Empirical Bayes prediction intervals in a normal regression model: higher order asymptotics

Ruma Basu[a], J.K. Ghosh[b,c], Rahul Mukerjee[d,*]

[a]*Department of Statistics, Columbia University, USA*
[b]*Purdue University, USA*
[c]*Indian Statistical Institute, India*
[d]*Indian Institute of Management Calcutta, Alipore Post Office, Diamond Harbour Road, P.O. Box 16757, Calcutta 700-027, India*

Abstract

We explore two proposals for finding empirical Bayes prediction intervals under a normal regression model. The coverage probabilities and expected lengths of such intervals are studied and compared via appropriate higher-order asymptotics.

Keywords: Coverage probability; Equal tail; Expected length

1. Introduction

Consider observable random variables $Y_1,\ldots,Y_k$ such that given $\theta_1,\ldots,\theta_k$, they are independent and Y_j is normal $\mathrm{N}(\theta_j,V)$, $1\leqslant j\leqslant k$. Also, the unobservable $\theta_1,\ldots,\theta_k$ are themselves independent and θ_j is normal $\mathrm{N}(z_j'\beta,A)$, $1\leqslant j\leqslant k$. Here $V(>0)$ is known, z_j' is a known $1\times r$ vector ($1\leqslant j\leqslant k$), β is an $r\times 1$ vector of unknown parameters, and $A(>0)$ is an unknown constant.

The present article studies two popular proposals, one due to Morris (1983) and the other discussed in Carlin and Louis (1996, p. 98), for obtaining empirical Bayes prediction intervals for θ_i on the basis of the observational vector $Y=(Y_1,\ldots,Y_k)'$. While the coverage probabilities of these intervals have so far been studied numerically, we develop the higher-order asymptotics on their coverage and expected lengths, as $k\to\infty$ keeping i and r fixed. The prediction interval proposed by Morris (1983) is considered in Section 2, where it is seen that the interval may not attain the target coverage

* Corresponding author.
E-mail address: rmuk1@hotmail.com (R. Mukerjee).

probability in all cases but that a simple modification thereof can achieve this with margin of error $o(k^{-1})$. Next, in Section 3 the idea of Carlin and Louis (1996) is explored analytically to obtain an explicit expression for another prediction interval with the same asymptotic coverage property as the modified version of Morris' interval. In Section 4, it is seen that these two intervals, both of which attain the target coverage probability with margin of error $o(k^{-1})$, have the same expected length too up to that order of approximation. The proofs are given in the appendix. It may be emphasized that our implementation of Carlin and Louis' (1996) idea in Section 3 for the normal regression model is analytic while their discussion seems to be motivated towards numerical derivation of intervals. We refer to Datta et al. (2000) for a related result on the choice of noninformative priors in a special case of the setup considered here.

Let Z_k be the $k \times r$ design matrix with rows $z_1', \ldots, z_k'$. As in Morris (1983), we assume that $r \leqslant k-3$ and that Z_k has full column rank. The following additional assumption is needed.

Assumption. Let $c_{ik} = k\{z_i'(Z_k'Z_k)^{-1}z_i\}$. For fixed i, the sequence $\{c_{ik}\}$ is bounded.

The above assumption holds quite commonly—e.g., it holds if the smallest eigenvalue of $k^{-1}Z_k'Z_k$ is bounded away from zero.

2. Morris' interval and modification thereof

Let $\xi = (\beta', A)'$, be the vector of unknown parameters and define

$$b = (Z_k'Z_k)^{-1}Z_k'Y, \qquad s^2 = (k-r)^{-1}\sum_{j=1}^{k}(Y_j - z_j'b)^2, \tag{2.1}$$

$$B = V/(V+A), \qquad \widehat{B} = \frac{k-r-2}{k-r}\,\frac{V}{\max(V, s^2)}, \tag{2.2}$$

$$\rho = B/(1-B), \qquad \widehat{\rho} = \widehat{B}/(1-\widehat{B}), \tag{2.3}$$

$$h_{ik}(u;\rho) = \tfrac{1}{2}(c_{ik} - r)\rho + \tfrac{1}{4}(u^2+1)\rho^2 \quad (-\infty < u < \infty), \tag{2.4}$$

$$\widehat{\theta_i} = (1-\widehat{B})Y_i + \widehat{B}z_i'b, \tag{2.5}$$

$$s_i^2 = V\left(1 - \frac{k-r}{k}\widehat{B}\right) + \frac{2}{k-r-2}\widehat{B}^2(Y_i - z_i'b)^2. \tag{2.6}$$

With a target coverage probability of at least $1-\alpha$, Morris (1983) proposed the prediction interval $\widehat{\theta_i} \pm z s_i$ for θ_i, where z is the upper $\alpha/2$ point of a standard normal variate. The following theorem is crucial in studying this interval.

Theorem 2.1. *For any convergent nonstochastic sequence $\{t_k\}$ of real numbers,*

$$P_\xi\{(\theta_i - \widehat{\theta_i})/s_i \leqslant t_k\} = \Phi(t_k) - k^{-1}t_k\phi(t_k)h_{ik}(t_k;\rho) + o(k^{-1}),$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and distribution functions, respectively.

For Morris' (1983) interval $\widehat{\theta_i} \pm zs_i$, by (2.4) and Theorem 2.1,

$$P_\xi(\widehat{\theta_i} - zs_i \leqslant \theta_i \leqslant \widehat{\theta_i} + zs_i) = 1 - \alpha - 2k^{-1}z\phi(z)h_{ik}(z;\rho) + o(k^{-1}). \tag{2.7}$$

Interestingly, $h_{ik}(z;\rho)$ can be positive and hence the term of order $O(k^{-1})$ in (2.7) can be negative. This happens, for example if $r = 1$ and Z_k equals the $k \times 1$ vector of 1's (cf. Datta et al., 2000). Then $c_{ik} = 1$ and by (2.4), $h_{ik}(z;\rho) = \frac{1}{4}(z^2+1)\rho^2$. Thus the coverage probability of the interval $\widehat{\theta_i} \pm zs_i$, as it stands, can fall short of $1-\alpha$ up to the order of approximation considered in (2.7).

We now indicate a simple modification of Morris' (1983) interval that attains a coverage probability $1-\alpha$ with margin of error $o(k^{-1})$. This is given by

$$I = [\widehat{\theta_i} - z\{1 + k^{-1}h_{ik}(z;\widehat{\rho})\}s_i, \hat{\theta}_i + z\{1 + k^{-1}h_{ik}(z,\widehat{\rho})\}s_i]. \tag{2.8}$$

By Eq. (A.7) in the appendix, $\widehat{B} = B + o(1)$ and hence $\widehat{\rho} = \rho + o(1)$, on a set with P_ξ-probability $1 + o(k^{-1})$ uniformly over compact ξ-sets. Hence by (2.8) and Theorem 2.1,

$$\begin{aligned} P_\xi(\theta_i \in I) &= P_\xi[-z\{1 + k^{-1}h_{ik}(z;\rho)\} \leqslant (\theta_i - \widehat{\theta_i})/s_i \leqslant z\{1 + k^{-1}h_{ik}(z;\rho)\}] + o(k^{-1}) \\ &= 2\Phi(z) - 1 + o(k^{-1}) = 1 - \alpha + o(k^{-1}). \end{aligned}$$

In fact, by Theorem 2.1, it can similarly be seen that

$$P_\xi[\theta_i > \widehat{\theta_i} + z\{1 + k^{-1}h_{ik}(z;\widehat{\rho})\}s_i] = \alpha/2 + o(k^{-1}),$$

$$P_\xi[\theta_i < \widehat{\theta_i} - z\{1 + k^{-1}h_{ik}(z;\widehat{\rho})\}s_i] = \alpha/2 + o(k^{-1}).$$

Hence, in addition to attaining a coverage probability $1-\alpha$ with margin of error $o(k^{-1})$, the interval I is equal tailed up to the same order of approximation.

3. Another interval

We now explore analytically the idea of Carlin and Louis (1996) to get another prediction interval for θ_i. Observe that given Y conditionally θ_i is normal $N((1-B)Y_i + Bz_i'\beta, V(1-B))$. Hence defining $\theta_i^* = (1-B^*)Y_i + B^*z_i'b$, where

$$B^* = V/\max\left(V, \frac{k-r}{k}s^2\right)$$

is the maximum likelihood estimator (MLE) of B, one can check that this approach yields a prediction interval for θ_i of the form

$$[\theta_i^* + t^{(1)}(\xi^*)\{V(1-B^*)\}^{1/2}, \theta_i^* + t^{(2)}(\xi^*)\{V(1-B^*)\}^{1/2}].$$

In the above, ξ^* is the MLE of ξ and $t^{(1)}(\cdot)$ and $t^{(2)}(\cdot)$ are such that

$$P_\xi[(\theta_i - \theta_i^*)/\{V(1-B^*)\}^{1/2} < t^{(1)}(\xi)] = \alpha/2, \tag{3.1}$$

$$P_\xi[(\theta_i - \theta_i^*)/\{V(1-B^*)\}^{1/2} > t^{(2)}(\xi)] = \alpha/2. \tag{3.2}$$

The following theorem helps in the approximate determination of $t^{(1)}(\cdot)$ and $t^{(2)}(\cdot)$ so that (3.1) and (3.2) hold with margin of error $o(k^{-1})$. Its proof is similar to that of Theorem 2.1 and hence omitted. In what follows, for any real u,

$$h_{ik}^*(u;\rho) = \tfrac{1}{2}(c_{ik} + r + 4)\rho + \tfrac{1}{4}(u^2+1)\rho^2. \tag{3.3}$$

Theorem 3.1. *For any convergent nonstochastic sequence* $\{t_k\}$ *of real numbers,*

$$P_\xi\{(\theta_i - \theta_i^*)/\{V(1-B^*)\}^{1/2} \leqslant t_k\} = \Phi(t_k) - k^{-1}t_k\phi(t_k)h_{ik}^*(t_k;\rho) + o(k^{-1}).$$

By Theorem 3.1, with

$$t^{(1)}(\xi) = -z\{1 + k^{-1}h_{ik}^*(z;\rho)\}, \quad t^{(2)}(\xi) = z\{1 + k^{-1}h_{ik}^*(z;\rho)\},$$

conditions (3.1) and (3.2) hold with margin of error $o(k^{-1})$. Hence the idea of Carlin and Louis (1996) yields the prediction interval

$$I^* = [\theta_i^* - z\{1 + k^{-1}h_{ik}^*(z;\rho^*)\}\{V(1-B^*)\}^{1/2},\ \theta_i^* + z\{1 + k^{-1}h_{ik}^*(z;\rho^*)\}\{V(1-B^*)\}^{1/2}]$$

for θ_i, where ρ^* equals $B^*/(1-B^*)$ if $B^* < 1$, and is defined arbitrarily if $B^* = 1$. As in Section 2, by Theorem 3.1, this interval is equal tailed up to $o(k^{-1})$ in addition to attaining a coverage probability $1-\alpha$ up to the same order of approximation.

4. Expected lengths

Let L and L^* be the lengths of the intervals I and I^*, respectively. Define

$$Q_{ik}(z;\rho) = (\tfrac{1}{2}c_{ik} + 1)\rho + \tfrac{1}{4}z^2\rho^2.$$

Then the following theorem holds.

Theorem 4.1. (a) $E_\xi(L) = 2z\{V(1-B)\}^{1/2}\{1 + k^{-1}Q_{ik}(z;\rho)\} + o(k^{-1})$.
(b) $E_\xi(L^*) = 2z\{V(1-B)\}^{1/2}\{1 + k^{-1}Q_{ik}(z;\rho)\} + o(k^{-1})$.

Thus the two prediction intervals I and I^*, based respectively on modification of Morris' (1983) proposal and analytic implementation of Carlin and Louis' (1996) idea, are at par with respect to both coverage probability and expected length, even under higher order asymptotics retaining terms of order $O(k^{-1})$.

Acknowledgements

We thank a referee for very constructive suggestions. This work was supported by a grant from the Center for Management and Development Studies, Indian Institute of Management, Calcutta.

Appendix

The following preliminaries will facilitate the presentation of the proofs. Let

$$\delta = (V + A)^{1/2}, \quad X_1 = k^{1/2}(z_i'b - z_i'\beta)/\delta, \quad X_2 = (Y_i - z_i'\beta)/\delta, \tag{A.1}$$

$$X_3 = k^{1/2}(\delta^{-2}s^2 - 1), \quad X_4 = X_3 - k^{-1/2}(X_3^2 - 2). \tag{A.2}$$

By (2.1), $c_{ik}^{-1/2}X_1$ is standard normal and hence the assumption made in Section 1 implies that X_1 is stochastically bounded. It is easy to see that so are X_2, X_3 and X_4; cf. (2.1).

Lemma. (i) $E_\xi(X_1) = 0$, (ii) $E_\xi(X_4) = o(k^{-1})$, (iii) $E_\xi(X_1^2) = c_{ik}$, (iv) $E_\xi(X_2^2) = 1$, (v) $E_\xi(X_4^2) = 2 + o(1)$, (vi) $E_\xi(X_1X_4) = 0$, (vii) $E_\xi(X_2X_4) = 0$, (viii) $E_\xi(X_2X_4^2) = 0$, (ix) $E_\xi(X_1X_2X_4) = o(1)$, (x) $E_\xi(X_2^2X_4^2) = 2 + o(1)$.

Proof. The proofs of parts (i)–(v) are either obvious or straightforward. To prove the remaining parts, define

$$U = (U_1, \ldots, U_k)' = \delta^{-1}(Y - Z_k\beta), \qquad w_i = Z_k(Z_k'Z_k)^{-1}z_i$$

and let M be the orthogonal projector on the orthocomplement of the column space of Z_k. Then

$$\operatorname{tr}(M) = k - r, \qquad Mw_i = 0 \tag{A.3}$$

and by (2.2), (A.1) and (A.2)

$$X_1 = k^{1/2}w_i'U, \quad X_2 = U_i, \quad X_3 = k^{1/2}\{(k-r)^{-1}U'MU - 1\}. \tag{A.4}$$

By (A.2) and (A.4), each of X_1X_4, X_2X_4 and $X_2X_4^2$ is an odd polynomial in U. Hence parts (vi)–(viii) follow noting that the elements of U are independently standard normal, a fact that is also used in the rest of the proof whenever necessary.

Next, by (A.2) and (A.4)

$$E_\xi(X_1X_2X_4) = kE_\xi[(w_i'U)U_i\{(k-r)^{-1}U'MU - 1\}] + o(1). \tag{A.5}$$

Write w_{ii} for the ith element of w_i and m_i for the ith column of M. Then by (A.3),

$$E_\xi\{(w_i'U)U_i(U'MU)\} = w_{ii}\operatorname{tr}(M) + 2w_i'm_i = w_{ii}(k-r).$$

Since $E_\xi\{(w_i'U)U_i\} = w_{ii}$, part (ix) is now evident from (A.5).

Finally, in order to prove (x), note that by (A.2) and (A.4)

$$E_\xi(X_2^2X_4^2) = kE_\xi[U_i^2\{(k-r)^{-1}U'MU - 1\}^2] + o(1). \tag{A.6}$$

Now, M is idempotent with (i,i)th element $1 - z_i'(Z_k'Z_k)^{-1}z_i (= 1 - k^{-1}c_{ik})$. Hence after some algebraic manipulation,

$$E_\xi\{U_i^2(U'MU)\} = k - r + 2(1 - k^{-1}c_{ik}),$$

$$E_\xi\{U_i^2(U'MU)^2\} = (k-r)^2 + 2(k-r) + 4(k-r+2)(1-k^{-1}c_{ik}).$$

Using the above in (A.6) and recalling the assumption on the boundedness of $\{c_{ik}\}$, part (x) of the lemma follows. □

We are now in a position to present the proofs. All stochastic expansions considered below are on a set with P_ξ-probability $1 + o(k^{-1})$ (uniformly over compact ξ-sets). In particular, $s^2 > V$ on this set, so that by (2.2), (A.1) and (A.2),

$$\widehat{B} = \frac{k-r-2}{k-r}(1 + k^{-1/2}X_3)^{-1}B = B(1 - k^{-1/2}X_4) + o(k^{-1}). \tag{A.7}$$

Proof of Theorem 2.1. Given Y, conditionally θ_i is normal $\mathrm{N}((1-B)Y_i + Bz_i'\beta, V(1-B))$. Hence

$$P_\xi\{(\theta_i - \widehat{\theta_i})/s_i \leqslant t_k\} = E_\xi\{\Phi(T)\}, \tag{A.8}$$

where

$$T = \{\widehat{\theta_i} + t_ks_i - (1-B)Y_i - Bz_i'\beta\}/\{V(1-B)\}^{1/2}. \tag{A.9}$$

Recalling the definitions of B, δ and ρ, by (2.5), (2.6), (A.1) and (A.7), after some algebra,

$$\widehat{\theta_i} - (1-B)Y_i - Bz_i'\beta = \delta B\{k^{-1/2}(X_1 + X_2X_4) - k^{-1}X_1X_4\} + o(k^{-1}), \tag{A.10}$$

$$s_i = \{V(1-B)\}^{1/2}[1 + \tfrac{1}{2}k^{-1/2}\rho X_4 + \tfrac{1}{2}k^{-1}\{(r + 2X_2^2)\rho - \tfrac{1}{4}\rho^2X_4^2\}] + o(k^{-1}). \tag{A.11}$$

By (A.9)–(A.11),

$$T = t_k + k^{-1/2}g_1 + k^{-1}g_2 + o(k^{-1}), \tag{A.12}$$

where

$$g_1 = \tfrac{1}{2}t_k\rho X_4 + \rho^{1/2}(X_1 + X_2X_4), \tag{A.13}$$

$$g_2 = \tfrac{1}{2}t_k\{(r + 2X_2^2)\rho - \tfrac{1}{4}\rho^2X_4^2\} - \rho^{1/2}X_1X_4. \tag{A.14}$$

By (A.12),

$$\Phi(T) = \Phi(t_k) + k^{-1/2}\phi(t_k)g_1 + k^{-1}\phi(t_k)(g_2 - \tfrac{1}{2}t_kg_1^2) + o(k^{-1}). \tag{A.15}$$

Using the lemma, from (A.13) and (A.14),

$$E_\xi(g_1) = \text{o}(k^{-1/2}), \quad E_\xi(g_2) = \tfrac{1}{2}\,t_k\{(r+2)\rho - \tfrac{1}{2}\,\rho^2\} + \text{o}(1), \tag{A.16}$$

$$E_\xi(g_1^2) = \tfrac{1}{2}\,t_k^2\rho^2 + (c_{ik}+2)\rho + \text{o}(1). \tag{A.17}$$

If one substitutes (A.15) in (A.8) and then employs (A.16) and (A.17) then the result follows. □

Proof of Theorem 4.1. By (2.3), (2.8) and (A.7), the length L of the interval I satisfies $L = 2z\{1 + k^{-1}h_{ik}(z;\rho)\}s_i + \text{o}(k^{-1})$. Hence by (A.11),

$$L = 2z\{V(1-B)\}^{1/2}[1 + \tfrac{1}{2}k^{-1/2}\rho X_4 + k^{-1}\{(\tfrac{1}{2}r + X_2^2)\rho - \tfrac{1}{8}\rho^2 X_4^2 + h_{ik}(z;\rho)\}] + \text{o}(k^{-1}).$$

Part (a) now follows using (2.4) and the lemma. The proof of (b) is similar. □

References

Carlin, B.P., Louis, T.A., 1996. Bayes and Empirical Bayes Methods for Data Analysis. Chapman & Hall, London.

Datta, G.S., Ghosh, M., Mukerjee, R., 2000. Some new results on probability matching priors. Calcutta Statist. Assoc. Bull. 50, 179–192.

Morris, C.N., 1983. Parametric empirical Bayes inference: theory and applications. J. Amer. Statist. Assoc. 78, 47–55.