Ann. Inst. Statist. Math.
Vol. 44, No. 4, 641–657 (1992)

ASSESSING THE PERFORMANCE OF EMPIRICAL BAYES ESTIMATORS

J. S. MARITZ[1]* AND T. LWIN[2]

[1]*Department of Statistics, La Trobe University, Bundoora, Victoria 3083, Australia*
[2]*Division of Mathematics and Statistics, CSIRO, Private Bag 10, Clayton, Victoria 3168, Australia*

(Received June 17, 1991; revised January 27, 1992)

Abstract. Methods for deriving empirical Bayes estimators are generally available. Corresponding general techniques for assessing the performance of these estimators are not widely developed yet, however. In this paper we provide a general procedure for assessing and comparing the performance of the empirical Bayes estimators and other estimators in a given data set.

Key words and phrases: Unequal component sample sizes, empirical Bayes, linear EB, assessment of performance, comparison with MLE.

1. Introduction

The empirical Bayes (EB) sampling scheme covering a fairly general class of situations can be described as follows. Let n independent past realizations of the random variable (r.v.) Θ be $\theta_1, \theta_2, \dots, \theta_n$. Each of these θ_i, $i = 1, 2, \dots, n$ acts as a parameter in the data distribution $F(x \mid \theta_i)$ of a random variable X on which a vector of m_i observations $\boldsymbol{x}_i = (x_{i1}, \dots, x_{im_i})$ is made. These observations are independent. The observed sequence of past observations $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$ is then used to construct empirical Bayes estimators (EBE's) of the current θ based on the current observation $\boldsymbol{x} = (x_1, \dots, x_m)$ from $F(x \mid \theta)$. The EB scheme which has received most attention is the special case when $m_i = m = 1$ ($i = 1, \dots, n$). Reduction of the general scheme is possible if all $m_i = m$ and if the observations at each stage can be summarized in a sufficient statistic.

In practice one can expect sampling schemes to be much less tidy. Certainly one may expect m_i to vary, and reduction by sufficiency may not be possible. Although particular cases of unequal m_i have been studied in some detail there has been surprisingly little systematic discussion of the unequal m_i scheme. Three matters clearly seem to need attention. The first two are, just how to construct

* Now at Department of Statistics, University of Stellenbosch, Stellenbosch 7600, Republic of South Africa.

EBE's and what to do about lack of sufficiency. The third is the somewhat neglected question of judging the goodness of EB estimation in a particular case, relative to, say, maximum likelihood (ML) estimation. For details of general EB developments and a discussion of unequal sample sizes see Maritz and Lwin (1989).

Attention in this paper will be confined to the case of estimating a single parameter θ. Extension to a parameter of dimension $d > 1$ is possible, but will not be considered here. We shall also not consider simple EB estimation, i.e. EB estimation in which prior distribution $G(\theta)$ of Θ is not estimated directly. There are published studies of the unequal m_i, simple EB problem, see for example O'Bryan (1976), O'Bryan and Susarla (1977). However it does appear that actual construction of simple EBE's will generally involve selection of subsets of observations at each stage, calculating the simple EBE for each subset and possibly averaging the individual EBE's over all such selections. Also, if the current m is greater than all past m_i, construction of a simple estimate of the actual Bayes estimate is impossible. Published accounts of the performance of simple EBE's indicate that EBE's based on direct estimation of G are better for finite n.

The performance of an estimate δ will be judged by the expected squared error $W(\delta)$. The Bayes point estimate δ_G minimizes $W(\delta)$ and is the mean of the posterior distribution. We shall from time to time use notations like W(Bayes), W(ML), W(EB) with obvious meanings. Well established practice is to judge the performance of an EB estimate by the expected value of W(EB) over n past realizations of Θ, $E_n W$(EB).

2. Construction of EB estimators

Construction of EBE's for the current parameter θ has been discussed for a general vector parameter case (Lwin and Maritz (1989)). In this section, we reproduce a summary of the results as specialized to the uniparameter case. General procedures for assessing the performance of EBE's constructed here are discussed in Sections 4 to 6 in detail.

2.1 *EBE with parametric priors of known form*

Let the prior distribution function of Θ be a continuous function $G(\theta; \boldsymbol{\alpha})$ with a known functional form depending on a vector of unknown parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)$. For notational convenience we shall express the likelihoods and other expressions in the following in terms of probability density function (p.d.f.), $f(x \mid \theta)$, of data distribution $F(x \mid \theta)$. Modifications for a discrete distribution are obvious.

The Bayes estimator of θ, based on the current observation vector $\boldsymbol{x}$ is

$$\hat{\theta}^*(\boldsymbol{x}; G, m) = \hat{\theta}^*(\boldsymbol{x}; \boldsymbol{\alpha}, m) = \int \theta dB(\theta \mid \boldsymbol{x}; G, m) \tag{2.1}$$

where

$$dB(\theta \mid \boldsymbol{x}; G, m) = p(\boldsymbol{x}; \theta, m) dG(\theta; \boldsymbol{\alpha}) / q(\boldsymbol{x}; \boldsymbol{\alpha}, m)$$

is the posterior d.f. of Θ and

$$p(\boldsymbol{x};\theta,m)=\prod_{j=1}^{m} f(x_j \mid \theta) \tag{2.2}$$

and

$$q(\boldsymbol{x};\boldsymbol{\alpha},m)=\int p(\boldsymbol{x};\theta,m)dG(\theta;\boldsymbol{\alpha}). \tag{2.3}$$

The unknown parameter vector $\boldsymbol{\alpha}$ is to be estimated from the previous data $\boldsymbol{x}_1,\dots,\boldsymbol{x}_n$ of the EB scheme.

The EBE of θ is straight forward and is given by $\hat{\theta}^*(\boldsymbol{x};\hat{\boldsymbol{\alpha}},m)$ where $\hat{\boldsymbol{\alpha}}$ is a "good" estimate of $\boldsymbol{\alpha}$ based on the previous data. Thus the crucial problem here is that of estimating $\boldsymbol{\alpha}$ using all the information available in the EB scheme. We consider the maximum likelihood (ML) estimation of $\boldsymbol{\alpha}$.

The likelihood function of $\boldsymbol{\alpha}$ based on the past observation vectors $\boldsymbol{x}_1,\dots,\boldsymbol{x}_n$ is

$$L(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n;\boldsymbol{\alpha})=\prod_{i=1}^{n} q(\boldsymbol{x}_i;\boldsymbol{\alpha},m_i) \tag{2.4}$$

where $q(\boldsymbol{x}_i;\alpha,m_i)$ is as defined in (2.3) with $\boldsymbol{x}_i$ and m_i in place of $\boldsymbol{x}$ and m. In fact $q(\boldsymbol{x}_i;\boldsymbol{\alpha},m_i)$ is the marginal p.d.f. of a r.v. $\boldsymbol{X}_i$ whose realization is $\boldsymbol{x}_i$ $(i=1,\dots,n)$.

The log-likelihood function $\ln L$ can be directly maximized. Alternatively, we can apply an EM algorithm (Dempster *et al.* (1977)), to obtain an ML estimate $\hat{\boldsymbol{\alpha}}_n$ of $\boldsymbol{\alpha}$. In either case one is concerned with the stationary points of the likelihood equations

$$(\partial/\partial\alpha_j)\ln L=S_j(\boldsymbol{x},\boldsymbol{\alpha})=0; \qquad j=1,\dots,k \tag{2.5}$$

where

$$S_j(\boldsymbol{x},\boldsymbol{\alpha})=\sum_{i=1}^{n} U_{ji}(\boldsymbol{\alpha})$$

with

$$U_{ji}(\boldsymbol{\alpha})=\partial \ln q(\boldsymbol{x}_i;\boldsymbol{\alpha},m_i)/\partial\alpha_j.$$

Under certain regularity conditions (see also Lwin and Maritz (1989)), the likelihood equations can be recast as

$$S_j(\boldsymbol{x},\boldsymbol{\alpha})=\sum_{i=1}^{n} E\left\{\frac{\partial \ln g(\Theta \mid \alpha)}{\partial\alpha_j}\,\middle|\, \boldsymbol{X}=\boldsymbol{x}_i\right\}=0 \qquad j=1,\dots,k \tag{2.6}$$

where the expectation is with respect to (w.r.t.) the posterior d.f. $B(\boldsymbol{x},\boldsymbol{\alpha},m)$. This may be solved directly if an explicit solution can be obtained. In general an iterative solution can be constructed as follows. Let S be a $k\times 1$ vector whose j-th element is

$$S_j(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n;\boldsymbol{\alpha}^{(i)})=\sum_{u=1}^{n} E\left[\left\{\frac{\partial \ln g(\Theta \mid \boldsymbol{\alpha})}{\partial\alpha_j}\right\}_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^{(i)}}\middle|\boldsymbol{X}=\boldsymbol{x}_u,\boldsymbol{\alpha}^{(i)}\right] \tag{2.7}$$

and J be a $k \times k$ matrix whose (j, t)-th element is

$$J_{jt}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n; \boldsymbol{\alpha}^{(i)}) = \sum_{u=1}^{n} E\left[\left\{\frac{\partial^2 \ln g(\Theta \mid \boldsymbol{\alpha})}{\partial \alpha_j \partial \alpha_t}\right\}_{\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}^{(i)}} \middle| \boldsymbol{X} = \boldsymbol{x}_u, \boldsymbol{\alpha}^{(i)}\right]. \tag{2.8}$$

In the above expressions $\boldsymbol{\alpha}^{(i)}$ is the i-th iterative step estimate of $\boldsymbol{\alpha}$ and $\alpha^{(1)}$ is taken as an initial estimate $\boldsymbol{\alpha}^{+}$. The estimates are updated by the equation:

$$\boldsymbol{\alpha}^{(i+1)} = \boldsymbol{\alpha}^{(i)} - J^{-1}S \quad i = 1, 2, \ldots. \tag{2.9}$$

This can be shown to be the result of the EM algorithm. When the iterative process converges, the resulting quantity $\hat{\boldsymbol{\alpha}}_n$ is taken to be the ML estimator of $\boldsymbol{\alpha}$.

An approximate covariance matrix of $\hat{\boldsymbol{\alpha}}_n$ is needed in evaluating the performance of the EBE

$$\hat{\theta}^*(\boldsymbol{x}; \hat{\boldsymbol{\alpha}}_n, m) \tag{2.10}$$

based on ML estimate $\hat{\boldsymbol{\alpha}}_n$. This will be discussed in more detail in Section 3. For now we only note that if an alternative route of using observed information matrix $\hat{I}_x$ is to be used, one can follow the lines of Louis (1982) to extract $\hat{I}_x$ as

$$\hat{I}_x = J - SS^T \tag{2.11}$$

where J and S are evaluated at $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}_n$. The use of (2.11) in an assessment of EBE is also discussed in Section 3.

An attractive feature of (2.11) is that if the EM iterative process (2.9) can be readily applied the quantities J and S are obtained as a by-product of the iterative process, when convergence is achieved. No extra computation was necessary. Furthermore the inverse of (2.11) can be used to estimate the covariance matrix C of $\hat{\boldsymbol{\alpha}}_n$. Using (2.6) an approximation to C is obtained as the inverse of V whose (i, j)-th element is $V_{ij} \simeq \operatorname{cov}(S_i(\boldsymbol{x}, \boldsymbol{\alpha}), S_j(\boldsymbol{x}, \alpha))$. This is a special case of the general result discussed in Section 3.

If $T_i = T_i(x_{i1}, \ldots, x_{im_i})$, an estimate of θ_i, is based on a sufficient statistic, the likelihood, apart from a multiplicative factor not involving $\boldsymbol{\alpha}$, becomes

$$L(\boldsymbol{t}; \boldsymbol{\alpha}) = \prod_{i=1}^{n} \int h(t_i \mid \theta_i, m_i) dG(\theta_i; \alpha) \tag{2.12}$$

where $h(t_i \mid \theta_i, m_i)$ is the density function of T_i. The above analysis can then be carried out equivalently by using $L(\boldsymbol{t}; \boldsymbol{\alpha})$ in place of $L(\boldsymbol{x}; \boldsymbol{\alpha})$.

If T_i is not a sufficient statistic, then a conventional estimator, say an MLE, $\hat{\theta}_i$, may be used to reduce the i-th sample. Such a reduction results in what we shall call pseudo-Bayes estimation. The use of an MLE $\hat{\theta}_i$ can be justified by the asymptotic sufficiency of MLE's (see Cox and Hinkley ((1974), p. 36)). In any case the EB procedure described above remains the same with $h(t_i \mid \theta_i, m_i)$ in place of $p(\boldsymbol{x}_i \mid \theta_i, m_i)$.

Estimation of $\boldsymbol{\alpha}$ need not be by the ML method. Sometimes other methods like the method of moments may be much more convenient. Thus it may be possible to write down tractable expressions, in terms of $\boldsymbol{\alpha}$ for the moments $E(T_{iG}^p)$, $p = 1, 2, \ldots, k$ of the distribution $h_G(t_i; m_i, \boldsymbol{\alpha})$. Then setting the observed moments, $(1/n)\sum t_i^p$, equal to the theoretical moments estimating equations

$$M_p(\boldsymbol{t}; \boldsymbol{\alpha}) = \sum_{i=1}^{n} W_{pi}(\boldsymbol{\alpha}) = 0, \qquad p = 1, 2, \ldots, k \tag{2.13}$$

are obtained whose solutions are the methods of moments (MM) estimates $\bar{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$. One can use (2.13) and the matrix V mentioned earlier to obtain approximate standard error of $\bar{\boldsymbol{\alpha}}$ satisfying (2.13). This is discussed in Section 3.

2.2 *EB estimation: Approximation of the prior distribution*

When the form of G is not known it is feasible to obtain a good approximation to the Bayes estimate through approximation of G by some means. One may use a member of some chosen family of distributions, such as the family of natural conjugate priors, where it exists. This would lead to an analysis with parametric priors as treated in Subsection 2.1 above. A more versatile approximation to G is a step function approximation of G as discussed in Maritz (1967).

Let G be approximated by a finite mixture G_k having jumps of size $\lambda_1, \ldots, \lambda_k$ $(0 \le \lambda_i \le 1, \sum_{i=1}^{k} \lambda_i = 1)$ at $\alpha_1, \alpha_2, \ldots, \alpha_k$. Then $L(\boldsymbol{x}, \boldsymbol{\alpha})$ can be approximated by

$$L_k(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \prod_{i=1}^{n} \left\{ \sum_{j=1}^{k} p(\boldsymbol{x}_i \mid \alpha_j, m_i)\lambda_i \right\}. \tag{2.14}$$

Two subapproaches are open. First, assume that $\lambda_i = 1/k$ and treat $\alpha_1, \ldots, \alpha_k$ as unknown parameters. One can then proceed with a likelihood procedure to obtain estimates of $\boldsymbol{\alpha}$. For identifiability one must impose a restriction $\alpha_1 \le \alpha_2 \le \cdots \le \alpha_k$ as pointed out in Maritz (1967). The estimating equations are the same as (2.9) with $L(\boldsymbol{x}, \boldsymbol{\alpha})$ replaced by $L_k(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\lambda})$.

Second, assume that the λ_i's are unknown, but the $\alpha_1, \alpha_2, \ldots, \alpha_k$ are known selected values. In this case too, one can proceed with a likelihood procedure using (2.14) with known or selected values of $\boldsymbol{\alpha}$. This is a typical problem of estimation of mixing proportions of a finite mixture of known components. Thus it can again be treated by an EM algorithm approach; see e.g. Dempster *et al.* (1977). In this case application of EM algorithm results in a set of iterative equations for λ_j's as follows:

$$\begin{aligned} &\hat{\lambda}_j^{(i+1)} = \sum_{u=1}^{n} \hat{\lambda}_j^{(i)}(\boldsymbol{x}_u)/n \qquad \text{with} \\ &\hat{\lambda}_j^{(i)}(\boldsymbol{x}_u) = \hat{\lambda}_j^{(i)} p(\boldsymbol{x}_u \mid \alpha_j, m_u) \Big/ \sum_{j=1}^{k} \hat{\lambda}_j^{(i)} p(\boldsymbol{x}_u \mid \alpha_j, m_u). \end{aligned} \tag{2.15}$$

In (2.15) $\hat{\lambda}_j^{(i)}$ is an estimate of λ_j at the i-th stage of iteration. At the initial stage one can take $\hat{\lambda}_j^{(0)} = 1/k$. As is well known (2.15) leads to the ML estimates of λ_j's based on the likelihood function (2.14). In the simpler case when sufficient statistics T_i $(i = 1, \ldots, n)$ exist, the above procedure is still valid if we replace $\prod_{u=1}^{m_i} f(x_{iu} \mid \alpha_j)$ by $h(t_i \mid \alpha_i, m_i)$; the same procedure serves as a pseudo-EB method when T_i is a summary statistic but does not possess the property of sufficiency.

Again following the lines of Louis (1982) the observed information matrix of $\hat{\boldsymbol{\lambda}}$ based on the likelihood $L_k(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\lambda})$ is obtained as $\hat{I}_x^{(k)}$, a $k \times k$ matrix whose (i, j)-th element is

$$\hat{I}_x^{(k)}(i, j) = -\sum_{u=1}^{n} \frac{p(\boldsymbol{x}_u \mid \alpha_i, m_u) p(\boldsymbol{x}_u \mid \alpha_j, m_u)}{r(\boldsymbol{x}_u)} \tag{2.16}$$

where

$$r(\boldsymbol{x}_u) = \left\{ \sum_{t=1}^{k} p(\boldsymbol{x}_u \mid \alpha_t, m_u) \lambda_t \right\}^2 .$$

As before the inverse of $\hat{I}_x^{(k)}$ can be used to estimate the covariance matrix C of estimates of the unknown parameters $\boldsymbol{\lambda}$.

2.3 *Linear EB estimation*

Suppose that T is an estimate of θ. In this section T need not necessarily be a sufficient statistic. Then a linear Bayes estimate of θ can be sought in the class:

$$\delta(T; \omega_0, \omega_1) = \omega_0 + \omega_1 T \tag{2.17}$$

as the one with ω_0, ω_1 chosen to minimize $W(\omega_0 + \omega_1 T)$. Except in special cases it is not the actual Bayes estimate, but in the Bayes sense, it is the optimal estimate in the class defined by (2.17). Estimates of this type have been discussed by many authors; e.g. Griffin and Krutchkoff (1971). In this section our concern is with the case when T is based on different sample sizes at different stages of the EB sampling scheme. We specialize to the case where we are concerned with the estimation of the expectation of T at the current stage.

Suppose that the distribution function $F(x \mid \theta)$ is parametrized such that

$$E(T \mid \theta) = \theta, \qquad E(T^2 \mid \theta) = \theta^2 + q(\theta)/m \tag{2.18}$$

at the current stage. Here $q(\theta)$ is a function taking positive values for all θ in the support of $F(x \mid \theta)$. When T_i is computed from $(X_{i1}, \ldots, X_{im_i})$ at the i-th stage, the same relationships (2.18) hold for T_i with m_i in place of m.

Now the choice of ω_0, ω_1 which minimizes $W(\omega_0 + \omega_1 T)$ at the current stage is given by

$$\begin{bmatrix} 1 & E(\Theta) \\ E(\Theta) & E(\Theta^2) + E\{q(\Theta)\}/m \end{bmatrix} \begin{bmatrix} \omega_0 \\ \omega_1 \end{bmatrix} = \begin{bmatrix} E(\Theta) \\ E(\Theta^2) \end{bmatrix} . \tag{2.19}$$

Let $\gamma_p = E(\Theta^p)$ and $\sigma_0^2 = \gamma_2 - \gamma_1$. Also let $\nu = E\{q(\Theta)\}$. Then

$$\begin{aligned} \omega_0 &= (\gamma_1 \nu/m)/(\sigma_0^2 + \nu/m), \\ \omega_1 &= 1 - (\nu/m)/(\sigma_0^2 + \nu/m). \end{aligned} \tag{2.20}$$

For constructing EB analogues of the linear Bayes estimator, we need to construct estimates of

$$\gamma_1 = E(\Theta), \qquad \sigma_0^2 = E(\Theta^2) - \{E(\Theta)\}^2, \qquad \nu = E\{q(\Theta)\} \tag{2.21}$$

based on the data $(T_1, \ldots, T_n)$.

An advantage of this approach is that estimation of the above expectations can be straightforward in certain important special cases. We demonstrate this by considering a special choice of $q(\theta)$ as

$$q(\theta) = k_0 + k_1\theta + k_2\theta^2. \tag{2.22}$$

This choice covers the special cases considered in more detail in the following sections.

Let

$$\begin{aligned} Y_i &= m_i(m_i + k_2)^{-1}(k_0 + k_1 T_i + k_2 T_i^2), \\ \bar{T} &= \sum_{i=1}^{k} T_i/n, \\ S_T^2 &= \sum_{i=1}^{n} (T_i - \bar{T})^2/(n-1). \end{aligned}$$

Then define the estimators for γ_1 and ν as

$$\begin{aligned} \hat{\gamma}_1 &= \bar{T}, \\ \hat{\nu} &= \sum_{i=1}^{n} Y_i/n. \end{aligned} \tag{2.23}$$

It can be readily shown that $\hat{\gamma}_1$ and $\hat{\nu}$ are unbiased for γ_1, ν. i.e.

$$\begin{aligned} E(\hat{\gamma}_1) &= \gamma_1, \\ E(\hat{\nu}) &= \nu. \end{aligned}$$

Further by noting that

$$E(S_T^2) = \sigma_0^2 + \nu n^{-1} \sum_{i=1}^{n} m_i^{-1}$$

we obtain an unbiased estimator of σ_0^2 as

$$\hat{\sigma}_0^2 = S_T^2 - \hat{\nu} n^{-1} \sum_{i=1}^{n} m_i^{-1}.$$

It should be noted however that $\hat{\nu}$ and $\hat{\sigma}_0^2$ could produce negative estimates for small n. However, we exclude this possibility by assuming that sufficient amount of previous data are available. Thus, under some conditions which allow Y_i to be defined, one can readily construct linear EB analogues of the linear Bayes estimator $\delta_G(T;\omega_0,\omega_1)$ by using estimates $\hat{\omega}_0$ and $\hat{\omega}_1$ based on $\hat{\gamma}_1$, $\hat{\nu}$ and $\hat{\sigma}_0^2$.

We acknowledge here that the method of construction of $\hat{\nu}$ and $\hat{\sigma}_0^2$ given above has been applied to the case where $F(x \mid \theta)$ is a binomial distribution with parameter (n,θ) by Southward and Van Ryzin (1972). Thus (2.23) gives an extension of their result to the case of data distribution with properties (2.18) and (2.22).

It is also possible to estimate directly $E(\Theta^p)$ for $p = 1,2$ using the moments $\sum_{i=1}^n T_i^p/n$ $(p = 1,2)$. In either case the procedure is the same as the method of moments described in (2.13) and hence produces the same estimates.

3. Assessing the performance of EB estimators

Parametric G case

The quantities to evaluate for assessing the performance of an EB estimator are $W(\mathrm{ML})$, $W(\mathrm{Bayes})$ and $E_nW(\mathrm{EB})$ mentioned in Section 1. In general each of these quantities is a function of the parameters $\boldsymbol{\alpha}$, and can, in principle, be estimated. The most troublesome is $E_nW(\mathrm{EB})$, because it requires estimation of the covariance matrix of the estimates of $\boldsymbol{\alpha}$. In the following we denote an estimate of $\boldsymbol{\alpha}$ based on a general method by $\breve{\boldsymbol{\alpha}}$ and in the case of a maximum likelihood estimate we use $\hat{\boldsymbol{\alpha}}$.

The general form of the estimating equations for $\boldsymbol{\alpha}$ is

$$S_p(\boldsymbol{x};\boldsymbol{\alpha}) = \sum_{i=1}^{n} U_{pi}(\boldsymbol{\alpha}) = 0, \qquad p = 1,2,\ldots,k$$

and a large n approximate formula for the covariance matrix C of the estimates is

$$\boldsymbol{C} = \boldsymbol{D}^{-1}\boldsymbol{V}(\boldsymbol{D}^T)^{-1} \tag{3.1}$$

where the elements of

$$\begin{aligned} &\boldsymbol{V} \text{ are } \operatorname{Cov}\{S_p(\boldsymbol{x};\boldsymbol{\alpha}), S_q(\boldsymbol{x};\boldsymbol{\alpha})\}, \\ &\boldsymbol{D} \text{ are } \left\{\frac{\partial E S_p(\boldsymbol{X};\boldsymbol{a})}{\partial a_q}\right\}_{\boldsymbol{a}=\boldsymbol{\alpha}} \end{aligned} \tag{3.2}$$

for $p,q = 1,2,\ldots,k$. See for example, Maritz ((1981), p. 12).

The general form (3.1) is useful in many contexts, for two reasons. The first, and more obvious one is that the reduction applying to the ML case does not always hold. An example of this sort occurs when an approximation to the true prior distribution is used. Other examples arose in alternative EB approaches, like linear EB estimators.

The other reason is that it is eminently realistic in some situations to suppose that the m_i vary randomly. In this case (3.1) is still applicable and we can estimate the elements of $\boldsymbol{C}$ by

$$\hat{C}_{pq} = \frac{1}{(n-1)} \sum_{i=1}^{n} \{U_{pi}(\breve{\boldsymbol{\alpha}}) - U_{p\cdot}(\breve{\boldsymbol{\alpha}})\}\{U_{qi}(\breve{\boldsymbol{\alpha}}) - U_{q\cdot}(\breve{\boldsymbol{\alpha}})\}, \quad p,q = 1,2,\ldots,k. \tag{3.3}$$

In the case of maximum likelihood estimation straightforward calculation shows that $\boldsymbol{D} = \boldsymbol{V}$ so that (3.1) reduces to the usual formula for the large sample covariance matrix of MLE's; $\boldsymbol{D}$ is the information matrix. The actual calculation of estimates of $\boldsymbol{C}$ can proceed either by finding expressions for $\boldsymbol{D}$ in terms of $\boldsymbol{\alpha}$ and then substituting the MLE $\hat{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}$, or simply by using the observed information matrix whose elements are

$$\left\{ \sum_{i=1}^{n} \partial U_{pi}(\boldsymbol{a})/\partial a_q \right\}_{\boldsymbol{a}=\hat{\boldsymbol{\alpha}}}, \qquad p,q = 1,2,\ldots,k, \tag{3.4}$$

see for example, Cox and Hinkley ((1974), p. 302). In fact (3.4) reduces to (2.11) when $\boldsymbol{\alpha}$ is estimated by ML approach, i.e. $\boldsymbol{C}$ can be estimated by the inverse of the observed information matrix in (2.11): $\hat{C} = \hat{I}_x^{-1}(\boldsymbol{\alpha})$.

Finite approximation to G

Next we consider the finite approximation to G as discussed in Subsection 2.3 where G is approximated by a finite step function G_k with jumps of size $\lambda, \ldots, \lambda_k$ at known steps $\alpha_1, \ldots, \alpha_k$. Here we have the observed information matrix for $\boldsymbol{\lambda}$ as given in (2.16), $\hat{I}_x^{(k)}$ whose inverse can be used to estimate the covariance matrix C of the estimate $\hat{\boldsymbol{\lambda}}$.

Applications

Details of the use of the estimated $\boldsymbol{C}$ in estimating $E_n W(\text{EB})$ vary with the forms of data distribution and prior distribution, and are illustrated in the following sections dealing with particular distributions.

It is often easier to evaluate the difference

$$E_n W(\text{EB}) - W(\text{Bayes}) = E_n E_Q\{\hat{\theta}^*(\boldsymbol{x}, \hat{\boldsymbol{\alpha}}, m) - \hat{\theta}^*(x, \boldsymbol{\alpha}, m)\}^2 \tag{3.5}$$

where E_Q is the expectation with respect to marginal distribution of $\boldsymbol{X}$, $Q(\boldsymbol{x}, G, m)$ whose p.d.f. is $q(\boldsymbol{x}; G, m)$ in (2.3).

4. The Poisson distribution

If the distribution of X for given θ is Poisson with mean θ, the statistic $T = (X_1 + \cdots + X_m)/m$ is sufficient for θ. The probability distribution of $Z = mT$ conditional on θ is Poisson with mean $(m\theta)$ and the moments of T are given by

$$E(T \mid \theta) = \theta, \qquad E(T^2 \mid \theta) = \theta^2 + \theta/m. \tag{4.1}$$

(i) *Parametric G case*

If the prior distribution G is assumed to have the Gamma (α_1, α_2) then the p.d.f. is

$$g(\theta \mid \boldsymbol{\alpha}) = \{\alpha_1^{\alpha_2}/\Gamma(\alpha_2)\}\theta^{\alpha_2 - 1}\exp(-\alpha_1\theta); \qquad \theta > 0,$$

and the posterior distribution of Θ given $T = t$ is also a Gamma$(\alpha_1 + m, \alpha_2 + mT)$. From (2.6),

$$\begin{aligned} U_{1i}(\boldsymbol{\alpha}) &= \alpha_2/\alpha_1 - E(\Theta \mid T = t_i) = \alpha_2/\alpha_1 - (\alpha_2 + m_i t_i)/(\alpha_1 + m_i), \\ U_{2i}(\boldsymbol{\alpha}) &= \ln\alpha_1 - \psi(\alpha_2) + E(\ln\Theta \mid T = t_i) \\ &= \ln\alpha_1 - \ln(\alpha_1 + m_i) - \psi(\alpha_2) + \psi(\alpha_2 + m_i t_i) \end{aligned} \tag{4.2}$$

which provide elements of S in (2.6). Also the matrix J^{-1} can be evaluated as

$$J^{-1} = \frac{-\alpha_1^2}{n\{\alpha_2\psi'(\alpha_2) - 1\}}\begin{bmatrix} \psi'(\alpha_2) & 1/\alpha_1 \\ 1/\alpha_1 & \alpha_2/\alpha_1^2 \end{bmatrix}.$$

Hence the iterative equations can be readily constructed. The observed information matrix can be estimated by

$$\hat{I}_t = n\begin{bmatrix} -\hat{\alpha}_2/\hat{\alpha}_1^2 & 1/\hat{\alpha}_1 \\ 1/\hat{\alpha}_1 & -\psi'(\hat{\alpha}_2) \end{bmatrix} - \begin{bmatrix} \sum_{i=1}^{n} U_{1i}^2(\hat{\boldsymbol{\alpha}}) & \sum_{i=1}^{n} U_{1i}(\hat{\boldsymbol{\alpha}})U_{2i}(\hat{\boldsymbol{\alpha}}) \\ \sum_{i=1}^{n} U_{1i}(\hat{\boldsymbol{\alpha}})U_{2i}(\hat{\boldsymbol{\alpha}}) & \sum_{i=1}^{n} U_{2i}^2(\hat{\boldsymbol{\alpha}}) \end{bmatrix}. \tag{4.3}$$

In (4.2), $\psi(\cdot)$ is the digamma function.

It should also be noted that the expressions for $U_{1i}(\boldsymbol{\alpha})$ and $U_{2i}(\boldsymbol{\alpha})$ and the I_t can be directly obtained from the marginal probability distribution of $Z = mT$ whose explicit expression is,

$$P_G(z) = m^z\alpha_1^{\alpha_2}[\Gamma(\alpha_2 + z)/\{\Gamma(\alpha_2)z!\}](\alpha_1 + m)^{\alpha_2 + z}.$$

Alternative method of estimating $\boldsymbol{\alpha}$ by MM technique may also be used by exploiting the moment expressions of a Gamma(α_1, α_2) distribution for θ. The resulting estimate $\tilde{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$ is obtained as the solutions of

$$\sum_{i=1}^{n} W_{ri}(\boldsymbol{\alpha}) = 0; \qquad r = 1, 2$$

where

$$\begin{aligned} W_{1i}(\boldsymbol{\alpha}) &= t_i - \alpha_2/\alpha_1, \\ W_{2i}(\boldsymbol{\alpha}) &= t_i^2 - \alpha_2(\alpha_2 + 1)/\alpha_1^2 - \alpha_2/(\alpha_1 m_i). \end{aligned}$$

The matrices D and V in (3.1) are now given by:

(4.4)
$$D = \begin{bmatrix} \dfrac{n\alpha_2}{\alpha_1^2} & -\dfrac{n}{\alpha_1} \\ \dfrac{2n\alpha_2(\alpha_2+1)}{\alpha_1^3} + \dfrac{\alpha_2}{\alpha_1^2}\sum m_i^{-1} & -n\dfrac{(2\alpha_2+1)}{\alpha_1^2} - \dfrac{1}{\alpha_1}\sum m_i^{-1} \end{bmatrix},$$
$$V = \begin{bmatrix} \sum_{i=1}^{n}(\mu'_{2i} - \mu'^2_{1i}) & \sum_{i=1}^{n}(\mu'_{3i} - \mu'_{2i}\mu'_{1i}) \\ \sum_{i=1}^{n}(\mu'_{3i} - \mu'_{2i}\mu'_{1i}) & \sum_{i=1}^{n}(\mu'_{4i} - \mu'^2_{2i}) \end{bmatrix} \quad \text{where} \quad \mu'_{ri} = E(T_i^r)$$

the expectation being taken with respect to the marginal p.d.f. of T_i.

For a given set of past data we can now obtain an EBE of the current θ by substituting $\hat{\boldsymbol{\alpha}}$ or $\bar{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}$ in the formula

$$\delta_G(t) = (\alpha_2 + mt)/(\alpha_1 + m)$$

for the Bayes estimate. Similarly, W(Bayes), given by

$$W(\text{Bayes}) = \alpha_2/\{\alpha_1(\alpha_1+m)\}, \tag{4.5}$$

is estimated by substituting $\hat{\boldsymbol{\alpha}}$ or $\bar{\boldsymbol{\alpha}}$.

The difference $E_nW(\text{EB}) - W(\text{Bayes})$ can be estimated by using (3.5).

For the EBE based on MLE's of $\boldsymbol{\alpha}$ we then have

$$W(\text{EB}) - W(\delta_G) = \left\{\frac{\hat{\alpha}_2 + m\alpha_2/\alpha_1}{\hat{\alpha}_1 + m} - \frac{\alpha_2 + m\alpha_2/\alpha_1}{\alpha_1 + m}\right\}^2 + \frac{(\alpha_1 - \hat{\alpha}_1)^2}{[(\hat{\alpha}_1 + m)(\alpha_1 + m)]^2}\operatorname{var}(Z) \tag{4.6}$$

where $\operatorname{var}(Z)$ is the variance of Z with respect to the marginal distribution $P_G(z)$. A simple approximation for $E_nW(\text{EB}) - W(\text{Bayes})$ is obtained by replacing $\hat{\alpha}_1$ in the denominators of the expression above by α in which case we obtain

$$\frac{\operatorname{var}(\hat{\alpha}_2) + (\alpha_2/\alpha_1)^2\operatorname{var}(\hat{\alpha}_1) - 2(\alpha_2/\alpha_1)\operatorname{cov}(\hat{\alpha}_1,\hat{\alpha}_2)}{(\alpha_1+m)^2} + \frac{\operatorname{var}(\hat{\alpha}_1)}{(\alpha_1+m)^4}\operatorname{var}(Z) \tag{4.7}$$

where $\operatorname{var}(Z) = (m\alpha_2/\alpha_1)(1 + m/\alpha_1)$ which can be estimated in the obvious way using the estimates of $\operatorname{var}(\hat{\alpha}_1)$, $\operatorname{var}(\hat{\alpha}_2)$ and $\operatorname{cov}(\hat{\alpha}_1,\hat{\alpha}_2)$ or through the use of the observed information matrix (4.3).

(ii) *Linear EB estimation*

For the Poisson data distribution comparing (4.1) and (2.18), we have $q(\theta) = \theta$. The solution of (2.19) for this case is

$$\begin{aligned} \omega_0 &= (\gamma_1^2/m)/(\gamma_2 - \gamma_1^2 + \gamma_1/m) = \omega_0(\gamma), \\ \omega_1 &= (\gamma_2 - \gamma_1^2)/(\gamma_2 - \gamma_1^2 + \gamma_1/m) = \omega_1(\gamma) \end{aligned} \tag{4.8}$$

giving the linear Bayes estimator,

$$\delta_G(T; \omega_0, \omega_1) = \delta_G(T; \gamma) = \omega_0(\gamma) + \omega_1(\gamma)T. \tag{4.9}$$

The linear EBE is obtained by substituting an estimate $\breve{\gamma}$ for γ in (4.9). One such estimate $\breve{\gamma}$ is obtained by estimating $\gamma_p = E(\theta^p)$, $p = 1, 2$, from the estimating equations

$$\sum_{i=1}^{n} W_{pi}^*(\gamma) = 0 \qquad p = 1, 2$$

where

$$\begin{aligned} W_{1i}^*(\gamma) &= t_i - \gamma_1, \\ W_{2i}^*(\gamma) &= t_i^2 - \gamma_1/m_i - \gamma_2. \end{aligned}$$

The covariance matrix for $\breve{\gamma}$ can be obtained using the method of (3.1). We have for this case

$$D = \begin{bmatrix} -n & 0 \\ -\sum \dfrac{1}{m_i} & -n \end{bmatrix},$$

while V is the same as the expression in (4.4).

The difference $E_n W(\text{linear EB}) - W(\text{linear Bayes})$ is given by

$$E_n\{(\breve{\omega}_0 - \omega_0)^2 + 2\gamma_1(\breve{\omega}_0 - \omega_0) + (\breve{\omega}_1 - \omega_1)^2\gamma_2\}$$

which can be evaluated to second order by expanding the above expression in terms of Taylor series in $\breve{\gamma}$ and using the covariance matrix C of $\breve{\gamma}$.

The quantity $W(\text{linear Bayes})$ can be directly evaluated as

$$W(\text{linear Bayes}) = \omega_0^2 + 2\omega_0(\omega_1 - 1)\gamma_1 + (\omega_1 - 1)^2\gamma_2 + \omega_1^2\gamma_1/m.$$

5. The binomial distribution

The binomial distribution has been previously studied; but even for the case when the parametric form is assumed to be known, say a conjugate beta prior d.f., the assessment of the standard errors has not been properly made except in some special cases. Simulation studies have been carried out sporadically, but by its nature these cannot provide a comprehensive study. The linear EBE for the binomial case has been proposed by Griffin and Krutchkoff (1971) who also gave the linear Bayes estimator for a general case. Martz and Lian (1974) compared various methods of EBE for the binomial case which contains the linear EBE. But except for a few methods the existing results are based on simulation studies.

The r.v. X is $\mathrm{Bin}(1, \theta)$ so that, in the current stage, $mT = (X_1 + \cdots + X_m)$ is $\mathrm{Bin}(m, \theta)$ and T is sufficient for θ. Also note that

$$E(T \mid \theta) = \theta, \qquad E(T^2 \mid \theta) = \frac{\theta(1-\theta)}{m} + \theta^2. \tag{5.1}$$

(i) *Parametric G case*

Consider the case when θ is distributed according to a $\mathrm{Beta}(\alpha_1, \alpha_2)$ distribution. Then

$$g(\theta \mid \boldsymbol{\alpha}) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \theta^{\alpha_1 - 1}(1-\theta)^{\alpha_2 - 1} \tag{5.2}$$

and the posterior distribution of Θ is given T is $\mathrm{Beta}(\alpha_1 + mT, \alpha_2 + m - Tm)$. From (2.5),

$$\begin{aligned} U_{1i}(\boldsymbol{\alpha}) &= \psi(\alpha_1+\alpha_2) - \psi(\alpha_1) + E(\ln\Theta \mid T = t_i) \\ &= \psi(\alpha_1+\alpha_2) - \psi(\alpha_1) - \psi(\alpha_1+\alpha_2+m_i) + \psi(\alpha_1 + m_i t_i), \\ U_{2i}(\boldsymbol{\alpha}) &= \psi(\alpha_1+\alpha_2) - \psi(\alpha_2) + E(\ln(1-\Theta) \mid T = t_i) \\ &= \psi(\alpha_1+\alpha_2) - \psi(\alpha_2) - \psi(\alpha_1+\alpha_2+m_i) + \psi(\alpha_2 + m_i - m_i t_i). \end{aligned} \tag{5.3}$$

These quantities define the vector S in (2.11). Also, we have

$$J^{-1} = \frac{1}{nd}\begin{bmatrix} \psi'(\alpha_1+\alpha_2) - \psi'(\alpha_2) & -\psi'(\alpha_1+\alpha_2) \\ -\psi'(\alpha_1+\alpha_2) & \psi'(\alpha_1+\alpha_2) - \psi'(\alpha_1) \end{bmatrix} \tag{5.4}$$

where

$$d = \psi'(\alpha_1)\psi'(\alpha_2) - \psi'(\alpha_1+\alpha_2)\{\psi'(\alpha_1) + \psi'(\alpha_2)\}.$$

Hence the iterative equation for updating a vector of initial estimates $(\hat{\alpha}_1^{(0)}, \hat{\alpha}_2^{(0)})$ can be readily written down. The covariance matrix of $\hat{\alpha}$ is obtained as the inverse of V whose elements are

$$\begin{cases} V_{11} = \displaystyle\sum_{i=1}^{n} \mathrm{var}\{\psi(\alpha_1 + m_i T_i)\}, \\ V_{12} = V_{21} = \displaystyle\sum_{i=1}^{n} \mathrm{cov}\{\psi(\alpha_1 + m_i T_i), \psi(\alpha_2 + m_i(1 - T_i))\}, \\ V_{22} = \displaystyle\sum_{i=1}^{n} \mathrm{var}\{\psi(\alpha_2 + m_i(1 - T_i))\}. \end{cases} \tag{5.5}$$

The Bayes estimator of θ is

$$\hat{\theta}^*(t; \boldsymbol{\alpha}, m) = (\alpha_1 + mt)/(\alpha_1 + \alpha_2 + m)$$

and the EBE of θ is obtained by replacing $\boldsymbol{\alpha}$ by $\hat{\boldsymbol{\alpha}}$. An approximation for the $E_n W(\mathrm{EB})$ is obtained from the approximate relation

$$(5.6)\qquad \begin{aligned} &E_n W(\mathrm{EB}) - W(\mathrm{Bayes}) \\ &\quad \simeq \frac{\alpha_2^2 \operatorname{var}(\hat{\alpha}_1) + \alpha_1^2 \operatorname{var}(\hat{\alpha}_2) + 2\alpha_1\alpha_2 \operatorname{cov}(\hat{\alpha}_1, \hat{\alpha}_2)}{(\alpha_1+\alpha_2)^2(\alpha_1+\alpha_2+m)^2} \\ &\qquad + \frac{\operatorname{var}(\hat{\alpha}_1) + \operatorname{var}(\hat{\alpha}_2) + 2\operatorname{cov}(\hat{\alpha}_1 + \hat{\alpha}_2)}{(\alpha_1+\alpha_2+m)^4} \operatorname{var}(mT) \end{aligned}$$

where

$$\operatorname{var}(mT) = \frac{m^2\alpha_1(\alpha_1+1) + m\alpha_1\alpha_2}{(\alpha_1+\alpha_2)(\alpha_1+\alpha_2+1)} - m^2 \frac{\alpha_1^2}{(\alpha_1+\alpha_2)^2}.$$

Approximate expressions for variances and covariance of $\hat{\alpha}_1$, $\hat{\alpha}_2$ are obtained from V^{-1} and the Bayes risk is

$$(5.7)\qquad W(\mathrm{Bayes}) = \alpha_1\alpha_2/\{(\alpha_1+\alpha_2)(\alpha_1+\alpha_2+1)(\alpha_1+\alpha_2+m)\}.$$

(ii) *Linear EB estimation*

For the Binomial distribution we have

$$E(T \mid \theta) = \theta,$$
$$E(T^2 \mid \theta) = \theta^2 + \theta(1-\theta)/m,$$

so, as in the Poisson case, we need estimates of $\gamma_p = E(\theta^p)$, $p = 1, 2$. They are obtained by solving

$$\sum_{i=1}^{n} W_{pi}^*(\gamma) = 0, \qquad p = 1, 2$$

where

$$(5.8)\qquad \begin{aligned} W_{1i}^*(\gamma) &= t_i - \gamma_1, \\ W_{2i}^*(\gamma) &= t_i^2 - \gamma_2 - (\gamma_1 - \gamma_2)/m_i \end{aligned}$$

and an estimated approximate covariance matrix of the estimates $\check{\gamma}$ of γ can be calculated as in the ML case. The solution of (2.19) for the binomial distribution is

$$\omega_0 = \{\gamma_1(\gamma_1 - \gamma_2)/m\}/\{\gamma_2 - \gamma_1^2 + (\gamma_1 - \gamma_2)/m\},$$
$$\omega_1 = (\gamma_2 - \gamma_1^2)/\{\gamma_2 - \gamma_1^2 + (\gamma_1 - \gamma_2)/m\}.$$

Also

$$W(\text{linear Bayes}) = \omega_0^2 + 2\omega_0(\omega_1 - 1)\gamma_1 + (\omega_1 - 1)^2\gamma_2 + \omega_1^2(\gamma_1 - \gamma_2)/m$$

and $E_n W(\text{linear EB}) - W(\text{linear Bayes})$ can be estimated as in the Poisson case, using the estimated covariance matrix of $\check{\omega}_0$, $\check{\omega}_1$ obtained from the estimated covariance matrix of $\check{\gamma}$ and the relation (5.8).

6. The normal distribution

We take the distribution of X to be $N(\theta, \sigma^2)$ with σ^2 known and fixed. Since we have $m > 1$ current observations, and $m_i > 1$ past observations at each stage it is possible to calculate an estimate of σ^2 and use it in EB estimation. It is also possible to develop EBE's in which σ is allowed to vary from one experiment to another so that we have the sequence (θ_i, σ_i^2) of realizations of the two-dimensional parameter random variable (Θ, Σ^2). However, we shall here consider only the fixed and known σ^2 case.

The estimate $T = (X_1 + \cdots + X_m)/m$ of θ is distributed $N(\theta, \sigma^2/m)$ and is, of course, a sufficient statistic. For the parametric G case we shall take the prior distribution to be $N(\alpha_1, \alpha_2)$; in our previous notations $\alpha_1 = \gamma_1$, $\alpha_2 = \gamma_2 - \gamma_1^2$.

(i) *Parametric G: Θ distributed $N(\alpha_1, \alpha_2)$*

The marginal distribution of T is $N(\alpha_1, \alpha_2 + \sigma^2/m)$ hence it follows simply that the MLE of α_1, α_2 are the solutions of $S_p(\boldsymbol{t}; \boldsymbol{\alpha}) = 0$, $p = 1, 2$ as in equations (2.5) or (2.6) with

$$\begin{aligned} U_{1i}(\boldsymbol{\alpha}) &= (t_i - \alpha_1)/(\alpha_2 + \sigma^2/m_i), \\ U_{2i}(\boldsymbol{\alpha}) &= -1/(\alpha_2 + \sigma^2/m_i) + (t_i - \alpha_1)^2/(\alpha_2 + \sigma^2/m_i)^2. \end{aligned} \tag{6.1}$$

The matrix J required for the iterative ML process is given by the elements

$$\begin{aligned} J_{11} &= -n/\alpha_2, \\ J_{12} &= -\sum_{i=1}^{n} \frac{m_i(t_i - \alpha_1)}{\alpha_2(\sigma^2 + m_i\alpha_2)} = J_{21}, \\ J_{22} &= \frac{n}{2\alpha_2^2} - \frac{1}{\alpha_2^3} \sum_{i=1}^{n} \frac{\sigma^2\alpha_2}{m_i\alpha_2 + \sigma^2} - \sum_{i=1}^{n} \frac{m_i^2(t_i - \alpha_1)^2}{\alpha_2(\sigma^2 + m_i\alpha_2)^2}. \end{aligned}$$

Now the elements of the approximate covariance matrix of $\hat{\boldsymbol{\alpha}}$ are given by V^{-1} where elements of V are given by

$$\begin{aligned} V_{11} &= \sum_{i=1}^{n} 1/(\alpha_2 + \sigma^2/m_i), \\ V_{12} &= V_{21} = 0, \\ V_{22} &= \sum_{i=1}^{n} 1/\{2(\alpha_2 + \sigma^2/m_i)^2\}. \end{aligned}$$

The covariance matrix of $\hat{\boldsymbol{\alpha}}$ is used, as before, in the following to assess the goodness of the EBE relative to the MLE. We have

$$W(\text{Bayes}) = 1/(m/\sigma^2 + 1/\alpha_2)$$

and after making approximations similar to those giving (4.7) and (5.6),

$$E_n W(\text{EB}) - W(\text{Bayes}) \simeq \frac{\sigma^4}{(\sigma^2 + m\alpha_2)^2} \text{var}(\hat{\alpha}_1) + \frac{\sigma^4}{(\sigma^2 + m\alpha_2)^3} \text{var}(\hat{\alpha}_2).$$

(ii) *Linear EB estimation*
For the normal distribution we have

$$E(T \mid \theta) = \theta,$$
$$E(T^2 \mid \theta) = \theta^2 + \sigma^2/m.$$

Again, we need estimates of $\gamma_p = E(\theta^p)$, $p = 1, 2$. These are obtained by solving

$$\sum_{i=1}^{n} W_{pi}^*(\gamma) = 0 \qquad p = 1, 2$$

where

$$W_{1i}^*(\gamma) = t_i - \gamma_1, \qquad W_{2i}^*(\gamma) = t_i^2 - \gamma_2 - \sigma^2/m_i \tag{6.2}$$

and the solution of (2.19) for the normal case is

$$\omega_0 = \gamma_1(\sigma^2/m)/\{\gamma_2 - \gamma_1^2 + \sigma^2/m\},$$
$$\omega_1 = (\gamma_2 - \gamma_1^2)/\{\gamma_2 - \gamma_1^2 + \sigma^2/m\}.$$

Also

$$W(\text{linear Bayes}) = \frac{(\gamma_2 - \gamma_1^2)\sigma^2}{\{\sigma^2 + m(\gamma_2 - \gamma_1^2)\}}.$$

The quantity $E_n W$(linear EB) can again be evaluated in terms of the covariance of $\breve{\omega}_0$ and $\breve{\omega}_1$.

7. Discussion

The main contribution of this paper is to bring the assessment of empirical Bayes estimators to the same level as the assessment of ML estimators in the usual theory at least for the case when the prior d.f. can be represented by a parametric d.f. This was made possible by the use of a second order result (3.1). A more rigorous proof of this result is recently given by Brown (1985); the examples used in this paper satisfy the required regularity conditions mentioned there.

References

Brown, B. M. (1985). Multiparameter linearization theorems, *J. Roy. Statist. Soc. Ser. B*, **47**, 323–331.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*, Chapman and Hall, London.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion), *J. Roy. Statist. Soc. Ser. B*, **39**, 1–38.

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

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm, *J. Roy. Statist. Soc. Ser. B*, **44**, 226–233.

Lwin, T. and Maritz, J. S. (1989). Empirical Bayes approach to multiparameter problems: with special application to the multinomial distribution, *Ann. Inst. Statist. Math.*, **41**, 81–99.

Maritz, J. S. (1967). Smooth empirical Bayes estimators for continuous distributions, *Biometrika*, **54**, 435–450.

Maritz, J. S. (1981). *Distribution-Free Statistical Methods*, Chapman and Hall, London.

Maritz, J. S. and Lwin, T. (1989). *Empirical Bayes Methods*, Chapman and Hall, London.

Martz, H. F., Jr. and Lian, M. G. (1974). A survey and comparison of several empirical Bayes estimators for the Binomial parameter, *J. Statist. Comput. Simulation*, **3**, 165–178.

O'Bryan, T. E. (1976). Some empirical Bayes results in the case of component problems with varying sample sizes for discrete exponential families, *Ann. Statist.*, **4**, 1290–1293.

O'Bryan, T. E. and Susarla, V. (1977). Empirical Bayes estimation with non-identical components: Continuous case, *Austral. J. Statist.*, **19**, 115–125.

Southward, G. M. and Van Ryzin, J. R. (1972). Estimating the mean of a random binomial parameter, *Proc. Sixth Berkeley Symp. on Math. Statist. Prob.*, Vol. IV, 249–263.