

**RANGES OF POSTERIOR PROBABILITIES FOR  
UNIMODAL PRIORS WITH SPECIFIED QUANTILES**

by

**James O. Berger  
Purdue University**

**Anthony O'Hagan  
University of Warwick**

**Technical Report #86-59**

**Department of Statistics  
Purdue University**

**Revised November 1987**

RANGES OF POSTERIOR PROBABILITIES FOR  
UNIMODAL PRIORS WITH SPECIFIED QUANTILES

by

James O. Berger

Purdue University

Anthony O'Hagan

University of Warwick

Abstract

Suppose several quantiles of the prior distribution for  $\theta$  are specified or, equivalently, the prior probabilities of a partitioning collection of intervals  $\{I_i\}$  are given. Suppose, in addition, that the prior distribution is assumed to be unimodal. Rather than selecting a single prior distribution to perform a Bayesian analysis, the class of all prior distributions compatible with these inputs will be considered. For this class and unimodal likelihood functions, the ranges of the posterior probabilities of the  $I_i$ , and of the posterior c.d.f. at the specified prior quantiles, are determined. Small ranges ensure robustness with respect to the exact functional form of the prior.

*Key Words:* Prior quantiles; Unimodal prior; Bayesian robustness; Ranges of posterior probabilities.

## 1. INTRODUCTION

In any practical Bayesian analysis a prior distribution for a continuous parameter cannot be specified in complete detail. To do so would imply infinitely many prior probability judgements. Instead, only a few judgements are actually made. The prior specification is then usually completed by assuming a reasonable, and preferably tractable, form of distribution which fits the judgements that have been made. Furthermore, the judgements that are required are often not prior probabilities but complex functions of them, such as prior means and variances. We consider here a scenario of prior specification that does not require the fitting of a specific distribution to the prior judgements. The user assigns his prior probabilities for the parameter lying in each of the intervals  $I_1, I_2, \dots, I_m$ , which are contiguous and partition the real line; this amounts to asserting  $m - 1$  points of the prior c.d.f. of the parameter. We then assume only that the prior distribution  $\pi$  lies in some set  $\Pi$  of distributions, all of which agree with the stated prior probabilities over the intervals  $\{I_i\}$ . The posterior distribution  $\pi^*$  then lies in a corresponding set  $\Pi^*$ , and we consider what bounds are thereby implied for relevant posterior probabilities.

*Example 1.* In Martz and Waller (1982), Example 5.1 supposes that two engineers are concerned with the mean life  $\theta$  of a proposed new industrial engine. The two engineers,  $A$  and  $B$ , quantify their beliefs about  $\theta$  in terms of the probabilities given in Table 1 for being in specified intervals. Note that  $A$  has substantially more precise beliefs than does  $B$ .

Table 1. Specified Prior Probabilities of Intervals

i	Interval $I_i$	$p_i^A = Pr(\theta \in I_i A)$	$p_i^B = Pr(\theta \in I_i B)$
1	[0, 1000)	0.01	0.15
2	[1000, 2000)	0.04	0.15
3	[2000, 3000)	0.20	0.20
4	[3000, 4000)	0.50	0.20
5	[4000, 5000)	0.15	0.15
6	[5000, $\infty$ )	0.10	0.15

Either of these probability specifications determines a class of prior distributions  $\pi$ , namely

$$\Pi_0 = \{ \pi : p_i = Pr(\theta \in I_i) = \int_{I_i} \pi(d\theta) \text{ for all } i \}. \quad (1.1)$$

Berliner and Goel (1986) determine the ranges of the posterior probabilities of the  $I_i$  when  $\Pi_0$  is the assumed class of priors. Earlier, DeRobertis (1978) had considered the related class of priors with  $Pr(\theta \in I_i) \geq \gamma_i$  for all  $i$ . For both this class and  $\Pi_0$  the ranges of the posterior probabilities are often quite large because the classes include discrete distributions concentrated at “least favorable” configurations. The engineers in Example 1 might well deny that such discrete priors are plausible reflections of their prior beliefs for a continuous parameter, and might insist that they actually have smooth prior densities. Indeed, it would not be uncommon to encounter the belief that  $\Pi$  is actually unimodal, leading to a class such as

$$\Pi_2 = \{\text{unimodal } \pi : p_i = Pr(\theta \in I_i) \text{ for all } i\}. \quad (1.2)$$

Use of this more realistic class can sharply reduce the variability in posterior answers, and is the subject of Section 4. It is shown that, although  $\Pi_2$  is a huge class of priors, maximizations and minimizations over this class can be reduced to low dimensional numerical optimization. Section 2 presents basic notation that will be needed, while Section 3 illustrates the type of answers obtained, through several examples.

Previous work on finding ranges of posterior quantities for classes of priors mainly dealt with conjugate priors (e.g. Leamer (1978, 1982) and Polasek (1985)). Huber (1973) was the first to explicitly consider a large “nonparametric” class of priors. He determined the range of the posterior probability of a set when  $\Pi$  is an  $\varepsilon$ -contamination class of priors having the form  $\pi = (1-\varepsilon)\pi_0 + \varepsilon q$ ; here  $\pi_0$  is a single elicited prior,  $\varepsilon$  reflects the uncertainty in  $\pi_0$ , and  $q$  is a “contamination”. Huber considered the case where all contaminations (even discrete) are allowed. Berger and Berliner (1986), Sivaganesan (1986a, 1986b), and Sivaganesan and Berger (1986) considered a variety of generalizations, to different classes of contaminations (e.g. unimodal) and different posterior criteria (e.g. the posterior mean and variance). DeRobertis and Hartigan (1981) considered a large class of priors specified by a type of upper and lower envelope on the prior density, and also find ranges of posterior quantities of interest. Each of these classes is plausible as a model of prior uncertainty. Classes such as  $\Pi_0$  and  $\Pi_2$  perhaps have the advantage of being the simplest to understand and elicit. Other work dealing with similar classes of priors includes Bierlein (1967), Kudō (1967) West (1979), Manski (1981), Lambert and Duncan (1986), Cano, Hernández, and

Moreno (1985), and Lehn and Rummel (1987). Related analyses for testing situations can be found in Edwards, Lindman, and Savage (1963), Berger and Sellke (1987), Berger and Delampady (1987), Casella and Berger (1987), and Delampady (1986). Other related works include Kadane and Chuang (1978), Wolfenson and Fine (1982), Berger (1984, 1985, 1987), and Walley (1986). These latter works of Berger and Walley also include general review and history of the subject.

## 2. NOTATION AND THE FORMAL PROBLEM

Prior information is to be stated for an unknown, continuous parameter  $\theta \in [a_0, a_m]$  by giving

$$p_i = Pr(\theta \in I_i) = Pr(a_{i-1} \leq \theta \leq a_i)$$

for  $i = 1, 2, \dots, m$ . The intervals partition the parameter space  $[a_0, a_m]$  and their endpoints  $a_1, a_2, \dots, a_{m-1}$  are arbitrary, possibly even specified by the user. Infinite parameter spaces are included by  $a_0 = -\infty$  and/or  $a_m = \infty$ . It is assumed that there is an underlying prior density  $\pi(\theta)$  on  $[a_0, a_m]$  constrained by

$$\int_{a_{i-1}}^{a_i} \pi(\theta) d\theta = p_i \quad (i = 1, 2, \dots, m). \quad (2.1)$$

Data are obtained, yielding a likelihood function  $l(\theta)$ . We will assume that  $l(\theta)$  is unimodal, with mode  $\theta_0$ . For an arbitrary prior density  $\pi$ , the posterior density  $\pi^*$  is, by Bayes theorem,

$$\pi^*(\theta) = \pi(\theta)l(\theta) / \int_{a_0}^{a_m} \pi(t)l(t)dt. \quad (2.2)$$

Of interest is some set

$$C = \cup_{i \in \Omega} I_i, \quad (2.3)$$

where  $\Omega$  is some subset of the indices  $\{1, 2, \dots, m\}$ . We will seek bounds on  $Pr^*(C)$ , the posterior probability of the set  $C$ . The two cases of most common interest will be  $C = I_i$  and  $C = \cup_{i \leq n} I_i$ ; for the latter case,  $Pr^*(C)$  is the posterior c.d.f. evaluated at  $a_i$ . Sets,  $C$ , more general than (2.3) can be considered (see Section 4.2), as could quantities such as the posterior mean, but the analyses then become substantially messier.

For a given  $C$  and class of priors,  $\Pi$ , we seek the range of the posterior probability of  $C$  as  $\pi$  ranges over  $\Pi$ . Specifically, we will calculate

$$\overline{P}_{\Pi}^*(C) = \sup_{\pi \in \Pi} Pr^*(C), \quad (2.4)$$

$$\underline{P}_{\Pi}^*(C) = \inf_{\pi \in \Pi} Pr^*(C). \quad (2.5)$$

The classes  $\Pi_0$  and  $\Pi_2$ , defined in (1.1) and (1.2), will be considered. Consideration of  $\Pi_2$  is sensible, of course, only when the specified  $p_i$  are compatible with the unimodality condition (so that  $\Pi_2$  is nonempty). This will be the case when the constants

$$q_i = p_i / (a_i - a_{i-1}), \quad i = 1, 2, \dots, m \quad (2.6)$$

(defined to be zero if the denominator is infinite) satisfy

$$q_1 \leq q_2 \leq \dots \leq q_k \geq q_{k+1} \geq q_{k+2} \geq \dots \geq q_m \quad (2.7)$$

for some  $k$ . Note that  $q_i$  is the uniform density on  $I_i$  which has mass  $p_i$ . We will henceforth assume that (2.7) holds, and that the prior mode is known to be in  $I_k$ .

In addition to  $\Pi_0$  and  $\Pi_2$ , two simplified versions of  $\Pi_2$  will be considered. The first (in section 4.1) is

$$\Pi'_2 = \{\pi \in \Pi_2 : a_k \text{ is the mode of } \pi \text{ and } \pi(a_k) \leq h^*\}. \quad (2.8)$$

The motivation behind consideration of this class is twofold. First, it substantially simplifies the analysis, there being no uncertainty in the location of the prior mode, and no need to allow for point masses at the prior mode. The additional inputs needed for  $\Pi'_2$  are, however, certainly ascertainable. Indeed, elicitation processes often begin by determining the “most likely” value of  $\theta$ , with the partitioning intervals  $I_i$  determined by working away from this mode. Allowing a maximum prior density of  $h^*$  is also very reasonable, and “safe” values of this upper bound are not hard to elicit. (Of course, setting  $h^*$  equal to a very large number will yield answers more or less equal to the answers with no constraint on this height.)

A reasonable value for this upper bound in many situations is

$$h^* = 3 \max_i \{q_i\}, \quad (2.9)$$

the reasoning being that it would typically be unreasonable for the prior density to exceed the largest “average” density on an interval by more than a factor of 3. In all examples, we will assume (2.9) as a “default” value of  $h^*$  for  $\Pi'_2$ .

While  $\Pi'_2$  results from applying additional restrictions to  $\Pi_2$ , a certain loosening of the unimodality condition leads to a class  $\Pi_1$ , defined in O'Hagan and Berger (1987), which involves only univariate optimizations, and is hence very easy to analyze. Furthermore, it will be seen to frequently give answers similar to those from the more complicated  $\Pi_2$ . (We do not repeat this analysis here but will include calculations for this class in our examples, for comparative purposes.) Note that

$$\Pi_0 \supset \Pi_1 \supset \Pi_2 \supset \Pi'_2,$$

so that the posterior ranges  $(\underline{P}_\Pi^*(C), \overline{P}_\Pi^*(C))$  will be nested in reverse order.

### 3. EXAMPLES

As the calculations are rather complex, we delay their development until after the presentation of several illustrative examples. In these examples we calculate the range of the posterior probabilities of the intervals  $I_i$ , and also of the c.d.f. evaluated at the  $a_i$ . The format used for each example is to present the range of the posterior probability of the relevant set  $C$  as the interval  $(\underline{P}_\Pi^*(C), \overline{P}_\Pi^*(C))$ . For each example, the classes  $\Pi_0$  and  $\Pi_2$  defined in Section 1, and the classes  $\Pi'_2$  and  $\Pi_1$  defined in Section 2, will be considered.

*Example 1 (continued).* Data becomes available in the form of two independent life-times which are exponentially distributed with mean  $\theta$ . The observed life-times are 2000 and 2500 hours, leading to a likelihood function

$$l(\theta) = \theta^{-2} \exp(-4500/\theta).$$

Tables 2 and 3 present the ranges of the posterior probabilities of the intervals  $I_i$  for engineers A and B, respectively. Tables 4 and 5 present the ranges of the posterior c.d.f.s evaluated at the  $a_i$ , for A and B respectively. For engineer B we make the natural assumption that the prior mode is specified to be  $a_k = 3000$  (see Table 1) in calculations with  $\Pi_1$ ,  $\Pi_2$ , and  $\Pi'_2$ . For engineer A, the probabilities in Table 1 lead to no natural

restriction on the prior mode, other than that it be in the interval [3000,4000). The calculations with  $\Pi_1$  and  $\Pi_2$  allow an arbitrary mode in this interval, but  $\Pi'_2$  admits only the choice of 3000 or 4000 as the mode. It somewhat surprisingly makes very little difference to the answer for  $\Pi'_2$  whether 3000 or 4000 is chosen as the mode; the tables present the answers for the choice 3000.

Table 2. Posterior Ranges for  $C = I_i$ , Engineer A

$I_i$	$p_i$	$\Pi_0$	$\Pi_1$	$\Pi_2$	$\Pi'_2$
[0,1000)	0.01	(0,0.006)	(0.001,0.004)	(0.001,0.004)	(0.001,0.004)
[1000,2000)	0.04	(0.019,0.057)	(0.037,0.049)	(0.037,0.049)	(0.038,0.049)
[2000,3000)	0.20	(0.214,0.291)	(0.222,0.260)	(0.225,0.260)	(0.229,0.260)
[3000,4000)	0.50	(0.476,0.613)	(0.512,0.584)	(0.517,0.584)	(0.517,0.579)
[4000,5000)	0.15	(0.106,0.164)	(0.121,0.147)	(0.121,0.147)	(0.122,0.146)
[5000, $\infty$ )	0.10	(0,0.083)	(0,0.071)	(0,0.071)	(0,0.071)

Table 3. Posterior Ranges for  $C = I_i$ , Engineer B

$I_i$	$p_i$	$\Pi_0$	$\Pi_1$	$\Pi_2$	$\Pi'_2$
[0,1000)	0.15	(0,0.111)	(0.020,0.023)	(0.020,0.023)	(0.020,0.023)
[1000,2000)	0.15	(0.088,0.255)	(0.171,0.197)	(0.172,0.197)	(0.172,0.197)
[2000,3000)	0.20	(0.235,0.391)	(0.282,0.327)	(0.283,0.327)	(0.284,0.327)
[3000,4000)	0.20	(0.197,0.349)	(0.247,0.288)	(0.248,0.288)	(0.248,0.288)
[4000,5000)	0.15	(0.125,0.233)	(0.149,0.175)	(0.149,0.175)	(0.149,0.175)
[5000, $\infty$ )	0.15	(0,0.146)	(0,0.121)	(0,0.121)	(0,0.121)

Table 4. Posterior Ranges for  $C = [0, a_i]$ , Engineer A

$a_i$	$\Pi_0$	$\Pi_1$	$\Pi_2$	$\Pi'_2$
1000	(0,0.006)	(0.001,0.004)	(0.001,0.004)	(0.001,0.004)
2000	(0.0194,0.062)	(0.038,0.053)	(0.039,0.050)	(0.039,0.050)
3000	(0.241,0.341)	(0.262,0.310)	(0.265,0.308)	(0.268,0.308)
4000	(0.769,0.886)	(0.794,0.871)	(0.800,0.870)	(0.801,0.869)
5000	(0.917,1)	(0.929,1)	(0.929,1)	(0.929,1)

Table 5. Posterior Ranges for  $C = [0, a_i]$ , Engineer B

$a_i$	$\Pi_0$	$\Pi_1$	$\Pi_2$	$\Pi'_2$
1000	(0,0.111)	(0.020,0.023)	(0.020,0.023)	(0.020,0.023)
2000	(0.096,0.327)	(0.191,0.221)	(0.192,0.221)	(0.192,0.221)
3000	(0.388,0.623)	(0.474,0.547)	(0.476,0.547)	(0.477,0.547)
4000	(0.659,0.860)	(0.725,0.830)	(0.728,0.830)	(0.728,0.830)
5000	(0.854,1)	(0.879,1)	(0.879,1)	(0.879,1)



Note first that  $\Pi_1$ ,  $\Pi_2$ , and  $\Pi'_2$  yield usefully small ranges of posterior probabilities, in all cases. For instance, if engineer A is willing to assume unimodality as well as the given  $p_i$ , then he knows that his posterior probability that  $\theta \in [3000, 4000)$  lies between 0.517 and 0.584, while his posterior probability that  $\theta \leq 2000$  lies between 0.039 and 0.050. For engineer B, the corresponding ranges are 0.248 to 0.288, and 0.192 to 0.221. These ranges are small enough that the engineers can probably make decisions on this basis, obviating the need for more detailed prior specification.

Note also that  $\Pi_1$ ,  $\Pi_2$ , and  $\Pi'_2$  tend to yield similar answers, so that the particular manner in which one chooses to implement unimodality does not seem to matter greatly. On the other hand,  $\Pi_0$  yields substantially broader intervals (typically 2 to 4 times larger than  $\Pi_2$ , say), indicating that the unimodality assumption has a pronounced effect.

A secondary point of interest is the very small interval of posterior probabilities that is obtained for interval  $I_1$  of engineer B when  $\Pi_1$ ,  $\Pi_2$ , or  $\Pi'_2$  is used. The reason serves as a warning about casual assumption of the unimodality constraint. It is easy to see that, when two adjacent intervals have equal  $q_i$  (as do  $I_1$  and  $I_2$  for engineer B), then any unimodal prior must have its mode in one of the intervals or be uniform over those intervals. In Table 3 the mode could only be between 2000 and 4000, so all priors in  $\Pi_1$ ,  $\Pi_2$ , and  $\Pi'_2$  are uniform over  $I_1$  and  $I_2$ . Thus there may be little variation in the prior (over  $\Pi$ ) under the unimodality assumption if certain of the adjacent  $q_i$  are nearly equal (the central intervals excepted.)

*Example 2.* As a second example, we illustrate the methodology on a standard type of Bayesian example. Suppose subjective elicitation yields the following intervals and corresponding prior probabilities,  $p_i$ , for a normal mean  $\theta$ .

Table 6. Intervals and Prior Probabilities: Normal Example

$I_i$	$(-\infty, -2)$	$(-2, -1)$	$(-1, 0)$	$(0, 1)$	$(1, 2)$	$(2, \infty)$
$p_i$	0.08	0.16	0.26	0.26	0.16	0.08

A “textbook” Bayesian analysis would be to notice that the  $p_i$  are a good match to a  $\mathcal{N}(0, 2)$  (normal, with mean 0 and variance 2) prior distribution. Suppose now that  $x = 1.5$  is observed from a  $\mathcal{N}(\theta, 1)$  experiment. Then usual conjugate prior Bayesian theory would be employed, resulting in a  $\mathcal{N}(1, 2/3)$  posterior distribution. The resulting posterior

probabilities of the  $I_i$  are listed in Table 7 as  $p_i^*$ .

As an indication of the robustness of the  $p_i^*$  to the prior normality assumption, we can calculate the ranges of the posterior probabilities of the  $I_i$  for the various classes of priors we are considering. These results are given in Table 7.

Table 7. Posterior Ranges for  $C = I_i$ , Normal Example

$I_i$	$p_i^*$	$\Pi_0$	$\Pi_1$	$\Pi_2$	$\Pi'_2$
$(-\infty, -2)$	.0001	(0,0.001)	(0,0.0002)	(0,0.0002)	(0,0.0002)
$(-2, -1)$	.007	(0.001,0.029)	(0.006,0.011)	(0.006,0.011)	(0.006,0.010)
$(-1, 0)$	.103	(0.024,0.272)	(0.095,0.166)	(0.095,0.166)	(0.095,0.155)
$(0, 1)$	.390	(0.208,0.600)	(0.320,0.447)	(0.322,0.447)	(0.332,0.447)
$(1, 2)$	.390	(0.265,0.625)	(0.355,0.475)	(0.357,0.473)	(0.360,0.467)
$(2, \infty)$	.110	(0,0.229)	(0,0.156)	(0,0.156)	(0,0.154)

The  $p_i^*$  are reasonably robust, except possibly for  $p_6^*$ . Also of interest is the now very dramatic difference between the  $\Pi_0$  ranges and the ranges for the unimodality classes; their sizes differ by roughly a factor of 4. This provides further evidence of the value of incorporating the unimodality assumption (if subjectively warranted). Of course, these are but two examples, and situations can be constructed where there is little difference between the results for  $\Pi_0$  and  $\Pi_1$ , but our general experience in looking at a variety of examples is that incorporation of unimodality typically has a substantial effect.

One final comment: the degrees of robustness in situations such as Example 2 will typically depend strongly on the data  $x$ . In particular, as  $x$  gets extreme, so that the likelihood and the prior clash, substantially less robustness will be observed (cf. Berger and Berliner (1986) and Sivaganesan and Berger (1986)).

#### 4. ANALYSIS UNDER STRICT UNIMODALITY CONSTRAINTS

In this section we present analyses for the classes  $\Pi_2$  and  $\Pi'_2$  defined in (1.2) and (2.8), respectively. Since  $\Pi'_2$  is easiest to analyze, it is discussed first. Section 4.2 considers  $\Pi_2$ .

##### 4.1 Solutions for $\Pi'_2$

Minimizations and maximizations of  $Pr^*(C)$  over  $\Pi'_2$  can be reduced to minimizations and maximizations over a small dimensional class of extreme points. We describe these

extreme points and the algorithm for calculating  $\overline{P}^*(C)$  and  $\underline{P}^*(C)$  (dropping the subscript  $\Pi'_2$ ) below, in a series of steps. Proofs of most assertions are given in the appendix.

*Step 1: Classification of Intervals*

There will be two relevant classifications of intervals. The first classification is as a “maximizing” interval (Max) or a “minimizing” interval (Min). Essentially, these are intervals in which it is desired to maximize or minimize the posterior probability, respectively. Table 8 presents this classification.

Table 8. Min-Max Interval Classification

Goal of analysis	Interval $i$	
	In $\Omega$	Not in $\Omega$
$\overline{P}^*(C)$	Max	Min
$\underline{P}^*(C)$	Min	Max

The second classification is according to the form of the “optimizing” prior in each interval. It will be shown that there is a prior distribution (possibly a subprobability distribution) at which  $\overline{P}^*(C)$  or  $\underline{P}^*(C)$  is actually achieved, and that in each interval this “optimizing” prior is one of the following types:

*Uniform* (denoted by U): on  $I_i$ ,

$$\pi(\theta) = q_i; \tag{4.1}$$

*Step* (denoted by S): on  $I_i$ ,

$$\pi(\theta) = \begin{cases} h_{i-1} & \text{if } a_{i-1} \leq \theta \leq s_i \\ h_i & \text{if } s_i < \theta \leq a_i, \end{cases} \tag{4.2}$$

where  $s_i$  is defined in (4.8) and makes the total probability in  $I_i$  equal to  $p_i$ . Each interval is classified as a U or an S in Table 9. In the table, “likelihood form” refers to whether  $l(\theta)$  is increasing, decreasing, or both (called modal) in the interval. In the modal interval (note that there can be only one), it is necessary to distinguish between three cases, depending on a comparison of the values of the likelihood at the endpoints with the “average” likelihood over the interval; in the table and in general, we define

$$L(x, y) = \int_x^y l(\theta) d\theta. \tag{4.3}$$

Table 9. U-S Interval Classification

Likelihood Form	Class	Intervals	
		$I_1$ to $I_k$	$I_{k+1}$ to $I_m$
Increasing	Max	S	U
	Min	U	S
Decreasing	Max	U	S
	Min	S	U
Modal			
(i) $L(a_{i-1}, a_i) \leq l(a_{i-1})(a_i - a_{i-1})$	Max	U	S
	Min	S	U
(ii) $L(a_{i-1}, a_i) \leq l(a_i)(a_i - a_{i-1})$	Max	S	U
	Min	U	S
(iii) otherwise	Max	S	S
	Min	S	S

*Step 2: Fill in Uniforms.*

After classifying each interval, choose  $\pi$  in each “U” interval according to (4.1). These will define the actual “optimizing”  $\pi$  in such intervals. Note that then

$$\int_{I_i} l(\theta)\pi(\theta)d\theta = q_i L(a_{i-1}, a_i). \quad (4.4)$$

*Step 3: Classify Chains*

A *chain* is a set of consecutive intervals among  $\{I_1, \dots, I_k\}$  or among  $\{I_{k+1}, \dots, I_m\}$  which consists only of “S” intervals and which is bordered on each side either by a “U” interval or by  $a_0$ ,  $a_{k-1}$ ,  $a_k$ , or  $a_m$ . Such a chain is *simple* if it consists only of “Max” or only of “Min” intervals; otherwise it is a *compound* chain.

It turns out that, within a chain, the “optimizing”  $\pi$  is one in which the adjoining step functions match up at the boundaries to form a stairway. This leads to the following algorithms for calculating the contributions of chains to  $\overline{P}^*(C)$  or  $\underline{P}^*(C)$ .

*Step 4: Evaluating Simple Chains*

Let  $\{I_l, I_{l+1}, \dots, I_n\}$  be a simple chain, and let  $h_i = \pi(a_i)$  for  $i = l - 1, \dots, n$ . Again, the “optimizing” prior over a chain is essentially a stairway ascending or descending from

$h_{l-1}$  to  $h_n$ , with one (or no) steps occurring in each interval (at  $s_i$ ), and  $h_i$  being the height of each step. (The special case of the chain consisting of only a single interval is separately treated, for convenience, as Case 4.) The step heights are constrained by the probability and unimodality requirements to satisfy

$$q_i \leq h_i \leq q_{i+1}, \quad i = l, l+1, \dots, n-1. \quad (4.5)$$

(Specification of  $h_{l-1}$  and  $h_n$  will be considered later; they are defined in terms of  $h_l, \dots, h_{n-1}$ .) Define

$$\Lambda = \{h = (h_l, \dots, h_{n-1}) : (4.5) \text{ holds}\}. \quad (4.6)$$

For any such "stairway" prior, the function determining the contribution of the chain to  $Pr^*(C)$  is

$$\begin{aligned} \Psi(h) &= \Psi((h_l, \dots, h_{n-1})) = \sum_{i=l}^n \int_{I_i} l(\theta) \pi(\theta) d\theta \\ &= h_{l-1} L(a_{l-1}, s_l) + \sum_{i=l}^{n-1} h_i L(s_i, s_{i+1}) + h_n L(s_n, a_n). \end{aligned} \quad (4.7)$$

The step locations,  $s_i$ , are defined by the constraint that  $I_i$  have total mass  $p_i$ ; (4.4) thus yields

$$s_i = [p_i + h_{i-1} a_{i-1} - h_i a_i] / [h_{i-1} - h_i] \quad (4.8)$$

(to be understood as "empty", i.e. there is no "step", if  $h_{i-1} = h_i$ ). We will need, depending on whether this is a "Max" or a "Min" chain, either

$$\bar{\Psi} = \sup_{h \in \Lambda} \Psi(h) \text{ or } \underline{\Psi} = \inf_{h \in \Lambda} \Psi(h). \quad (4.9)$$

Numerical maximization or minimization is generally needed to find  $\bar{\Psi}$  or  $\underline{\Psi}$ . This is typically easy, since  $\Psi(h)$  turns out to be concave on  $\Lambda$  if it is a "Max" chain and convex if it is a "Min" chain. Additional comments about this calculation are given in the appendix.

It remains only to determine the boundary values,  $h_{l-1}$  and  $h_n$ , for each simple chain. Defining

$$\underline{h} = \begin{cases} 0 & \text{if } l = 1 \\ h^* & \text{if } l = k + 1 \\ q_{l-1} & \text{otherwise,} \end{cases} \quad (4.10)$$

$$\bar{h} = \begin{cases} h^* & \text{if } n = k \\ 0 & \text{if } n = m \\ q_{n+1} & \text{otherwise,} \end{cases} \quad (4.11)$$

three cases need to be considered.

*Case 1. (No Modal Interval).*

If the chain does *not* contain the modal interval, then set

$$h_{l-1} = \underline{h}, \quad h_n = \bar{h}. \quad (4.12)$$

*Case 2. (Modal; Min-Left or Max-Right)*

If the chain contains the modal interval and is a “Min” chain lying to the left of  $a_k$  or a “Max” chain lying to the right of  $a_k$ , then it must be the case that the first interval in the chain,  $I_l$ , is the modal interval. To solve this case, find the unique  $v > \theta_0$  in  $I_l$  such that

$$(v - a_{l-1})l(v) = L(a_{l-1}, v), \quad (4.13)$$

define

$$f = [p_l - h_l(a_l - v)] / (v - a_{l-1}), \quad (4.14)$$

and set

$$h_n = \bar{h}, \text{ and } h_{l-1} = \max\{f, \underline{h}\} \text{ (left case) } \underline{\text{or}} \text{ } h_{l-1} = \min\{f, \underline{h}\} \text{ (right case)}. \quad (4.15)$$

Note that  $h_{l-1}$  depends on  $h_l$ , one of the variables over which optimization is performed.

*Case 3. (Modal; Max-Left or Min-Right)*

If the chain contains the modal interval, and is a “Max” chain lying to the left of  $a_k$  or a “Min” chain lying to the right of  $a_k$ , then it must be the case that the last interval,  $I_n$ , is the modal interval. To solve this case, find the unique  $v < \theta_0$  in  $I_n$  such that

$$(a_n - v)l(v) = L(v, a_n), \quad (4.16)$$

define

$$f = [p_n - h_{n-1}(v - a_{n-1})] / (a_n - v), \quad (4.17)$$

and set

$$h_{l-1} = \underline{h}, \text{ and } h_n = \min\{f, \bar{h}\} \text{ (left case) or } h_n = \max\{f, \bar{h}\} \text{ (right case).} \quad (4.18)$$

Note that  $h_n$  depends on  $h_{n-1}$ , one of the variables over which optimization is performed.

*Case 4. (Single Interval)*

The preceding formulas are formally correct when  $l = n$  (i.e., there is a single interval in the chain), but the notation might become garbled in a computer program; also no numerical optimization is needed. Thus we present the formulas for this special case separately. Labelling the single  $S$ -interval  $I_l$ , it is clear that

$$\Psi(h) = h_{l-1}L(a_{l-1}, s_l) + h_lL(s_l, a_l), \quad (4.19)$$

and that  $\bar{\Psi}$  or  $\underline{\Psi}$  are achieved at the following choices of  $h_{l-1}$  and  $h_l$ , analogous to Cases 1 through 3. Here  $\underline{h}$  is as in (4.10), and  $\bar{h}$  is as in (4.11) with  $n$  replaced by  $l$ .

*No Modal Interval:*  $h_{l-1} = \underline{h}$  and  $h_l = \bar{h}$ .

*Modal; Min-Left or Max-Right:*  $h_{l-1} = \max\{f, \underline{h}\}$  and  $h_l = \bar{h}$ , where  $v > \theta_0$  satisfies (4.13) and  $f$  is defined by  $f = [p_l - \bar{h}(a_l - v)]/(v - a_{l-1})$ .

*Modal; Max-Left or Min-Right:*  $h_{l-1} = \underline{h}$  and  $h_l = \min\{f, \bar{h}\}$ , where  $v < \theta_0$  satisfies (4.16) and  $f$  is defined by  $f = [p_l - \underline{h}(v - a_{l-1})]/(a_l - v)$ .

*Step 5. (The Solution If No Compound Chain is Present)*

The contributions of the "U" intervals to  $Pr^*(C)$  are determined by

$$\begin{aligned} K_1 &= \sum_{i \in \Omega: I_i \text{ is a "U"}} q_i L(a_{i-1}, a_i), \\ K_2 &= \sum_{i \notin \Omega: I_i \text{ is a "U"}} q_i L(a_{i-1}, a_i). \end{aligned} \quad (4.20)$$

Next, suppose that there are  $r_1$  simple chains consisting of “Max” intervals, and  $r_2$  simple chains consisting of “Min” intervals. Denote the corresponding  $\bar{\Psi}$  or  $\underline{\Psi}$  by  $\bar{\Psi}_1, \dots, \bar{\Psi}_{r_1}$  and  $\underline{\Psi}_1, \dots, \underline{\Psi}_{r_2}$ . Then

$$\bar{P}^*(C) = \left[ 1 + \frac{(K_2 + \sum_{i=1}^{r_2} \underline{\Psi}_i)}{(K_1 + \sum_{i=1}^{r_1} \bar{\Psi}_i)} \right]^{-1}, \quad (4.21)$$

$$\underline{P}^*(C) = \left[ 1 + \frac{(K_2 + \sum_{i=1}^{r_1} \bar{\Psi}_i)}{(K_1 + \sum_{i=1}^{r_2} \underline{\Psi}_i)} \right]^{-1}. \quad (4.22)$$

(These are *not* meant to be the same  $\underline{\Psi}_i$  or  $\bar{\Psi}_i$  in the two formulas; one generally has to start again at the beginning for each separate calculation.)

*Step 6. (Solution If a Compound Chain Exists)*

First calculate  $K_1, K_2, \bar{\Psi}_1, \dots, \bar{\Psi}_{r_1}$ , and  $\underline{\Psi}_1, \dots, \underline{\Psi}_{r_2}$  as in Step 5. Now at most one compound chain can exist, and it must be a simple chain of “Max” intervals followed by a simple chain of “Min” intervals, or vice versa. Furthermore, the modal interval must be at the boundary of the Max chain and the Min chain. Let  $a_t$  be this boundary, and label the simple chains to the left and right by  $S_L$  and  $S_R$ , respectively. We present the solution only for the compound chain being to the left of the prior mode  $a_k$ ; if it is to the right, use the obvious reflection symmetry to reduce it to the left-side case.

A compound chain on the left must have  $S_L$  being a “Max” simple chain and  $S_R$  being a “Min” simple chain. The formulas differ slightly depending on whether  $a_t$  is at the left or right boundary of the modal interval.

*Case 1:  $a_t$  is the right boundary of the modal interval*

Let  $v \leq \theta_0$  be the unique solution to

$$(a_t - v)l(v) = L(v, a_t), \quad (4.23)$$

and define

$$\tilde{h} = \min \left\{ q_{t+1}, \frac{p_t - q(v - a_{t-1})}{(a_t - v)} \right\}, \quad (4.24)$$

where  $q$  equals zero if  $t = 1$  and equals  $q_{t-1}$  otherwise. For fixed  $h_t$ ,  $q_t \leq h_t \leq \tilde{h}$ , let

$$\varphi_L(h_t) = \bar{\Psi},$$



where  $\bar{\Psi}$  is the supremum of (4.7) for the simple chain  $S_L$ ; here, of course,  $n = t$ . The supremum should be calculated over  $\Lambda$  in (4.6), with the slight modification that the range for  $h_{n-1} = h_{t-1}$  be changed to

$$q_{t-1} \leq h_{t-1} \leq \frac{p_t - h_t(a_t - v)}{v - a_{t-1}}. \quad (4.25)$$

Also, set  $h_{l-1} = \underline{h}$  (see (4.10)) and  $h_n = h_t$  (the fixed value above). Note that if  $S_L$  consists only of the single interval  $I_t$ , then  $\varphi_L(h_t)$  is given by

$$\varphi_L(h_t) = \underline{h}L(a_{t-1}, s_t) + h_tL(s_t, a_t). \quad (4.26)$$

Similarly, optimize over the right chain  $S_R$  for fixed  $h_t$ , finding

$$\varphi_R(h_t) = \underline{\Psi},$$

where  $\underline{\Psi}$  is the infimum of (4.7) for the simple chain  $S_R$ ; here, of course,  $l = t + 1$ . The infimum should be calculated over  $\Lambda$  as in (4.6), with  $h_{l-1}$  set equal to the fixed  $h_t$  and  $h_n = \bar{h}$ . Note that if  $S_R$  consists only of the single interval  $I_{t+1}$ , then  $\varphi_R(h_t)$  is given by

$$\varphi_R(h_t) = h_tL(a_t, s_{t+1}) + \bar{h}L(s_{t+1}, a_{t+1}). \quad (4.27)$$

Finally,

$$\bar{P}^*(C) = \sup_{q_t \leq h_t \leq \bar{h}} \left[ 1 + \frac{K_2 + \sum_{i=1}^{r_2} \underline{\Psi}_i + \varphi_R(h_t)}{K_1 + \sum_{i=1}^{r_1} \underline{\Psi}_i + \varphi_L(h_t)} \right]^{-1} \quad (4.28)$$

and

$$\underline{P}^*(C) = \inf_{q_t \leq h_t \leq \bar{h}} \left[ 1 + \frac{K_2 + \sum_{i=1}^{r_1} \underline{\Psi}_i + \varphi_L(h_t)}{K_1 + \sum_{i=1}^{r_2} \underline{\Psi}_i + \varphi_R(h_t)} \right]^{-1}. \quad (4.29)$$

*Case 2:  $a_t$  is the left boundary of the modal interval*

Let  $v \geq \theta_0$  be the unique solution to

$$(v - a_t)l(v) = L(a_t, v), \quad (4.30)$$

and define

$$\underline{h} = \max \left\{ q_t, \frac{p_{t+1} - \tilde{q}(a_{t+1} - v)}{(v - a_t)} \right\}, \quad (4.31)$$

where  $\tilde{q}$  equals  $h^*$  if  $t = k + 1$  and equals  $q_{t+2}$  otherwise. For fixed  $h_t$ ,  $\underline{h} \leq h_t \leq q_{t+1}$ , let  $\varphi_L(h_t)$  be  $\overline{\Psi}$ , the supremum of (4.7) for the simple chain  $S_L$ ; here  $n = t$ . The supremum should be calculated over  $\Lambda$  in (4.6), with  $h_{l-1} = \underline{h}$  and  $h_n$  equal to the fixed  $h_t$ . If  $S_L$  consists only of the single interval  $I_t$ , then  $\varphi_L(h_t)$  is given by (4.26).

For fixed  $h_t$ , let  $\varphi_R(h_t)$  be  $\underline{\Psi}$ , the infimum of (4.7) for the simple chain  $S_R$ ; here  $l = t + 1$ . The infimum should be calculated over  $\Lambda$  as in (4.6), with the range of  $h_l = h_{t+1}$  being changed to

$$\frac{p_{t+1} - h_t(v - a_t)}{(a_{t+1} - v)} \leq h_{t+1} \leq q_{t+2}, \quad (4.32)$$

and with  $h_{l-1}$  equal to the fixed  $h_s$  and  $h_n = \overline{h}$ . If  $S_R$  consists only of the single interval  $I_{t+1}$ , then  $\varphi_R(h_t)$  is given by (4.27).

Equation (4.28) or (4.29) can be used to complete the calculation.

#### 4.2 Solution for $\Pi_2$ and for Arbitrary Intervals $C$

The algorithm in section 4.1 for  $\Pi'_2$  can be modified to solve the problem for  $\Pi_2$ . Let  $a^*$  be the location of the prior mode in the interval  $I_k$ . It is easy to show that  $a^*$  must be one of the endpoints of the interval, unless the interval is the modal interval (of the likelihood); then  $a^*$  could also be at the mode of the likelihood. In the former case the analysis is as in Section 4.1, while in the latter case the original  $I_k$  can be considered to be two intervals  $I^* = [a_{k-1}, a^*)$  and  $I^{**} = [a^*, a_k)$ . Let  $p^*$  be the prior probability assigned to  $I^*$ , so that  $p_k - p^*$  is that assigned to  $I^{**}$ . To ensure that the unimodality and probability constraints are satisfied, it is easy to check that  $p^*$  must satisfy

$$(a^* - a_{k-1})q_{k-1} \leq p^* \leq p_k - q_{k+1}(a_k - a^*).$$

Finally, let  $h^*$  be the maximum allowed prior density at  $a^*$  (as before).

For specified  $a^*$ ,  $p^*$ , and  $h^*$ , one has a  $\Pi'_2$  problem with  $a_k = a^*$ . Hence the algorithm in the preceding section can be applied to find  $\overline{P}(C^*)$  or  $\underline{P}(C^*)$  for this  $\Pi'_2$ . One can then optimize these quantities over  $p^*$ . (Note that the optimizing  $h^*$  will be  $\infty$ , i.e. a point mass will creep in. A reasonable practical way to deal with this is just to choose a very large fixed  $h^*$ , such as  $10^4 q_k$ ; this will produce essentially the same answer as  $h^* = \infty$ .)

The analysis above suggests the manner in which one can handle an arbitrary interval  $C = [c_1, c_2]$ . Simply create new intervals in each of the  $I_i$  where a  $c_j$  occurs, with each  $c_j$  becoming an endpoint. Of course, two new parameters,  $p_1^*$  and  $p_2^*$  (analogous to  $p^*$ ), might then be introduced, and require additional optimizations.

## APPENDIX

The purpose of the appendix is to outline a proof of the results in Section 4. For simplicity, we consider only the calculation of

$$\begin{aligned}
\bar{P}^*(C) &= \sup Pr^*(C) \\
&= \sup \frac{\sum_{i \in \Omega} \int_{I_i} l(\theta) \pi(\theta) d\theta}{\sum_{\text{all } i} \int_{I_i} l(\theta) \pi(\theta) d\theta} \\
&= \left[ 1 + \inf \frac{\sum_{i \notin \Omega} \int_{I_i} l(\theta) \pi(\theta) d\theta}{\sum_{i \in \Omega} \int_{I_i} l(\theta) \pi(\theta) d\theta} \right]^{-1}. \tag{A1}
\end{aligned}$$

### *Verification of Step 1*

Expression (A1) makes it clear that one wants to choose  $\pi$  to make the  $\int_{I_i} l(\theta) \pi(\theta) d\theta$  small for  $i \notin \Omega$  and large for  $i \in \Omega$ . This is the basis for the classification of an interval as a “Max” or “Min” in Step 1.

The first key to the proof is establishing that, on each  $I_i$ , the optimizing  $\pi$  should be either a U or an S, as in Table 9. For this purpose, consider the interval  $(a_{l-1}, a_l)$ , and consider arbitrary allowable fixed values,  $h_{l-1}$  and  $h_l$ , for  $\pi(a_{l-1})$  and  $\pi(a_l)$ . (The allowable values must satisfy  $h_{l-1} \leq q_l \leq h_l$ ; if this is violated it is easy to see that  $\pi$  cannot be unimodal.) We are done if we can show that a U or an S in the interval is optimal for any allowable  $h_{l-1}$  and  $h_l$ . (There is the technical point here that uniform segments violate unimodality; the uniform segments arise as the extreme point limits of strictly unimodal priors, however, so that we can ignore the distinction.)

Consider first the case where  $h_{l-1} < h_l$  and  $l(\theta)$  is increasing on  $I_l$ . By unimodality,  $\pi(\theta)$  must be nondecreasing on  $I_l$ . Depending on whether  $I_l$  is a “Max” or a “Min” interval, we thus seek to minimize  $\int_{I_l} l(\theta) \pi(\theta) d\theta$  over

$$\Gamma = \left\{ \pi : \pi(\theta) \text{ is nondecreasing on } I_l, \pi(a_{l-1}) = h_{l-1}, \pi(a_l) = h_l, \text{ and } \int_{I_l} \pi(\theta) d\theta = p_l \right\}.$$

It is trivial that the minimum is attained by a uniform segment on  $I_l$  (since  $\pi(\theta)$  and  $l(\theta)$  are both increasing); the probability constraint then implies that  $\pi(\theta) = q_l$  on  $I_l$ . That the maximum is obtained by a step distribution follows from observing that as much of the mass,  $p_l$ , should be shifted towards high values of  $l(\theta)$  as is possible. Since  $l(\theta)$  is increasing, this means that as much mass should be given to large  $\theta$  as is possible and, correspondingly, as little mass to small  $\theta$  as is possible. This is clearly achieved, subject to being in  $\Gamma$ , by a step density as in (4.2). Noting that the intervals  $I_1$  to  $I_k$  have increasing  $\pi(\theta)$ , while those from  $I_{k+1}$  to  $I_m$  have decreasing  $\pi(\theta)$ , variants of these arguments can be used to fill in the “Increasing” and “Decreasing” Likelihood Form sections of Table 9.

The analysis for the modal interval follows by breaking up the modal interval into the two intervals  $[a_{l-1}, \theta_0)$  and  $[\theta_0, a_l)$ . Since (i) we will either want to maximize or minimize the relevant integrals over both intervals jointly, (ii)  $\pi(\theta)$  is either increasing or decreasing over both intervals, and (iii)  $l(\theta)$  is increasing on  $[a_{l-1}, \theta_0)$  but decreasing on  $[\theta_0, a_l)$ , the previous argument indicates that the optimizing  $\pi$  will be uniform on one interval and a step density on the other; for illustration, suppose it is uniform with density  $h$  on  $[\theta_0, a_l)$  and is a step density on  $[a_{l-1}, \theta_0)$ , with  $\pi(a_{l-1}) = h_{l-1}$ . But the argument of the preceding paragraph applies to show that  $\pi$  must be of the form (on  $[a_{l-1}, \theta_0)$ )

$$\pi(\theta) = \begin{cases} h_{l-1} & \text{if } a_{l-1} \leq \theta < s \\ h & \text{if } s \leq \theta < \theta_0, \end{cases}$$

for some point  $s$ . Hence we have that the optimizing  $\pi$  is actually a single step function on  $I_l$ , given by

$$\pi(\theta) = \begin{cases} h_{l-1} & \text{if } a_{l-1} \leq \theta < s \\ h & \text{if } s \leq \theta < a_l, \end{cases} \quad (A2)$$

where  $a_{l-1} \leq s \leq \theta_0$ .

To complete the argument for Table 9, it must be shown that, under the conditions indicated therein, the step function in the modal interval is actually a uniform density. We illustrate the argument in the case where  $I_l$  is a “Max” interval to the left of  $a_k$  (which is consistent with our earlier choice of  $[\theta_0, a_l)$  as the uniform subinterval, since  $\pi(\theta)$  is then decreasing on  $[\theta_0, a_l)$ ).

Consider the condition (from Table 9)

$$L(a_{l-1}, a_l) \leq l(a_{l-1})(a_l - a_{l-1}).$$

A consequence of this condition is that

$$L(s, a_l) < l(s)(a_l - s) \quad (A3)$$

for  $a_{l-1} < s \leq \theta_0$  (since  $L(s, a_l) - l(s)(a_l - s)$  is a decreasing function of  $s$  in this range).

Now

$$\begin{aligned} \int_{I_l} l(\theta)\pi(\theta)d\theta &= h_{l-1}L(a_{l-1}, s) + hL(s, a_l) \\ &= h_{l-1}L(a_{l-1}, a_l) + (h - h_{l-1})L(s, a_l) \\ &= h_{l-1}L(a_{l-1}, a_l) + (q_l - h_{l-1})(a_l - a_{l-1})\frac{L(s, a_l)}{(a_l - s)}, \end{aligned} \quad (A4)$$

the last equality following from the facts that

$$h_{l-1}(s - a_{l-1}) + h(a_l - s) = p = q_l(a_l - a_{l-1}).$$

Consider (A4) as a function of  $s$ , for fixed  $h_{l-1}$ . Since

$$\begin{aligned} \frac{d}{ds} \frac{L(s, a_l)}{(a_l - s)} &= \frac{-l(s)}{(a_l - s)} + \frac{L(s, a_l)}{(a_l - s)^2} \\ &= (a_l - s)^{-2}[-l(s)(a_l - s) + L(s, a_l)], \end{aligned} \quad (A5)$$

which is negative for  $a_{l-1} \leq s \leq \theta_0$  by (A3), it follows that  $\int_{I_l} l(\theta)\pi(\theta)d\theta$  is decreasing in  $s$ . The conclusion is that  $s = a_{l-1}$  is optimal, i.e.  $\pi$  is uniform on  $I_l$ . This proves the validity of the relevant entry in Table 9.

### *Verification of Step 2.*

We demonstrated that, for any allowable  $h_{l-1}, h_l$ , the “local” optimizing  $\pi$  in a U interval,  $I_l$ , is  $\pi(\theta) = q_l$ . Since  $h_{l-1} \leq q_l \leq h_l$ , this satisfies the global unimodality condition. Furthermore, it does not depend on the particular choice of  $h_{l-1}$  and  $h_l$ , so it must define the global optimizing  $\pi$  on  $I_l$ .

### *Verification of Steps 3 and 4.*

The key feature which greatly simplifies the problem is that two intervals with adjoining step densities will have the steps match at the boundary. The reason for this is

that at least one of the intervals must be a non-modal interval, and hence have monotonic  $l(\theta)$  in the interval. As argued before, any such interval, if constrained at its boundaries by values  $h_{l-1}$  and  $h_l$  for  $\pi$ , will be optimized by a step function attaining these heights. Thus  $\pi$  must be continuous at the boundary of two step functions. This fact verifies the nature of chains, as described in Step 3. The formula (4.7) follows immediately.

The formula (4.12), for the boundary values when both end intervals of the chain are non-modal, follows from the previous observation that the step heights for a non-modal interval will always seek the extremes allowed. The possible extremes in (4.10) and (4.11) correspond to the possibilities that an end interval of the chain is either  $I_1$ ,  $I_k$ ,  $I_{k+1}$ , or  $I_m$  (with corresponding extreme allowed boundary values of 0,  $h^*$ ,  $h^*$ , and 0), or that the end interval adjoins to a U interval  $I_r$  (with the corresponding extreme allowed boundary value of  $h$  to be  $q_r$ ).

Cases 2 and 3 are more complex in that, if the end of a chain is a modal interval, its outer step height need not match up with the density in the adjoining interval. As an illustration of the argument, consider the case where the modal interval is a “Max” interval to the left of  $a_k$ ; this can only occur when the modal interval is  $I_n$ , the rightmost interval of the chain, and when (see Table 9)

$$L(a_{n-1}, a_n) > l(a_{n-1})(a_n - a_{n-1}). \quad (\text{A6})$$

Differentiating with respect to  $s$  shows that

$$L(s, a_n) - l(s)(a_n - s)$$

is a decreasing function of  $s$  on  $(a_{n-1}, \theta_0)$  (since  $l(\theta)$  is increasing there), so that (by (A6) and (4.16)),

$$L(s, a_n) - l(s)(a_n - s) > 0$$

for  $a_{n-1} < s < v$ , with the inequality reversed for  $s > v$ . Finally, using (A4) and (A5) (with  $I_n$  instead of  $I_l$ ), it can be concluded that  $\int_{I_n} l(\theta)\pi(\theta)d\theta$  has a unique maximum at  $s = v$ , providing this leads to an allowable step height at the right boundary. The correct step height, corresponding to a step at  $v$ , is  $f$  in (4.17), but (as before) the extreme allowable height is  $\bar{h}$ . This is the reason for the restriction in (4.18). (If  $f > \bar{h}$ , the

maximizing  $s$  will be the point in  $(a_{n-1}, v)$  which results in a step height of  $\bar{h}$ , but this is automatically taken care of by the formulas.)

The arguments for Case 4 are entirely analogous, though now they lead to a deterministic solution (i.e., optimization over interval step heights of the chain is not necessary.)

#### *Verification of Step 5.*

This is just bookkeeping, adding up the contributions of all the separately maximized or minimized components in (A1). It is important to realize that the decomposition into U intervals and chains decomposed the global problem into a set of local problems which could be analyzed separately and then combined to yield the global solution.

#### *Verification of Step 6.*

The reason only one compound chain can exist, and that it must be of the indicated form, follows from examining Table 9 and realizing that a chain on one side of  $a_k$  can have adjacent Max-S and Min-S intervals only if one corresponds to a modal interval. Furthermore, on one side of this modal interval the chain must be all "Max" intervals, and on the other side it must be all "Min" intervals. All subsequent comments about the nature of a compound chain follow from similar examination of Table 9.

The remaining analysis in Step 6 is very similar to that in Cases 2 through 4 of Step 4, and will be omitted. The essential difference here is that the modal interval cannot have a step height which is less than the adjoining step height. This leads to the possible constraints on the heights  $h_{t-1}$  in (4.25) and  $h_{t+1}$  in (4.32).

#### *The Numerical Calculation in Step 4*

For a simple chain  $\{I_l, I_{l+1}, \dots, I_n\}$ ,  $l \neq n$ ,  $\Psi$  in (4.7) turns out to be a convex function of  $h_l, \dots, h_{n-1}$  if the chain is a "Min" chain, and a concave function if the chain is a "Max" chain. To verify this, note that first and second partial derivatives of  $\Psi$  are as follows (using (4.8) to simplify the expressions):

I. *First Order Partial*s: For  $i = l, \dots, n - 1$ ,

$$\frac{\partial}{\partial h_i} \Psi = (s_i - a_i)l(s_i) + (a_i - s_{i+1})l(s_{i+1}) + L(s_i, s_{i+1}). \quad (A7)$$

II. *Second Order Partial*s: For  $i = l, \dots, n - 1$ ,

$$\frac{\partial^2}{\partial h_i^2} \Psi = \frac{(s_i - a_i)^2}{(h_{i-1} - h_i)} l'(s_i) + \frac{(a_i - s_{i+1})^2}{(h_i - h_{i+1})} l'(s_{i+1}); \quad (A8)$$

except, if Case 2 of Step 4 applies and  $h_{l-1} = f$ , then

$$\frac{\partial^2}{\partial h_l^2} \Psi = \frac{(a_l - s_{l+1})^2}{(h_l - h_{l+1})^2} l'(s_{l+1}); \quad (A9)$$

or, if Case 3 of Step 4 applies and  $h_n = f$ , then

$$\frac{\partial^2}{\partial h_{n-1}^2} \Psi = \frac{(s_{n-1} - a_{n-1})^2}{(h_{n-2} - h_{n-1})} l'(s_{n-1}); \quad (A10)$$

here  $l'(s) = \frac{d}{ds} l(s)$ .

III. *Second Order Mixed Partial*s: For  $i = l + 1, \dots, n - 1$ ,

$$\frac{\partial^2}{\partial h_i \partial h_{i-1}} \Psi = \frac{(s_i - a_i)(a_{i-1} - s_i)}{(h_{i-1} - h_i)} l'(s_i),$$

and

$$\frac{\partial^2}{\partial h_i \partial h_j} \Psi = 0 \text{ for } |i - j| > 1.$$

Note: It can happen that  $h_i = h_{i-1}$  for some  $i$  in a chain (i.e., the interval has collapsed to a U interval), and this needs to be considered in a numerical program so that formulas such as those above are not ill-defined.

### *Convexity or Concavity of $\Psi$*

The matrix of mixed second partial derivatives of  $\Psi$  is tri-diagonal, and an induction argument shows that, in Case 1 of Step 4, the upper  $r \times r$  determinant of this matrix is

$$D_r = (-1)^r \sum_{i=1}^{(r+1)} \left[ \left( \prod_{j=1}^r \{c_j^2 \chi_{(r \geq i+j-1)} + d_j^2 \chi_{(r < i+j-1)}\} \right) \frac{\prod_{j=1}^{r+1} l'(s_j)/w_j}{l'(s_i)/w_i} \right],$$



where  $c_j = s_j - a_j$ ,  $d_j = a_j - s_{j+1}$ ,  $w_j = h_j - h_{j-1}$ , and  $\chi$  denotes the indicator function, as usual. (Slightly different formulas hold in Cases 2 and 3 when  $h_{l-1} = f$  or  $h_n = f$ , but the conclusions are identical.)

Consider the case where the simple chain is a “Max” chain to the left of  $a_k$ . Then each  $w_j > 0$ . From Table 9, it further follows that  $l(\theta)$  must be increasing over the chain, so that each  $l'(s_j) > 0$ . Thus the sign of  $D_r$  is  $(-1)^r$ , establishing the concavity of  $\Psi$  in this case. All other cases are handled similarly.

### *The Optimization of $\Psi$*

Because of the convexity or concavity of  $\Psi$ , no special problems are encountered in its maximization or minimization (other than the reduction of dimension caused by possible equality of adjacent  $h_i$ ). Almost any maximization program should work well; note the availability of analytic derivatives in (A7) through (A10).

The maximization (or minimization) is over an  $n - l - 1$  dimensional rectangle. There is a way to reduce the problem to almost a two dimensional optimization. The idea is to fix  $\{h_l, h_{l+2}, \dots\}$ , and then use the deterministic relationships in Step 4 to calculate the optimal corresponding  $\{h_{l+1}, h_{l+3}, \dots\}$ . Then consider these heights fixed, and recalculate the  $\{h_l, h_{l+2}, \dots\}$ . Continue iterating between these two “dimensions” until convergence. Note that one has to pay careful attention to  $h_{l-1}$  and  $h_n$  in Step 4 at each stage of the iteration. Also, one must “fill in” a sequence from the outside in. Although this is, in a sense, just a two-dimensional problem, the number of iterations needed may be much larger than a good general-purpose  $(n - l - 1)$  dimensional algorithm, unless  $(n - l - 1)$  is quite large.

### ACKNOWLEDGMENT

The authors are grateful to Mr. Kun-Liang Lu for performing the numerical work, and to the National Science Foundation (U.S.) Grant #DMS-8401996 and the Science and Engineering Research Council (U.K.) for support of the research.

## REFERENCES

- BERGER, J. (1984), "The Robust Bayesian Viewpoint (With Discussion)", in *Robustness of Bayesian Analyses*, ed. J. Kadane, Amsterdam: North-Holland.
- BERGER, J. (1985), *Statistical Decision Theory and Bayesian Analysis*, New York: Springer-Verlag.
- BERGER, J. (1987), "Robust Bayesian Analysis: Sensitivity to the Prior," Technical Report #87-10, Department of Statistics, Purdue University.
- BERGER, J. and BERLINER, L.M. (1986), "Robust Bayes And Empirical Bayes Analysis With  $\epsilon$ -Contaminated Priors", *Annals of Statistics*, **14**, 461-486.
- BERGER, J. and DELAMPADY, M. (1987), "Testing Precise Hypotheses," To appear in *Statistical Science*.
- BERGER, J., and SELLKE, T. (1987), "Testing a Point Null Hypothesis: The Irreconcilability of Significance Levels and Evidence", *Journal of the American Statistical Association* **82**, 112-122.
- BERLINER, L.M., and GOEL, P. (1986), "Incorporating Partial Prior Information: Ranges of Posterior Probabilities", Technical Report No. 357, Department of Statistics, Ohio State University.
- BIERLEIN, D. (1967), "Zur Einbeziehung der Erfahrung in Spieltheoretische Modelle," *Z. Oper. Res. Verfahren* **III**, 29-54.
- CANO, J.A., HERNÁNDEZ, A., and MORENO, E. (1985), "Posterior Measure Under Partial Prior Information," *Statistica* **2**, 219-230.
- CASELLA, G., and BERGER, R. (1987), "Reconciling Bayesian And Frequentist Evidence in the One-Sided Testing Problem", *Journal of the American Statistical Association* **82**, 106-111.
- DELAMPADY, M. (1986), "Testing a Precise Hypothesis: Interpreting  $P$ -Values From a Robust Bayesian Perspective", *Ph.D. Thesis*, Purdue University, West Lafayette.
- DE ROBERTIS, L. (1978), "The Use of Partial Prior Knowledge In Bayesian Inference",

- Ph.D. Thesis*, Yale University, New Haven.
- DE ROBERTIS, L., and HARTIGAN, J.A. (1981), "Bayesian Inference Using Intervals of Measures", *Annals of Statistics*, **1**, 235-244.
- EDWARDS, W., LINDMAN, H., and SAVAGE, L.J. (1963), "Bayesian Statistical Inference for Psychological Research," *Psychological Review*, **70**, 193-242.
- GOOD, I.J. (1983), *Good Thinking: The Foundations of Probability and Its Applications*, Minneapolis: University of Minnesota Press.
- HUBER, P.J. (1973), "The Use of Choquet Capacities in Statistics", *Bulletin of the International Statistical Institute*, **45**, 181-191.
- KADANE, J.B., and CHUANG, D.T. (1978), "Stable Decision Problems", *Annals of Statistics*, **6**, 1095-1110.
- KUDŌ, H. (1967), "On Partial Prior Information and the Property of Parametric Sufficiency," *Proceedings of Fifth Berkeley Symposium on Statistics and Probability*, **1**, Berkeley: University of California Press.
- LAMBERT, D. and DUNCAN, G. T. (1986), "Single-Parameter Inference Based on Partial Prior Information," *Canadian Journal of Statistics* **14**, 297-305.
- LEAMER, E.E. (1978), *Specification Searches*, New York: Wiley.
- LEAMER, E.E. (1982), "Sets of Posterior Means With Bounded Variance Prior", *Econometrica*, **50**, 725-736.
- LEHN, J. and RUMMEL, F. (1987), "Gamma Minimax Estimation of a Binomial Probability Under Squared Error Loss," *Statistics and Decisions*, **5**.
- MANSKI, C.F. (1981), "Learning and Decision Making When Subjective Probabilities Have Subjective Domains," *Annals of Statistics*, **9**, 59-65.
- MARTZ, H.F., and WALLER, R.A. (1982), *Bayesian Reliability Analysis*, New York: Wiley.
- O'HAGAN, A. and BERGER, J. (1987), "Ranges of Posterior Probabilities for Quasi-Unimodal Priors with Specified Quantiles," to appear in the *Journal of the American*

*Statistical Association.*

- POLASEK, W. (1985), "Sensitivity Analysis for General and Hierarchical Linear Regression Models", in *Bayesian Inference and Decision Techniques with Applications*, eds. P.K. Goel and A. Zellner, Amsterdam: North-Holland.
- SIVAGANESAN, S. (1986a), "Robust Bayes Analysis with Arbitrary Contaminations", Technical Report No. SMU-DS-TR-198, Department of Statistical Science, Southern Methodist University.
- SIVAGANESAN, S. (1986b), "Sensitivity Of The Posterior Mean To Unimodality Preserving Contaminations", Technical Report No. SMU-DS-TR-199, Department of Statistical Science, Southern Methodist University.
- SIVAGANESAN, S., and BERGER, J. (1986), "Ranges Of Posterior Measures For Priors With Unimodal Contaminations", Technical Report #86-41, Department of Statistics, Purdue University.
- WALLEY, P. (1986), *Rationality and Vagueness*, Manuscript in preparation.
- WEST, S. (1979), "Upper and Lower Probability Inferences for the Logistic Function," *Annals of Statistics*, **7**, 490-413.
- WOLFENSON, M., and FINE, T.L. (1982), "Bayes-Like Decision Making With Upper And Lower Probabilities", *Journal of the American Statistical Association*, **77**, 80-88.