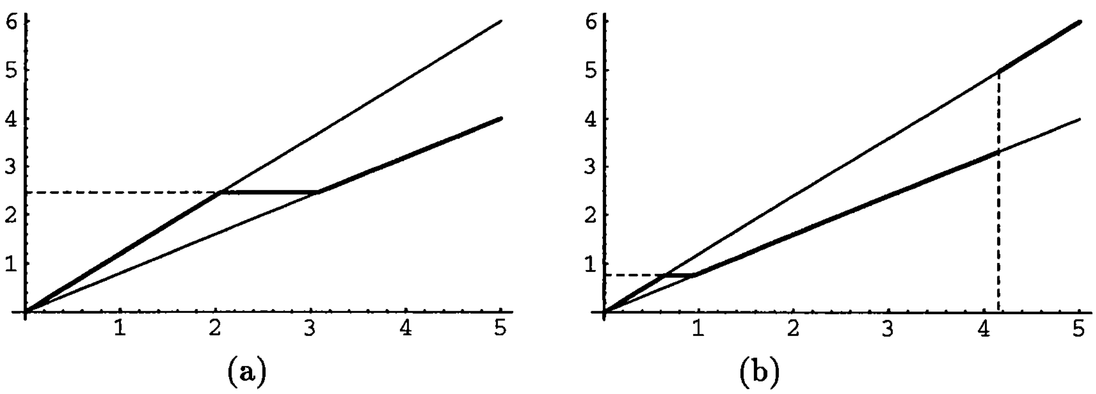


Figure 3. Graphs of $\underline{w}(x)$ and $\overline{w}(x)$ (dark lines in (a) and (b), respectively), together with $w_1(x)$ and $w_2(x)$ for Example 4.

If the "standard" length bias $w(x) = x$ were used, then the posterior distribution for $\theta$ would be *Inverse Gamma* $(2n, (n\overline{x})^{-1})$. For $n = 10$ and $\overline{x} = 2$, the posterior mean plus or minus one posterior standard deviation would be the interval $I = (0.805, 1.301)$, and $\Pr(\theta \in I|\text{data}) = 0.714$. We wish to study the robustness of this posterior coverage as $w$ varies over $\mathcal{W}_2$. To do so, we set $\psi(\theta) = 1_I(\theta)$, and apply Theorem 3. Note that $\psi(\cdot)$ is $\uparrow\downarrow$, so that Theorem 3 asserts that $\underline{w}$ is of form III and $\overline{w}$ is of form IV. Numerical computation reveals that $\underline{w}$ is of form III, but with $a = 0$ and $c = 2.458$; thus the solution is actually of form II, illustrating the need to consider limiting cases. The maximizer, $\overline{w}$, is of form IV with $a = 4.150$ and $c = 0.768$. Figures 3a and 3b graph $\underline{w}$ and $\overline{w}$, respectively (the dark lines); the lighter lines are $w_1$ and $w_2$; and the dashed lines are $c$ (Figure 3a) and $c$ and $a$ (Figure 3b). The corresponding bounds $(\underline{H}_\psi, \overline{H}_\psi)$ are (0.647, 0.755).

## Appendix: Proof of Theorem 3.

Because the numerator and denominator in (1.4) are bounded above and below by $w_1$ and $w_2$, it is straightforward to show that $\underline{H}_\psi$ and $\overline{H}_\psi$ exist. To show that $\underline{w}$ and $\overline{w}$ exist, note first that $\mathcal{W}_2$ is compact in the topology of pointwise convergence. Hence to prove existence of $\underline{w}$ and $\overline{w}$ we need only show that $H_\psi(\nu_w)$ is a continuous function of $w$ (under pointwise convergence), i.e., that

$$\lim_{i\to\infty} H_\psi(\nu_{w_{(i)}}) = H_\psi(\nu_{w^*}) \text{ if } \lim_{i\to\infty} w_{(i)}(x) = w^*(x). \tag{A1}$$

But since we require $\nu_{w_2}(\theta) < \infty$ and $w(\cdot) \geq 0$, the Lebesgue dominated convergence theorem yields

$$\lim_{i\to\infty} \nu_{w_{(i)}}(\theta) = \nu_{w^*}(\theta) \text{ if } \lim_{i\to\infty} w_{(i)}(x) = w^*(x),$$

from which it is clear that

$$\lim_{i\to\infty} [\nu_{w_{(i)}}(\theta)]^{-n} \to [\nu_{w^*}(\theta)]^{-n}$$

(using also $\nu_{w_1}(\theta) > 0$). But since $[\nu_{w_{(i)}}(\theta)]^{-n} \leq [\nu_{w_1}(\theta)]^{-n}$, and the integrals in (1.4) are assumed to exist for $w_1$, the dominated convergence theorem can again be applied to establish (A1).

We proceed with determination of the form of $\overline{w}$; the proof for $\underline{w}$ is similar. For a function $g(x)$, we will use the notation

$$g^+(x) = \lim_{\epsilon\to 0} g(x+\epsilon), \quad g^-(x) = \lim_{\epsilon\to 0} g(x-\varepsilon).$$

The first step of the proof is to apply the usual linearization argument (cf., Lavine, Wasserman and Wolpert, 1993), rewriting

$$\overline{H}_\psi = \sup_{w\in\mathcal{W}_2} H_\psi(w) = H_\psi(\overline{w})$$

as

$$0 = \sup_{w\in\mathcal{W}_2} G_\psi(w) = G_\psi(\overline{w}), \tag{A2}$$

where

$$G_\psi(w) = \int (\psi(\theta) - \overline{H}_\psi) l(\theta)[\nu_w(\theta)]^{-n}\pi(\theta)d\theta.$$

Define (existence guaranteed by Assumption 2)

$$\overline{\Lambda}(x) = \int (\psi(\theta) - \overline{H}_\psi) l(\theta)[\nu_{\overline{w}}(\theta)]^{-(n+1)}\pi(\theta) f(x|\theta)d\theta.$$

**Lemma 1.** *Let $I \subset (r,s)$ be an interval, and*

$$\tilde{w}(x) = \overline{w}(x) + \delta(x) 1_I(x),$$

*where $\delta(\cdot)$ is of constant sign (or zero) on $I$. Assume $\overline{\Lambda}(x)$ is of constant (nonzero) sign on $I$ and, for all $\theta \in \Theta$,*

$$|\int_I \delta(x) f(x|\theta)dx \;\; | \;\; /\nu_{\overline{w}}(\theta) \leq \epsilon. \tag{A3}$$

*Then,*

$$G_\psi(\tilde{w}) = -n(1+0(\epsilon)) \int_I \delta(x)\overline{\Lambda}(x)dx.$$

**Proof.** Clearly

$$\nu_{\tilde{w}}(\theta) = \nu_{\overline{w}}(\theta) + \int_I \delta(x) f(x|\theta)dx.$$

Hence, using (A3) and the constant sign of $\delta(\cdot)$ on $I$,

$$[\nu_{\tilde{w}}(\theta)]^{-n} = [\nu_{\overline{w}}(\theta)]^{-n}(1 - \frac{n(1+0(\epsilon))}{\nu_{\overline{w}}(\theta)} \int_I \delta(x) f(x|\theta)dx),$$

and

$$\begin{aligned} G_\psi(\tilde{w}) &= \int (\psi(\theta) - \overline{H}_\psi) l(\theta)[\nu_{\overline{w}}(\theta)]^{-n} \left(1 - \frac{n(1+0(\epsilon))}{\nu_{\overline{w}}(\theta)} \int_I \delta(x) f(x|\theta)dx\right) \pi(\theta)d\theta \\ &= G_\psi(\overline{w}) - n(1+0(\epsilon)) \int (\psi(\theta) - \overline{H}_\psi) l(\theta)[\nu_{\overline{w}}(\theta)]^{-(n+1)} \left(\int_I \delta(x) f(x|\theta)dx\right) \pi(\theta)d\theta \\ &= -n(1+0(\epsilon)) \int_I \delta(x)\overline{\Lambda}(x)dx, \end{aligned}$$

using (A2), Assumption 2 (to justify interchanging the order of integration), and the assumption that $\overline{\Lambda}(x)$ is nonzero and of constant sign on $I$. □

**Lemma 2.** *The number of sign changes of $\overline{\Lambda}(\cdot)$ (counting any zero as a sign change) is less than or equal to one when $\psi(\theta)$ is $\uparrow$ or is $\downarrow$, and is less than or equal to two when $\psi(\theta)$ is $\uparrow\downarrow$ or is $\downarrow\uparrow$. And, if the number of sign changes is the maximal one or two, respectively, then the sign changes must occur in the order given below:*

| Shape of $\psi$ | $\uparrow$ | $\downarrow$ | $\uparrow\downarrow$ | $\downarrow\uparrow$ |
|---|---|---|---|---|
| Sign Changes of $\overline{\Lambda}$ | $-$ to $+$ | $+$ to $-$ | $-$ to $+$ to $-$ | $+$ to $-$ to $+$ |

**Proof.** This is a direct consequence of the variation reducing properties of $f(x|\theta)$, specified in Assumption 1, when applied to

$$(\psi(\theta) - \overline{H}_\psi) l(\theta)[\nu_{\overline{w}}(\theta)]^{-(n+1)} \pi(\theta),$$

the sign changes of which are clearly determined by those of $(\psi(\theta) - \overline{H}_\psi)$, which occur as indicated in the lemma. $\square$

**Lemma 3.** *Let $x^*$ be a jump point of $\overline{w}$.*

*(i) Suppose $\epsilon > 0$ and $\delta > 0$ can be found for which $\overline{w}(x) > 0$ on $(x^* - \epsilon, x^*)$ and*

$$\tilde{w}_-(x) = \overline{w}(x) + \delta 1_{(x^*-\epsilon, x^*)}(x) \in \mathcal{W}_2. \tag{A4}$$

*Then $\overline{\Lambda}(x) > 0$ for $x \in (x^* - \epsilon, x^*)$ and small enough $\epsilon$.*

*(ii) Suppose $\epsilon > 0$ and $\delta > 0$ can be found for which*

$$\tilde{w}_+(x) = \overline{w}(x) - \delta 1_{(x^*, x^*+\epsilon)}(x) \in \mathcal{W}_2. \tag{A5}$$

*Then $\overline{\Lambda}(x) < 0$ for $x \in (x^*, x^* + \epsilon)$ and small enough $\epsilon$.*

**Proof.** Condition (A3) must first be verified. In case $(i)$, one can choose $\epsilon$ small enough so $\overline{w}(x^* - \epsilon) > 0$. Then

$$\frac{|\int_{x^*-\epsilon}^{x^*} \delta(x) f(x|\theta) dx|}{\nu_{\overline{w}}(\theta)} \leq \frac{\delta \int_{x^*-\epsilon}^{x^*} f(x|\theta) dx}{\overline{w}(x^* - \epsilon) \int_{x^*-\epsilon}^{x^*} f(x|\theta) dx} = \frac{\delta}{\overline{w}(x^* - \epsilon)}.$$

Since $\delta$ can be made arbitrarily small, (A3) is satisfied. Verification of (A3) for case $(ii)$ is similar.

Application of Lemma 1 thus yields

$$G_\psi(\tilde{w}_-) = -n(1 + 0(\delta))\delta \int_{x^*-\epsilon}^{x^*} \overline{\Lambda}(x) dx,$$
$$G_\psi(\tilde{w}_+) = n(1 + 0(\delta))\delta \int_{x^*}^{x^*+\epsilon} \overline{\Lambda}(x) dx.$$

For (A2) to be true, $G_\psi(\tilde{w}_-)$ and/or $G_\psi(\tilde{w}_+)$ must be less than or equal to zero. Since $\overline{\Lambda}(x)$ has at most 2 sign changes (including zeroes) and is continuous, $\epsilon$ can be chosen small enough so that $\overline{\Lambda}(x)$ is nonzero and has constant sign on $(x^* - \epsilon, x^*)$ or on $(x^*, x^* + \epsilon)$. The conclusions of the lemma are immediate. $\square$

**Lemma 4.** *If $\overline{\Lambda}(x) > 0$ for $r < x < x_0$, then $\overline{w}(x) = w_1(x)$ for $r < x < x_0$. If $\overline{\Lambda}(x) < 0$ for $x_0 < x < s$, then $\overline{w}(x) = w_2(x)$ for $x_0 < x < s$.*

**Proof.** To prove the first part, suppose not. Then let

$$\tilde{w}(x) = \begin{cases} \overline{w}(x) & \text{for } x_0 < x < s \\ \max\{w_1(x), (1-\epsilon)\overline{w}(x)\} & \text{for } r < x < x_0 \end{cases}$$
$$= \overline{w}(x) + \delta(x)1_{(r,x_0)}(x),$$

where

$$\delta(x) = -\min\{\overline{w}(x) - w_1(x),\ \epsilon\overline{w}(x)\}.$$

Note that $\delta(x) \leq 0$ on $(r, x_0)$ with positive measure (else $\overline{w}(x) = w_1(x)$). It is easy to see that $\tilde{w} \in \mathcal{W}_2$. Also,

$$\frac{|\int_r^{x_0} \delta(x)f(x|\theta)dx|}{\nu_{\overline{w}}(\theta)} \leq \frac{\varepsilon \int_r^{x_0} \overline{w}(x)f(x|\theta)dx}{\nu_{\overline{w}}(\theta)} \leq \epsilon,$$

so (A3) is satisfied. Lemma 1 then yields

$$G_\psi(\tilde{w}) = -n(1 + 0(\epsilon)) \int_r^{x_0} \delta(x)\overline{\Lambda}(x)dx.$$

Since $\overline{\Lambda}(x) > 0$, it follows that, for small enough $\epsilon$, $G_\psi(\tilde{w}) > 0$, which contradicts the maximality of $\overline{w}$.

To prove the second part of the lemma, define

$$\tilde{w}(x) = \overline{w}(x) + \delta(x)1_{(x_0,s)}(x),$$

where $\delta(x) = \min\{w_2(x) - \overline{w}(x), \epsilon\overline{w}(x)\}$. Again Lemma 1 applies, and yields a contradiction unless $\delta(x) = 0$ almost everywhere. Since $\overline{w}$ cannot be identically zero ($w_1$ isn't), this last can be true only if $\overline{w}(x) = 1_{(x_1,s)}(x)w_2(x)$ for some $x_0 < x_1 < s$. But $w_2(x) > \lambda > 0$ on a small enough interval $(x_1 - \varepsilon, x_1)$ and for small enough $\lambda$; else, $w_2(x) = 0$ for $x < x_1$ and $\overline{w} = w_2$ as claimed. Hence (A4) in Lemma 3 could be satisfied, so part ($i$) of that lemma would contradict the assumption that $\overline{\Lambda}(x) < 0$ for $x > x_0$. □

**Lemma 5.** *If $w_1(x) < \overline{w}(x) < w_2(x)$ for $a < x < b$, then $\overline{w}(x)$ is constant on $(a, b)$.*

**Proof.** First consider the case where $\overline{w}(x)$ is strictly increasing or strictly decreasing on some subinterval $(a', b')$, without jumps. Then there is a subinterval $[c, d]$ of $(a', b')$ in which, for some $\varepsilon > 0$, $w_1(x) + \varepsilon < \overline{w}(x) < w_2(x) - \epsilon$, and in which $\overline{\Lambda}(\cdot)$ is always positive or always negative (using, again, the fact that $\overline{\Lambda}(\cdot)$ has at most two sign changes, including zeroes). It follows directly, from Lemma 1, that such a $\overline{w}$ is not maximal; simply shift $\overline{w}$

on $[c, d]$ towards either $\overline{w}(c)$ or $\overline{w}(d)$ (depending on the sign of $\overline{\Lambda}(\cdot)$); the details of proof are similar to those in Lemma 3.

It follows that $\overline{w}$ must be a step function on $(a, b)$. Let $x^* \in (a, b)$ be a jump point. Note that $\overline{w}(x) < w_2(x) + \lambda$ for $x \in (x^* - \varepsilon, x^*)$ and small enough $\lambda > 0$ and $\epsilon > 0$; this follows from $\overline{w}$ being constant over a small enough interval, $w_2$ being nondecreasing, and $\overline{w} < w_2$ on $(a, b)$. It can be concluded from Lemma 3 $(i)$ that $\overline{\Lambda}(x) > 0$ on $(x^* - \varepsilon, x^*)$, for small enough $\epsilon$. A similar argument shows that $\overline{\Lambda}(x) < 0$ on $(x^*, x^* + \varepsilon)$, for small enough $\epsilon$. But then, either $\overline{\Lambda}(x) > 0$ for $x < x^*$ or $\overline{\Lambda}(x) < 0$ for $x > x^*$; if neither of these were true, there would have to be at least three sign changes of $\overline{\Lambda}(\cdot)$, which is not possible. But then Lemma 4 would yield that either $\overline{w}(x) = w_1(x)$ for $x < x^*$ or $\overline{w}(x) = w_2(x)$ for $x > x^*$, contradicting the assumptions here. Hence $\overline{w}$ cannot have a jump at $x^*$, and the lemma is proved. □

**Lemma 6.** *(i) If $\overline{\Lambda}(x) \leq 0$ for $r < x \leq x^*$ and $\overline{w}(x) > 0$, then there exists $r \leq x_0 \leq x^*$ such that $\overline{w}(x) = w_2(x)$ for $r \leq x < x_0$, and $\overline{w}(x) = c \geq w_1^+(x^*)$ for $x_0 < x < x^*$.*

*(ii) If $\overline{\Lambda}(x) \geq 0$ for $x^* \leq x < s$, then there exists $x^* \leq x_1 \leq s$ such that $\overline{w}(x) = w_1(x)$ for $x_1 < x < s$, and $\overline{w}(x) = c \leq w_2^-(x^*)$ for $x^* < x < x_1$.*

**Proof.** To prove part $(i)$, let $x_0 = \sup\{x \leq x^* : \overline{w}(x) = w_2(x)\}$, defined to be $x_0 = r$ if equality is never satisfied. Then $\overline{w}(x) < w_2(x)$ for $x_0 < x \leq x^*$. Also, Lemma 1 can be used to show that $\overline{w}(x) = w_2(x)$ for $r \leq x < x_0$ (the argument is similar to the second part of the proof of Lemma 4). Now, if $\overline{w}(x) < w_1^+(x^*)$ for a subinterval of $(x_0, x^*)$, Lemma 1 could again be used to contradict the maximality of $\overline{w}$; simply shift $\overline{w}$ over this range towards $\min\{w_2(x), w_1^+(x^*)\}$.

It remains to show that $\overline{w}$ is constant over $(x_0, x^*)$. To prove this, let $(x_0, x_2)$ be the (possibly empty) interval over which $\overline{w}(x) = w_1^+(x^*)$. (It is an interval because $\overline{w}$ is nondecreasing.) On the interval $(x_2, x^*)$ we have $w_1(x) \leq w_1^+(x^*) < \overline{w}(x) < w_2(x)$, so that Lemma 5 implies that $\overline{w}$ is a constant, $c$, over this interval. But this constant cannot differ from that over $(x_0, x_2)$; else, Lemma 1 could be used to contradict the maximality of $\overline{w}$, by shifting $\overline{w}$ on $(x_0, x_2)$ towards $\min\{w_2(x), c\}$. The proof of part $(ii)$ is similar. □

**Completion of Proof:** Assume first that $w_1(x) > 0$ for $x > r$, so that $\underline{w} > 0$ and $\overline{w} > 0$.

*Case 1:* $\psi(\cdot) \downarrow$

By Lemma 2, either $\overline{\Lambda}(\cdot)$ has no sign change, or it has one sign change going from $+$ to $-$. In the latter case, Lemma 4 applies, showing that $\overline{w}$ is of Form I from (3.3). If $\overline{\Lambda}(\cdot)$ has no sign change, Lemma 4 still applies with $x_0 = s$ or $x_0 = r$, with $\overline{w}$ being $w_1$ or $w_2$, respectively. Since these are special cases of Form I, the proof is complete.

*Case 2:* $\psi(\cdot) \uparrow$

By Lemma 2, either $\overline{\Lambda}(\cdot)$ has no sign change or $\overline{\Lambda}(\cdot)$ has one sign change going from $-$ to $+$, with the change occurring at, say, $x^*$. In the latter case, Lemma 6 asserts that there exist $r \leq x_0 \leq x^*$ and $x^* \leq x_1 \leq s$ such that $\overline{w}(x) = w_2(x)$ for $r < x < x_0$ and $\overline{w}(x) = w_1(x)$ for $x_1 < x < s$, with $\overline{w}(x)$ being constant over the intervals $(x_0, x^*)$ and $(x^*, x_1)$. It is easy to see that these constants must be the same; again apply Lemma 1, and use the assumption (3.1). This form of solution can be written as Form II from (3.3).

If $\overline{\Lambda}(\cdot)$ has no sign changes, the argument proceeds similarly to Case I, yielding the conclusion that $\overline{w}$ is either $w_1$ or $w_2$. Either of these is also a limiting case of Form II, completing the proof.

*Case 3.* $\psi(\cdot) \downarrow\uparrow$

Now $\overline{\Lambda}(\cdot)$ has at most two sign changes. The cases of zero and one sign change are handled as before. If there are two sign changes, Lemma 2 shows that the order must be $+$ to $-$ to $+$. Letting $a$ be the location of the first sign change, Lemma 4 shows that $\overline{w}(x) = w_1(x)$ for $x < a$. But the problem for $x > a$ is then identical to the Case 2 problem. Combining the solutions over these domains with the constraint that $\overline{w}(\cdot)$ be nondecreasing results in a solution of Form III of (3.3).

*Case 4.* $\psi(\cdot) \downarrow\uparrow$

The only new case is that in which $\overline{\Lambda}(\cdot)$ has two sign changes from $-$ to $+$ to $-$. Letting $a$ be the point of the second sign change, Lemma 4 shows that $\overline{w}(x) = w_2(x)$ for $x > a$. The problem for $x < a$ is then a Case 2 problem. Combining the solutions over these domains with the constraint that $\overline{w}(\cdot)$ be nondecreasing results in a solution of Form IV of (3.3).

To deal with the case $w_1(x) = 0$ on $(r, c)$, define a sequence $w_{1,i}$ of strictly increasing functions such that $w_{1,i}(x) \to w_1(x)$ pointwise and $w_{1,i}(x) = w_1(x)$ on $(c', s)$. Let $\underline{w}_i$ and $\overline{w}_i$ be the minimizers and maximizers of $H_\psi(\nu_w)$ over the $\mathcal{W}_{2,i} = \{\text{nondecreasing } w : w_{1,i}(x) \leq w(x) \leq w_2(x)\}$; these solutions exist by the above results.

From the compactness of $\mathcal{W}_{2,i}$, it follows that $\{\underline{w}_i\}$ and $\{\overline{w}_i\}$ have subsequences $\{\underline{w}_{i_n}\}$

and $\{\overline{w}_{i_n}\}$ with limits $\underline{w}^*$ and $\overline{w}^*$, respectively. It is straightforward to verify that the class of functions of form II or the class of form IV in (3.3) are closed under pointwise limits. Hence, if the $\{\underline{w}_{i_n}\}$ (or $\{\overline{w}_{i_n}\}$) are of form II or IV (the only cases we need worry about), then so is $\underline{w}^*$ (or $\overline{w}^*$).

Next, define $H_\psi^i$ as in (1.4) with $\mathcal{W} = \mathcal{W}_{2,i}$. An argument similar to that proving (A1) yields

$$\lim_{n\to\infty} H_\psi^{i_n}(\nu_{\underline{w}_{i_n}}) = H_\psi(\nu_{\underline{w}^*}), \quad \lim_{n\to\infty} H_\psi^{i_n}(\nu_{\overline{w}_{i_n}}) = H_\psi(\nu_{\overline{w}^*}). \tag{A6}$$

Finally, since

$$H_\psi^{i_n}(\nu_{\underline{w}_{i_n}}) \le H_\psi^{i_n}(\nu_{\underline{w}}), \quad H_\psi^{i_n}(\nu_{\overline{w}_{i_n}}) \ge H_\psi^{i_n}(\nu_{\overline{w}}),$$

it can easily be shown from (A6) that

$$H_\psi(\nu_{\underline{w}^*}) \le H_\psi(\nu_{\underline{w}}), \quad H_\psi(\nu_{\overline{w}^*}) \ge H_\psi(\nu_{\overline{w}}).$$

Hence $\underline{w}^*$ is a minimizer of $H_\psi(\nu_w)$ and $\overline{w}^*$ is a maximizer, proving that solutions of the form II and/or IV exist. (Note that we did not rule out the existence of solutions with an initial zero segment; we just showed that a solution of form II or IV always exists.)

## REFERENCES


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

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

Bose, S. (1990). *Bayesian Robustness With Shape — Constrained Priors and Mixture Priors.* Ph.D. Thesis, Purdue University, West Lafayette.

Brown, L. D., Johnstone, I. M., and MacGibbon, K. B. (1981). Variation diminishing transformations: a direct approach to total positivity and its statistical applications. *J. Amer. Statist. Assoc.* **76**, 824–832.

DeRobertis, L. and Hartigan, J. (1981). Bayesian inference using intervals of measures. *Ann. Statist.* **9**, 235–244.

Karlin, S. (1968). *Total Positivity*. Stanford University Press, Stanford.

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

Lavine, M., Wasserman, L. and Wolpert, R. (1993). Linearization of Bayesian robustness problems. *J. Statist. Planning and Inference*, To appear.

Patil, G. P., and Rao, C. R. (1977). The weighted distributions: a survey of their applications. In *Applications of Statistics*, P. R. Krishnaiah (Ed.), 383–405. North-Holland, Amsterdam.

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

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