Robust Bayesian Analysis of the Binomial Empirical Bayes Problem

by

S. Sivaganesan and James Berger
University of Cincinnati, Cincinnati, OH
Purdue University, West Lafayette, IN

Technical Report # 90-49C

Department of Statistics
Purdue University

September, 1990

△

ROBUST BAYESIAN ANALYSIS OF THE BINOMIAL EMPIRICAL BAYES PROBLEM*

by

S. Sivaganesan
University of Cincinnati
Cincinnati, OH

and

James Berger
Purdue University
West Lafayette, IN

ABSTRACT

Empirical Bayes estimation is considered for an i.i.d. sequence of binomial parameters θ_i arising from an unknown prior distribution $G(\cdot)$. This problem typically arises in industrial sampling where samples from lots are routinely used to estimate the lot fraction defective of each lot. Two related issues are explored. The first concerns the fact that only the first few moments of G are typically estimable from the data. This suggests consideration of the interval of estimates (e.g., posterior means) corresponding to the different possible G with the specified moments. Such intervals can be obtained by application of well-known moment theory. The second development concerns the need to acknowledge the uncertainty in the estimation of the first few moments of G. Our proposal is to determine a credible set for the moments, and then find the range of estimates (e.g., posterior means) corresponding to the different possible G with moments in the credible set.

AMS 1980 Subject Classification: Primary 62A15, Secondary 62F15.

*Research supported by the National Science Foundation, Grants DMS–8923071 and DMS–8717799.

1. Introduction

We consider the nonparametric empirical Bayes estimation problem of observing $X_i \sim \text{Binomial}(n, \theta_i)$, $i = 1,\ldots,N$, the X_i being conditionally independent given the θ_i, and the θ_i being i.i.d. from some completely unknown prior distribution G (i.e., the θ_i are known only to be infinitely exchangeable). Of interest will be estimation of the "current" θ_N , utilizing the current X_N and "past" $X_1,\ldots,X_{N-1}$.

There is an extensive literature on this problem, starting with Robbins (1955). A few of the many references include Maritz(1966), Atchison and Martz (1969), Jackson, O'Donovan, Zimmer and Deely(1970), Maritz(1970), Griffin and Krutchkoff(1971), Copas(1972), Lin(1972), Martz and Lian(1974), and Albert (1984). Most of these present developments of specific empirical Bayes estimators and proofs of convergence properties of these estimators as $N \longrightarrow \infty$.

The binomial problem is an atypical empirical Bayes problem in the sense that only the first n moments of G can be determined as $N \longrightarrow \infty$ (see Section 2). Hence substantial uncertainty about G can remain no matter how large an N is available. This uncertainty can be quantified in a robust Bayesian fashion. As $N \longrightarrow \infty$, the first n moments, $c_1,\ldots,c_n$, of G become essentially known, so that it becomes known that

$$G \in \mathcal{G} = \left\{ \text{all distributions } G : \; c_i = \int_0^1 \theta^i dG(\theta), \; i = 1,\ldots,n \right\}.$$

One can then determine the range of the Bayes estimate (typically, the range of the posterior mean) as G varies over $\mathcal{G}$. The analysis here is based on well known moment theory (from, e.g., Karlin and Studden(1966)), and is reviewed in the appendix. Section 2 applies this theory to determination of the range of the posterior mean. Related references to robust Bayesian estimation include Berger(1984), Berger(1985), Berger and Berliner(1986), Berger and O'Hagan(1988), Berliner(1984), DasGupta and Studden(1988), DeRobertis and Hartigan (1981), Lavine(1987), O'Hagan and Berger(1988), Sivaganesan and Berger(1989) and Wasserman(1989).

When N is small, one can still, of course, estimate the first n moments of G. The most common estimate would be the ML-II estimate (see Section 3), based on maximizing the effective likelihood for the moments. One could then find the range of the posterior mean over all G with these specified moments, but this clearly is too optimistic because it does not incorporate the error in estimation of the moments. Indeed, it often happens that the ML-II estimate of the first n moments of G lies on the boundary of the moment space, in which case it can be shown that there is no variation whatsoever in the posterior mean. Thus the ML-II approach here provides a very inadequate reflection of prior uncertainty.

An interesting possibility to address the inadequacy of the ML-II approach is to develop a credible set for the moments of G, and determine the range of the posterior mean over all G which have moments in the credible set. This is done in Section 3, where it is proposed that the range of the posterior mean actually be presented as a function of the probability, γ, of the credible set. The credible sets considered are those with respect to a constant prior on a convenient parameterization of the moment space, a parameterization chosen so as to allow comparatively easy calculation of credible probabilities via simulation. (The credible sets could also be considered likelihood based confidence sets, though we calculate their probabilities in a Bayesian fashion.) Rather surprisingly, the answers do not seem to depend too dramatically on the choice of (moderately large) γ.

The purpose behind this approach is to attempt to reflect the uncertainty that arises out of uncertainties in G (the basic difficulty with empirical Bayes analysis) without requiring either parametric modelling of G or the full Bayesian solution of introducing a (nonparametric) second stage prior distribution on G. The hope is that, in a given practical situation, the range of the posterior mean will be small enough, for reasonable credible levels γ, that such knowledge about G need not be elicited. (This is a major goal of this approach to robustness: one seeks to avoid difficult subjective elicitation by arguing that the answer is robust over any reasonable prior.)

2. Empirical Bayes Estimation of a Binomial Parameter

Consider the more general empirical Bayes situation where (i) at stage i ($i=1,2,\ldots$) we observe X_i which, conditional on θ_i, has a $\mathrm{Bin}(n_i,\theta_i)$ distribution, the different X_i's being independent; (ii) the θ_i's are independently distributed according to a common prior distribution G. Our goal, at stage i, is to estimate θ_i without any assumption about G.

At stage i, X_i can take any one of the values $0,1,\ldots,n_i$, and the (marginal) probability, say $p_i(k)$, that $X_i=k$ is given by

$$p_i(k) = \int \binom{n_i}{k}\theta^k(1-\theta)^{n_i-k}\,dG(\theta)$$

$$= \sum_{j=0}^{n_i} a_i(k,j)\int\theta^j dG(\theta)$$

where $a_i(k,j)=(-1)^{j-k}n_i!/(k!(j-k)!(n_i-j)!)$ if $j\geq k$, and $=0$ if $j<k$. Thus, letting c_j denote the jth moment of the distribution G, we have

$$p_i(k) = \sum_{j=0}^{n_i} a_i(k,j)c_j \qquad \text{for} \quad k=0,1,\ldots,n_i. \tag{2.1}$$

We will assume that there will only be a finite number of distinct n_i's as $i\longrightarrow\infty$, i.e.: the set of all n_i's will be a finite set. Moreover, we will assume that n_i will be equal to $\max\{n_i\}\overset{\text{def}}{=}\nu$ infinitely often. Consequently, as the number of stages increases, one could use the relative frequencies from the collected data to accurately estimate the marginal probabilities, denoted

$p(k)$, corresponding to a stage with $n_i=\nu$. But, as in (2.1), we can write

$$p(k) \;=\; \sum_{j=k}^{\nu} a(k,j)c_j \qquad \text{for } k=0,\ldots,\nu, \tag{2.2}$$

where $a(k,j)$ is obtained by replacing n_i by ν in $a_i(k,j)$ of (2.1). Since the relation between the $p(k)$'s and the c_j's given above is one-to-one, accurate estimates could then be obtained for the first ν moments, $c_1,\ldots,c_\nu$, of the prior distribution G by simply solving the linear system of equations (2.2). In the remainder of this section we assume that $c_1,\ldots,c_\nu$ have been so determined.

The Bayes estimate of θ_i given $X_i=k$ (assuming quadratic loss) is given by the posterior mean

$$\hat{\theta}_i(k) \;=\; \frac{\int \theta\binom{n_i}{k}\theta^k(1-\theta)^{n_i-k}\,dG(\theta)}{\int \binom{n_i}{k}\theta^k(1-\theta)^{n_i-k}\,dG(\theta)}$$

$$= \sum_{j=0}^{n_i} a_i(k,j)c_{j+1}/p_i(k).$$

Writing this as

$$\hat{\theta}_i(k) \;=\; \alpha(c_1,\ldots,c_{n_i})\; c_{n_i+1} \;+\; \beta(c_1,\ldots,c_{n_i}), \tag{2.3}$$

where

$$\alpha(c_1,\dots,c_{n_i}) = \frac{a_i(k,n_i)}{\sum_{j=0}^{n_i} a_i(k,j)c_j} \quad \text{and} \quad \beta(c_1,\dots,c_{n_i}) = \frac{\sum_{j=0}^{n_i-1} a_i(k,j)c_{j+1}}{\sum_{j=0}^{n_i} a_i(k,j)c_j}, \tag{2.4}$$

it is clear that, when the first ν moments of G have become known, the Bayes estimate, at stage i, is completely determined when $n_i < \nu$, and is determined up to the $(\nu+1)$st moment, $c_{\nu+1}$, of G when $n_i = \nu$. We, therefore, need pursue only the case where $n_i = \nu$.

Although the value of $c_{\nu+1}$ cannot be determined from the data, upper and lower bounds on $c_{\nu+1}$ can be obtained using Theorem A.2. This, therefore, enables us to find bounds on the Bayes estimate without making any assumptions, functional or otherwise, about G. Letting c_u and c_l, respectively, denote the upper and lower bounds on $c_{\nu+1}$, the bounds on the Bayes estimate are given as below:

$$\inf \hat{\theta}_i(k) = \alpha(c_1,\dots,c_\nu)\, c_* + \beta(c_1,\dots,c_\nu) \tag{2.5}$$

and

$$\sup \hat{\theta}_i(k) = \alpha(c_1,\dots,c_\nu)\, c^* + \beta(c_1,\dots,c_\nu) \tag{2.6}$$

where $c_* = c_l$ (c_u) and $c^* = c_u$ (c_l) when $(\nu-k)$ is even (odd).

Example 2.1: Let $n_i=5$ for all i, so that $\nu=5$. Furthermore, suppose that, after N-1 stages (N sufficiently large) have passed by, the probabilities p(k) (k=0,1,...,5) are accurately estimated to be

$$p(0)= .10,\ p(1)= .15,\ p(2)= .25,\ p(3)= .25,\ p(4)= .15 \text{ and } p(5)= .10 \ .$$

Then, the first five moments of the prior distribution can be obtained by solving the equation (2.2), and are given below:

$$c_1= .500,\ c_2= .290,\ c_3= .185,\ c_4= .130 \text{ and } c_5= .100 \ .$$

Using Theorem A.2, we find the bounds for the 6th moment of the prior distribution G to be

$$c_l = .0827 \quad \text{and} \quad c_u = .0831 \ .$$

Now, suppose that, at stage N, we observe X_N which has a $\text{Bin}(\nu,\theta_N)$ dist., and that $X_N=k$. Then, the bounds for the Bayes estimate of θ_N, viz: $\hat{\theta}_N(k)$, can be found using (2.5) and (2.6), and are given in Table 2.1 for various values of k.

Table 2.1

k	lower bound of $\hat{\theta}_N(k)$	upper bound of $\hat{\theta}_N(k)$
0	.17	.17
1	.42	.44
2	.48	.49
3	.51	.52
4	.56	.58
5	.83	.83

Note, in Example 2.1, that the upper and lower bounds of the estimate, for each k, are very similar. This is due to the fact that the values of c_l and c_u are very close. This phenomenon is to be expected for 'large' ν. In fact, from Theorem 4.5 of Karlin and Studden(1966), the largest value of $c_u - c_l$, for fixed ν, is $2^{-2\nu}$. (Moreover, when $\underline{c} = (c_1, \ldots, c_\nu) \in$ Boundary of $\mathcal{M}$, the difference is always 0.)

3. Expanded Ranges for the Estimate

Since we are looking for answers that are robust w.r.t G, it would be desirable, when N is not large, to give a range of values (for the estimate) allowing for possible uncertainty about the moments $c_1, \ldots, c_\nu$, rather than give a single value. Moreover, when N is not large, the c_i's obtained by solving (2.2) (for the observed p(k)'s) may even fail to be in the moment space $\mathcal{M}$ - thus making the method described in the previous section inapplicable. We address these concerns in this section by developing credible regions for $\underline{c}$ and calculating the ranges of values (for the estimate) as $\underline{c}$ varies over the credible regions.

3.1 Credible Regions for $\underline{c}$

Assume that we are at stage N, and let x_i be the data on $X_i \sim \text{Bin}(n_i, \theta_i)$ for $i=1,\ldots,N$. Then, letting $p_i(x_i)$ be the probability that $X_i = x_i$, the likelihood at stage N is

$$\prod_{i=1}^{N} p_i(x_i) = \prod_{i=1}^{N} \binom{n_i}{x_i} \int \theta^{x_i}(1-\theta)^{n_i-x_i} dG(\theta).$$

It will happen that, without loss of generality, we only need to consider $\underline{c} \in \text{int}(\mathcal{M})$. Then, using (2.1) and Theorem A.1 of the appendix, the likelihood at stage N is proportional to

$$\prod_{i=1}^{N} \left(\sum_{j=0}^{n_i} b_{ij} c_j \right) I_{\Omega^*}(\underline{c}). \tag{3.1}$$

Here, b_{ij} is obtained by replacing k by x_i in $a_i(k,j)$ of (2.1), $I_A(\cdot)$ is the indicator function of the set A, and (see A.1 and A.2 for definitions)

$$\begin{aligned} \Omega^* &= \text{interior of } \mathcal{M} \\ &= \{ \underline{c} = (c_0, c_1, \dots, c_\nu) : A_{m+1} \text{ and } B_{m+1} \text{ are p.d. if} \\ &\qquad \nu=2m+1 \text{ and, } D_{m+1} \text{ and } E_m \text{ are p.d. if } \nu=2m \}. \end{aligned}$$

It will be necessary to simulate in the moment space in order to develop credible regions for $\underline{c}$. Unfortunately, because of the structure of Ω^*, direct simulation proved to be very difficult to carry out. However, it turns out that there is a simple transformation from Ω^* to a space which is numerically convenient to work in. We describe this transformation in the following theorem, proof of which can be seen in pp44-48 of Karlin and Studden(1966).

Theorem 3.1 (Karlin and Studden): Let $\underline{c} \in \Omega^*$. Then, $\underline{c}$ has a unique representation given by

$$c_i = \sum_{j=1}^{m} \lambda_j t_j^i \qquad \text{for } i=0,1,\ldots,\nu, \tag{3.2}$$

where $\underline{t}=(t_1,\ldots,t_m)$ and $\underline{\lambda}=(\lambda_1,\ldots,\lambda_m)$ satisfy

$$(\underline{t},\underline{\lambda}) \in \Omega = \left\{ (\underline{t},\underline{\lambda}) : 0 < t_1 < \cdots < t_m < 1,\ 0 < \lambda_i < 1, \text{ and } \sum_{j=1}^{m} \lambda_j = 1 \text{ if } \nu=2m-1,\ \sum_{j=1}^{m} \lambda_j \le 1 \text{ if } \nu=2m. \right\}$$

Using the above representation, the likelihood in (3.1) can be written as

$$\prod_{i=1}^{N} \left[\sum_{j=1}^{m} \lambda_j \binom{n_i}{x_i} t_j^{x_i} (1-t_j)^{n_i-x_i} \right] I_\Omega((\underline{t},\underline{\lambda})). \tag{3.3}$$

Note that Ω has simpler structure than Ω^* of (3.1) - this makes the likelihood in (3.3) more tractable than that in (3.1).

To reflect the uncertainty in $\underline{c}$ or, equivalently, in $(\underline{t},\underline{\lambda})$, we will consider credible sets for $(\underline{t},\underline{\lambda})$ based on the constant noninformative prior

$$\pi(\underline{t},\underline{\lambda}) = I_\Omega((\underline{t},\underline{\lambda})).$$

Then the posterior distribution is proportional to

$$\pi^*(\underline{t},\underline{\lambda}) = \prod_{i=1}^{N} \left[\sum_{j=1}^{m} \lambda_j \binom{n_i}{x_i} t_j^{x_i} (1-t_j)^{n_i-x_i} \right] I_\Omega((\underline{t},\underline{\lambda})).$$

Thus, the $100\gamma\%$ HPD-credible region, say $R(s_\gamma)$, of $(\underline{t},\underline{\lambda})$ is given by

$$R(s_\gamma) = \{ (\underline{t},\underline{\lambda}) \in \Omega : \pi^*(\underline{t},\underline{\lambda}) > s_\gamma \},$$

where s_γ is chosen to satisfy

$$\int_{R(s_\gamma)} \pi^*(\underline{t},\underline{\lambda})d\underline{t}d\underline{\lambda} = \gamma \int_{\Omega} \pi^*(\underline{t},\underline{\lambda})d\underline{t}d\underline{\lambda} .$$

That the posterior distribution is unimodal with unique mode, and (hence) the HPD region $R(s_\gamma)$ is unique and connected are easy to verify. Note, however, that numerical solution for s_γ is hard due to the difficulty in integrating over sets of the form $R(s_\gamma)$. In this paper, we find the value of s_γ using Monte Carlo integration as described below. First we generate an "importance" sample of ℓ=10,000 values $(\underline{t}_i,\underline{\lambda}_i)$, $i=1,\ldots,\ell$, of $(\underline{t},\underline{\lambda}) \in \Omega$ by generating each $\underline{t}_i$ as an ordered sample from the uniform distribution over (0,1), and (independently) generating each $\underline{\lambda}_i$ from the Dirichlet(1,...,1) distribution. Then, for a given s, we approximate the posterior probability of R(s) by

$$\rho(s) = \frac{\sum_{i=1}^{\ell} \pi^*(\underline{t}_i,\underline{\lambda}_i)\ \delta_s(i)}{\sum_{i=1}^{\ell} \pi^*(\underline{t}_i,\underline{\lambda}_i)} ,$$

where $\delta_s(i) = 1$ if $\pi^*(\underline{t}_i,\underline{\lambda}_i) > s$, and $= 0$ otherwise. Then, the value of s_γ, for given γ, is approximated by choosing that value of s for which $\rho(s)$ is as close as possible to γ.

The results of this simple Monte Carlo integration proved to be accurate, probably due to the fact that we are integrating a smooth unimodal function over a small region. Thus we did not attempt to obtain a more sophisticated importance function. We again emphasize, however, the value of simulating the reparametrized $(\underline{t},\underline{\lambda})$ rather than the original moments $\underline{c}$; early attempts at direct simulation of $\underline{c}$ proved unsuccesful.

3.2 Bounds for the Bayes estimate

Let the sample size at (the current) stage N be denoted by n^*, i.e.: $n_N=n^*$. Here, we describe how we calculate bounds for the Bayes estimate $\hat{\theta}_N(k)$ when the uncertainty in $(\underline{t},\underline{\lambda})$ or, equivalently, in $\underline{c}$ is given by $R(s_\gamma)$. We first consider the case where $n^*=\nu$.

For fixed $(\underline{t},\underline{\lambda}) \in R(s_\gamma)$, let $c_i(\underline{t},\underline{\lambda})$ be the $i^{th}(i=0,1,\ldots,\nu)$ moment given by (3.2). Also, let the bounds on the $(\nu+1)^{st}$ moment, when the first ν moments are fixed at these values, be $c_l(\underline{t},\underline{\lambda})$ and $c_u(\underline{t},\underline{\lambda})$. (Note that $c_l(\underline{t},\underline{\lambda})$ and $c_u(\underline{t},\underline{\lambda})$ can each be written as a function of $(\underline{t},\underline{\lambda})$ using Theorem A.2 and equation (3.2).) Now, let

$$\alpha(\underline{t},\underline{\lambda})= \alpha(c_1(\underline{t},\underline{\lambda}),\ldots,c_\nu(\underline{t},\underline{\lambda})) \quad \text{and} \quad \beta(\underline{t},\underline{\lambda})= \beta(c_1(\underline{t},\underline{\lambda}),\ldots,c_\nu(\underline{t},\underline{\lambda})) \qquad (3.4)$$

where $\alpha(c_1,\ldots,c_\nu)$ and $\beta(c_1,\ldots,c_\nu)$ are as in (2.4).

Finally, using (2.5) and (2.6), the lower and upper bounds over all priors with moments in $R(s_\gamma)$ for the Bayes estimate $\hat{\theta}_N(k)$ of θ_N, when $X_N=k$, are given, respectively, by

$$\underset{\sim}{\theta}_N(k) = \inf\{ \alpha(\underline{t},\underline{\lambda})c_*(\underline{t},\underline{\lambda}) + \beta(\underline{t},\underline{\lambda}) \quad : \quad \pi^*(\underline{t},\underline{\lambda}) > s_\gamma \} \tag{3.5}$$

and

$$\tilde{\theta}_N(k) = \sup\{ \alpha(\underline{t},\underline{\lambda})c^*(\underline{t},\underline{\lambda}) + \beta(\underline{t},\underline{\lambda}) \quad : \quad \pi^*(\underline{t},\underline{\lambda}) > s_\gamma \}. \tag{3.6}$$

The case $\gamma=0$ is of special interest, since $R(s_0)$ is then the Type-II maximum likelihood estimate of $(\underline{t},\underline{\lambda})$, i.e.: the point that maximizes the likelihood in (3.3). The ensuing range is thus the range of the possible estimates corresponding to the "most likely" moments of G.

When $n^*<\nu$, the corresponding bounds on $\hat{\theta}_N(k)$ are given by the inf and sup of

$$\{ \alpha(\underline{t},\underline{\lambda})c_{n^*+1}(\underline{t},\underline{\lambda}) + \beta(\underline{t},\underline{\lambda}) \quad : \quad \pi^*(\underline{t},\underline{\lambda}) > s_\gamma \}$$

where $\alpha(\underline{t},\underline{\lambda})$ and $\beta(\underline{t},\underline{\lambda})$ are as in (3.4) with ν now replaced by n^*.

<u>Example 3.1:</u> Suppose N=41 and that $n_i=5$ for $i=1,\dots,N$. Suppose the past N-1=40 data values $x_1,x_2,\dots,x_{40}$ are summarized by the following:

k	0	1	2	3	4	5
Freq.	4	6	10	10	6	4

The values of the bounds $\underset{\sim}{\theta}_N(k)$ and $\tilde{\theta}_N(k)$ on the Bayes estimate are displayed in Figures 3.1 and 3.2 for k=2 and 4, as γ varies from 0 to 0.99. (In these calculations, the likelihood of $(\underline{t},\underline{\lambda})$, see (3.3), has been obtained using all

N data values, both past and current.). Note that the bounds corresponding to the ML-II moment estimate (i.e., corresponding to γ=0) are very tight, quickly spreading apart as γ increases to 0.1, and then remaining reasonably stable until γ reaches to 0.9. This clearly shows the unsuitability of using the ML-II moment estimate to determine robustness, and perhaps, surprisingly indicates that any moderately large γ should provide a reasonable indicator of robustness.

Example 3.2: Suppose N=42 and that n_i=5 for all i. Suppose the past N-1=41 observations are summarized as below:

k	0	1	2	3	4	5
Freq.	0	3	6	9	14	9

Note that this data is skewed as opposed to the (symmetric) data in the previous example. The bounds on the Bayes estimate for the cases k=0 and 4 are displayed in Figures 3.3 and 3.4. The extreme nonrobustness when k=0 probably relates to the clash of the "current" observation (k=0) with the past data.

Example 3.3 (Martz and Lian(1974)):

The Portsmouth Naval Shipyard, Portsmouth, NH, must routinely assess the quality of submitted lots of vendor produced material. The data consist of the number of defects of a specified type in samples of size n=5 from the past N-1=5 lots and from the current (sixth) lot. The past data are summarized by the following.

k	0	1	2	3	4	5
Freq.	3	1	0	0	0	1

The number of defects in the current lot is 0. It is desired to estimate the lot fraction defective, θ, in the current lot. The bounds on the Bayes estimate of θ are displayed in Figure 3.5.

Finding the values of $\underset{\sim}{\theta}_N(k)$ and $\tilde{\theta}_N(k)$ in the above examples required maximizing/minimizing functions of 5 variables, with the variables being subject to constraints. The calculations were done using the IMSL routine NCONF. With the value of s_γ given (for fixed γ), it took, in most cases, about 2.5 seconds CPU time on an AMDAHL/470 V7 (VM/CMS) to compute a single bound. For the routine NCONF to work, we had to shrink the range, viz: (0,1), of each variable by 0.01 at both end points.

APPENDIX

Let $c_0=1$ and c_i denote the i^{th} moment of a prob. distribution on the interval [0,1]. Thus, we define the (n+1)-moment space $\mathcal{M}$ by

$$\mathcal{M} = \{ (c_0, c_1, \ldots, c_n) : c_i = \int \theta^i dG(\theta) \text{ for some prob. dist. G on } [0,1] \}. \quad (A.1)$$

For use in this section, we also define square matrices A_{m+1}, B_{m+1}, D_{m+1} and E_m (where the suffix indicates the order of the matrix) as follows:

$$A_{m+1} = (c_{i+j-1}) \quad , \quad B_{m+1} = (c_{i+j-2} - c_{i+j-1}) ,$$
$$\qquad (A.2)$$
$$D_{m+1} = (c_{i+j-2}) \text{ and } E_m = (c_{i+j-1} - c_{i+j}).$$

Moreover, let A^0_{m+1} and B^0_{m+1} respectively be the matrices obtained by substituting $c_{2m+1} = 0$ in A_{m+1} and B_{m+1}, and let D^0_{m+1} and E^0_{m+1} respectively be the matrices obtained by substituting $c_{2m} = 0$ in D_{m+1} and E_m.

A necessary and sufficient condition for an (n+1)-vector $(c_0,\ldots,c_n)$ to be in the moment space $\mathcal{M}$ is given in the following theorem. Proof can be seen in, e.g: Karlin and Studden(1966).

<u>Theorem A.1 (Karlin and Studden)</u>:

If n=2m+1 (n=2m), then $\underline{c} \in \mathcal{M}$ if and only if the matrices A_{m+1} and B_{m+1} (the matrices D_{m+1} and E_m) are positive semi definite. In addition, $\underline{c} \in \text{int } \mathcal{M}$ if and only if these two matrices are positive definite.

It is of interest to consider the range of the (n+1)st moment c_{n+1} subject to the first n moments $c_0, c_1, \ldots, c_n$ being fixed. Thus, let c_l and c_u respectively be the lower and upper bounds of c_{n+1} defined by

$$c_l = \inf \{ \int \theta^{n+1} dG(\theta) : \int \theta^i dG(\theta) = c_i \text{ for } i = 0,1,\ldots,n \}$$

and

$$c_u = \sup \{ \int \theta^{n+1} dG(\theta) : \int \theta^i dG(\theta) = c_i \text{ for } i = 0,1,\ldots,n \}.$$

<u>Theorem A.2 (Karlin and Studden)</u>: Let $\underline{c} = (c_0,\ldots,c_n)$ be in the interior of $\mathcal{M}$. Then, c_l and c_u are given as follows. When n=2m,

$$c_l = -\frac{\det(A^0_{m+1})}{\det(A_{m+1})} \quad \text{and} \quad c_u = c_n + \frac{\det(B^0_{m+1})}{\det(B_m)}.$$

When n=2m+1,

$$c_l = - \frac{\det(D^0_{m+2})}{\det(D_{m+1})} \quad \text{and} \quad c_u = c_n + \frac{\det(E^0_{m+1})}{\det(E_m)} .$$

Proof: Follows from Corollary 2.2b on p112 of Karlin and Studden(1966).

REFERENCES

Albert, J.H.(1984). Empirical Bayes estimation of a set of binomial probabilities. *J. Statist. Comput. Simul.* 20, 129-144.

Atchison, T.A. and Martz, H.F.(Eds.) (1969). *Proceedings of the Symposium on Empirical Bayes Estimation and Computing in Statistics.* Series No. 6, Texas Technical University, Lubbock.

Berger, J.(1984). The robust Bayesian viewpoint (with discussion). In *Robustness of Bayesian Analysis*, J. Kadane(Ed.). North Holland, Amsterdam.

Berger, J.(1985). *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, New York.

Berger, J. Berliner, L.M.(1986). Robust Bayes and empirical Bayes analysis analysis with ε-contaminated priors. *Ann. Statist.* 14, 461--486.

Berger, J. and O'Hagan, A.(1988). Ranges of posterior probabilities for unimodal priors with specified quantiles. In *Bayesian Statistics III*, J.M.Bernado, M.H.DeGroot, D.V.Lindley, and A.F.M. Smith (Eds.). Oxford University Press, Oxford.

Berliner, L.M.(1984). A decision-theoretic structure for robust Bayesian analysis with applications to the estimation of a multivariate normal mean. In

Bayesian Statistics II, J.M.Bernado, M.H.DeGroot, D.V.Lindley, and A.F.M. Smith (Eds.). North Holland, Amsterdam.

Copas, J.B.(1972). Empirical Bayes methods and the repeated use of a standard. Biometrica 59, 349-360.

DasGupta, A. and Studden, W.J.(1988). Robust Bayesian analysis and optimal experimental designs in normal linear models with many parameters. Technical Report #88-14, Department of Statistics, Purdue University, W.Lafayette, IN.

DeRobertis, L. and Hartigan, J.A.(1981). Bayesian inference using intervals of measures. *Ann. Statist.* 1, 235-244.

Griffin, B.S. and Krutchkoff, R.G.(1971). Optimal linear estimators: an empirical Bayes version with application to the binomial distribution. *Biometrica* 58, 195-201.

Jackson, D.A., O'Donovan, T.M. Zimmer, W.J. and Deely, J.J.(1970). $\mathcal{G}_2$-minimax estimators in the exponential family. *Biometrica* 57, 439-443.

Karlin, S. and Studden, W.J.(1966) *Tchebycheff Systems: With Applications in Analysis and Statistics*. John Wiley & Sons, New York.

Lin, P.E.(1972). Rates of convergence in empirical Bayes estimation problem: Discrete Case. *Ann. of Inst. of Statist. Math.* 24, 319-325.

Maritz, J.S.(1966). Smooth empirical Bayes estimation for one-parameter discrete distributions. *Biometrica* 53, 417-429.

Maritz, J.S.(1970). *Empirical Bayes Methods*. Methuen, London.

Martz, H.F. and Lian, M.G.(1974). Empirical Bayes estimation of the binomial parameter. *Biometrica* 61, 517-523.

O'Hagan, A. and Berger, J.(1988). Range of posterior probabilities for quasi-unimodal priors with specified quantiles. *Journal Amer. Statist. Assoc.* 83, 503-508.

Robbins, H.(1955). An empirical Bayes approach to statistics. *Proc. Third Berkely Symp.* 1, 157-164.

Sivaganesan, S. and Berger, J.(1989). Range of posterior measures for priors with unimodal contaminations. *Ann. Statist.* 17, 868-889.

Wasserman, L. A. (1989). A robust Bayesian interpretation of likelihood regions. *Ann. Statist.* 17, 1387-1393.

Department of Mathematical Sciences
University of Cincinnati
Cincinnati, OH 45221-0025.

Department of Statistics
Purdue University
W. Lafayette, IN 47907.

Figure 3.1: Lower and Upper Bounds on the Bayes Estimate $\hat{\theta}_N$ for varying γ when k=2 and data are as in Example 3.1.

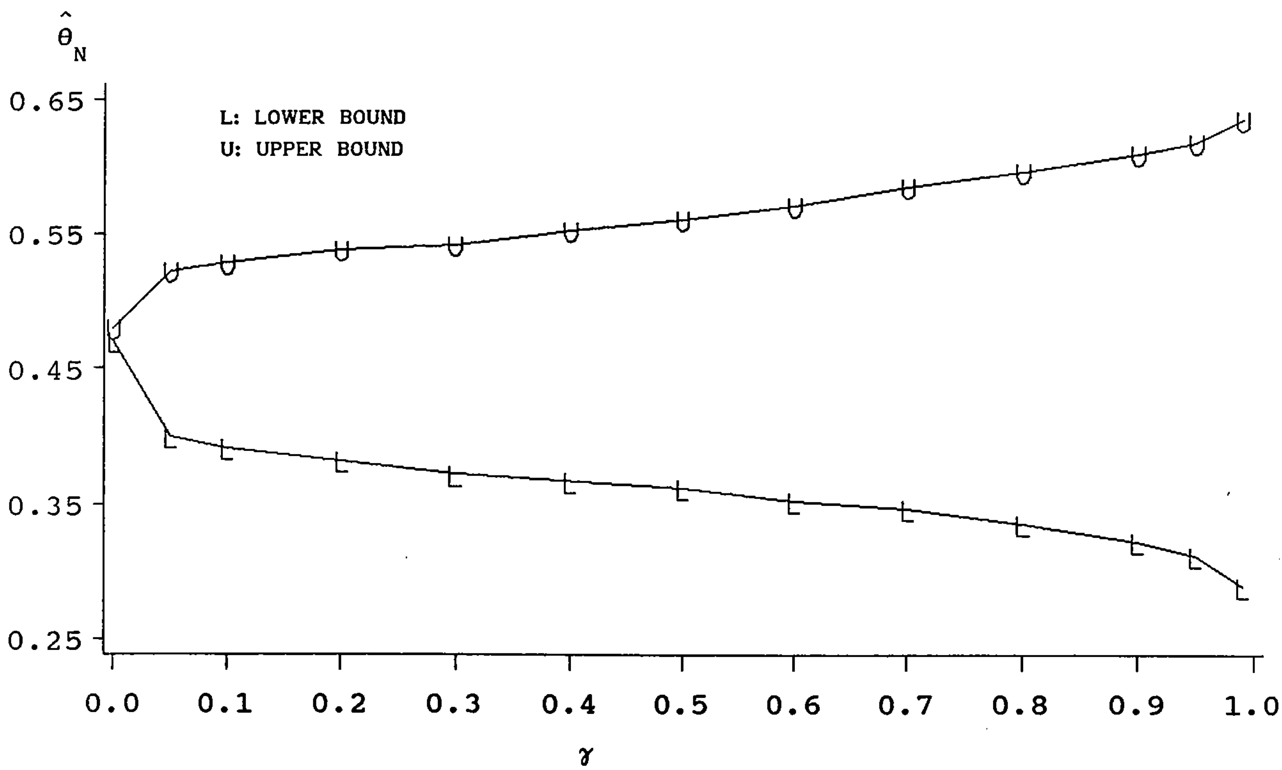

Figure 3.2: Lower and Upper Bounds on the Bayes Estimate $\hat{\theta}_N$ for varying γ when k=4 and data are as in Example 3.1.

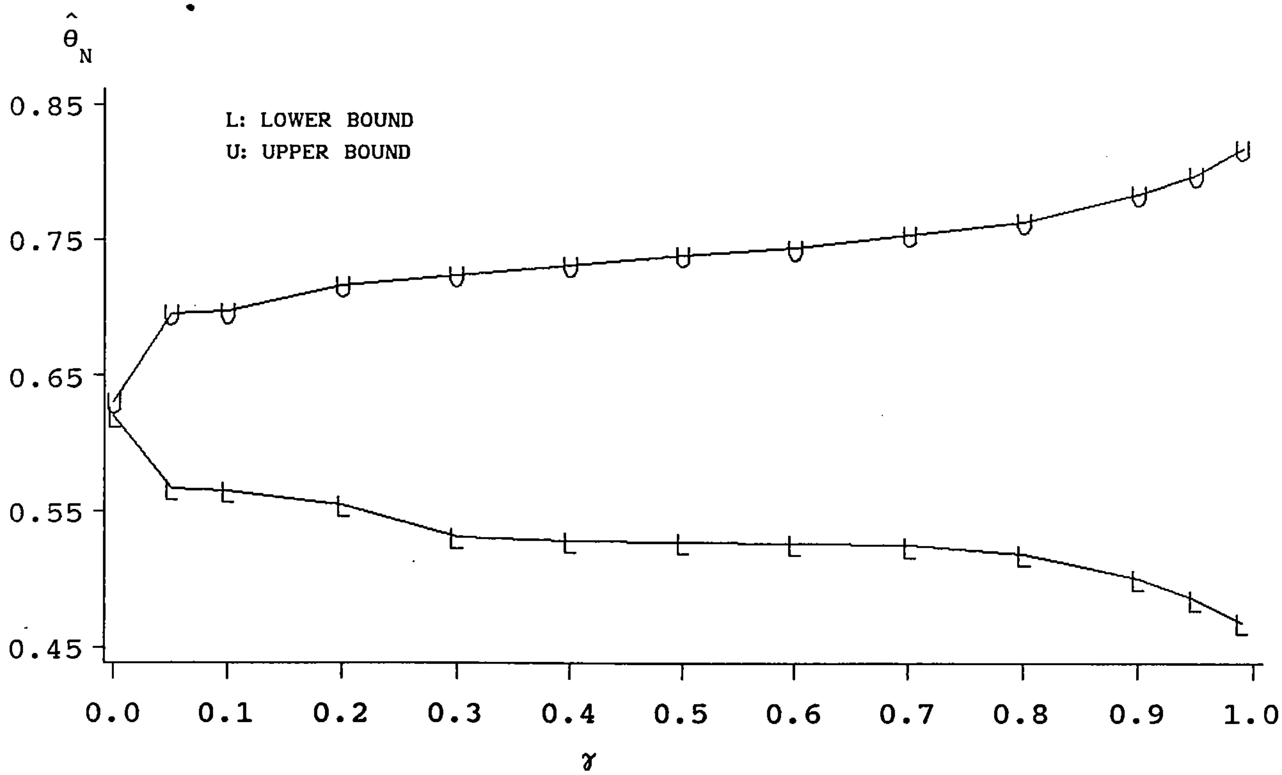

Figure 3.3: Lower and Upper Bounds on the Bayes Estimate $\hat{\theta}_N$ for varying γ when k=0 and data are as in Example 3.2.

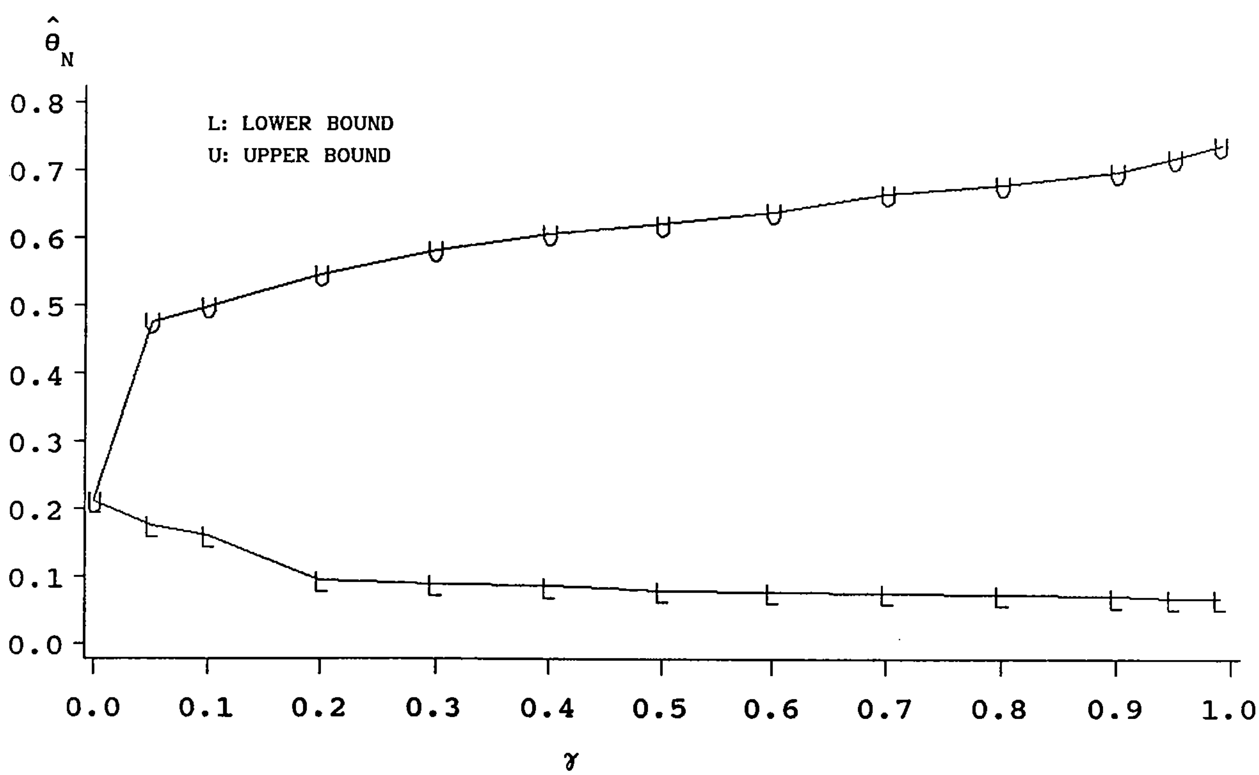

Figure 3.4: Lower and Upper Bounds on the Bayes Estimate $\hat{\theta}_N$ for varying γ when k=4 and data are as in Example 3.2.

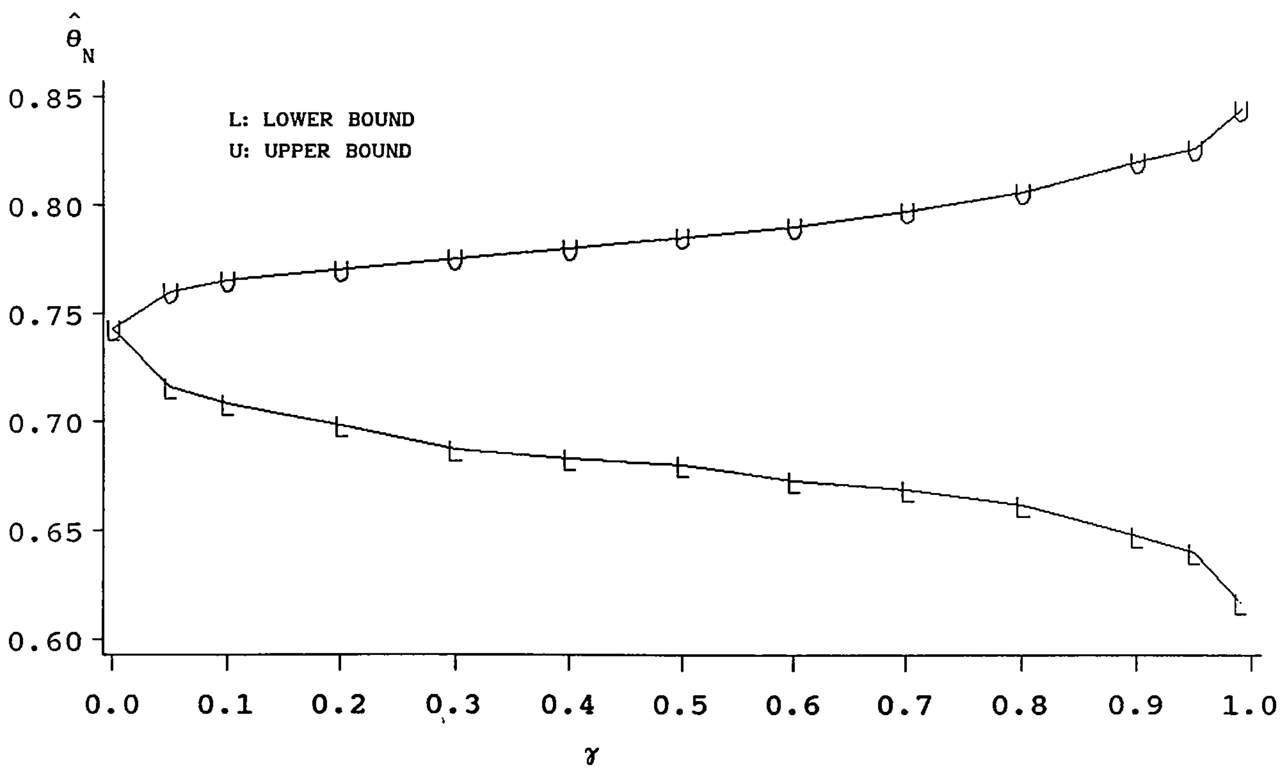

Figure 3.5: Lower and Upper Bounds on the Bayes Estimate $\hat{\theta}_N$ for varying γ when the data are as in Example 3.3.