BAYESIAN ANALYSIS WITH LIMITED COMMUNICATION*

by

James O. Berger and Julia Mortera
Purdue University University of Rome

Technical Report #89–38C

Department of Statistics
Purdue University

November 1989
Revised May 1990

* Research supported by the National Science Foundation, Grant DMS–8702620 and by MPI projects and CNR funds.

◇

Abstract

The ith member of a group of m individuals (or stations) observes a random quantity X_i, where $\underset{\sim}{X} = (X_1, \ldots, X_m)$ has density $g(\underset{\sim}{x}|\theta)$. Each individual can report only $y_i = h_i(x_i)$, because of a limitation on the amount of information that can be communicated. Based on $\underset{\sim}{y} = (y_1, \ldots, y_m)$ and a prior distribution $\pi(\theta)$, Bayesian inference or decision concerning θ is to be undertaken.

The first version of this problem that will be studied is the "team" problem, where the m individuals form a team with common prior π and the reports, y_i, are the posterior distributions of each team member. We compare the optimal Bayesian posterior for this problem ($\pi(\theta|\underset{\sim}{y})$) with previous suggestions, such as the optimal linear opinion pool.

The second facet of the problem that is explored is that of choosing $\underset{\sim}{y}$ to optimize the information communicated, subject to a constraint on the amount of information that can be communicated. In particular, we will consider the dichotomous case, in which each y_i can be only 0 or 1, and will illustrate the optimal choice of y_i for both inference and decision criteria. The inference criterion considered will be closeness of the posteriors $\pi(\theta|\underset{\sim}{x})$ and $\pi(\theta|\underset{\sim}{y})$, in an expected Kullback–Leibler sense, while the decision criterion considered will be usual optimality with respect to overall expected loss. Examples are presented, including discussion of a situation that arises in reliability demonstration.

◇

1. INTRODUCTION

1.1. Limited Communication

Consider a group of $m(m \geq 1)$ individuals, each observing a random quantity X_i, where $\underset{\sim}{X} = (X_1, \ldots, X_m)$ has density $g(\underset{\sim}{x}|\theta)$ and θ is an unknown parameter. Suppose that there is a constraint on the amount of information that can be communicated in that individual i, $i = 1, \ldots, m$, can only report $y_i = h_i(x_i)$. There are a variety of situations in which such limited communication can arise. One such is the "team problem" (see Section 1.2) in which each individual reports only his posterior density for θ based on x_i. Another such is when communication is expensive, and each individual can, say, report only one bit of information ($y_i = 0$ or 1); see Section 1.3. A third such scenario is when, for reasons of confidentiality or secretivity, only the limited y_i can be reported. We will denote the joint density of $\underset{\sim}{Y} = (Y_1, \ldots, Y_m)$ (induced from $g(\underset{\sim}{x}|\theta)$ by the transformation $(h_1, \ldots, h_m)$) by $f(\underset{\sim}{y}|\theta)$. Based on $\underset{\sim}{y} = (y_1, \ldots, y_m)$ and a prior distribution $\pi(\theta)$, Bayesian inference or decision concerning θ is to be undertaken.

Interest will focus on comparing the reported posterior distribution under the available limited information, $\pi(\theta|\underset{\sim}{y})$, and the full posterior distribution under all information, $\pi(\theta|\underset{\sim}{x})$. When the parameter of interest is $\eta = \psi(\theta)$, we will compare the limited information posterior distribution, $\pi(\eta|\underset{\sim}{y})$, with the full posterior distribution $\pi(\eta|\underset{\sim}{x})$. (An example of this that is related to reliability demonstration is given in Section 3.3.)

1.2. The Team Problem.

The first limited information scenario that will be considered is that of combining individual reports expressed as posterior distributions, when the group of individuals forms a team. The concept of a team was introduced, from an economics viewpoint, by Marschak and Radner (1971) and developed from a statistical viewpoint in DeGroot and Mortera (1988). A team is a group of m individuals who each report their posterior distribution, $\pi(\theta|x_i)$, based on a common prior distribution $\pi(\theta)$, a common joint model $g(\underset{\sim}{x}|\theta)$, $\underset{\sim}{x} = (x_1, \ldots, x_m)$, but on different data sets x_i, $i = 1, \ldots, m$. A single final distribution for θ must be chosen by the team by combining the individual posterior distributions, $\pi(\theta|x_i)$.

This can be formulated in the framework of Section 1.1 by defining y_i as $\pi(\theta|x_i)$. (The implicitly defined h_i will be described in Section 2.) The suggested final pooled report is then $\pi(\theta|\underset{\sim}{y})$, which is clearly the optimal pooled report. This approach will be illustrated in Section 2, with two examples, and compared with previous suggestions for pooling.

1.3. Optimal Limited Communication.

If the amount of information that can be communicated by each individual is limited, it is natural to seek the optimal choice of information to communicate: i.e., the optimal choice of the h_i for the reports $\underset{\sim}{y} = (y_1, \ldots, y_m) = (h_1(x_1), \ldots, h_m(x_m))$, subject to the constraints on the amount of information that can be reported. This will be considered in Section 3 for the case where each y_i can only be one bit of information, that is $y_i = 0$ or 1. Such y_i arise as

$$y_i = h_i(x_i) = \begin{cases} 1 & \text{if } x_i \in C_i \\ 0 & \text{if } x_i \notin C_i, \end{cases}$$

$i = 1, \ldots, m$, so that optimal choice of the h_i is equivalent to optimally choosing $\underset{\sim}{C} = (C_1, \ldots, C_m)$.

For general inference about θ, it is natural to define "optimal" in terms of closeness between the limited information posterior density, $\pi(\theta|\underset{\sim}{y})$, and the full (but unobtainable) posterior density $\pi(\theta|\underset{\sim}{x})$. The Kullback–Leibler measure of distance between $\pi(\theta|\underset{\sim}{y})$ and $\pi(\theta|\underset{\sim}{x})$ will be used to measure closeness. Since neither $\underset{\sim}{X}$ nor $\underset{\sim}{Y}$ will be known when $\underset{\sim}{C}$ is chosen, it is necessary to consider the expected Kullback–Leibler distance, the expectation being over $\underset{\sim}{X}$ as well as θ. Thus we seek to choose $\underset{\sim}{C}$ to minimize

$$E^{\underset{\sim}{X},\theta}\left\{\log\left[\frac{\pi(\theta|\underset{\sim}{X})}{\pi(\theta|\underset{\sim}{Y})}\right]\right\}.$$

This is equivalent to choosing $\underset{\sim}{C}$ so as to minimize

$$\Lambda(\underset{\sim}{C}) = -E^{\underset{\sim}{X},\theta}\left[\log\left(\frac{\pi(\theta|\underset{\sim}{Y})}{\pi(\theta)}\right)\right]. \tag{1.1}$$

(Only the term involving $\underset{\sim}{Y}$ depends on C; $\pi(\theta)$ has been included for later notational convenience. For related uses of similar measures see Bernardo (1979a,b) and Zellner (1977). At times, we will refer to $\Lambda(\underset{\sim}{C})$ as the shifted Kullback-Leibler distance.) This will be considered in Section 3.2. A version applicable to inference about $\eta = \psi(\theta)$ will be considered in Section 3.3.

Instead of inference, one may face a decision problem, with loss function $L(\theta, a)$ for action $a \in \mathcal{A}$. For specified $\underset{\sim}{C}$, a Bayes decision rule $\delta_{\underset{\sim}{C}}(\underset{\sim}{y})$, is given by any action that minimizes the posterior expected loss. Before observing $\underset{\sim}{Y}$ (or $\underset{\sim}{X}$), the overall expected loss for a given choice of $\underset{\sim}{C}$ is thus the frequentist Bayes risk

$$r(\underset{\sim}{C}) = E^{\underset{\sim}{X},\theta}[L(\theta, \delta_{\underset{\sim}{C}}(\underset{\sim}{Y})]. \tag{1.2}$$

In a decision problem, therefore, an optimal $\underset{\sim}{C}$ will be one which minimizes $r(\underset{\sim}{C})$. An example will be given in Section 3.4.

Note that there is a certain similarity here to optimal Bayesian design, which seeks to optimally allocate a limited number of possible observations among possible design points. Indeed a formal analogy could be made by defining $\underset{\sim}{X}$ to be the set of all observations at all possible allocations, and having the limited information constraint be that only the observations from a single allocation can be reported. Literature on Bayesian design includes Smith and Verdinelli (1980), Pilz (1983), Chaloner (1984), and DeGroot and Goel (1988). Example 4 in Section 3.3 is also related to design.

2. THE TEAM PROBLEM

In DeGroot and Mortera (1988), the optimal rule for combining the team members individual posterior reports, $\pi(\theta|x_i)$, $i = 1, \ldots, m$, was given when the X_i are conditionally independent given θ. This is equivalent (for the conditionally independent case) to our $\pi(\theta|\underset{\sim}{y})$ defined below. When the X_i are not conditionally independent given θ, DeGroot and Mortera (1988) considered use of a linear opinion pool to combine the $\pi(\theta|x_i)$, and determined the optimal weights for the pool. Here we derive the optimal pooled report, $\pi(\theta|\underset{\sim}{y})$, and compare it with the optimal linear opinion pool. The optimal pooled report can be dramatically better. (Note, however, that there might be computational and other reasons to use the suboptimal linear opinion pool approach. For general reviews of pooling, see French (1985) and Genest and Zidek (1986).)

Analogously to Winkler (1968), French (1980), Lindley (1983), and Lindley and Singpurwalla (1986), the optimal Bayesian pool of $\pi(\theta|x_1), \ldots, \pi(\theta|x_m)$ is found by treating these as the "data" $\underset{\sim}{y} = (y_1, \ldots, y_m)$, and then determining $\pi(\theta|\underset{\sim}{y})$. More formally, define Y_i as a *minimal Bayes sufficient* statistic corresponding to X_i and the prior $\pi(\theta)$. In other words, $Y_i = h_i(X_i)$ is a statistic such that $\pi(\theta|X_i) = \pi(\theta|Y_i)$ with probability one (i.e., Y_i is Bayes sufficient with respect to π) and, if $y_i \neq y_i'$, then $\pi(\theta|y_i)$ and $\pi(\theta|y_i')$ differ (i.e., Y_i is minimal). The information actually conveyed by $\pi(\theta|x_i)$ is thus y_i. Note that, because of minimality, y_i can be retrieved from $\pi(\theta|x_i)$ (whereas x_i may not be retrievable). Note also that there is no need for each individual to have the same prior π; if each has a *known* prior $\pi_i(\theta)$, and Y_i is minimal Bayes sufficient with respect to X_i and π_i, then the following analysis still holds. (In this case, it is assumed that there is a central decision maker with prior $\pi(\theta)$ that is doing the analysis.)

Since $\underset{\sim}{Y} = (Y_1, \ldots, Y_m)$ is a statistic, one can determine its density $f(\underset{\sim}{y}|\theta)$, and then calculate the posterior density $\pi(\theta|\underset{\sim}{y}) \propto f(\underset{\sim}{y}|\theta)\pi(\theta)$. This will be the optimal pooling of $\pi(\theta|x_1), \ldots, \pi(\theta|x_m)$.

This approach will be illustrated on two examples in which $\underset{\sim}{Y}$ has lost some of the information in $\underset{\sim}{X}$. In Example 1, the loss of information arises because the team members have some common data, while in Example 2 the loss of information results from the team members separately eliminating a relevant nuisance parameter in determining $\pi(\theta|x_i)$, $i = 1, \ldots, m$. Thus the information sources are dependent (cf., Winkler (1981), Clemen (1987)).

Example 1. Overlapping normal samples.

As in DeGroot and Mortera (1988), consider a 2 person team where both members observe c common observations, $\underset{\sim}{u} = (u_1, \ldots, u_c)$, and n_1 and n_2 "private" observations, $\underset{\sim}{v} = (v_1, \ldots, v_{n_1})$ for the first member and $\underset{\sim}{z} = (z_1, \ldots, z_{n_2})$ for the second member. All observations are iid $N(\theta, \frac{1}{\tau})$, τ known. The common prior density for θ is $N(\alpha, \frac{1}{h})$. Let $\overline{x}_1 = (c\overline{u} + n_1\overline{v})/(c + n_1)$ and $\overline{x}_2 = (c\overline{u} + n_2\overline{z})/(c + n_2)$ denote each team member's sample mean, respectively.

The individual posterior reports are

$$\pi(\theta|x_i) = N\left[\frac{h\alpha + \nu_i \overline{x}_i}{h+\nu_i},\ \frac{1}{h+\nu_i}\right], \tag{2.1}$$

where $\nu_i = \tau(c+n_i),\ i = 1,2$. Clearly, the minimal Bayes sufficient statistics are (up to one–to–one transformations) the sample means $\overline{X}_1$ and $\overline{X}_2$. For notational simplicity we set $Y_i = (c+n_i)\overline{X}_i,\ i = 1,2$.

The joint density of Y_1 and Y_2 is

$$f(y_1, y_2|\theta) = N\left(\begin{bmatrix}(c+n_1)\theta \\ (c+n_2)\theta\end{bmatrix},\ \frac{1}{\tau}\begin{bmatrix}(c+n_1) & c \\ c & (c+n_2)\end{bmatrix}\right).$$

The team's pooled posterior report,

$$\pi_L(\theta) = \pi(\theta|y_1, y_2)\ \propto\ f(y_1, y_2|\theta)\pi(\theta),$$

can easily be shown to be

$$\pi_L(\theta) = N[\mu_L,\ \frac{1}{\nu_L}], \tag{2.2}$$

where

$$\begin{aligned} \nu_L &= h + \frac{\tau(n_1+n_2)(c+n_1)(c+n_2)}{c(n_1+n_2)+n_1n_2}, \\ \mu_L &= \frac{1}{\nu_L}\left[\alpha h + \frac{\tau n_1(c+n_2)y_1 + \tau n_2(c+n_1)y_2}{c(n_1+n_2)+n_1n_2}\right]. \end{aligned} \tag{2.3}$$

The full, but unobtainable, posterior distribution, given all the information $\underset{\sim}{u}$, $\underset{\sim}{v}$ and $\underset{\sim}{z}$, is

$$\pi_F(\theta) = \pi(\theta|\underset{\sim}{u},\ \underset{\sim}{v},\ \underset{\sim}{z}) = N[\mu_F,\ \frac{1}{\nu_F}], \tag{2.4}$$

where

$$\begin{aligned} \nu_F &= h + (n_1+n_2+c)\tau, \\ \mu_F &= \frac{1}{\nu_F}[\alpha h + \tau(c\overline{u} + n_1\overline{v} + n_2\overline{z})]. \end{aligned} \tag{2.5}$$

One can easily see, from (2.3) and (2.5), that, if $c > 0$, the posterior variance given the full information, $1/\nu_F$, is always smaller than the posterior variance based on the individual reports, $1/\nu_L$.

The optimal linear opinion pool is given (see DeGroot and Mortera, 1988) by

$$\pi_{LP}(\theta) = w\pi(\theta|x_1) + (1-w)\pi(\theta|x_2), \tag{2.6}$$

where the $\pi(\theta|x_i)$, $i = 1, 2$, are given by (2.1) and

$$w = \frac{[h+\tau(c+n_1)]^{1/2} - \left[\frac{(h+\tau(c+n_1))(h+\tau(c+n_2))}{h+\tau(c+n_1+n_2)}\right]^{1/2}}{[h+\tau(c+n_1)]^{1/2} + [h+\tau(c+n_2)]^{1/2} - 2\left[\frac{(h+\tau(c+n_1))(h+\tau(c+n_2))}{h+\tau(c+n_1+n_2)}\right]^{1/2}}.$$

A variety of numerical comparisons of π_L, π_F, and π_{LP} were performed. Figure 1 is typical, presenting the three posteriors (labelled L, F, and LP, respectively) for randomly generated data when $\theta = 2$, $c = 1$, $n_1 = 2$, $n_2 = 1$, $\alpha = 0$, and $\tau = h = 1$. The variance associated with π_{LP} is greater than that for π_L which, in turn, is greater than that for π_F. Also, π_L is clearly closer to π_F than is π_{LP}, but the latter is not greatly different. These conclusions held for most of the cases analyzed.

Example 2. Normal Variance.

In this example, the team members' reported posteriors result is a loss of information because individual elimination of nuisance parameters yields Y_i's that are not jointly sufficient for the full parameter. Assume a 2 person team where each member observes n_i iid $X_{ij} \sim N(\mu, \tau)$, $j = 1, \ldots, n_i$, $i = 1, 2$. The common prior distribution on μ and τ is given by the improper prior density

$$\pi(\mu, \tau) \propto \tau^{\alpha-1}e^{-\beta\tau};$$

that is, τ has a Gamma distribution with parameters α and β (i.e., $\tau \sim \Gamma(\alpha, \beta)$) and $\mu|\tau \sim$ constant.

Suppose that μ is a nuisance parameter, so that the individual posterior reports for the parameter of interest, τ, after integrating out μ, are

$$\pi(\tau|\underset{\sim}{x}_i) = \Gamma(\alpha + (n_i-1)/2,\ \beta + s_i^2/2), \tag{2.7}$$

where $\underset{\sim}{x}_i = (x_{i1}, \ldots, x_{in_i})$, $s_i^2 = \sum_{j=1}^{n_i}(x_{ij} - \overline{x}_i)^2$ and $\overline{x}_i = \left(\sum_{j=1}^{n_i} x_{ij}\right)/n_i$, $i = 1, 2$. Thus the minimal Bayes sufficient statistics (for each individual) are $Y_i = S_i^2$, $i = 1, 2$. Computation yields (noting that the Y_i/τ are independently $\chi^2_{(n_i-1)}$)

$$\pi_L = \pi(\tau|y_1, y_2) = \Gamma(\alpha_L, \beta_L), \tag{2.8}$$

where $\alpha_L = \alpha + \frac{1}{2}(n_1 + n_2 - 2)$, $\beta_L = \beta + \frac{1}{2}(s_1^2 + s_2^2)$. Finally, the (unobtainable) full posterior distribution can be shown to be

$$\pi_F = \pi(\tau|\overline{x}_1, \overline{x}_2, s_1^2, s_2^2) = \Gamma(\alpha_F, \beta_F), \tag{2.9}$$

where

$$\alpha_F = \alpha_L + \frac{1}{2}, \; \beta_F = \beta_L + \frac{n_1 n_2}{2(n_1 + n_2)}(\overline{x}_1 - \overline{x}_2)^2.$$

A variety of numerical comparisons of π_L, π_F, and π_{LP} were performed. Figure 2 is typical, presenting the three posteriors (labelled L, F, and LP, respectively) for randomly generated data when $\tau = 2$, $n_1 = n_2 = 20$, and $\alpha = \beta = 1$. (Because $n_1 = n_2$ and $\alpha = \beta$, it can be shown that the optimal linear opinion pool is $\pi_{LP} = \frac{1}{2}\pi(\tau|\underset{\sim}{x}_1) + \frac{1}{2}\pi(\tau|\underset{\sim}{x}_2)$.) In this example, π_{LP} is markedly inferior to π_L, and indeed is almost bimodal. In contrast, π_L is a very accurate approximation to π_F.

3. OPTIMAL LIMITED COMMUNICATION

In this section we assume that only one bit of information can be reported by each individual, so that the reports are

$$y_i = \begin{cases} 1 & \text{if } x_i \in C_i \\ 0 & \text{if } x_i \notin C_i, \end{cases}$$

for $i = 1, \ldots, m$. The goal is to choose $\underset{\sim}{C} = (C_1, \ldots, C_m)$ optimally: i.e., to minimize expected Kullback–Leibler distance between $\pi(\theta|\underset{\sim}{Y})$ and $\pi(\theta|\underset{\sim}{X})$ for inference problems, and to minimize expected Bayes risk for decision problems. These are discussed in Sections 3.2 and 3.4, respectively. Section 3.1 presents needed expressions involving $\pi(\theta|\underset{\sim}{y})$. Section 3.3 considers inference for a function $\eta = \psi(\theta)$.

3.1. Preliminaries

It will be assumed throughout the section that the X_i's, and hence the Y_i's, are conditionally independent given θ, so that

$$f(\underset{\sim}{y}|\theta) = \prod_{i=1}^{m} f(y_i|\theta),$$

$$f(y_i|\theta) = \begin{cases} p_\theta(C_i) = Pr(X_i \in C_i|\theta) & \text{if } y_i = 1 \\ 1 - p_\theta(C_i) = Pr(X_i \notin C_i|\theta) & \text{if } y_i = 0. \end{cases}$$

Thus $f(\underset{\sim}{y}|\theta)$ can be written

$$f(\underset{\sim}{y}|\theta) = \prod_{i=1}^{m} p_\theta^{y_i}(C_i)(1 - p_\theta(C_i))^{1-y_i}. \tag{3.1}$$

The marginal density of $\underset{\sim}{Y}$, for a prior density π, is given by

$$f(\underset{\sim}{y}, \underset{\sim}{C}) = \int f(\underset{\sim}{y}|\theta)\pi(\theta)d\theta. \tag{3.2}$$

(We include $\underset{\sim}{C}$ as an argument here, and in the following, because the ultimate problem will involve optimation over $\underset{\sim}{C}$.) The final posterior distribution is thus

$$\pi(\theta|\underset{\sim}{y}) = K(\underset{\sim}{y}, \underset{\sim}{C})\pi(\theta) \prod_{i=1}^{m} p_\theta^{y_i}(C_i)(1 - p_\theta(C_i))^{1-y_i}, \tag{3.3}$$

where the normalizing constant is

$$K(\underset{\sim}{y}, \underset{\sim}{C}) = [f(\underset{\sim}{y}, \underset{\sim}{C})]^{-1}. \tag{3.4}$$

When the X_i's are identically distributed and a common choice $C_i = C$, for $i = 1, \ldots, m$, is made, (3.3) reduces to

$$\pi(\theta|\underset{\sim}{y}) = K(t(C), C)\pi(\theta)[p_\theta(C)]^{t(C)}[1 - p_\theta(C)]^{m-t(C)}, \tag{3.5}$$

where

$$p_\theta(C) = Pr(X_i \in C|\theta), \; t(C) = \sum_{i=1}^{m} y_i$$

and, abusing notation,

$$K(t(C),C) = \left[\int \pi(\theta)[p_\theta(C)]^{t(C)}[1-p_\theta(C)]^{m-t(C)}d\theta\right]^{-1}. \tag{3.6}$$

The marginal density of $t(C)$, in this case, is

$$f(t(C),C) = \binom{m}{t}/K(t(C),C). \tag{3.7}$$

It will prove convenient to use the notation

$$\mathcal{E}n(\underset{\sim}{v}) = -\sum_{i=1}^{k} v_i \log v_i,$$

when $\underset{\sim}{v} = (v_1,\ldots,v_k)$. This will be used when the v_i are the probabilities associated with a discrete random V, so that $\mathcal{E}n(\underset{\sim}{v})$ is then just the entropy of V. We will also use the notation

$$1_\Omega(w) = \begin{cases} 1 & \text{if } w \in \Omega \\ 0 & \text{if } w \notin \Omega. \end{cases}$$

3.2. Inference About θ.

When inference about θ is of interest, an optimal choice of $\underset{\sim}{C}$, according to expected Kullback–Leibler distance, is any $\underset{\sim}{C}$ that minimizes $\Lambda(\underset{\sim}{C})$ of (1.1).

Lemma 3.1. For the situation defined by (3.1) through (3.3),

$$\begin{aligned}\Lambda(\underset{\sim}{C}) &= -E^{\theta,\underset{\sim}{X}}[\log(\pi(\theta|\underset{\sim}{Y})/\pi(\theta))] \\ &= -\sum_{i=1}^{m} E^\theta\{p_\theta(C_i)\log p_\theta(C_i) + [1-p_\theta(C_i)]\log[1-p_\theta(C_i)]\} - \mathcal{E}n[f(\underset{\sim}{y},\underset{\sim}{C})]. \end{aligned} \tag{3.8}$$

Proof. From (3.3), one has that

$$\Lambda(\underset{\sim}{C}) = -E^{\underset{\sim}{Y}}[\log K(\underset{\sim}{Y},\underset{\sim}{C})] - \sum_{i=1}^{m} E^\theta\{E^{Y_i|\theta}(Y_i)\log p_\theta(C_i) + E^{Y_i|\theta}(1-Y_i)\log[1-p_\theta(C_i)]\}.$$

From (3.4),

$$\begin{aligned} E^{\underset{\sim}{Y}}[\log K(\underset{\sim}{Y},\underset{\sim}{C})] &= -\sum_{\text{all } \underset{\sim}{y}} [K(\underset{\sim}{y},\underset{\sim}{C})]^{-1}\log K(\underset{\sim}{y},\underset{\sim}{C}) \\ &= \sum_{\text{all } \underset{\sim}{y}} f(\underset{\sim}{y},\underset{\sim}{C})\log f(\underset{\sim}{y},\underset{\sim}{C}) \\ &= -\mathcal{E}n[f(\underset{\sim}{y},\underset{\sim}{C})], \end{aligned}$$

and $E^{Y_i|\theta}(Y_i) = p_\theta(C_i)$, thus proving (3.8). □

Corollary 3.1. If $C_i = C$ for all i,

$$\Lambda(\underset{\sim}{C}) = -mE^{\theta}[p_\theta(C)\log p_\theta(C) + (1-p_\theta(C))\log(1-p_\theta(C))] - E^t[\log(K(t(C),C)]. \quad (3.9)$$

Also,

$$\begin{aligned} E^t[\log(K(t(C),C)] &= \sum_{t=0}^{m} \binom{m}{t} \frac{\log K(t,C)}{K(t,C)} \\ &= \mathcal{E}n(f(t(C),C)) + E^t[\log \binom{m}{t}]. \end{aligned} \quad (3.10)$$

Proof: Equation (3.9) follows immediately from Lemma 3.1, and (3.10) follows directly from (3.7). □

Example 3. Uniform distribution.

Suppose that X_i, $i = 1,\ldots,m$, are i.i.d. $U[0,\theta]$ random variables, and consider the choice $C_i = C = [c,\infty)$. (Note that this is equivalent to the choice $C_i = C = [0,c)$ because of symmetry in (3.9).) Here

$$p_\theta(C) = Pr(X_i \in C|\theta) = \begin{cases} 1 - c/\theta & \text{for } c < \theta \\ 0 & \text{for } c \geq \theta. \end{cases}$$

Assuming the natural conjugate Pareto prior density for θ,

$$\pi(\theta) = \alpha\theta^{-(\alpha+1)} w_0^{\alpha}\, 1_{[w_0,\infty)}(\theta), \quad \alpha > 0 \text{ and } 0 < w_0 \leq \theta,$$

the final posterior density (3.5) is

$$\pi(\theta|\underset{\sim}{y}) = K(t(c),c)\frac{\alpha w_0^{\alpha}}{\theta^{\alpha+1}}\left(1-\frac{c}{\theta}\right)^{t(c)}\left(\frac{c}{\theta}\right)^{m-t(c)} 1_{[w_0,\infty)}(\theta),$$

where $K(t(c),c)$ is the normalizing constant defined in (3.6) and given in equation (A.1) of Appendix A1.

The following lemma gives a useful expression for $\Lambda(\underset{\sim}{C})$ in this example. For use in the lemma, define

$$r = \frac{c}{w_0} \quad \text{and} \quad a_{ht} = (-1)^{t-h}\binom{t}{h}\frac{\alpha}{(m-h+\alpha)}. \tag{3.11}$$

Lemma 3.2. For the uniform example,

$$\Lambda(C) = \begin{cases} \Lambda_1(c) & \text{for } r \geq 1 \\ \Lambda_2(c) & \text{for } r < 1, \end{cases} \tag{3.12}$$

where

$$\begin{aligned} \Lambda_1(c) = & -r^{-\alpha}\sum_{t=1}^{m}\binom{m}{t}\left[\sum_{h=0}^{t} a_{ht}\right]\left[\alpha \log r - \log(\sum_{h=0}^{t} a_{ht})\right] \\ & + [1 - r^{-\alpha}(\frac{m}{m+\alpha})]\log\left[1 - r^{-\alpha}(\frac{m}{m+\alpha})\right] - r^{-\alpha}\frac{m}{(\alpha+1)}\sum_{i=1}^{\alpha}\frac{1}{i}, \end{aligned} \tag{3.13}$$

and

$$\begin{aligned} \Lambda_2(c) = & -\frac{m}{(\alpha+1)}\left\{(\alpha + 1 - r^{-\alpha})\log(1-r) + \alpha r[\log r - \log(1-r)] + \sum_{i=0}^{\alpha-1}\frac{r^{-i}}{(i-\alpha)}\right\} \\ & + \sum_{t=0}^{m}\binom{m}{t}\left(\sum_{h=0}^{t} a_{ht} r^{m-h}\right)\log\left(\sum_{h=0}^{t} a_{ht} r^{m-h}\right). \end{aligned} \tag{3.14}$$

Proof. See Appendix A1. □

Lemma 3.3. For $c > w_0$ (i.e., $r > 1$), the minimum of $\Lambda_1(c)$ is attained at

$$c_{\min} = w_0[\frac{m}{(m+\alpha)} + \exp\{(1 + \frac{\alpha}{m})A_m\}]^{1/\alpha}, \tag{3.15}$$

where

$$A_m = \frac{m}{(\alpha+1)}\sum_{i=1}^{\alpha}\frac{1}{i} + \sum_{t=1}^{m}\binom{m}{t}\left(\sum_{h=0}^{t} a_{ht}\right)\log\left(\sum_{h=0}^{t} a_{ht}\right). \tag{3.16}$$

Proof: Differentiating $\Lambda_1(c)$ in (3.13) with respect to c and setting it equal to 0, after some algebra one obtains (3.15). □

It seems likely that c_{min} in (3.15) is actually the global minimum, because $\Lambda_2(c)$ appears to be monotonically decreasing in c. This is proved in Appendix A2 for the case $m = 1$. In all of the many numerical studies we performed for $m > 1$, this was also true.

Figure 3 presents a typical numerical example of $\Lambda(C)$ for various values of m. The graphs are presented as functions of $r = c/w_0$, for convenience. Observe that, as m increases (i.e., more individuals report information), $r_{\text{min}} = c_{\text{min}}/w_0$ moves closer to 1. It was also observed in the numerical studies that, as α increases (i.e., the prior becomes more concentrated), $\Lambda(C)$ becomes more sharply peaked, implying a greater sensitivity to the choice of C.

3.3. Inference About a Function of θ

Often a function $\eta = \psi(\theta)$, and not θ itself, is of primary interest. We illustrate this possibility here for the situation of testing H_0: $\theta \in A$ vs. H_1: $\theta \notin A$. The quantity of interest is then $\eta = 1_A(\theta)$, where $1_A(\theta)$ is the indicator function of the set A. In this case, (1.1) and the development in Section 3.1 should be applied to η, rather than to θ.

For simplicity, we further restrict consideration to the case where the reports are, independently, $Y_i = 1_C(X_i)$. Then (3.5) and (3.6) give the posterior for θ, so that the posterior for η is given by

$$\begin{aligned}
&\pi(0|\underset{\sim}{y}) = 1 - \pi(1|\underset{\sim}{y}), \\
&\pi(1|\underset{\sim}{y}) = \pi(1|t(C)) = K(t(C), C) \int_A \pi(\theta)[p_\theta(C)]^{t(C)}[1 - p_\theta(C)]^{m-t(C)} d\theta. \qquad (3.17)
\end{aligned}$$

Thus (1.1) becomes (ignoring, for convenience, the constant term $E^\eta[\log \pi(\eta)]$)

$$\begin{aligned}
\Lambda(C) &= -E^{\eta,\underset{\sim}{Y}}[\log \pi(\eta|\underset{\sim}{Y})] \\
&= -E^t[E^{\eta|t}[\log \pi(\eta|\underset{\sim}{Y})]] \\
&= -\sum_{t=0}^{m} \binom{m}{t} [K(t,c)]^{-1} [\pi(0|t) \log \pi(0|t) + \pi(1|t) \log \pi(1|t)] \\
&= \sum_{t=0}^{m} \binom{m}{t} [K(t,c)]^{-1} \mathcal{E}n[\pi(\eta|t)]. \qquad (3.18)
\end{aligned}$$

Again, the goal is to minimize $\Lambda(C)$ over the choice of C. An example follows.

Example 4. Exponential Distribution.

Suppose X_i, $i = 1, \ldots, m$, are i.i.d. $\Gamma(1, \theta)$ and $C_i = C = [0, c)$. Then

$$f(y_i|\theta) = \begin{cases} 1 - e^{-c\theta} & \text{if } y_i = 1 \\ e^{-c\theta} & \text{if } y_i = 0. \end{cases}$$

Suppose further that $\eta = 1_A(\theta)$, with $A = [a, \infty)$, $a > 0$. Finally, consider the natural conjugate prior density for θ, $\pi(\theta) = \Gamma(\alpha, \beta)$ with $\alpha > 0$, $\beta > 0$.

This situation arises in reliability demonstration (cf. Mann, Schafer and Singpurwalla (1974)) where c is the termination (truncation) time of life tests X_i, only "failure" or "nonfailure" ($Y_i = 1$ or 0, respectively) are reported, and a^{-1} is the desired level of reliability. Our problem corresponds to the optimal design choice, c.

Lemma 3.4. The posterior distribution of η is defined by $\pi(1|t) = 1 - \pi(0|t)$ and

$$\pi(0|t) = K(t, c) \sum_{h=0}^{t} a_{h,t}(c) \Gamma_\alpha(a[c(m-h) + \beta]), \tag{3.19}$$

where $t = \sum_{i=1}^{m} y_i$,

$$\Gamma_\alpha(v) = \frac{1}{\Gamma(\alpha)} \int_0^v \theta^{\alpha-1} e^{-\theta} d\theta,$$

$$a_{h,t}(c) = \binom{t}{h} (-1)^{t-h} [\frac{c}{\beta}(m-h) + 1]^{-\alpha},$$

$$K(t, c) = \left[\sum_{h=0}^{t} a_{h,t}(c)\right]^{-1}. \tag{3.20}$$

Proof. See Appendix A3. □

Using (3.19) and (3.20), $\Lambda(C)$ in (3.18) can easily be calculated. Numerous cases were investigated; Figure 4 presents a typical graph of $\Lambda(C)$ for various m. It is interesting that increasing m (i.e., increasing the number of reports y_i) has almost no effect on the optimal choice of c ($c_{\min}$). Also, as c increases, $\Lambda(C)$ decreases only by a relatively small amount, indicating that each reported Y_i carries (in this example) relatively little information about η.

It was observed in other numerical examples (not given here), that $c_{\min}$ was strongly affected by the prior mean α/β, but was only slightly affected by the prior variance α/β^2. Also, $\Lambda(C)$ became slightly more peaked as the prior variance decreased.

A natural question to consider in this example is whether the report of an interval such as (c_1, c_2) would be superior to the report of $[0, c)$. Unfortunately, we could not answer this question in general, but the following theorem shows that $[0, c)$ suffices for the case $m = 1$, $\alpha = 1$.

Theorem 3.1. If $m = 1$ and $\alpha = 1$, $\inf_{(c_1, c_2)} \Lambda((c_1, c_2))$ is attained at an interval for which $c_2 = \infty$ or, equivalently by symmetry, for which $c_1 = 0$.

Proof. See Appendix A4. □

3.4. Decision Analysis.

As stated in Section 1.4, the optimal $\underset{\sim}{C}$ in a decision problem will be the $\underset{\sim}{C}$ that minimizes the frequentist Bayes risk given in (1.2). An example follows.

Example 5. Suppose $m = 2$, where the X_i, $i = 1, 2$, are i.i.d. $\Gamma(1, \theta)$. Suppose that the reports are $Y_i = 1_{[0, c_i)}(\theta)$, where c_1 and c_2 are allowed to differ. Define $\underset{\sim}{c} = (c_1, c_2)$. Note that, for $i = 1, 2$,

$$f(y_i|\theta) = \begin{cases} 1 - e^{-c_i\theta} & \text{if } y_i = 1 \\ e^{-c_i\theta} & \text{if } y_i = 0. \end{cases} \tag{3.21}$$

Suppose that the prior for θ is the natural conjugate prior $\pi(\theta) = \Gamma(1, \beta)$, and consider the decision problem of estimating θ under the quadratic loss $L(\theta, a) = (\theta - a)^2$. The Bayes decision rule, $\delta_{\underset{\sim}{c}}(\underset{\sim}{y})$, is simply the posterior mean and the frequentist Bayes risk can be written

$$r(\underset{\sim}{c}) = \sum_{\text{all } \underset{\sim}{y}} m(\underset{\sim}{y}) \text{ Var } (\theta|\underset{\sim}{y}), \tag{3.22}$$

where $m(\underset{\sim}{y})$ is the marginal probability of $\underset{\sim}{Y}$ and $\text{Var}(\theta|\underset{\sim}{y})$ is the posterior variance. (This formula for $r(\underset{\sim}{c})$ follows easily from the observation that $r(\underset{\sim}{c})$ is the expectation over $\underset{\sim}{Y}$ of the posterior expected loss given $\underset{\sim}{Y}$, and the posterior expected loss here is simply the posterior variance.)

To simplify notation, we will make the transformations $\theta' = \beta\theta$, $X_i' = X_i/\beta$, and define $r_1 = c_1/\beta$, $r_2 = c_2/\beta$. Then the X_i' are i.i.d. $\Gamma(1,\theta')$, $\pi(\theta') = \Gamma(1,1)$, and the reports are $Y_i = 1_{[0,r_i)}(X_i')$. The posterior distributions, $\pi(\theta'|\underset{\sim}{y})$, and the marginal probabilities, $m(\underset{\sim}{y})$, are given in Appendix 5. The posterior means and variances for θ' are therein shown to be

$$\delta'_{\underset{\sim}{c}}(\underset{\sim}{y}) = \begin{cases} 1 + (1+r_1)^{-1} + (1+r_2)^{-1} + (1+r_1+r_2)^{-1} - 2(2+r_1+r_2)^{-1} & \text{for } \underset{\sim}{y} = (1,1) \\ (1+r_1)^{-1} + (1+r_1+r_2)^{-1} & \text{for } \underset{\sim}{y} = (0,1) \\ (1+r_2)^{-1} + (1+r_1+r_2)^{-1} & \text{for } \underset{\sim}{y} = (1,0) \\ (1+r_1+r_2)^{-1} & \text{for } \underset{\sim}{y} = (0,0) \end{cases}$$

and, for the corresponding $\underset{\sim}{y}$,

$$\text{Var}(\theta'|\underset{\sim}{y}) = \begin{cases} 1 + (1+r_1)^{-2} + (1+r_2)^{-2} + (1+r_1+r_2)^{-2} - 4(2+r_1+r_2)^{-2} \\ (1+r_1)^{-2} + (1+r_1+r_2)^{-2} \\ (1+r_2)^{-2} + (1+r_1+r_2)^{-2} \\ (1+r_1+r_2)^{-2}. \end{cases}$$

Using this in (3.22), together with the definition of $m(\underset{\sim}{y})$ from Appendix A5, yields after lengthy algebra

$$r(\underset{\sim}{c}) = 1 + \frac{1}{(1+r_1)^2} + \frac{1}{(1+r_2)^2} + \frac{(2+r_1+r_2)}{(1+r_1+r_2)^2} - \frac{4}{(2+r_1+r_2)(1+r_1+r_2)} - \frac{(2+r_1+r_2)}{(1+r_1)(1+r_2)} - \frac{(r_1-r_2)^2}{(1+r_1)^2(1+r_2)^2(2+r_1+r_2)}. \tag{3.23}$$

Numerical minimization of $r(\underset{\sim}{c})$ reveals that the minimum occurs at $r_1 = r_2 \cong 0.9004$, which corresponds to $c_1 = c_2 = (.9004)\beta$ as the optimal choice of $\underset{\sim}{c} = (c_1, c_2)$. It is interesting to note that the optimal c_i's are equal and are near the predictive mean $E^{\theta}[E^{X_i|\theta}(X_i)] = \beta$.

The posterior means and variances at the optimal $\underset{\sim}{c}^* = ((.9004)\beta,\ (.9004)\beta)$ are

$$\delta_{\underset{\sim}{c}^*}(\underset{\sim}{y}) = \begin{cases} (1.883)\beta & \text{for } \underset{\sim}{y} = (1,1) \\ (0.8832)\beta & \text{for } \underset{\sim}{y} = (0,1) \text{ or } \underset{\sim}{y} = (1,0) \\ (0.3570)\beta & \text{for } \underset{\sim}{y} = (0,0), \end{cases}$$

$$\text{Var}(\theta|\underset{\sim}{y}) = \begin{cases} (1.4044)\beta^2 & \text{for } \underset{\sim}{y} = (1,1) \\ (0.4044)\beta^2 & \text{for } \underset{\sim}{y} = (0,1) \text{ or } \underset{\sim}{y} = (1,0) \\ (0.1274)\beta^2 & \text{for } \underset{\sim}{y} = (0,0). \end{cases}$$

Acknowledgment

We are grateful to Nozer Singpurwalla for valuable suggestions that significantly improved the article.

References

Bernardo, J. M. (1979a). Reference posterior distributions for Bayesian inference (with discussion). *J. Roy. Statist. Soc.* **41**, 113–147.

Bernardo, J. M. (1979b). Expected information as expected utility. *Ann. Statist.* **7**, 686–690.

Chaloner, K. (1984). Optimal Bayesian experimental design for linear models. *Ann. Statist.* **4**, 283–300.

Clemen, R. (1987). Combining overlapping information. *Management Science* **33**, 373–380.

DeGroot, M. H. and Goel P. (1988). Bayesian design and analysis of accelerated life testing with step stress. In *Acceletated Life Testing and Experts' Opinion in Reliability* (C. A. Clarotti and D. V. Lindley, eds.). North-Holland, Amsterdam, 193–202.

DeGroot, M. H. and Mortera, J. (1988). Optimal opinion pools. Technical Report No. 443, Department of Statistics, Carnegie-Mellon University, Pittsburgh.

French, S. G. (1980). Updating of belief in the light of someone else's opinion. *J. Roy. Statist. Soc. A* **143**, 43–48.

French, S. (1985). On pooling group consensus probability distributions: a critical survey. In *Bayesian Statistics 2* (J. M. Bernardo, et. al., eds.). North-Holland, Amsterdam, 183–201.

Genest, C. and Zidek, J. V. (1986). Combining probability distributions: a critique and annotated bibliography. *Statistical Science* **1**, 114–148.

Lindley, D. V. (1983). Reconciliation of probability distributions. *Operations Research* **31**, 866–880.

Lindley, D. V. and Singpurwalla, N. D. (1986). Reliability (and fault tree) analysis using expert opinions. *J. Amer. Statist. Assoc.* **81**, 87–90.

Mann, N. R., Schafer, R. E., and Singpurwalla, N. D. (1974). *Methods of Statistical Analysis of Reliability and Life Data*, pp. 379–418, Wiley and Son, New York.

Marschak, J. and Radner, R. (1971). *The Economic Theory of Teams.* Yale University Press, New Haven.

Pilz, J. (1983). *Bayesian Estimation and Experimental Design in Linear Regression Models*. Teubner-Texte, Leipzig.

Smith, A. F. M. and Verdinelli, I. (1980). A note on Bayes designs for inference using a hierarchical linear model. *Biometrika* **67**, 613–619.

Winkler, R. L. (1968). The consensus of subjective probability distributions. *Management Sci.* **15**, B61–B75.

Winkler, R. L. (1981). Combining probability distributions from dependent information sources. *Management Sci.* **27**, 479–488.

Zellner, A. (1977). Maximal data information prior distributions. In *New Methods in the Applications of Bayesian Methods* (A. Aykac and C. Brumat, eds.). North-Holland, Amsterdam.

Appendix A1

a) Derivation of the normalizing constant $K(t(c), c)$ for Example 3:

$$[K(t,c)]^{-1} = \begin{cases} \int_{w_0 \wedge c}^{\infty} \pi(\theta)(\frac{c}{\theta})^m d\theta + \int_{w_0}^{c} \pi(\theta) d\theta, & \text{for } t = 0 \\ \int_{w_0 \wedge c}^{\infty} \pi(\theta)(\frac{\theta - c}{\theta})^t (\frac{c}{\theta})^{m-t} d\theta, & \text{for } t > 0 \end{cases}$$

$$= \begin{cases} \int_{w_0 \wedge c}^{\infty} \pi(\theta)(\frac{c}{\theta})^m d\theta + [1 - (\frac{w_0}{c})^\alpha]^+, & \text{for } t = 0, \\ \sum_{h=0}^{t} \binom{t}{h}(-1)^{t-h} \alpha w_0^\alpha c^{m-h} \int_{w_0 \wedge c}^{\infty} \theta^{-(m-h+\alpha+1)} d\theta, & \text{for } t > 0 \end{cases}$$

$$= \begin{cases} 1 - r^{-\alpha} \frac{m}{(m+\alpha)} & \text{for } r \geq 1 \text{ and } t = 0 \\ r^{-\alpha} \sum_{h=0}^{t} a_{ht} & \text{for } r \geq 1 \text{ and } t > 0 \\ \sum_{h=0}^{t} a_{ht}\, r^{m-h} & \text{for } r < 1, \end{cases} \tag{A.1}$$

where $w_0 \wedge c = \max\{w_0, c\}$, $[x]^+$ denotes the positive part of x, and r and a_{ht} are defined in (3.11).

b) Derivation of equation (3.12): Substituting $p_\theta(C) = (1 - c/\theta)$ in the first term of (3.9) yields

$$\begin{aligned} &- m\{E^\theta[p_\theta(C) \log p_\theta(C)] + E^\theta[(1 - p_\theta(C)) \log(1 - p_\theta(C))]\} \\ &= -m\{E^\theta[(1 - \frac{c}{\theta}) \log(\frac{\theta - c}{\theta})] + E^\theta[\frac{c}{\theta} \log \frac{c}{\theta}]\} \\ &= -m\{E^\theta[\log(\theta - c) - \frac{c}{\theta} \log(\theta - c) - \log\theta + \frac{c}{\theta} \log c]\}. \end{aligned} \tag{A.2}$$

Now

$$\begin{aligned} E^\theta[\log(\theta - c)] &= w_0^\alpha \alpha \int_{w_0 \wedge c}^{\infty} \theta^{-(\alpha+1)} \log(\theta - c) d\theta \\ &= w_0^\alpha[-\theta^{\alpha} \log(\theta - c)]_{w_0 \wedge c}^{\infty} + w_0^\alpha \int_{w_0 \wedge c}^{\infty} \frac{1}{(\theta - c)\theta^\alpha} d\theta. \end{aligned}$$

Integration by partial fractions yields

$$\int \frac{1}{(\theta - c)\theta^\alpha} d\theta = c^{-\alpha} \log(\theta - c) - \sum_{i=1}^{\alpha-1} \frac{1}{c^i} \frac{\theta^{i-\alpha}}{(i - \alpha)} - c^{-\alpha} \log\theta,$$

so that

$$E^\theta[\log(\theta - c)] = w_0^\alpha \left[(c^{-\alpha} - \theta^{-\alpha}) \log(\theta - c) - c^{-\alpha} \log\theta - \sum_{i=1}^{\alpha-1} \frac{1}{c^i} \frac{\theta^{i-\alpha}}{(i - \alpha)} \right]_{w_0 \wedge c}^{\infty}. \tag{A.3}$$

Similarly,

$$\begin{aligned} E^\theta[-\frac{c}{\theta} \log(\theta - c)] = \frac{\alpha}{(\alpha + 1)} w_0^\alpha \Bigg[&[c^{-\alpha} - c\theta^{-(\alpha+1)}][\log(\frac{1}{\theta - c})] \\ &+ c^{-\alpha} \log\theta + \sum_{i=1}^{\alpha} \frac{1}{c^{i-1}} \frac{\theta^{i-(\alpha+1)}}{(i - (\alpha + 1))} \Bigg]_{w_0 \wedge c}^{\infty}. \end{aligned} \tag{A.4}$$

Simple computations yield

$$E^{\theta}[-\log\theta] = w_0^{\alpha}\left[\theta^{-\alpha}\log\theta + \frac{\theta^{-\alpha}}{\alpha}\right]_{w_0\wedge c}^{\infty} \tag{A.5}$$

and

$$E^{\theta}[\frac{c}{\theta}\log c] = \frac{\alpha}{(\alpha+1)}w_0^{\alpha}c\log c[-\theta^{-(\alpha+1)}]_{w_0\wedge c}^{\infty}. \tag{A.6}$$

Since

$$\frac{\alpha}{(\alpha+1)}\sum_{i=1}^{\alpha}\frac{1}{c^{i-1}}\frac{\theta^{i-(\alpha+1)}}{(i-(\alpha+1))} - \sum_{i=1}^{\alpha-1}\frac{1}{c^i}\frac{\theta^{i-\alpha}}{(i-\alpha)} + \frac{\theta^{-\alpha}}{\alpha} = \sum_{i=0}^{\alpha-1}\frac{1}{c^i}\frac{\theta^{i-\alpha}}{(i-\alpha)}, \tag{A.7}$$

substituting (A.3) to (A.7) in equation (A.2) yields, after some simple algebra,

$$-\frac{m}{(\alpha+1)}\left[-\left(\frac{w_0}{c}\right)^{\alpha}\log\left(\frac{\theta}{\theta-c}\right) - \alpha c w_0^{\alpha}\theta^{-(\alpha+1)}\log\left(\frac{c}{\theta-c}\right)\right.$$
$$\left. + (\alpha+1)(\frac{w_0}{\theta})^{\alpha}\log\left(\frac{\theta}{\theta-c}\right) - \sum_{i=0}^{\alpha-1}\left(\frac{w_0}{\theta}\right)^{\alpha}\left(\frac{\theta}{c}\right)^{i}\frac{1}{(i-\alpha)}\right]_{w_0\wedge c}^{\infty}.$$

For $r = c/w_0 \geq 1$, this reduces to

$$-\frac{m}{(\alpha+1)}r^{-\alpha}\sum_{i=1}^{\alpha}\frac{1}{i}, \tag{A.8}$$

and, for $r = c/w_0 < 1$, this reduces to

$$-\frac{m}{(\alpha+1)}\{(\alpha+1-r^{-\alpha})\log(1-r) + \alpha r[\log r - \log(1-r)] + \sum_{i=0}^{\alpha-1}\frac{r^{-i}}{(i-\alpha)}\}. \tag{A.9}$$

From (3.10), the second term in equation (3.9) is given by

$$\sum_{t=0}^{m}\binom{m}{t}[K(t,c)]^{-1}\log[K(t,c)]^{-1}.$$

Substituting (A.1) for $[K(t,c)]^{-1}$ and using (A.2), (A.8) and (A.9) in (3.9), yields (3.13) and (3.14). □

Appendix A2

Lemma A1. For $m = 1$, $\Lambda_2(c)$ is a monotone decreasing function of $r = c/w_0$ for $r < 1$.

Proof. For convenience, we will slightly abuse notation in the proof by writing $\Lambda_2(r)$ instead of $\Lambda_2(c)$. For $m = 1$,

$$\Lambda_2(r) = -\frac{1}{(\alpha+1)}\Big\{(\alpha+1-r^{-\alpha}-\alpha r)\log(1-r) - \alpha r\log\alpha$$
$$+ (\alpha r - \alpha - 1)\log(\alpha+1-\alpha r) + (\alpha+1)\log(\alpha+1) + \sum_{i=0}^{\alpha-1}\frac{r^{-i}}{(i-\alpha)}\Big\}.$$

Thus

$$\frac{d}{dr}\Lambda_2(r) = -\frac{1}{(\alpha+1)}\Big\{\alpha(r^{-\alpha-1}-1)\log(1-r) - \frac{(\alpha(1-r)+1-r^{-\alpha-1})}{1-r}$$
$$+ \alpha(1-\log\alpha) + \alpha\log(\alpha+1-\alpha r) + \sum_{i=0}^{\alpha-1}\frac{i}{(\alpha-i)}r^{-i-1}\Big\}.$$

Note that

$$\log(1-r) > -r + \frac{1}{2}r^2 - \frac{1}{3}r^3,$$
$$\frac{\alpha(1-r)+1-r^{-\alpha}}{1-r} = \alpha - \sum_{i=0}^{\alpha-1} r^{i-\alpha},$$

and $\alpha\log(\alpha+1-\alpha r) > 0$, so that

$$\frac{d}{dr}\Lambda_2(r) < -\frac{1}{(\alpha+1)}\Big\{\alpha(-r^{-\alpha} + \frac{1}{2}r^{-\alpha+1} - \frac{1}{3}r^{-\alpha+2} + r - \frac{1}{2}r^2 + \frac{1}{3}r^3)$$
$$+ \sum_{i=0}^{\alpha-1} r^{i-\alpha} - \alpha\log\alpha + \sum_{i=0}^{\alpha-1}\frac{i}{(\alpha-i)}r^{-i-1}\Big\}. \qquad (A.10)$$

a) For $\alpha = 1$, and since $0 < r < 1$,

$$\frac{d}{dr}\Lambda_2(r) < -\frac{1}{2}\left[\frac{1}{2} + \frac{2}{3}r - \frac{1}{2}r^2 + \frac{1}{3}r^3\right] < 0,$$

so that $\Lambda_2(r)$ is decreasing.

b) For $\alpha = 2$,

$$\frac{d}{dr}\Lambda_2(r) < -\frac{1}{3}\left[\frac{2}{r} - \frac{2}{3} + 2r - r^2 + \frac{2}{3}r^3 - 2\log 2\right].$$

It is easy to show, by minimizing a quadratic over (0,1), that

$$2r - r^2 + \frac{2}{3}r^3 > \frac{13}{8}r,$$

so that

$$\frac{d}{dr}\Lambda_2(r) < -\frac{1}{3r}[2 - (\frac{2}{3} + 2\log 2)r + \frac{13}{8}r^2].$$

It is easy to verify that the quadratic in brackets is always positive; hence $\Lambda_2(r)$ is decreasing.

c) For $\alpha \geq 3$, equation (A.10) can be rewritten as

$$\frac{d}{dr}\Lambda_2(r) < -\frac{1}{(\alpha+1)}\Big\{\alpha(r^{1-\alpha} + r - \frac{1}{2}r^2 + \frac{1}{3}r^3) + \sum_{i=3}^{\alpha-1} r^{i-\alpha} + \sum_{i=0}^{\alpha-4} \frac{i}{(\alpha-i)} r^{-(i+1)} - \alpha\log\alpha\Big\}.$$

Note that, for $0 < r < 1$,

$$r - \frac{1}{2}r^2 + \frac{1}{3}r^3 > (\frac{13}{16})r,$$

so that to prove that $\Lambda_2(r)$ is decreasing it suffices to show that

$$\alpha r^{1-\alpha} + \frac{13}{16}\alpha r + \sum_{i=3}^{\alpha-1} r^{i-\alpha} + \sum_{i=0}^{\alpha-4} \frac{i}{(\alpha-i)} r^{-(i+1)} > \alpha\log\alpha. \tag{A.11}$$

Equation (A.11) is trivially true for $r^{(1-\alpha)} > \log\alpha$. For $r^{(1-\alpha)} < \log\alpha$ or, equivalently,

$$(\log\alpha)^{1/(1-\alpha)} < r < 1,$$

note that (A.11) is implied by

$$\alpha + \frac{13}{16}\alpha(\log\alpha)^{1/(1-\alpha)} + (\alpha-4)^+ + \sum_{i=1}^{\alpha-4} \frac{i}{(\alpha-i)} > \alpha\log\alpha, \tag{A.12}$$

where the summation is defined to be 0 if $\alpha = 3$ or $\alpha = 4$. Simple substitution verifies (A.12) for $\alpha = 3$, 4, and 5.

It remains to verify (A.12) for $\alpha \geq 6$. Defining

$$\psi_\alpha = \alpha + \frac{13}{16}\alpha(\log\alpha)^{1/(1-\alpha)} + (\alpha-4) + \sum_{i=1}^{\alpha-4}\frac{i}{(\alpha-i)} - \alpha\log\alpha,$$

it is clear, by induction, that (A.12) will be true if we can show that $\psi_{\alpha+1} - \psi_\alpha > 0$ for $\alpha \geq 5$. This reduces to showing that

$$1+\frac{\alpha}{3}+\alpha\log\frac{\alpha}{(\alpha+1)}-\log(\alpha+1)+\frac{13}{16}[(\alpha+1)[\log(\alpha+1)]^{-1/\alpha}-\alpha(\log\alpha)^{1/(1-\alpha)}] > 0 \quad \text{(A.13)}$$

for $\alpha \geq 5$. Note that $\log[\alpha/(\alpha+1)] > -1/(\alpha+1)$, so that

$$1+\alpha\log[\alpha/(\alpha+1)] > 1/(\alpha+1). \quad \text{(A.14)}$$

Also note that, for $\alpha \geq 5$,

$$\frac{\alpha}{3}+\frac{1}{(\alpha+1)}-\log(\alpha+1) > 0. \quad \text{(A.15)}$$

Finally, since $\log(1+\alpha) < 2\log\alpha$,

$$\frac{[\log(\alpha+1)]^{1/\alpha}}{[\log\alpha]^{1/(\alpha-1)}} < \frac{[2\log\alpha]^{1/\alpha}}{[\log\alpha]^{1/(\alpha-1)}} < 2^{1/\alpha} < 1+\frac{1}{\alpha},$$

which can be rewritten

$$(\alpha+1)[\log(\alpha+1)]^{-1/\alpha} > \alpha[\log\alpha]^{1/(1-\alpha)}. \quad \text{(A.16)}$$

Equation (A.13) follows directly from (A.14), (A.15), and (A.16), completing the proof. □

Appendix A3

a) Proof of (3.19): Equation (3.17) with $p_\theta(C) = 1 - e^{-\theta c}$ yields

$$\pi(0|t) = K(t,c)\int_0^a \pi(\theta)(1-e^{-\theta c})^t e^{-\theta[c(m-t)]}d\theta,$$

where $K(t,c)$ is the normalizing constant (derived below). Since $\pi(\theta) = \Gamma(\alpha,\beta)$ and since

$$(1-e^{-\theta c})^t = \sum_{h=0}^{t}(-1)^{t-h}\binom{t}{h}e^{-\theta c(t-h)},$$

one has that

$$\pi(0|t) = K(t,c)\sum_{h=0}^{t}(-1)^{t-h}\binom{t}{h}\frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^a \theta^{\alpha-1}e^{\theta[c(m-h)+\beta]}d\theta.$$

The conclusion follows by a change of variables from θ to $\theta[c(m-h)+\beta]$. □

b) Proof of (3.20): The normalizing constant is

$$\begin{aligned} K(t,c) &= \left[\sum_{h=0}^{t}(-1)^{t-h}\binom{t}{h}\frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty \theta^{\alpha-1}e^{-\theta[c(m-h)+\beta]}d\theta\right]^{-1} \\ &= \left[\sum_{h=0}^{t}(-1)^{t-h}\binom{t}{h}[1+\frac{c}{\beta}(m-h)]^{-\alpha}\right]^{-1} \\ &= \left[\sum_{h=0}^{t} a_{h,t}(c)\right]^{-1}. \end{aligned}$$

Appendix A4.

Proof of Theorem 3.1: For notational convenience, define $r_1 = c_1/\beta$, $r_2 = c_2/\beta$, and $b = a\beta$. Note that, for $C = (c_1, c_2)$,

$$p_\theta(C) = e^{-\theta c_1} - e^{-\theta c_2}.$$

For $m = 1$ and $\alpha = 1$, (3.17) and (3.6) yield (noting that $y = t = 1$ or 0 are the only possible reports)

$$\begin{aligned} 1-\pi(1|1) &= \pi(0|1) = 1 - g^{-1}(r_1,r_2)h(r_1,r_2), \\ 1-\pi(1|0) &= \pi(0|0) = 1 + [1-g(r_1,r_2)]^{-1}[h(r_1,r_2) - e^{-b}], \\ 1-K^{-1}(1,C) &= K^{-1}(0,C) = g(r_1,r_2), \end{aligned}$$

where

$$g(r_1,r_2) = (1+r_1)^{-1} - (1+r_2)^{-1},$$

$$h(r_1,r_2) = (1+r_1)^{-1}e^{-(1+r_1)b} - (1+r_2)^{-1}e^{-(1+r_2)b}.$$

Writing $g = g(r_1, r_2)$ and $h = h(r_1, r_2)$, for convenience, (3.18) becomes

$$\Lambda(C) = \{(1-g)\log(1-g) + g\log g\} - (1-g+h-e^{-b})\log(1-g+h-e^{-b})$$
$$- (g-h)\log(g-h) - (e^{-b}-h)\log(e^{-b}-h) - h\log h. \qquad \text{(A.17)}$$

Calculation yields

$$(1+r_1)^2 \frac{\partial}{\partial r_1}\Lambda(C) = A(r_1,r_2) + [1+(1+r_1)b]e^{-(1+r_1)b}[A(r_1,r_2)+B(r_1,r_2)], \quad \text{(A.18)}$$

$$(1+r_2)^2 \frac{\partial}{\partial r_2}\Lambda(C) = -A(r_1,r_2) + [1+(1+r_2)b]e^{-(1+r_2)b}[-A(r_1,r_2)-B(r_1,r_2)], \quad \text{(A.19)}$$

where

$$A(r_1,r_2) = \log\left(\frac{\pi(0|1)}{\pi(0|0)}\right), \quad B(r_1,r_2) = \log\left(\frac{1-\pi(0|1)}{1-\pi(0|0)}\right).$$

Case 1. $r_1 = 0$: It is clear that, when $r_1 = 0$, $\Lambda(C)$ cannot be minimized at $r_2 = 0$ or $r_2 = \infty$. Hence the minimizing r_2 must be a zero of (A.19). Setting (A.19) equal to zero yields

$$A(r_1,r_2) = -B(r_1,r_2)\left[1+(1+(1+r_2)b)^{-1}e^{(1+r_2)b}\right]^{-1}.$$

Substituting this into (A.18) yields

$$(1+r_2)^2 \frac{\partial}{\partial r_1}\Lambda(c) = \qquad \text{(A.20)}$$
$$\frac{B(r_1,r_2)}{[1+(1+(1+r_2)b)e^{-(1+r_2b)}]}\left\{[1+(1+r_1)b]e^{-(1+r_2)b} - [1+(1+r_2)b]e^{-(1+r_2)b}\right\}.$$

Note that $(1+x)e^{-x}$ is a decreasing function and $r_1 < r_2$, so that the term in curly brackets in (A.20) is positive. Also, at $r_1 = 0$,

$$B(0,r_2) = \log((1+r_2)e^{r_2 b} - 1) - \log r_2 > 0,$$

since $r_2 < (1+r_2)e^{r_2 b} - 1$. Thus $\frac{\partial}{\partial r_1}\Lambda(C)$ is positive at $r_1 = 0$ and the corresponding maximizing r_2.

Case 2. $r_2 = \infty$: As has been stated before, the problem is identical whether one considers the interval $(0, r_1)$ or the interval (r_1, ∞). Hence it can be stated that, at $r_2 = \infty$, $\frac{\partial}{\partial r_2}\Lambda(C)$ is negative for the corresponding minimizing r_1.

Case 3. $0 < r_1 < r_2 < \infty$: Any minimizing or maximizing r_1, r_2 in this range must be zeros of (A.18) and (A.19), and hence zeros of (A.20). But since the term in curly brackets in (A.20) is nonzero, it must be the case that $B(r_1, r_2) = 0$ or, equivalently, that $\pi(0|0) = \pi(0|1)$. Algebra shows that this, in turn, implies that $\pi(0|0) = \pi(0|1) = 1 - e^{-b}$. Hence, the only possible zeros of (A.18) and (A.19) yield (using (3.18))

$$\Lambda(C) = -(1 - e^{-b}) \log(1 - e^{-b}) + be^{-b}.$$

But this is clearly a maximum, since it is easily seen to be the value of $\Lambda(C)$ in (A.17) when $r_1 = r_2$, corresponding to a noninformative report. Thus the minimizing interval must be of the form $(0, r_2)$ or, equivalently, (r_1, ∞). □

Appendix A5

The posterior distribution of θ', given $\underset{\sim}{y} = (y_1, y_2)$, can be calculated to be

$$\pi(\theta'|\underset{\sim}{y}) = \begin{cases} \frac{1}{m(1,1)} \left[e^{-\theta'} - e^{-(1+r_1)\theta'} - e^{-(1+r_2)\theta'} + e^{-(1+r_1+r_2)\theta'} \right] & \text{for } \underset{\sim}{y} = (1,1) \\ \frac{1}{m(0,1)} \left[e^{-(1+r_1)\theta'} - e^{-(1+r_1+r_2)\theta'} \right] & \text{for } \underset{\sim}{y} = (0,1) \\ \frac{1}{m(1,0)} \left[e^{-(1+r_2)\theta'} - e^{-(1+r_1+r_2)\theta'} \right] & \text{for } \underset{\sim}{y} = (1,0) \\ \frac{1}{m(0,0)} e^{-(1+r_1+r_2)\theta'} & \text{for } \underset{\sim}{y} = (0,0), \end{cases}$$

where the marginal probabilities are

$$m(\underset{\sim}{y}) = \begin{cases} 1 - (1+r_1)^{-1} - (1+r_2)^{-1} + (1+r_1+r_2)^{-1} & \text{for } \underset{\sim}{y} = (1,1) \\ (1+r_1)^{-1} - (1+r_1+r_2)^{-1} & \text{for } \underset{\sim}{y} = (0,1) \\ (1+r_2)^{-1} - (1+r_1+r_2)^{-1} & \text{for } \underset{\sim}{y} = (1,0) \\ (1+r_1+r_2)^{-1} & \text{for } \underset{\sim}{y} = (0,0). \end{cases}$$

As an example of the posterior mean calculation, observe that

$$\begin{aligned} \delta_{\underset{\sim}{C}}((1,0)) &= \int_0^\infty \theta' \pi(\theta'|(1,0)) d\theta' \\ &= \frac{1}{m(1,0)} \left[\int_0^\infty \theta' e^{-(1+r_2)\theta'} d\theta' - \int_0^\infty \theta' e^{-(1+r_1+r_2)\theta'} d\theta' \right] \\ &= [(1+r_2)^{-1} - (1+r_1+r_2)^{-1}]^{-1} [(1+r_2)^{-2} - (1+r_1+r_2)^{-2}] \\ &= (1+r_2)^{-1} + (1+r_1+r_2)^{-1}. \end{aligned}$$

As an example of the posterior variance calculation, observe that

$$\begin{aligned}
\mathrm{Var}(\theta'|(1,0)) &= \int_0^\infty \theta'^2 \pi(\theta'|(1,0))d\theta' - \delta^2_{\underset{\sim}{c}}((1,0)) \\
&= \frac{1}{m(1,0)}\left[\int_0^\infty \theta'^2 e^{-(1+r_2)\theta'} d\theta' - \int \theta'^2 e^{-(1+r_1+r_2)\theta'} d\theta'\right] - \delta^2_{\underset{\sim}{c}}((1,0)) \\
&= \frac{2[(1+r_2)^{-3} - (1+r_1+r_2)^{-3}]}{[(1+r_2)^{-1} - (1+r_1+r_2)^{-1}]} - [(1+r_2)^{-1} + (1+r_1+r_2)^{-1}]^2 \\
&= (1+r_2)^{-2} + (1+r_1+r_2)^{-2}.
\end{aligned}$$

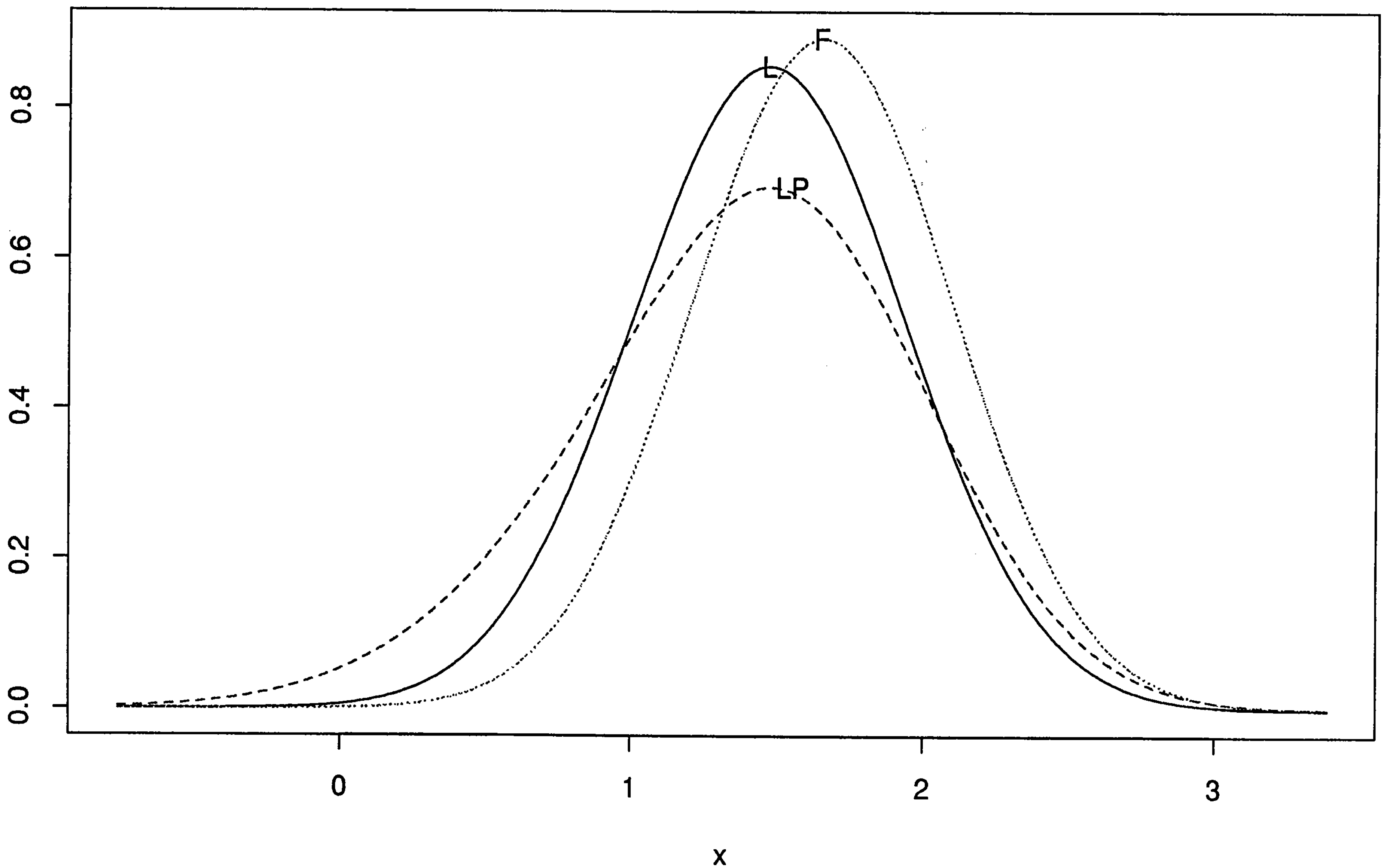

Figure 1. F,L and LP for overlapping normal samples with c=1,n1=2,n2=1, and particular random data generated from theta=2. Note w=.711

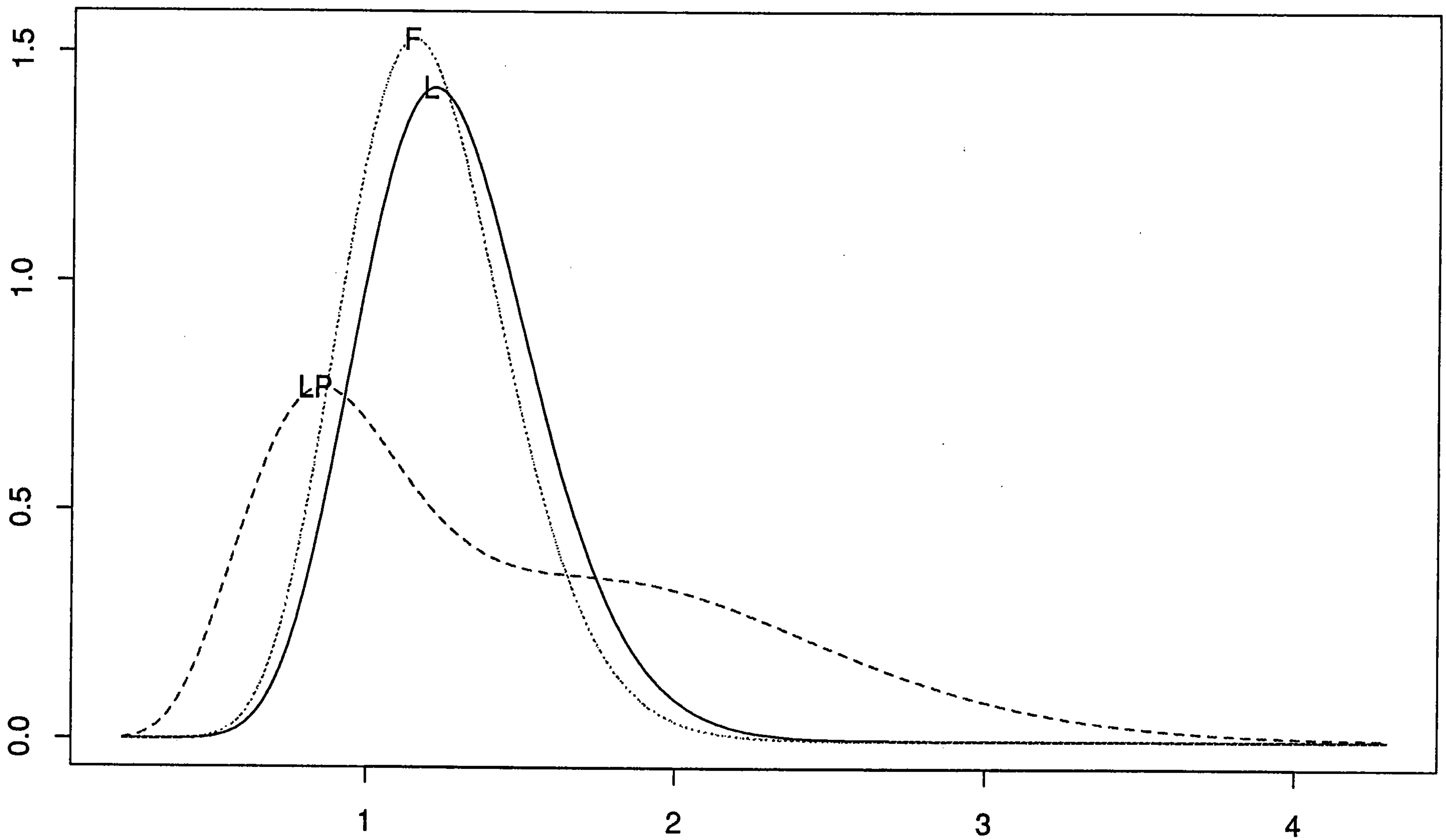

Figure 2. F,L,and LP for the normal variance example with n1=n2=20 and particular random data generated when tau=2. Note that w=.5

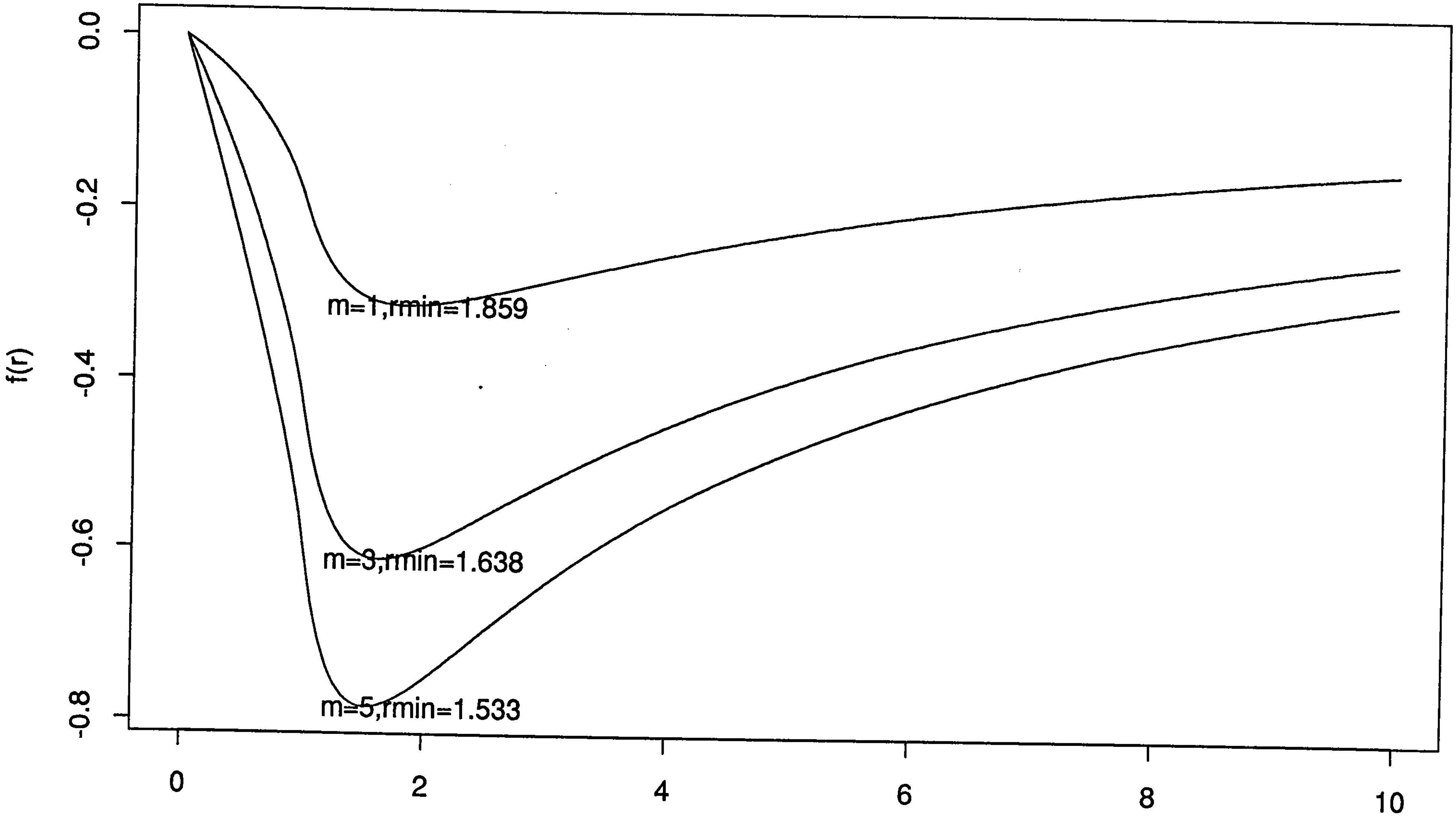

Figure 3. Shifted Kullback-Leibler distance between the posterior given x and the posterior given y for alpha=1, r=c/w0, and various m

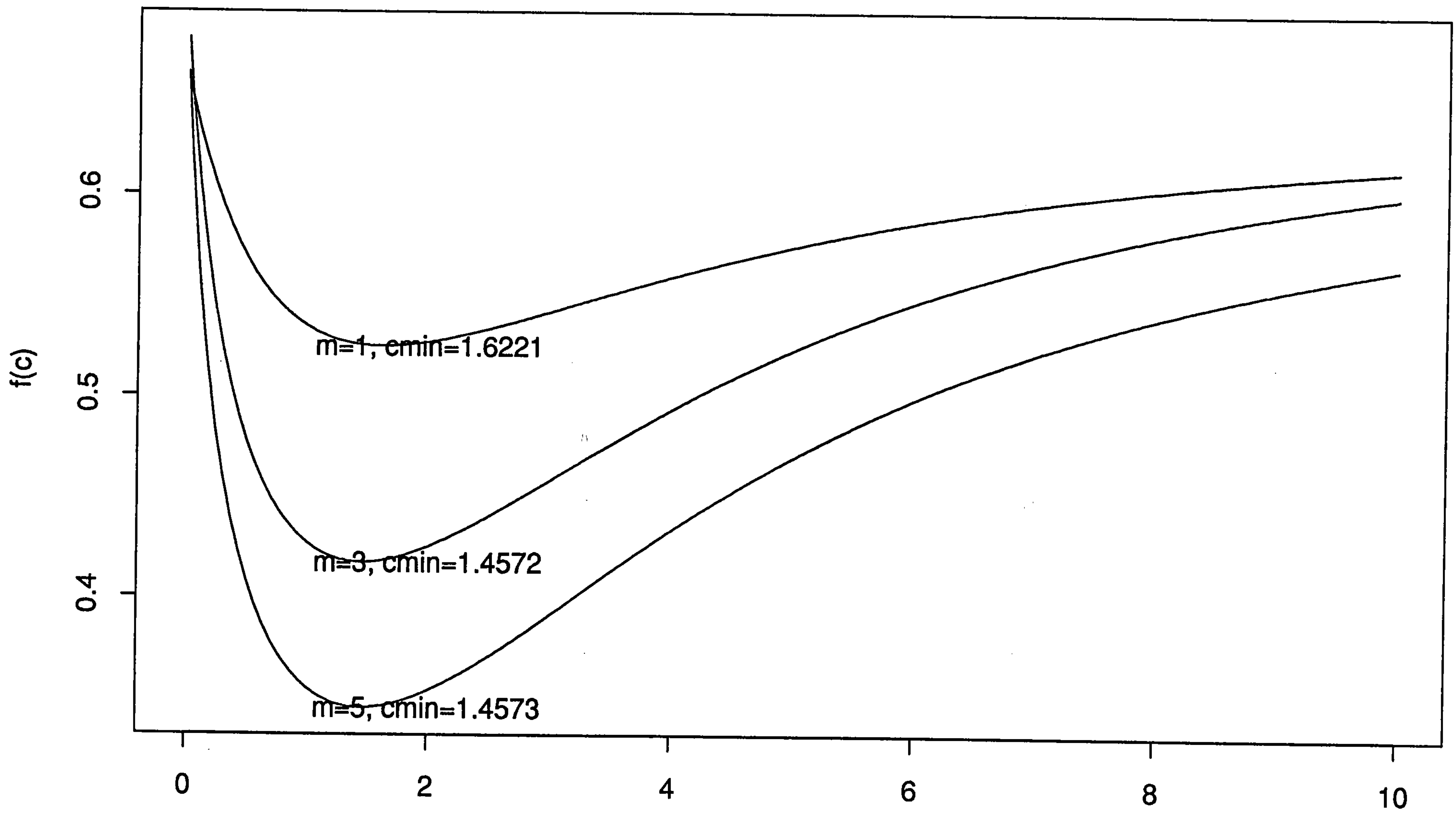

Figure 4. Shifted Kullback-Leibler distance between the posterior given x and the posterior given y for alpha=beta=1,a=1,and various m.