EMPIRICAL BAYES ESTIMATION OF THE
BINOMIAL PARAMETER

by

L. Billard\(^1\) and M. Meshkani

FSU Statistics Report M429

August, 1977
The Florida State University
Department of Statistics
Tallahassee, Florida

\(^1\)Research supported by the National Science Foundation Grant
Number MCS76-10453.
EMPIRICAL BAYES ESTIMATION OF THE
BINOMIAL PARAMETER

by

L. Billard and M. Meshkani
The Florida State University

ABSTRACT

A study of empirical Bayes estimation procedures for the parameter of a
binomial distribution is undertaken. Various procedures are derived. Properties
of the resulting estimators are investigated. In particular, it is shown that
some of these estimators are asymptotically optimal and are asymptotically normally
distributed.
1. INTRODUCTION

This work is concerned with empirical Bayes estimation of the parameter $\lambda$ of a binomial distribution $B(\lambda, N)$. Mark and Lian [6] recently undertook a simulation study of some estimators of $\lambda$. Though they provided some very useful insights into the estimators considered, the scope of their study can in fact be expanded somewhat. Therefore, additional estimators are derived. Further, an investigation of limiting properties of some of the previous and present estimators is undertaken.

The justification to consider empirical Bayes procedures as an alternative to conventional and pure Bayesian approaches to statistical inference has been well documented by Neyman [7]. The general theory has been ably expounded by Robbins [9, 10, 11] and Moritz [5]. Applications of the binomial distribution abound. However, another application of ultimate interest pertains to estimation in two-state Markov chains. Estimation theory in these situations is unfortunately relatively sparse. Since the very nature of such data sets suggests empirical procedures may be appropriate, it seems desirable to investigate more fully the present and proposed estimators.

The empirical Bayes estimators of $\lambda$ are presented in Section 2. In Section 3, asymptotic properties of the estimators are considered. Finally, results of a numerical study are given in Section 4.
2. EMPIRICAL BAYES ESTIMATION OF $\lambda$

2.1 Preliminary Results.

Let $X$ be a binomially distributed random variable, $B(\lambda, N)$, on the Euclidean state space $X$. Let the probability of success on a single trial be $\lambda$, a particular realization of the random variable $\Lambda$ defined on the Euclidean state space $L$. Let the action space be denoted by $A$ with particular element $a$. The loss function is given by

$$L(a, \lambda) = (a - \lambda)^2.$$

Let $x_i$, $i = 1, \ldots, n+1$, be the number of successes among $N$ trials in the $i$th of $(n + 1)$ independent experiments. The first $n$ experiments are deemed to have occurred in the past with the $x_{n+1}$ being the observation of the current experiment. Two prior distributions for $\Lambda$ are to be considered, viz., the general unspecified prior $G$ and the natural conjugate prior, namely the beta distribution

$$g(\lambda) = \lambda^{r-1}(1 - \lambda)^{s-1}/\beta(r, s), \quad 0 \leq \lambda \leq 1, \quad r > 0, \quad s > 0$$

(2.1)

where $r$ and $s$ are unspecified. Hence, the parametric probability function is

$$p_{G}(x) = \binom{N}{x} \frac{\beta(r+x, N+s-x)}{\beta(r+s)}.$$  (2.2)

From Robbins [9], the Bayes estimator based on $G$ is

$$\delta_{G}(x) = \frac{x + 1}{N} \frac{p_{G,N}(x + 1)}{p_{G,N-1}(x)}$$  (2.3)

where

$$p_{G,N}(x) = \Pr(X_G = x \text{ in } N \text{ trials})$$

with $X_G$ is the random variable associated with the marginal distribution of $X$ with respect to $G$. 
Simple estimators of these probabilities can be found in terms of \( G_{N,n}(x) \) defined as the absolute frequency of \( x \) among \( n \) observations on experiments each of \( N \) trials. Thus, the so-called simple empirical Bayes estimator of \( \lambda \) is given by

\[
\delta_1 = \delta_n(x) = \frac{x + 1}{N} \frac{G_{N,n}(x + 1)}{[1 + G_{N-1,n}(x)]};
\]

(2.3a)

if \( G_{N,n}(x + 1) = 0 \), the maximum likelihood estimator is used, that is,

\[
\delta_1 \equiv \delta_0 = x/N.
\]

(2.3b)

The Bayes estimator based on the beta prior is

\[
\delta_\beta(x) = \frac{x + r}{N + r + s}.
\]

(2.4)

The problem then is to find estimators for \( r \) and \( s \) in (2.4). Note there are \( n \) observations on \( p_\beta(x) \). Hence, conventional methods of estimation of the \( r \) and \( s \) can be implemented. This operation produces smoothed Bayes estimators of \( \lambda \). Maritz [5, Ch. 3] provides a variety of general methods for estimation of the prior distribution. This general theory as it pertains to the binomial parameter can be applied.

2.2 Smoothed empirical Bayes estimators.

2.2.1 Smoothing by method of moments.

Using (2.2), it is readily shown that

\[
E(\delta_\beta) = \frac{Nr}{r + s}, \quad E(\delta_\beta^2) = \frac{Nr(N(r + 1) + s)}{(r + s)(r + s + 1)}.
\]
The first and second sample moments are defined by

\[ m'_1 = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad m'_2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2. \]

Equating sample and theoretical moments of like order, estimators for \( r \) and \( s \) are obtained as

\[ \tilde{r} = \frac{m'_2 - N m'_1}{(N-1)(m'_1)^2 + N(m'_1 - m'_2)} \]

and

\[ \tilde{s} = \frac{(N - m'_1)(m'_2 - N m'_1)}{(N-1)(m'_1)^2 + N(m'_1 - m'_2)} \]

Thus, the empirical Bayes estimator of \( \lambda \) is

\[ \delta_2 = \begin{cases} 
\frac{(x + \tilde{r})}{N + \tilde{r} + \tilde{s}}, & \tilde{r}, \tilde{s} > 0, \\
x/N, & \text{otherwise.}
\end{cases} \quad (2.5) \]

It is sometimes deemed more desirable to equate the variances rather than the second moments. Thus, it is readily shown that

\[ \text{Var}(X_\beta) = \frac{Nrs(N + r + s)}{(r + s)^2(1 + r + s)}. \]

It is convenient to reparameterize to

\[ \gamma = r/(r + s) \quad \text{and} \quad \theta = 1/(r + s). \]

Then, equating the corresponding sample and theoretical values, it can be shown that (Griffiths [2])

\[ \tilde{\gamma} = m'_1/N \]
and
\[ \tilde{\theta} = \frac{S^2 N \tilde{\gamma}(1-\tilde{\gamma})}{N^2 \tilde{\gamma}(1-\tilde{\gamma}) - s^2} \]

where
\[ s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - m_1)^2 \]

Hence the empirical Bayes estimator of \( \lambda \) is
\[ \hat{\theta}_3 = \begin{cases} 
  (\tilde{\gamma} + x \tilde{\theta})/(1 + N \tilde{\theta}), & \tilde{\theta} > 0, \\
  x/N, & \text{otherwise.} 
\end{cases} \quad (2.6) \]

2.2.2 Smoothing by maximum likelihood.

The likelihood function is
\[ L = L(r, s|x) = \prod_{x=0}^{n} \{ p_\beta(x) \}^n \]

Due to the special nature of \( p_\beta(x) \), maximization of \( L \), or \( \log L \), requires an iterative technique. One such technique is the method of scoring described by Rao [8, p. 366] using the moment estimators as starting values.

Define
\[ R = \frac{\partial \log L}{\partial r} = \sum_{x=0}^{N} \chi_{N,n}^n(x) \frac{\partial \log p_\beta(x)}{\partial r} \]

and
\[ S = \frac{\partial \log L}{\partial s} = \sum_{x=0}^{N} \chi_{N,n}^n(x) \frac{\partial \log p_\beta(x)}{\partial s} . \]

The information matrix is defined by
\[ I = \begin{pmatrix} ER^2 & ERS \\ ERS & ES^2 \end{pmatrix} \]
where

\[ \text{ERS} = \sum_{x=0}^{N} \frac{3 \log p_\beta(x)}{\partial r} \cdot \frac{3 \log p_\beta(x)}{\partial s} \cdot p_\beta(x) \]

and where \( \text{ER}^2 \) and \( \text{ES}^2 \) are defined similarly. Let the index zero denote the initial value for each iteration. Let \( \Delta r = r - r_0 \) and \( \Delta s = s - s_0 \). Then,

\[
\begin{pmatrix}
\text{ER}_0^2 & \text{ER}_0 \text{ES}_0 \\
\text{ER}_0 \text{ES}_0 & \text{ES}_0^2
\end{pmatrix}
\begin{pmatrix}
\Delta r \\
\Delta s
\end{pmatrix} = 
\begin{pmatrix}
R_0 \\
S_0
\end{pmatrix}.
\]

Evaluation of \( \frac{\partial \log p_\beta(x)}{\partial r} \) is accomplished via the digamma function which is tabulated in, for example, Abramowitz and Stegun [1, p. 258]. When \( x \) is sufficiently large the functional approximation \( \frac{d \log \Gamma(x)}{dx} = \log x - x/2 \) may be used (Kale [3]).

Iteration is continued until stable values, \( \hat{r} \) and \( \hat{s} \), are obtained for \( r \) and \( s \), respectively. Hence, the corresponding smoothed empirical Bayes estimator of \( \lambda \) is

\[
\delta_4 = \begin{cases} 
(x + \hat{r})/(N + \hat{r} + \hat{s}), & \hat{r}, \hat{s} > 0, \\
x/N, & \text{otherwise}.
\end{cases}
\] (2.7)

2.2.3 Direct smoothing.

The Bayes estimator of (2.4) can be written as a linear equation in \( x \), viz.,

\[
\delta_\beta \beta (x) = \frac{r}{N + r + s} + \frac{1}{N + r + s} \cdot x
\]

\[= A + Bx, \quad A, B > 0.\]
This suggests an empirical Bayes estimator of the form
\[ \delta^* = \delta^*(x) = A^* + B^* x, \quad A^*, B^* > 0. \]

Now, the risk function, \( W(\delta) \), may be written as
\[
W(\delta) = \iint (\delta(x) - \lambda)^2 \, d\Phi(x|\lambda) \, dG(\lambda)
\]
\[ = W(\delta_G) + K(\delta, \delta_G) \]

where
\[
K(\delta, \delta_G) = \int (\delta(x) - \delta_G(x))^2 \, d\Phi(x).
\]

The estimator \( \delta^* \) converges to \( \delta_\beta \) as \( K(\delta^*, \delta_\beta) \) converges to zero. Suppose \( \delta_n = \delta_n(x) \) is any other empirical Bayes estimator. Then,
\[
K(\delta_n, \delta_\beta) = K(\delta^*, \delta_\beta) + \sum_{x=0}^{N} (\delta_n - \delta^*)^2 \, p_\beta(x)
\]
\[ + 2 \sum_{x=0}^{N} (\delta_n - A^* - B^* x)(A^* - A) + (B^* - B)x \, p_\beta(x). \]

If the cross-product term is zero for all \( x \), it follows that for all \( x \),
\[
K(\delta^*, \delta_\beta) \leq K(\delta_n, \delta_\beta).
\]

Hence, setting that cross-product term at zero, we have the system
\[
\begin{bmatrix}
1 & \sum_x x p_\beta(x) \\
\sum_x x p_\beta(x) & \sum_x x^2 p_\beta(x)
\end{bmatrix}
\begin{bmatrix}
A^* \\
B^*
\end{bmatrix}
= \begin{bmatrix}
\sum_x \delta_n p_\beta(x) \\
\sum_x x \delta_n p_\beta(x)
\end{bmatrix}
\]

(2.6)
which is solved iteratively. A straight line is fitted to the plot of \( \delta_n(x) \) against \( x \) where \( \delta_n(x) \) is found from (2.3), that is,

\[
\delta_n(x) = \frac{x + 1}{N} \frac{p_\beta(x + 1)}{p_\beta(x)}.
\]

Hence, for example,

\[
p_\beta(1) = \delta^0(0) p_\beta(0) N,
\]

and likewise for \( x = 2, 3, \ldots, N \). Thus new values of \( p_\beta(x) \) to be used at each stage of the iteration are found.

Iteration of (2.6) is continued until stable solutions \( A^* \) and \( B^* \) are found. Hence, the empirical Bayes estimator is

\[
\delta_e = \begin{cases} 
A^* + B^* x, & A^*, B^* > 0, \\
x/N, & \text{otherwise.}
\end{cases}
\]

(2.7)

3. ASYMPTOTIC RESULTS

3.1 Asymptotic optimality.

Asymptotic optimality (a.o.) in a strong and weak sense has been defined by Robbins [11] and Maritz [5], respectively, as follows.

Definition 3.1 Strong asymptotic optimality.

The estimator \( \delta_n(x) \) of \( \lambda \) is a.o. (E) if
\[
\lim_{n \to \infty} E \{ W(\delta_n) \} = W(\delta_G)
\]

where the expectation is with respect to the marginal distribution of \((X_1, \ldots, X_n)\).

**Definition 3.2**  Weak asymptotic optimality.

The estimator \(\delta_n(x)\) of \(\lambda\) is a.o. (P) if \(W(\delta_n)\) converges in probability to \(W(\delta_G)\).

Suppose \(\delta_n(x)\) is truncated at the finite values \(L\) and \(U\). Maritz [5] proves the following result.

**Theorem 3.1**

Let

\[
\int \{ \delta_G(x) \}^2 dF_G(x) < \infty
\]

(3.1)

and

\[
\delta_n(x) \xrightarrow{P} \delta_G(x),
\]

for all \(x\), such that \(\delta_G(x) \in (L, U)\). Then,

\[
W(\delta_n) \to W(\delta_G) + \epsilon', \quad \text{as} \quad n \to \infty,
\]

where \(\epsilon'\) can be made arbitrarily small. That is, \(\delta_n(x)\) is a.o. (P).

In our situation, there are natural bounds on \(\delta_n(x)\), viz., \((L, U) = (0, 1)\). Therefore, it is easy to show that a.o. (P) is equivalent to a.o. (E).

The asymptotic optimality of some of the above estimators of \(\lambda\) can now be proved. It is assumed that the appropriate conditions ensuring the empirical Bayes estimator is not equal to the maximum likelihood estimator are satisfied.

**Theorem 3.2**

The empirical Bayes estimators \(\delta_1, \delta_2, \text{ and } \delta_3\) of equations (2.3), (2.5) and (2.6), respectively, are asymptotically optimal.
Proof:

(i) To prove \( \delta_1 \) is a.o. -- For fixed \( x \), let

\[
y_i = y_{i,N}(x + 1) = \begin{cases} 
1, & \text{if } X_{G,i} = x + 1 \text{ in } N \text{ trials}, \\
0, & \text{otherwise};
\end{cases}
\]

and

\[
z_i = z_{i,N-1}(x) = \begin{cases} 
1, & \text{if } X_{G,i} = x \text{ in first } N - 1 \text{ trials}, \\
0, & \text{otherwise},
\end{cases}
\]

for \( i = 1, \ldots, n \). Then, \( (y_i, z_i) \) are independent identically distributed random variables with

\[
E(y_i, z_i) = \begin{pmatrix} p_{G,N}(x + 1), & p_{G,N-1}(x) \end{pmatrix}^T.
\]

It is readily seen that

\[
n^{-1} g_{N,n}(x + 1) = n^{-1} \sum_{i=1}^{n} y_i \xrightarrow{P} p_{G,N}(x + 1)
\]

and

\[
n^{-1} g_{N-1,n}(x) \xrightarrow{P} p_{G,N-1}(x).
\]

Since \( \delta_1 \) is a continuous function of \( g_{N,n}(x + 1) \) and \( g_{N-1,n}(x) \), it follows that (see, for example, Serfling [12, p. 25])

\[
\delta_n(x) \xrightarrow{P} \delta_G(x)
\]

where \( \delta_G(x) \) is as defined in (2.3) and

\[
g_{N-1,n}(x) + 1 \neq 0, \quad \text{for all } x.
\]
Further, the condition (3.1) is seen to hold as
\[
\sum_{x=0}^{N-1} \left( \frac{(x+1)}{N} \frac{p_{G,N}(x+1)}{p_{G,N-1}(x)} \right)^2 p_{G,N}(x) < \infty
\]
since \( p_{G,N-1}(x) \neq 0 \), for all \( x \). Therefore, \( \delta_1 \) is asymptotically optimal.

(ii) To prove \( \delta_2 \) is a.o. -- Let us write \( M^\ast = (m_1^\ast, m_2^\ast) \) and \( \mu^\ast = (\mu_1^\ast, \mu_2^\ast) \).

Since \( \|\mu\| < \infty \), the strong law of large numbers implies
\[
M \xrightarrow{a.s.} \mu
\]
and since \( \tilde{r} \) and \( \tilde{s} \) are continuous functions, it follows that (Serfling [12])
\[
\tilde{r} \xrightarrow{a.s.} r \quad \text{and} \quad \tilde{s} \xrightarrow{a.s.} s.
\]

Thus,
\[
\delta_2 \xrightarrow{p} \delta_B(x).
\]

Further, the condition (3.1) holds since there is a finite sum of finite summands.

Hence, \( \delta_2 \) is asymptotically optimal.

(iii) To prove \( \delta_3 \) is a.o. -- Following the arguments used in (ii), it can be shown that
\[
\tilde{\gamma} \xrightarrow{a.s.} \gamma.
\]

In addition using Serfling [12, p. 70], it follows that
\[
s^2 \xrightarrow{a.s.} \text{Var}(X_B).
\]

Then using Slutsky's Theorem twice,
\[
\delta_3 \xrightarrow{p} \delta_B(x).
\]
Since \( \delta_8(x) \) is finite and since it can be shown that (3.1) holds, it follows that \( \delta_3 \) is asymptotically optimal.

The estimators \( \delta_4 \) and \( \delta_5 \) result from iterative solutions. Hence, it is not possible to verify their asymptotic optimality analytically. However, using Monte Carlo studies, Maritz [5, p. 65] has shown that comparable methods for the Poisson and Gamma distributions do lead to highly efficient estimators. Hence, it is hypothesised that our estimators \( \delta_4 \) and \( \delta_5 \) are indeed asymptotically optimal. The verification of this conjecture is left to those interested.

3.2 Asymptotic normality.

In this section, it is shown that the estimators \( \delta_1, \delta_2 \) and \( \delta_3 \) are asymptotically normally distributed. To prove this result on \( \delta_1 \), it is necessary for the frequencies to assume a particular representation in terms of a sum of independent and identically distributed indicator-vectors. In order to gain access to \( g_{N,n}(x + 1) \) and \( g_{N-1,n}(x) \), the following three methods of recording the results of the consecutive binomial experiments in the past can be visualised.

(i) There are two independent sequences of experiments, one with \( N \) trials and one with \( N-1 \) trials;

(ii) There is one sequence only but the results for the first \( N-1 \) trials have been recorded separately from the result of the \( N^{\text{th}} \) trial; or

(iii) There is just one sequence of \( N \) trials with corresponding records.

It is clear that (i) is ideal but from an estimation point of view, it is unrealistic. The case (ii) is practical and (iii) is the minimum information required for an empirical Bayes procedure. The following theorems can now be proved.

**Theorem 3.3**

The simple empirical Bayes estimator \( \delta_1 \) has an asymptotically normal distribution. When the experimental results are recorded as in (ii) above,
\[ E(\delta_1) = (x + 1) \frac{\alpha_2}{(N\alpha_1)}, \]  

and

\[ \text{Var}(\delta_1) = (x+1)^2 \frac{\alpha_2}{N} \left\{ \frac{1}{2} - \frac{2(x+1)}{N} \alpha_2 - \alpha_1 \right\}/(N^2 \alpha_1^3) \]  

where

\[ \alpha_1 = p_{G,N-1}(x) \quad \text{and} \quad \alpha_2 = p_{G,N}(x) \]

and when the experimental results are recorded as in (iii) above,

\[ E(\delta_1) = (x + 1)\gamma_2/(N(1 + \gamma_1)) \]  

and

\[ \text{Var}(\delta_1) = (x+1)^2\gamma_2 \left\{ \gamma_1 \gamma_2 (3 - \gamma_1) + (1 + \gamma_1)^2 (1 - \gamma_2) \right\}/(n+1)N^2(1+\gamma_1)^4 \]  

where

\[ \gamma_1 = p_{G,N}(x) \quad \text{and} \quad \gamma_2 = p_{G,N}(x + 1). \]

**Proof.**

Let \( 0 \leq x \leq N-1 \) be a fixed integer. Define

\[ I_{G,1}(x) = I_1(x) = \begin{bmatrix} I_{1,1}(X_{G,N-1} = x) \\ I_{1,2}(X_{G,N} = x+1) \end{bmatrix} \equiv \begin{bmatrix} I_{1,1} \\ I_{1,2} \end{bmatrix}, \; i = 1, \ldots, n \]

where

\[ I_{1,1} = \begin{cases} 1, \text{if } i^\text{th} \text{ experiment with } N-1 \text{ trials ends in } x, \\ 0, \text{otherwise,} \end{cases} \]
and

\[ I_{i,2} = \begin{cases} 
1, & \text{if } i^{th} \text{ experiment with } N \text{ trials ends in } x+1, \\
0, & \text{otherwise}. 
\end{cases} \]

Hence, \( I_1(x), i = 1, \ldots, n, \) are independent and identically distributed with

\[ E(I_1(x)) = g = (\alpha_1, \alpha_2)^T \]

and

\[ E(I_1(x) - g)(I_1(x) - g)^T = \sum \begin{bmatrix} \alpha_1(1-\alpha_1) & \alpha_2 \left( \frac{x+1}{N} - \alpha_1 \right) \\ \alpha_2 \left( \frac{x+1}{N} - \alpha_1 \right) & \alpha_2(1-\alpha_2) \end{bmatrix} \]

Hence, the observed relative frequencies of \( x \) and \( (x+1) \) are elements of

\[ n^{-1} \mathbf{g}_n = n^{-1} \begin{bmatrix} g_{N-1,n}(x) \\ g_{N,n}(x+1) \end{bmatrix} = n^{-1} \sum_{i=1}^{n} I_1(x) \]  \hspace{1cm} (3.6)

From the multivariate central limit theorem, \( n^{-1} \mathbf{g}_n \) is asymptotically multivariate normal in distribution. Since \( \delta_1 \) is continuous function of the elements of \( n^{-1} \mathbf{g}_n \) with non-zero differential at \( E(n^{-1} \mathbf{g}_n) \), \( \delta_1 \) is asymptotically normal. It follows from (3.6) that the mean and variance are as given in (3.2) and (3.3), respectively.

When the data is recorded according to (iii), the \( p_{G,N}(x+1) \) and \( p_{G,N-1}(x) \) are estimated from the same sequence of observed frequencies. To reduce the amount of error introduced through this approximation, consider the relation

\[ p_{G,N-1}(x) = w(x) p_{G,N}(x) \]

where

\[ w(x) = \frac{(N-x) \int \lambda^x (1-\lambda)^{N-x-1} dG(\lambda)}{N \int \lambda^x (1-\lambda)^{N-x} dG(\lambda)} \]
The frequencies based on \( N \) trials can then be adjusted to, say,

\[ g_{N-1,n}(x) = w(x) g_{N,n}(x). \]

A difficulty with this is the unknown nature of \( G(\lambda) \). Therefore, the crude approximation \( w(x) = 1 \) under the assumption that \( N \) is reasonably large, is used.

Define

\[ J_i(x) = (J_{X_i;N}(x), J_{X_i;N}(x + 1))^\prime, \quad i = 1, \ldots, n. \]

The vectors \( J_i(x), i = 1, \ldots, n, \) are independent and identically distributed with

\[ E(J_i(x)) = \chi = (\gamma_1, \gamma_2)^\prime, \]

and

\[ E(J_i(x) - \chi) (J_i(x) - \chi)^\prime = \Sigma = \begin{pmatrix} \gamma_1(1-\gamma_1) & -\gamma_1\gamma_2 \\ -\gamma_1\gamma_2 & \gamma_2(1-\gamma_2) \end{pmatrix}. \]

Hence, the observed relative frequencies of \( x \) and \( (x + 1) \) are elements of

\[ n^{-1} g_n = n^{-1} \sum_{i=1}^{n} J_i(x). \]

Using the same arguments as for the case (ii) above, it can be shown that \( \delta_1 \) is asymptotically normally distributed with mean and variance as given in (3.4) and (3.5), respectively.

\[ \square \]

**Theorem 3.4**

The smoothed empirical Bayes estimator \( \delta_2 \) is asymptotically normally distributed with

\[ E(\delta_2) = (x + r)/(N + r + s) \]
\[
\text{Var}(\delta_2) = \left\{ \nu_1 (1-\nu_1) + \nu_2 \right\} \left\{ 2(N+x)\nu_1 (\nu_1 + \nu_2) - (Nx+\nu_1 - \nu_2) \right\}^2
\]

\[
+ \nu_1 (\nu_1 - N)(\nu_1 - x)\{ \nu_2 (7 - 2\nu_1 - \nu_2) + 6\nu_3 + \nu_4 - \nu_1 \} \right\} \right/ \{ nN^2 (N-1)^2 \nu_2 \} \quad (3.8)
\]

where

\[
\nu_k = N! \beta(r+k,s)/(N-k)! \beta(r,s), \quad k = 1, \ldots, 4.
\]

**Proof.**

Let

\[
\mathbf{Y}_i = (\mathbf{x}_{\beta i}, \mathbf{x}_{\beta i}^2)^\top, \quad i = 1, \ldots, n.
\]

Then, the \(\mathbf{Y}_i, i = 1, \ldots, n,\) are independent and identically distributed with

\[
E(\mathbf{Y}_1) = \mu = (\mu_1, \mu_2)^\top
\]

and

\[
E(\mathbf{Y}_1 - \mu)(\mathbf{Y}_1 - \mu)^\top = \mathbf{I}.
\]

The elements of \(\mathbf{I}\) are found by using the first four factorial moments of \(\mathbf{x}_\beta;\)

given by \(\nu_k, k = 1, \ldots, 4,\) respectively. It is readily shown that

\[
\mathbf{I} = \begin{pmatrix}
\nu_2 + \nu_1 (1-\nu_1) & \nu_3 + (3-\nu_1)\nu_2 + (1-\nu_1)\nu_1 \\
0 & \nu_4 + 6\nu_3 + \nu_2 (7-\nu_2 - 2\nu_1) + \nu_1 (1-\nu_1)
\end{pmatrix}.
\]

Then, by the multivariate central limit theorem,

\[
\overline{\mathbf{Y}} = n^{-1} \sum_{i=1}^{n} \mathbf{Y}_i
\]

is multivariate normally distributed with mean \(\overline{\mu}\) and variance \((n^{-1} \mathbf{I})\). Since
\(\delta_2\) is a continuous function of \(\overline{X}\) with non-zero differential at \(\mu\), it follows that \(\delta_2\) is asymptotically normal with mean and variance as given in (3.7) and (3.8), respectively.

\[\]

**Theorem 3.5**

The smoothed empirical Bayes estimator \(\delta_3\) is asymptotically normally distributed with

\[
E(\delta_3) = (\gamma + \theta x)/(1 + \theta)
\]

(3.9)

and

\[
\text{Var}(\delta_3) = n^{-1} A' \overline{X} \overline{A}
\]

(3.10)

where

\[
A' = \begin{bmatrix}
\frac{2\nu_1(N+\mu) - 3\nu_1^2 - Nx - v}{(N-1)v}, & \frac{Nxv_1 + \nu_1^3 - (N+\mu)v_1^2}{N(N-1)v^2}
\end{bmatrix}
\]

and

\[
\overline{X} = \begin{bmatrix}
\nu_2 + \nu_3(1-\nu_1)/N^2, & n[\nu_3 + (3\nu_2 + \nu_1(1-2\nu_1))(1-\nu_1)]/(N(n-1))

0, & \nu_4 + 2\nu_3(3-\nu_1) + \nu_2(7 - \nu_2 - 2\nu_1(4-\nu_1)) + \nu_1(1-\nu_1)(1-2\nu_1)
\end{bmatrix}
\]

with \(\nu_i, i = 1, \ldots, 4\), defined as for Theorem 3.4 and with \(v = \text{Var}(X_\theta)\).

**Proof.**

Let \(X_i, i = 1, \ldots, n\) and \(\overline{X}\) be defined as for Theorem 3.4 and define

\[
Z = (m_1/N, n(m_2^2 - (m_1^2)/(n-1))^\prime \equiv (\overline{X}, S^2)\prime.
\]

Then, using Serfling [12, p. 129] and an argument similar to that used in the proof of Theorem 3.4, it can be shown that \(Z\) is asymptotically multivariate normally distributed.
distributed with mean \((v_1/N, v)\) and variance \((n^{-1} y)\). Since \(\delta_3\) is a continuous function of \(Z\), it follows that \(\delta_3\) is asymptotically normally distributed with mean and variance as given by (3.9) and (3.10) respectively.

Finally, although no attempt is made here to establish any distributional properties of the estimators \(\delta_4\) and \(\delta_5\), it seems reasonable to conjecture that \(\delta_4\) at least is asymptotically normal since it is a function of maximum likelihood estimators.

4. A NUMERICAL STUDY

In order to gain insights into the comparative merits of the proposed estimators, a numerical analysis was undertaken. In this study, binomial data were generated for \(N = 5\) with the parameter \(\lambda\) forming realizations from a beta distribution with \(r = 2\) and \(s = 4\). Here, \(n = 10\). Table 4.1 presents the average value obtained for the estimate of \(\lambda\) arising from 50 repetitions. Also shown is the average risk for each estimator. The estimators \(\delta_i, i = 0, \ldots, 4\) are found using their formulae as presented in Section 2. The true value of the unknown parameter \(\lambda\) is in fact \(0.3\). If, in the use of the estimator \(\delta_1\), the data were recorded according to (iii) (see Section 3), then setting \(w(x) = 1\) gives the estimator

\[
\delta_1^* = \frac{(x + 1)}{N} \frac{g_{N,n}(x + 1)}{\{1 + g_{N,n}(x)\}}.
\]

Comparison with the results for \(\delta_1\) clearly shows that this crude approximation does not provide results "as good as" those for \(\delta_1\) which should be expected.

Comparing all estimators, it is interesting to note that the maximum likelihood estimator \(\delta_0\) is closest to the real value of \(0.3\). However, \(\delta_4\) obtained by using maximum likelihood estimates for \(r\) and \(s\) is also very close. Note,
too, that the average risk for the estimators (excluding $\delta_1^*$) is less than that for $\delta_0$, lending further justification to the use of empirical Bayes procedures. However, with the estimator $\delta_4$ having the smallest average risk coupled with its "close" estimate, it suggests that the combination of the concepts of the classical maximum likelihood approach and the empirical Bayes approach could provide the best estimation procedure.

Table 4.1

Results of a numerical study to estimate $\lambda$ in a Binomial $B(\lambda, N)$ distribution.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$\delta_0$</th>
<th>$\delta_1$</th>
<th>$\delta_1^*$</th>
<th>$\delta_2$</th>
<th>$\delta_3$</th>
<th>$\delta_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average value for $\lambda$</td>
<td>.3040</td>
<td>.3562</td>
<td>.3906</td>
<td>.3280</td>
<td>.3212</td>
<td>.3056</td>
</tr>
<tr>
<td>Average risk</td>
<td>.0644</td>
<td>.0554</td>
<td>.0713</td>
<td>.0387</td>
<td>.0374</td>
<td>.0083</td>
</tr>
</tbody>
</table>
REFERENCES


A study of empirical Bayes estimation procedures for the parameter of a binomial distribution is undertaken. Various procedures are derived. Properties of the resulting estimators are investigated. In particular, it is shown that some of these estimators are asymptotically optimal and are asymptotically normally distributed.