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
Revised December 1996

◇

ROBUST BAYESIAN ANALYSIS
OF SELECTION MODELS

by

M. J. Bayarri and James Berger
Universitat de Valencia — Duke University and 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 $X \in \mathbb{R}^1$ is distributed over some population of interest according to the density (with respect to Lebesgue measure) $f(x|\theta)$, $\theta \in \Theta =$ some interval (possibly infinite) in $\mathbb{R}^1$, but that, when $X = x$, the probability of recording x (or the probability that x is *selected* to enter the sample) is $w(x)$. Then the true density of an actual observation is

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

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 θ. 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 Generalitat Valenciana, under Grant GV-1081/93. We are grateful to a referee for valuable suggestions and the Associate Editor for suggesting a considerably more transparent and powerful line of proof.

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

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

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

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.

From (1.1), note that multiplying w by a constant has no effect on the density. Hence, for $i = 1$ or 2, $\mathcal{W}_i$ could be replaced by $\mathcal{W}_i{}^* = \{kw : k > 0 \text{ and } w \in \mathcal{W}_i\}$, without affecting the conclusions. Interestingly, $\mathcal{W}_1^*$ can be rewritten as

$$\mathcal{W}_1^* = \{w\ :\ w(x)/w(y) \leq w_2(x)/w_1(y)\},$$

which is the class considered in DeRobertis and Hartigan (1981); $\mathcal{W}_2^*$ can be similarly rewritten. Some may find it more natural to elicit w_1 and w_2 by considering $\mathcal{W}_1^*$ or $\mathcal{W}_2^*$ but, again, the answers will not change once w_1 and w_2 have been elicited.

The robust Bayesian 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 θ 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 θ (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 π 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 π^* than might the prior π. Hence we will treat $\pi(\theta)$ as given here; for instance, it might be chosen to be a noninformative prior for the base model $f(x_i|\theta)$. Section 2 in the paper generalizes trivially to the scenario in which π is also specified only up to a class Γ, 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 ψ of interest include $\psi(\theta) = \theta$ (yielding the posterior mean, μ), $\psi(\theta) = (\theta - \mu)^2$ (yielding the posterior variance corresponding to μ), 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, for $i = 1, 2$,

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

As usual in Bayesian robustness, if $(\underline{H}_\psi^{(i)}, \overline{H}_\psi^{(i)})$ 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 π). For general discussion and references concerning this type of Bayesian robustness, see Berger (1990, 1994) and Wasserman (1992). Note that virtually the entire literature considers robustness w.r.t. the prior — not the likelihood, as here. (Lavine, 1991, and Shyamalkumar, 1996, are exceptions.)

Section 2 exploits rather trivial inequalities to obtain a simple lower bound on $\underline{H}_\psi^{(i)}$ and upper bound on $\overline{H}_\psi^{(i)}$, using the technique of DeRobertis and Hartigan (1981). Unfortunately, these simple bounds yield too broad a range to be of much use (unless n is quite small). Hence, in Section 3 we tackle the variational problem of finding $\underline{H}_\psi^{(i)}$ and $\overline{H}_\psi^{(i)}$ 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

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

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

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

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

$$\underset{\sim}{H}_\psi^{(i)} = \inf_{\nu \in \Gamma_i} H_\psi(\nu), \quad \tilde{H}_\psi^{(i)} = \sup_{\nu \in \Gamma_i} H_\psi(\nu).$$

Since $w \in \mathcal{W}_i \Rightarrow \nu_w \in \Gamma_i$, it is clear that $\underset{\sim}{H}_\psi^{(i)} \le \underline{H}_\psi^{(i)}$ and $\overline{H}_\psi^{(i)} \le \tilde{H}_\psi^{(i)}$. Thus the bounds obtained by employing the Γ_i are conservative, in that they contain the desired $\mathcal{W}_i$ bounds. The reason for considering the Γ_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 $a^* = a^*(a) \equiv \inf\{y \ge a{:}\, U(a) = L(y)\}$; note that h_a is defined only for those values of a for which a^* is well-defined.

Theorem 1. *If* Γ_1 *is considered, then*

$$\underset{\sim}{H}_\psi^{(1)} = \inf_a \frac{\int_{\Omega_a} \psi(\theta)L(\theta)d\theta + \int_{\Omega_a^c} \psi(\theta)U(\theta)d\theta}{\int_{\Omega_a} L(\theta)d\theta + \int_{\Omega_a^c} U(\theta)d\theta};$$

$$\underset{\sim}{H}_{\psi}^{(1)} = \sup_{a} \frac{\int_{\Omega_a} \psi(\theta)U(\theta)d\theta + \int_{\Omega_a^c} \psi(\theta)L(\theta)d\theta}{\int_{\Omega_a} U(\theta)d\theta + \int_{\Omega_a^c} L(\theta)d\theta}.$$

Proof. This is essentially just Theorem 4.1 of DeRobertis and Hartigan (1981). $\square$

Theorem 2. *If $\psi(\theta)$ is nondecreasing and Γ_2 is considered, then*

$$\underset{\sim}{H}_{\psi}^{(2)} = \inf_{a} \ [\int \psi(\theta)h_a(\theta)d\theta \ / \int h_a(\theta)d\theta],$$

$$\tilde{H}_{\psi}^{(2)} = \sup_{a} \ [\int \psi(\theta)g_a(\theta)d\theta \ / \int g_a(\theta)d\theta].$$

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). $\square$

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 $\underline{H}_{\psi}^{(i)}$ and $\overline{H}_{\psi}^{(i)}$.

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 Γ_1 and Γ_2 are quite simple. In all following illustrations we assume $\overline{x} = 2$.

Let $\psi(\theta) = \theta$, so that $H_{\psi}(\nu_w)$ is the posterior mean of θ. 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 $\underset{\sim}{H}_{\psi}^{(1)}$ and $\tilde{H}_{\psi}^{(1)}$, the minimum and maximum of the posterior mean as ν_w ranges over Γ_1.

Similarly, for Γ_2 we can numerically compute $\underset{\sim}{H}_{\psi}^{(2)}$ and $\tilde{H}_{\psi}^{(2)}$ 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 $\tilde{H}_{\psi}^{(i)}$ is the same for Γ_1 and Γ_2, but that the lower bounds, $\underset{\sim}{H}_{\psi}^{(i)}$, 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 Γ_1, and the dotted line is the lower bound corresponding to Γ_2. Since the upper bound is unchanged for Γ_2, it is clear that imposing monotonicity on w in Γ_2 provides only a slight improvement over the Γ_1 bounds.

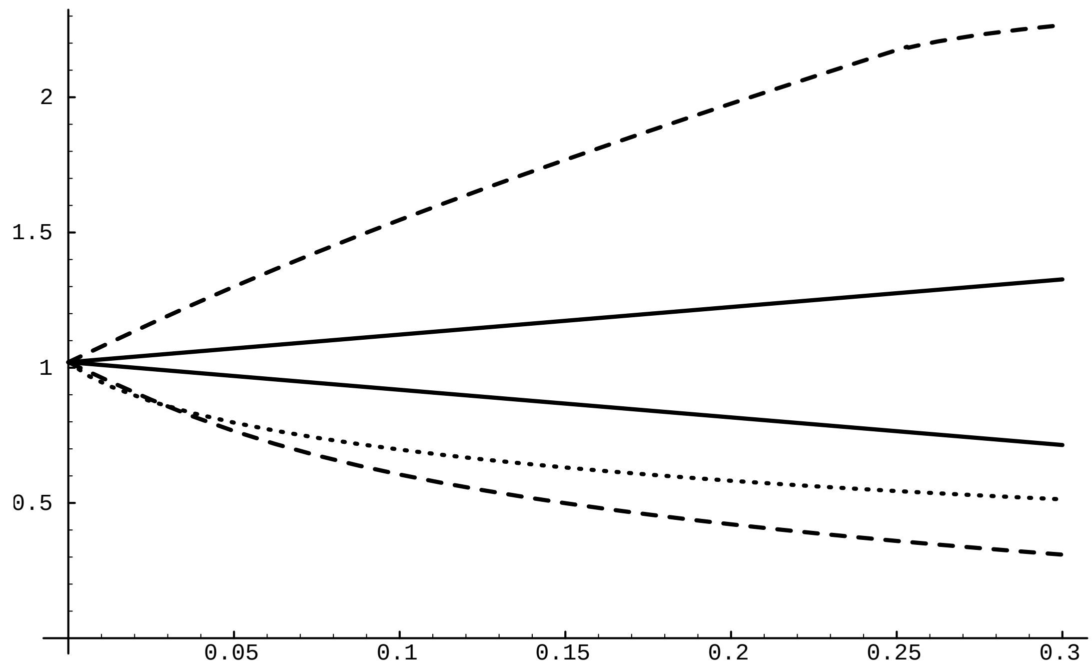
It is of considerable interest to study the effect of the sample size, n. This is done in Table 1, for the case $\tau_1 = 0.8$ and $\tau_2 = 1.2$. The startling feature of the results is that the range of the posterior mean *increases* with n; thus larger sample sizes result in less robustness. This is a clear indication that replacing $\mathcal{W}$ by Γ_1 or Γ_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, $\underline{H}_\psi^{(2)}$ and $\overline{H}_\psi^{(2)}$, 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_τ, 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 $\underline{H}_\psi^{(2)} = n(\overline{x}-\tau_1)/(n-1)$ and $\overline{H}_\psi^{(2)} = n(\overline{x}-\tau_2)/(n-1)$.

Besides being available in closed form, these exact bounds are considerably tighter than the $\underset{\sim}{H}_\psi^{(i)}$ and $\tilde{H}_\psi^{(i)}$. 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 Γ_1, Γ_2, and $\mathcal{W}_2$ in the exponential example with sample size n.

	$n = 5$	$n = 10$	$n = 50$
$(\underset{\sim}{H}_\psi^{(1)}, \tilde{H}_\psi^{(1)})$	(0.813,1.711)	(0.610, 1.681)	(0.421, 1.976)
$(\underset{\sim}{H}_\psi^{(2)}, \tilde{H}_\psi^{(2)})$	(1.117, 1.711)	(0.864, 1.681)	(0.582, 1.976)
$(\underline{H}_\psi^{(2)}, \overline{H}_\psi^{(2)})$	(1.0, 1.5)	(0.889, 1.333)	(0.816, 1.224)



d

Figure 1. Ranges of the posterior mean over Γ_1 (dashed lines), Γ_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 $n = 50$.

3. DETERMINING THE POSTERIOR BOUNDS

Example 2 in Section 2 demonstrated the need for exact calculation of $\underline{H}_{\psi}^{(i)}$ and $\overline{H}_{\psi}^{(i)}$ in (1.4). In this section, we indicate how this can be done.

The following assumptions will be used in the optimization proof. The first assumption utilizes the concept of variation diminishing transformations (see Brown, Johnstone, and MacGibbon, 1981, for precise definitions and discussion). We will require that $f(x|\theta)$ be SVR_2 or SVR_3 (strictly variation reducing of order 2 or 3, respectively). Note that being SVR_2 is equivalent to having strict monotone likelihood ratio (decreasing, by convention). SVR_3 means that, if $g(x)$ has at most two sign changes (ignoring zeros), then $h(\theta) = E_\theta[g(X)]$ has no more sign changes (counting zeros) than $g(x)$ does (ignoring zeros); furthermore, if these numbers of sign changes of g and h are equal, then the changes occur in the same order. Any distribution in the exponential family is SVR_3 (indeed, is SVR_∞); so is the noncentral t, noncentral χ^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 SVR_3.

Assumption 2. Let $f(x|\theta)$ be a density with respect to Lebesgue measure on the interval (r, s), where r and s could be infinite. For all $r < x < s$ and some small $\epsilon > 0$, assume that

$$\int (1 + |\psi(\theta)|) l(\theta) [\nu_{w_1}(\theta)]^{-(n+1)} \pi(\theta) f(x|\theta) d\theta < \infty, \tag{3.1}$$

$$\int (1 + |\psi(\theta)|) l(\theta) [\nu_{w_1}(\theta)]^{-(n+1+\epsilon)} [\nu_{w_2}(\theta)]^{(1+\epsilon)} \pi(\theta) d\theta < \infty. \tag{3.2}$$

Note that (3.1) and (3.2) then hold with w_1 and w_2 replaced by any $w \in \mathcal{W}_i$.

It will be seen that $\overline{H}_\psi^{(i)}$ and $\underline{H}_\psi^{(i)}$ are achieved at a w which has one of the four following forms. Define $h_1(c) = \inf\{x: w_1(x) \geq c\}$, and $h_2(c) = \sup\{x: w_2(x) \leq c\}$. 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:

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

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. We also define three limiting cases, II* , III*, and IV*, as the versions of II, III, and IV, with $h_1(c) = h_2(c)$ (i.e., the "uniform at c" piece is absent).

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. *$\underline{H}_\psi^{(i)}$ and $\overline{H}_\psi^{(i)}$ exist, and are attained, respectively, at some $\underline{w}$ and $\overline{w}$ in $\mathcal{W}_i$. If Assumptions 1 and 2 hold and $\mathcal{W} = \mathcal{W}_2$, then $\underline{w}$ and $\overline{w}$ can be chosen to be of the form indicated in the following table. If, instead, $\mathcal{W} = \mathcal{W}_1$, the solution forms are as in the table but with* II, III, *and* IV *replaced by* II*, III*, and IV*.

Extreme Points for Optimization

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

Proof. See the Appendix, which also contains more general results. □

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

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

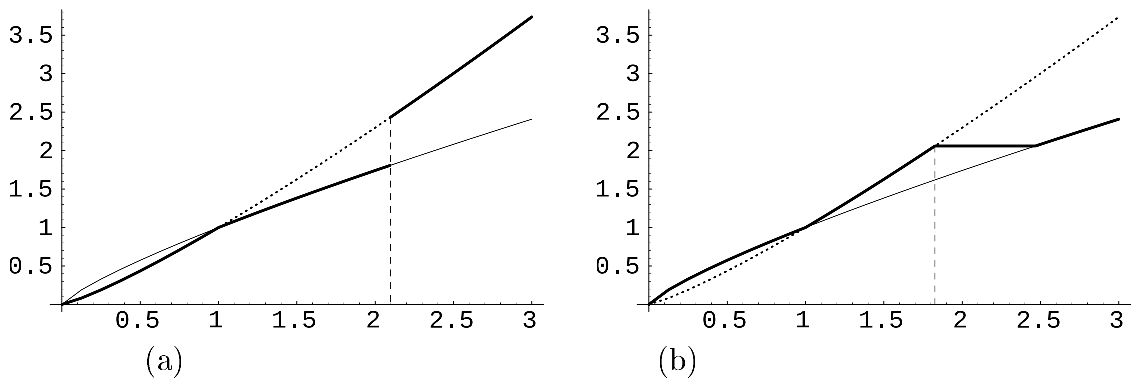


Figure 2. Graphs of $\underline{w}(x)$ and $\overline{w}(x)$ (dark lines in (a) and (b), respectively), together with $w_1(x)$ and $w_2(x)$ for Example 3.

If we are again interested in the posterior mean, so that $\psi(\theta) = \theta$ which is increasing, Theorem 3 states that $\underline{H}_\psi^{(2)}$ is achieved at $\underline{w}$ of form I while $\overline{H}_\psi^{(2)}$ 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 vertical lines mark a and $h_2(c)$, respectively. The corresponding bounds $(\underline{H}_\psi^{(2)}, \overline{H}_\psi^{(2)})$ 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 $\mathcal{W}_2$.

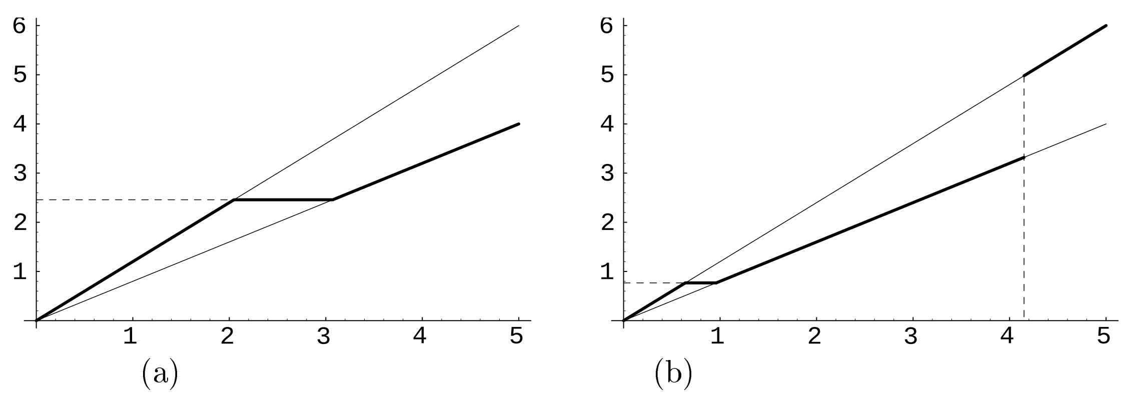


Figure 3. Graphs of $\underline{w}(x)$ and $\overline{w}(x)$ (dark lines in (a) and (b), respectively), together with $w_1(x)$ and $w_2(x)$ for Example 4.

If the "standard" length bias $w(x) = x$ were used, then the posterior distribution for θ 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 mark c (Figure 3a) and c and a (Figure 3b). The corresponding bounds $(\underline{H}_\psi^{(2)}, \overline{H}_\psi^{(2)})$ are (0.647, 0.755).

Appendix: Existence and Forms of $\overline{w}$ and $\underline{w}$

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 $\underline{H}_\psi^{(i)}$ and $\overline{H}_\psi^{(i)}$ exist. To show that $\underline{w}$ and $\overline{w}$ exist, note first that $\mathcal{W}_i$ 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 begin the determination of the extreme points with some needed definitions. Much of the following applies to quite general classes, $\mathcal{W}$, so define $\underline{H}_\psi$ and $\overline{H}_\psi$ as in (1.5) but for arbitrary $\mathcal{W}$. Next define (existence guaranteed by (3.1))

$$\overline{\Lambda}(x) = \int (\psi(\theta) - \overline{H}_\psi) l(\theta) [\nu_{\overline{w}}(\theta)]^{-(n+1)} \pi(\theta) f(x|\theta) d\theta, \tag{A2}$$

$$\overline{A}^{\,+} = \{x: \overline{\Lambda}(x) > 0\} \text{ and } \overline{A}^{\,-} = \{x: \overline{\Lambda}(x) < 0\}. \tag{A3}$$

Similarly, define $\underline{\Lambda}(x)$, $\underline{A}^+$, and $\underline{A}^-$ by replacing $\overline{H}_\psi$ and $\overline{w}$ with $\underline{H}_\psi$ and $\underline{w}$ in (A2). We will use the notation

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

Finally, we need the following definition.

Definition. For every open set $A \subseteq (r,s)$ and $w_0 \in \mathcal{W}$, define the weight functions

$$w^*(x|w_0, A) = \sup\ \{w(x): w \in \mathcal{W} \text{ and } w(x) = w_0(x)\ \forall x \in A^c\}, \tag{A4}$$

$$w_*(x|w_0, A) = \inf\ \{w(x): w \in \mathcal{W} \text{ and } w(x) = w_0(x)\ \forall x \in A^c\}.$$

The family $\mathcal{W}$ will be said to be *closed under conditional supremum and infimum operations* if $w^*(x|w_0, A) \in \mathcal{W}$ and $w_*(x|w_0, A) \in \mathcal{W}$ for any w_0 and A.

Note: It is easy to see that $\mathcal{W}_1$ and $\mathcal{W}_2$ are closed under conditional supremum and infimum operations.

Linearization: The first step in establishing the form of $\overline{w}$ and $\underline{w}$ is to apply the usual linearization argument (cf., Lavine, Wasserman, and Wolpert, 1993) rewriting, for $\overline{w}$ say,

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

as

$$0 = \sup_{w\in\mathcal{W}}\ \overline{G}_\psi(w) = \overline{G}_\psi(\overline{w}), \tag{A5}$$

where

$$\overline{G}_\psi(w) \ = \ \int (\psi(\theta) - \overline{H}_\psi) l(\theta)[\nu_w(\theta)]^{-n}\pi(\theta)d\theta. \tag{A6}$$

For $\underline{w}$, the analogous expressions hold with $\overline{H}_\psi$, sup, and $\overline{w}$ replaced by $\underline{H}_\psi$, inf, and $\underline{w}$; we denote the resulting variant of (A6) by $\underline{G}_\psi(w)$. In dealing with (A5), the following lemma will be needed.

Lemma 1. *Suppose that $\mathcal{W}$ is convex and $w \in \mathcal{W}$ satisfies $w_1(x) \le w(x) \le w_2(x)$, where w_1 and w_2 satisfy Assumption 2. Then*

$$n\int \overline{\Lambda}(x)[\overline{w}(x) - w(x)]dx = \lim_{t\to 0+} t^{-1}\overline{G}_\psi((1-t)\overline{w} + tw) \le 0. \tag{A7}$$

An analogous expression holds for $\underline{\Lambda}(x)$, $\underline{w}(x)$, and $\underline{G}_\psi$, but with the inequality reversed.

Proof. Note that $(1-t)\overline{w}+tw \in \mathcal{W}$ by convexity. The inequality in (A7) follows directly from (A5). To prove the first equality in (A7), let $\tilde{w}_t = (1-t)\overline{w}+tw$ and note that $\overline{G}_\psi(\overline{w})=0$ implies that

$$\overline{G}_\psi(\tilde{w}_t) = \int (\psi(\theta)-\overline{H}_\psi) l(\theta)[\nu_{\overline{w}}(\theta)]^{-n}\left(\left[\frac{\nu_{\overline{w}}(\theta)}{\nu_{\tilde{w}_t}(\theta)}\right]^n - 1\right)\pi(\theta)d\theta.$$

Break this integral up into integrals over

$$\Omega_t = \{\theta\colon \nu_w(\theta)/\nu_{\overline{w}}(\theta) \le t^{-(1-\epsilon/2)}\}$$

and Ω_t^c, where ϵ is from Assumption 2.

Dealing first with the integral over Ω_t^c, a Chebyshev argument yields

$$\begin{aligned}
&t^{-1}\int_{\Omega_t^c} |\psi(\theta)-\overline{H}_\psi \mid l(\theta)\ (\ [\nu_{\tilde{w}_t}(\theta)]^{-n} - \ [\nu_{\overline{w}}(\theta)]^{-n})\pi\ (\theta)d\theta \\
&\le t^{-1}t^{(1-\epsilon/2)(1+\epsilon)}\int |\psi(\theta)-\overline{H}_\psi|\ l(\theta)\ ([\nu_{\tilde{w}_t}(\theta)]^{-n} + [\nu_{\overline{w}}(\theta)]^{-n})(\nu_w(\theta)/\nu_{\overline{w}}(\theta))^{1+\epsilon}\pi\ (\theta)d\theta.
\end{aligned}$$

This last integral is clearly bounded by a multiple of the finite (3.2), and (since ϵ is small) the leading factor goes to zero as $t\to 0$.

For $\theta\in\Omega_t$, note that

$$\begin{aligned}
[\nu_{\overline{w}}(\theta)/\nu_{\tilde{w}_t}(\theta)]^n - 1 &= [1-t(1-\nu_w(\theta)/\nu_{\overline{w}}(\theta))]^{-n} - 1 \\
&= nt(1-\nu_w(\theta)/\nu_{\overline{w}}(\theta)) + O(t^{2-(1-\epsilon/2)^2}),
\end{aligned}$$

where $O(.)$ is uniform in θ. Thus, using also (3.2) and the dominated convergence theorem,

$$\begin{aligned}
&\lim_{t\to 0+} t^{-1}\int_{\Omega_t} (\psi(\theta)-\overline{H}_\psi) l(\theta)[\nu_{\overline{w}}(\theta)]^{-n}\left(\left[\frac{\nu_{\overline{w}}(\theta)}{\nu_{\tilde{w}_t}(\theta)}\right]^n - 1\right)\pi(\theta)d\theta \\
&= \lim_{t\to 0+} n\int_{\Omega_t} (\psi(\theta)-\overline{H}_\psi) l(\theta)[\nu_{\overline{w}}(\theta)]^{-n}\left(1-\frac{\nu_w(\theta)}{\nu_{\overline{w}}(\theta)} + O(t^{(1+\epsilon-\epsilon^2/4)})\right)\pi(\theta)d\theta \\
&= n\int (\psi(\theta)-\overline{H}_\psi) l(\theta)[\nu_{\overline{w}}(\theta)]^{-(n+1)}(\nu_{\overline{w}}(\theta)-\nu_w(\theta))\pi(\theta)d\theta.
\end{aligned}$$

Recalling the definitions of $\nu_{\overline{w}}(\theta)$ and $\nu_w(\theta)$ and using (3.2) and the dominated convergence theorem to reverse orders of integration, the result follows. The proof of the analogous result for $\underline{w}$ is similar. $\square$

We are now ready to prove two theorems, which can be considered to be generalizations of Theorem 3. Because these theorems are considerably harder to apply than Theorem 3, we leave them in the Appendix.

Theorem 4. *If $\mathcal{W} = \mathcal{W}_1$ and Assumption 2 holds, then $\overline{w} = w_1$ on $\overline{A}^{\,+}$ and $\overline{w} = w_2$ on $\overline{A}^{\,-}$. Also, $\underline{w} = w_2$ on $\underline{A}^+$ and $\underline{w} = w_1$ on $\underline{A}^-$.*

Proof. We only prove the result for $\overline{w}$. Since $\mathcal{W}_1$ is closed under conditional supremum and infimum operations, $w_*(x|\overline{w}, \overline{A}^{\,+})$ and $w^*(x|\overline{w}, \overline{A}^{\,-})$ are in $\mathcal{W}_1$. Setting $w = w_*(x|\overline{w}, \overline{A}^{\,+})$ in (A7) yields

$$n \int_{\overline{A}^{\,+}} \overline{\Lambda}(x)(\overline{w}(x) - w_*(x|\overline{w}, \overline{A}^{\,+}))dx \leq 0. \tag{A8}$$

But, from its definition, it is clear that $w_*(x|\overline{w}, \overline{A}^{\,+}) \leq \overline{w}(x)$ so that, for (A8) to be true, $\overline{w}$ must equal $w^*(x|\overline{w}, \overline{A}^{\,+})$, except possibly on a set of measure zero. Likewise, $\overline{w}$ must equal $w^*(x|\overline{w}, \overline{A}^{\,-})$. Again recalling the definitions of w_* and w^*, it is clear that $\overline{w}$ must thus take the smallest possible value in $\overline{A}^{\,+}$ and the largest possible value in $\overline{A}^{\,-}$. The result is immediate. □

Theorem 5. *If $\mathcal{W} = \mathcal{W}_2$ and Assumption 2 holds, then*

(a) for any interval (a,b) $\subseteq \overline{A}^{\,+}$ (or $\overline{A}^{\,-}$), $\overline{w}$ satisfies $\overline{w}(x) =$ max$\{w_1(x), \overline{w}(a-)\}$ (or min$\{w_2(x), \overline{w}(b+)\}$);

(b) for any interval $(a, b) \subseteq \underline{A}^+$ (or $\underline{A}^-$), $\underline{w}$ satisfies $\underline{w}(x) =$ min$\{w_2(x), \overline{w}(b+)\}$ (or max$\{w_1(x), \overline{w}(a-)\}$).

Proof. We only prove the first part of (a). The proof proceeds as in the proof of Theorem 4, with the conclusion still being that $\overline{w}$ must take the smallest possible value in $\overline{A}^{\,+}$ and the largest possible value in $\overline{A}^{\,-}$. In an interval $(a, b) \subseteq \overline{A}^{\,+}$, the smallest possible value would be $w_1(x)$, except that only nondecreasing w are allowed. To be nondecreasing, $\overline{w}(x)$ must be smaller than $\overline{w}(a-)$ on (a, b). The result is immediate. □

Proof of Theorem 3. There are many different cases to check, but they are all similar to the following two cases that we will verify: (i) finding $\overline{w}$ when $\mathcal{W} = \mathcal{W}_1, \psi(\theta) \uparrow$, and $f(x|\theta)$ has MLR; (ii) finding $\overline{w}$ when $\mathcal{W} = \mathcal{W}_2, \psi \uparrow\downarrow$, and $f(x|\theta)$ is SVR$_3$.

For case (i), note that the number of sign changes of

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

is zero or one and, in the latter situation, the sign changes from $-$ to $+$. Hence the variation reducing property of an MLR density implies that the number of sign changes of $\overline{\Lambda}(x) = \int T(\theta) f(x|\theta) d\theta$ is zero or one and, in the latter situation, the sign changes from $-$ to $+$. Thus $\overline{A}^{\,+} = (r, a)$ and $\overline{A}^{\,-} = (a, s)$ for some $r \le a \le s$ (note that the variation reducing property of strict MLR implies that $\overline{\Lambda}(x) = 0$ only on a set of measure zero), and Theorem 4 implies that $\overline{w}(x) = w_1(x)$ for $x \in (r, a)$ and $\overline{w}(x) = w_2(x)$ for $x \in (a, s)$, as was to be proved.

For case (ii), note that the number of sign changes of $T(\theta)$ is $0, 1$, or 2 and, in the latter situation, the sign changes from $-$ to $+$ to $-$. Hence the variation reducing property of an SVR_3 density implies that $\overline{\Lambda}$ has $0, 1$, or 2 sign changes and, in the latter situation, the sign changes from $-$ to $+$ to $-$. When $\overline{\Lambda}$ has zero or one sign change, the proof proceeds exactly as in case (i). Thus suppose $\overline{\Lambda}$ has two sign changes, from $-$ to $+$ to $-$. Then $\overline{A}^{\,+} = (a, b)$ and $\overline{A}^{\,-} = (r, a) \cup (b, s)$, where $r \le a \le b \le s$. (The SVR_3 property implies that $\overline{\Lambda}(x) = 0$ only on a set of measure zero.) Theorem 2 then yields that a maximizing w must satisfy

$$\overline{w} = \begin{cases} \min\{w_2(x), \overline{w}(a+)\} & r < x < a \\ \max\{w_1(x), \overline{w}(a-)\} & a < x < b \\ \min\{w_2(x), \overline{w}(s+)\} & b < x < s. \end{cases}$$

It follows directly that a maximizer must be of form IV, completing the proof. □

REFERENCES

Bayarri, M. J. and DeGroot, M. (1992). A "BAD" view of weighted distributions and selection models. In *Bayesian Statistics 4*, J. M. Bernardo, et. al. (Eds.), pp 17–33. Oxford University Press, London.

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

Berger, J. (1994). An overview of robust Bayesian analysis. *Test* **3**, 5–124.

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*, **37**, 307–316.

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.

Shyamalkumar, N. D. (1996). Bayesian robustness with respect to the likelihood. Technical Report 96-23, Department of Statistics, Purdue University.

Wasserman, L. (1992). Recent methodological advances in robust Bayesian inference. In *Bayesian Statistics 4*, J. M. Bernardo, et. al. (Eds.), pp 483–502. Oxford University Press, London.