Approaches for Empirical Bayes Confidence Intervals

Author(s): Bradley P. Carlin and Alan E. Gelfand

Source: *Journal of the American Statistical Association*, Mar., 1990, Vol. 85, No. 409 (Mar., 1990), pp. 105-114

Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association

Stable URL: https://www.jstor.org/stable/2289531

Approaches for Empirical Bayes Confidence Intervals

BRADLEY P. CARLIN and ALAN E. GELFAND*

Parametric empirical Bayes (EB) methods of point estimation date to the landmark paper by James and Stein (1961). Interval estimation through parametric empirical Bayes techniques has a somewhat shorter history, which was summarized by Laird and Louis (1987). In the exchangeable case, one obtains a "naive" EB confidence interval by simply taking appropriate percentiles of the estimated posterior distribution of the parameter, where the estimation of the prior parameters ("hyperparameters") is accomplished through the marginal distribution of the data. Unfortunately, these "naive" intervals tend to be too short, since they fail to account for the variability in the estimation of the hyperparameters. That is, they do not attain the desired coverage probability in the EB sense defined by Morris (1983a,b). They also provide no statement of conditional calibration (Rubin 1984). In this article we propose a conditional bias correction method for developing EM intervals that corrects these deficiencies in the naive intervals. As an alternative, several authors have suggested use of the marginal posterior in this regard. We attempt to clarify its role in achieving EB coverage. Results of extensive simulation of coverage probability and interval length for these approaches are presented in the context of several illustrative examples.

KEY WORDS: Bias correction; Conditional calibration; Parametric bootstrap.

1. INTRODUCTION

Consider the usual exchangeable Bayesian formulation; that is, given θ_i the data Y_{ij} ($j = 1, \ldots, n_i$) are independent, having probability density function $f(y \mid \theta)$, $i = 1, \ldots, p$, and the θ_i's are iid from some continuous prior distribution having density $\pi(\theta \mid \eta)$ over Θ. Our ensuing development assumes θ_i is a scalar; however, extension to θ_i, a vector, is illustrated in Example 2.4. We shall work in the parametric empirical Bayes (EB) setting (Morris 1983a) and let η index the members of the family π, although η could be viewed as indexing all distributions, producing the nonparametric empirical Bayes of Robbins (1983). By construction, the $Y_i = (Y_{i1}, \ldots, Y_{in_i})$ are marginally independent with distribution $m(y \mid \eta)$, although within i, Y_{ij} and $Y_{ij'}$ are not independent. The joint marginal distribution of all the data, $\mathbf{Y} = (Y_1, \ldots, Y_p)$ is thus $m(\mathbf{Y} \mid \eta) = \prod_{i=1}^{p} m(Y_i \mid \eta)$. Finally, let $f(\theta_i \mid y_i, \eta)$ denote the posterior distribution of θ_i.

In the fully Bayesian setting, one chooses a value for η (based on subjective information or prior knowledge) and then bases all inference about θ_i on $f(\theta_i \mid y_i, \eta)$. Familiar confidence intervals for θ_i based upon this posterior distribution include the following:

- Equal tail, where we take the upper and lower $\alpha/2$ points of $f(\theta_i \mid y_i, \eta)$, respectively, as our interval. If we let $q_\alpha(y_i, \eta)$ be the αth quantile of $f(\theta_i \mid y_i, \eta)$, we may write this interval as

$$(q_{\alpha/2}(y_i, \eta), q_{1-\alpha/2}(y_i, \eta)). \tag{1.1}$$

- Highest posterior density (see Berger 1985), where we take all $\theta_i \in S$ such that $f(\theta_i \mid y_i, \eta) \geq c(\alpha)$ and $\Pr(\theta_i \in S) = 1 - \alpha$. If our posterior is unimodal we obtain an interval

$$(q^*_{\alpha}(y_i, \eta), q^{**}_{1-\alpha}(y_i, \eta)), \quad \alpha^* + \alpha^{**} = \alpha. \tag{1.2}$$

In the EB setting, we view η as unknown and use $m(\mathbf{Y} \mid \eta)$ to obtain an estimator $\hat{\eta}(\mathbf{Y})$. EB point estimation based upon the resulting "estimated posterior," $f(\theta_i \mid y_i, \hat{\eta})$, has been well discussed (see Berger 1985). Best choices of $\hat{\eta}$ [e.g., maximum likelihood estimate (MLE), uniformly minimum variance unbiased estimate (UMVUE), moments estimate] in a decision-theoretic sense usually require case-by-case investigation. This same problem arises in developing EB confidence intervals. The "naive" EB confidence intervals based upon $f(\theta_i \mid y_i, \hat{\eta})$ corresponding to (1.1) and (1.2) are, respectively,

$$(q_{\alpha/2}(y_i, \hat{\eta}), q_{1-\alpha/2}(y_i, \hat{\eta})) \tag{1.3}$$

and

$$(q^*_{\alpha}(y_i, \hat{\eta}), q^{**}_{1-\alpha}(y_i, \hat{\eta})). \tag{1.4}$$

These intervals are called "naive" because they ignore randomness in $\hat{\eta}$. Though relatively easy to compute, they are often too short, inappropriately centered, or both. More precisely, for θ_i Morris (1983a,b) defined an EB confidence set of size $1 - \alpha$ as a subset $t_\alpha(\mathbf{Y})$ of Θ such that $P_\eta(\theta_i \in t_\alpha(\mathbf{Y})) \geq 1 - \alpha$, where the probability is calculated over the joint distribution of θ_i and $\mathbf{Y}$. This definition becomes more appealing if the inequality is replaced by approximate equality. Hence we shall say that $t_\alpha(\mathbf{Y})$ is an unconditional $1 - \alpha$ EB confidence set for θ_i if and only if for each η,

$$P_\eta(\theta_i \in t_\alpha(\mathbf{Y})) \approx 1 - \alpha. \tag{1.5}$$

Rubin (1984) observed that (1.5) is "a fairly weak statement in the absence of statements about calibration conditional on characteristics of the data" (p. 1163). We concur and hence modify (1.5) to an approximately conditional statement given a suitable summary of the data, $b(\mathbf{Y})$. That is, $t_\alpha(\mathbf{Y})$ is a conditional $1 - \alpha$ EB confidence

* Bradley P. Carlin is Visiting Assistant Professor, Department of Statistics, Carnegie-Mellon University, Pittsburgh, PA 15213. Alan E. Gelfand is Professor, Department of Statistics, University of Connecticut, Storrs, CT 06269. The authors acknowledge Nan Laird and Joe Hill for valuable discussions. In particular, Laird's ideas led to the argument surrounding (3.4) and (3.5) and Hill encouraged addressing conditional coverage.

Journal of the American Statistical Association
March 1990, Vol. 85, No. 409, Theory and Methods

set for θ_i given $b(\mathbf{Y})$ if and only if for each η and $b(\mathbf{Y}) = b$,

$$P_\eta(\theta_i \in t_\alpha(\mathbf{Y}) \mid b(\mathbf{Y}) = b) \approx 1 - \alpha. \qquad (1.6)$$

The naive intervals (1.3) and (1.4) generally fail to satisfy both (1.5) and (1.6). In Section 2, we introduce a method for correcting the naive interval (1.3) to meet (1.6), where correction is made conditionally on $b(\mathbf{Y}) = Y_i$, the sufficient statistic for the posterior. Theoretical results and empirical work show, in response to Rubin, that roughly nominal coverage conditionally on Y_i ensues. This in turn insures that unconditional nominal coverage (1.5) will be roughly achieved. The method is applied to several examples, including simultaneous simple linear regression.

Several authors (Deely and Lindley 1981; Laird and Louis 1987; Morris 1983a,b, 1987; Pepple 1988; Rubin 1982) have employed a hyperprior on η to adjust confidence intervals based upon the estimated posterior to reflect the uncertainty in $\hat{\eta}$. The proposal is to use corresponding quantiles of the resulting "marginal posterior" in place of those of the estimated posterior. This additional integration (mixing) produces a distribution with more spread than the estimated posterior; hence it produces intervals longer than the naive ones. In Section 3 we explore the link between using the marginal posterior and satisfying (1.5). In Section 4 we present simulation results of coverage probabilities and interval lengths for these approaches in the context of the aforementioned examples. We summarize our findings in Section 5.

2. THE BIAS-CORRECTED NAIVE APPROACH

Efron (1987) proposed a general framework for correcting the bias in naive EB intervals. In the exchangeable case, a direct conditional bias correction may be developed as follows. We consider confidence sets for θ_i given $b(\mathbf{Y}) = Y_i$. Taking $i = 1$ w.l.o.g., recall that $q_\alpha(y_1, \eta)$ is such that

$$\Pr(\theta_1 \le q_\alpha(y_1, \eta) \mid \theta_1 \sim f(\theta_1 \mid y_1, \eta)) = \alpha. \qquad (2.1)$$

Define

$$r(\hat{\eta}, \eta, y_1, \alpha) = \Pr(\theta_1 \le q_\alpha(y_1, \hat{\eta}) \mid \theta_1 \sim f(\theta_1 \mid y_1, \eta)), \qquad (2.2)$$

and finally,

$$R(\eta, y_1, \alpha) = E_{\hat{\eta}|y_1,\eta}(r(\hat{\eta}, \eta, y_1, \alpha)), \qquad (2.3)$$

where the expectation is taken over $g(\hat{\eta} \mid y_1, \eta)$, a density with respect to Lebesgue measure. Note that R depends upon the dimensionality p of the problem as well, but this is suppressed. Since (2.3) need not be close to α, we can see why (1.3) and (1.4) usually fail to meet (1.6) for $b(\mathbf{Y}) = Y_i$. Suppose we solve

$$R(\eta, y_1, \alpha') = \alpha \qquad (2.4)$$

for α'. This α' would conditionally "correct the bias" in using $\hat{\eta}$ in our naive procedure. Applying (2.4) to each tail would produce intervals that meet (1.6) exactly. But of course we cannot solve (2.4), since η is unknown. Instead, we propose to solve

$$R(\hat{\eta}, y_1, \alpha') = \alpha \qquad (2.5)$$

to obtain $\alpha' = \alpha'(\hat{\eta}, y_1, \alpha)$. Then we take as our bias-corrected naive EB confidence interval (1.3) [or (1.4)] with α replaced by α'. In this article we confine ourselves to the case where the density $g(\hat{\eta} \mid y_1, \eta)$ is available in closed form. Calculating the left side of (2.5) in this case is called a conditional *parametric bootstrap* (Laird and Louis 1987). When $g(\hat{\eta} \mid y_1, \eta)$ is not tractable a conditional Type III parametric bootstrap (terminology again attributed to Laird and Louis; see also Sec. 3 here) estimator of the left side of (2.4) may be used in (2.5). We will detail such estimation in a subsequent paper. Note that to effect unconditional bias correction (1.5) we would replace (2.3) with $R(\eta, \alpha) = E_{\hat{\eta},y_1|\eta}(r(\hat{\eta}, \eta, y_1, \alpha))$ and solve $R(\hat{\eta}, \alpha') = \alpha$.

Under mild regularity conditions, our procedure gives a unique confidence interval.

Lemma 2.1. If $\partial r/\partial\alpha$ exists, then the bias-corrected confidence interval is unique.

Proof. From (2.1) we see $q_\alpha(\hat{\eta}, y_1) \uparrow \alpha$; hence $r(\hat{\eta}, \eta, y_1, \alpha) \uparrow \alpha$. But $\partial R/\partial\alpha = (\partial/\partial\alpha) \int r(\hat{\eta}, \eta, y_1, \alpha)\, dG(\hat{\eta} \mid y_1, \eta) = \int \partial r(\hat{\eta}, \eta, y_1, \alpha)/\partial\alpha\, dG(\hat{\eta} \mid y_1, \eta) > 0$. Thus $R \uparrow \alpha$, and (2.5) has a unique solution.

Conditional coverage given Y_1 is consistent with the Bayesian view given in (1.1) and (1.2), since in the exchangeable case Y_1 is sufficient for θ_1 in the posterior family; that is, $f(\theta_1 \mid \mathbf{Y}, \eta) = f(\theta_1 \mid Y_1, \eta)$. Typically when $\eta_1 > 1$ we condition on a minimal sufficient function of Y_1 (see Examples 2.3 and 2.4). Moreover, Theorem 2.1 and our empirical work show that our conditional bias correction approach for suitable $\hat{\eta}$ in fact roughly achieves (1.6) with $b(\mathbf{Y}) = Y_1$.

Implementation of (2.3)–(2.5) may be easier if $\hat{\eta}$ is independent of Y_1, for example, if $\hat{\eta}$ is based on $Y_2, \ldots, Y_p$. The integration in (2.3) is now over the usually more accessible distribution of $\hat{\eta} \mid \eta$, but correction is still conditional given Y_1 (see Case 2 in Sec. 4).

Again for θ_1 scalar, suppose there exists a function ξ_1 of θ_1 and y_1 monotone in θ_1 for fixed y_1 such that the conditional distribution of ξ_1 given y_1 is the same as the unconditional distribution of ξ_1. Then ξ_1 may be called a "pivotal" (see Cox and Hinkley 1974). Bias correction of (1.3) or (1.4) is equivalent to bias correction of the corresponding quantiles of ξ_1's distribution. Expressions (2.1) and (2.2) may now be replaced by corresponding ones with y_1 deleted.

If unconditional EB coverage is the objective, the pivotal is helpful. We may integrate trivially over $Y_1 \mid \hat{\eta}, \eta$ and then numerically over $\hat{\eta} \mid \eta$. A corresponding version of Lemma 2.1 holds, and a corresponding version of Theorem 2.1 will go through if $\xi_1 \mid \eta$ and $\hat{\eta} \mid \eta$ are stochastically ordered in η. Bounds on the unconditional expected tail probability result. To illustrate, we turn to Examples 2.1

and 2.2, where a pivotal is available, enabling simple bias correction to satisfy (1.5).

Example 2.1: Exponential/Inverse Gamma (IG). First, suppose $n_i = 1$ for all i. Let $Y_1, \ldots, Y_p \sim \exp(\theta_i)$, $i = 1, \ldots, p$, independent, and let $\theta_1, \ldots, \theta_p \overset{\text{iid}}{\sim} \mathrm{IG}(\eta, b)$, $\eta, b > 0$. Thus $f(y_i \mid \theta_i) = \theta_i^{-1}\exp(-y_i/\theta_i)$, $y_i > 0$, $\theta_i > 0$, $i = 1, \ldots, p$, and $\pi(\theta_i \mid \eta, b) = \exp(-1/\theta_i b)/(\Gamma(\eta)b^\eta\theta_i^{\eta+1})$, $\eta, b > 0$, $i = 1, \ldots, p$. Hence the marginal distribution of Y_i is

$$m(y_i \mid \eta, b) = \eta b/(by_i + 1)^{\eta+1}, \qquad y_i > 0, \quad (2.6)$$

and the posterior distribution of θ_i is

$$f(\theta_i \mid y_i, \eta, b) = \frac{\exp(-(y_i + 1/b)/\theta_i)(y_i + 1/b)^{\eta+1}}{\Gamma(\eta + 1)\theta_i^{\eta+2}}; \quad (2.7)$$

that is, (2.7) is $\mathrm{IG}(\eta + 1, (y_i + 1/b)^{-1})$. Taking $b = 1$, from (2.7) we have the pivotal $\xi_i = \theta_i/(y_i + 1) \sim \mathrm{IG}(\eta + 1, 1)$. From (2.6) the MLE of η is $\hat{\eta} = p/\sum_{i=1}^p \log(y_i + 1)$ and (2.2) becomes $r(\hat{\eta}, \eta, \alpha) = 1 - D_{2(\eta+1)}(D^{-1}_{2(\hat{\eta}+1)}(1 - \alpha))$, where D_k is the χ^2 cdf with k degrees of freedom, k now necessarily an integer. For unconditional coverage we need the distribution of $\hat{\eta} \mid \eta$, which is $\mathrm{IG}(p, 1/(\eta p))$. We solve $R(\hat{\eta}, \alpha') = \alpha$ using a one-dimensional numerical integration [transforming the IG to the interval (0, 1) and using 16-point Gaussian integration; see Abramowitz and Stegun (1967)] with one root-finder (using false position). As an illustration, Figure 1 plots $\alpha'(\eta, \alpha)$ versus η for nominal upper and lower tail areas $\alpha = .01, .025, .05, .1$, with $p = 10$. For conditional coverage we need the conditional distribution of $\hat{\eta} \mid \mathbf{y}_1, \eta$. This may be obtained by routine transformation after noting that, given η, $\hat{\eta}$ and $a = \hat{\eta}\log(Y_1 + 1)/p$ are independent, the latter having a beta$(1, p - 1)$ distribution. We omit the details.

Example 2.2. We can extend Example 2.1 to the gamma/IG problem; that is, $Y_i \overset{\text{ind}}{\sim} \mathrm{gamma}(\nu_i, \theta_i)$, where ν_i are known and not necessarily all equal (for example, ν_i might be n_i) and $\theta_i \overset{\text{iid}}{\sim} \mathrm{IG}(\eta, b)$, $i = 1, \ldots, p$. Again we take $b = 1$. (Note that this case includes the χ^2 scale problem.) One can show that $Y_i \mid \eta \sim \Gamma(\nu_i + \eta)/(\Gamma(\nu_i)\Gamma(\eta)) \times y_i^{\nu_i-1}/(y_i + 1)^{\nu_i+\eta}$, a Pearson Type IV distribution (Johnson and Kotz 1970). Again $\xi_1 = \theta_1/(y_1 + 1)$ is a pivotal, which is now distributed as $\mathrm{IG}(\nu_1 + \eta, 1)$. Though the MLE $\hat{\eta}$ is no longer available in closed form, we can show

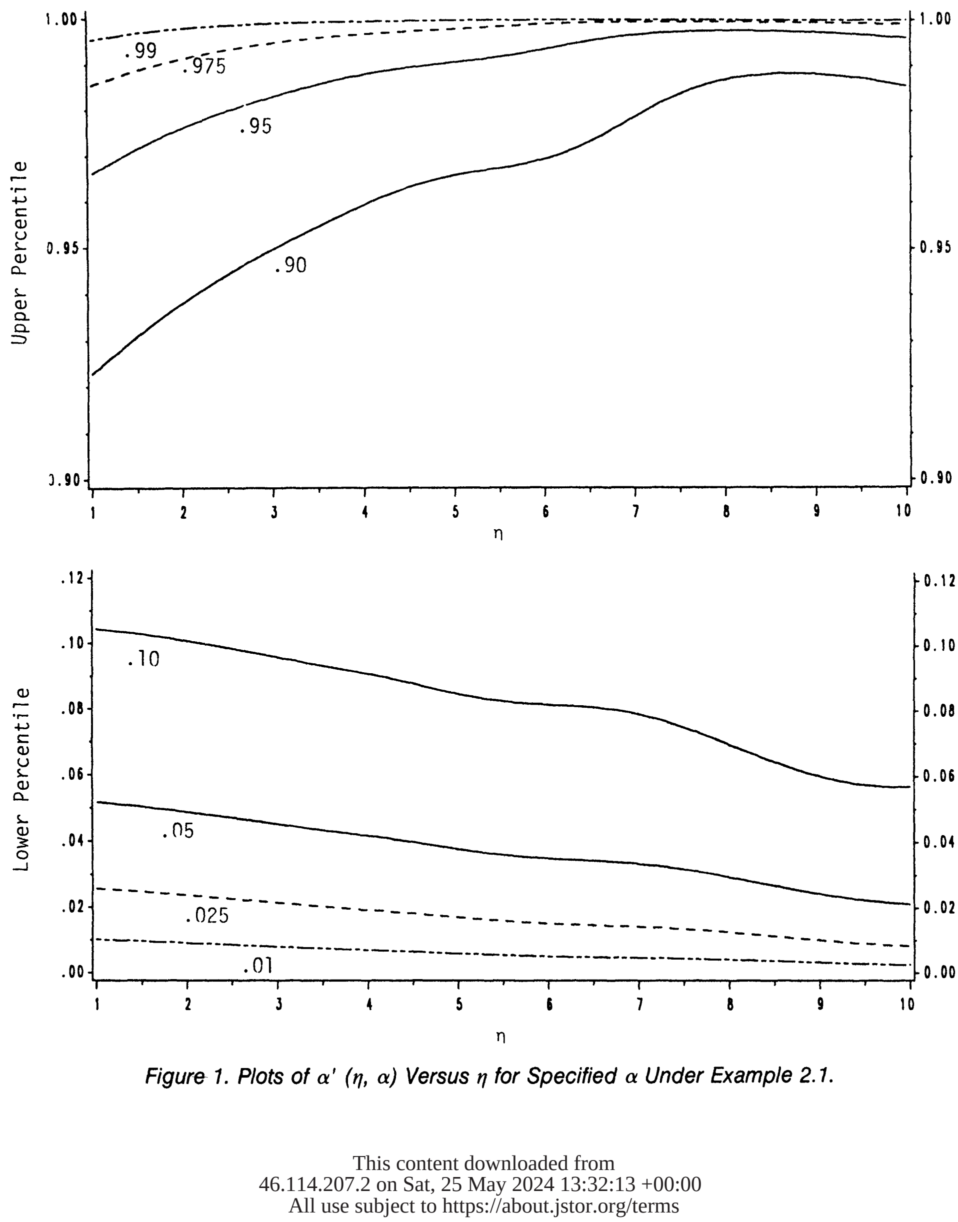

Figure 1. Plots of α' (η, α) Versus η for Specified α Under Example 2.1.

that $T(\hat{\eta}) = \Sigma_1^p \log(y_i + 1)$ is decreasing in $\hat{\eta}$; thus we can use $T(\hat{\eta})$ to implement bias correction.

Remark 1. With a pivotal, unconditional correction will automatically conditionally correct for bias, given any $T(\mathbf{Y})$ independent of $\hat{\eta}$, since integration over $\hat{\eta} \mid T, \eta$ is the same as over $\hat{\eta} \mid \eta$. This means that if $\hat{\eta}$ is chosen independent of Y_1, unconditional bias correction will achieve conditional bias correction given Y_1 (see Ex. 2.4). If $\hat{\eta}$ and Y_1 are not independent the pivotal is not helpful, since the integration in (2.3) is still with respect to $g(\hat{\eta} \mid y_1, \eta)$ even if r is free of y_1. Bias correction in the absence of a pivotal can often be implemented using a Type III parametric bootstrap procedure. In particular, Carlin and Gelfand (1989) considered the cases where the Y_i are binomial or Poisson random variables and the priors are conjugate. Nonconjugate priors could theoretically be handled using this bootstrap procedure, but the attendant computational burden becomes quite heavy.

Examples 2.3 and 2.4 offer a class of problems where (2.4) is free of η, as well as y_1. This means α' can be obtained from α without having to estimate η, and nominal unconditional coverage is exactly achieved. Empirical work in Section 4 shows that such unconditional intervals demonstrate good conditional behavior given Y_1 as well.

Example 2.3: The Normal/Normal Problem Where We Assume $n_i = 1$ for All i. We have $Y_i \overset{\text{ind}}{\sim} N(\theta_i, \sigma^2)$ and $\theta_i \overset{\text{iid}}{\sim} N(\mu, \tau^2)$, with $i = 1, \ldots, p$. Let σ^2 be known and equal to 1 w.l.o.g. Then

$$f(\theta_1 \mid y_1, \mu) = N(B\mu + (1 - B)y_1, 1 - B), \quad (2.8)$$

where $B = 1/(1 + \tau^2)$. If we assume τ^2 known, $\xi_1 = \theta_1 - (1 - B)Y_1$ is a pivotal distributed as $N(B\mu, 1 - B)$. If $q_\alpha(\mu)$ denotes the αth quantile of this distribution, $q_\alpha(\mu) = B\mu + (1 - B)^{1/2}\Phi^{-1}(\alpha)$ and (2.2) becomes

$$r(\hat{\mu}, \mu, \alpha) = \Phi\{B(\hat{\mu} - \mu)/\sqrt{1 - B} + \Phi^{-1}(\alpha)\}, \quad (2.9)$$

where $\hat{\mu} = \overline{Y}$. For EB coverage we integrate (2.9) with respect to the distribution of $\hat{\mu} \mid \mu$, which is $N(\mu, 1/Bp)$. Clearly the resulting R is free of μ; α' depends only on α. Hence exact unconditional bias correction can be achieved and exact EB coverage attained (see Cox 1975, sec. 6). Conditional bias correction requires integration with respect to the distribution of $\hat{\mu} \mid y_1, \mu$, which is $N((\mu(p - 1) + Y_1)/p, (p - 1)/Bp^2)$.

Alternatively, if we assume μ known but τ^2 (hence B) unknown, then no pivotal from (2.8) is possible. For conditional bias correction, assuming $\hat{B}$ is a function of $T = \Sigma_{i=1}^p (Y_i - \mu)^2$ we need the distribution of $T \mid Y_1, B$, which is immediate from the fact that $T - (Y_1 - \mu)^2 \mid y_1, B \sim B^{-1}\chi^2_{p-1}$ (see Case 1 in Sec. 4).

If both μ and τ^2 are assumed unknown, conditional bias correction requires the joint distribution of $\overline{Y}, \Sigma (Y_i - \overline{Y})^2 \mid Y_1, \mu, B$, which can be attacked through a Helmert transformation on $\mathbf{Y}$. If $\overline{Y}$ and $\Sigma (Y_i - \overline{Y})^2$ are based only on $Y_2, \ldots, Y_p$, matters are simpler.

Example 2.4. The previous example can be extended to the case of p simultaneous regressions. Let $Y_i \mid \theta_i \sim N(X_i\theta_i, \sigma_i^2)$, $i = 1, \ldots, p$, where Y_i is $n_i \times 1$, X_i is $n_i \times k$ full rank, and σ_i^2 is assumed known. In practice we would use an independent estimator of σ_i^2 based upon Y_i in what follows. When n_i is at least moderate there is evidence (Lawless 1981, pp. 463–464) that the resulting coverage will differ little from that with σ_i^2 known. Suppose $\theta_i \overset{\text{iid}}{\sim} N(\mu_\theta, \tau^2 I)$. This prior is perhaps most reasonable if the columns of the X_i are centered and scaled. For convenience we, in fact, assume that $X_i^T X_i = I_{k\times k}$. Routine calculation shows that $\theta_i \mid Y_i, \mu_\theta, \tau^2 \sim N(B_i\mu_\theta + (1 - B_i)X_i^T Y_i, \sigma_i^2(1 - B_i)I)$, where $B_i = \sigma_i^2/(\sigma_i^2 + \tau^2)$, and $\mathbf{Y} \mid \mu_\theta, \tau^2 \sim N(\mathbf{X}\mu_\theta, \Sigma_Y)$, where $\mathbf{X}^T = (X_1^T, \ldots, X_p^T)$ and Σ_Y is block diagonal with the ith block being $B_i/\sigma_i^2 \times I_{n_i\times n_i}$. If τ^2 is assumed known then $\xi_1 = \theta_1 - (1 - B_1)X_1^T Y_1$ is a pivotal having distribution $N(B_1\mu_\theta, \sigma_1^2(1 - B_1)I)$ and $\hat{\mu}_\theta = (X^T \Sigma_Y^{-1} X)^{-1}X^T \Sigma_Y^{-1} Y \sim N(\mu_\theta, p^{-1}I)$. The independence of the coordinates of ξ_1 combined with the argument at the beginning of Example 2.3 enables construction of a simultaneous k-dimensional confidence rectangle attaining exactly nominal EB coverage. A simultaneous EB confidence ellipsoid can be developed by noting that $\xi_1^T\xi_1 \sim \sigma_1^2(1 - B_1)\chi^2_{p,\lambda_1}$, where $\lambda_1 = (B_1^2\mu_\theta^T\mu_\theta)/(2\sigma_1^2(1 - B_1))$, and then bias correcting $r(\hat{\lambda}_1, \lambda_1, \alpha) = \Pr(\xi_1^T\xi_1 \le q_\alpha(\hat{\lambda}_1) \mid \xi_1^T\xi_1 \sim \sigma_1^2(1 - B_1)\chi^2_{p,\lambda_1})$. Conditional EB coverage could be attempted through the distribution of $\hat{\mu}_\theta \mid Y_1$; however, if $\hat{\mu}_\theta$ is calculated deleting Y_1, then by Remark 1, exact conditional EB coverage given Y_1 can be achieved. If τ^2 is assumed unknown matters become much more complicated. No pivotal exists, $\hat{\mu}_\theta$ and $\hat{\tau}^2$ will be unavailable in closed form unless all σ_i^2 are equal, and the conditional distribution of $\hat{\mu}_\theta, \hat{\tau}^2 \mid Y_1$ is intractable. A bootstrapping method will be the only feasible approach.

Does the conditional bias correction method actually produce approximate conditional coverage given Y_i? Again taking $i = 1$ to answer the question, we need to see how close the expectation

$$\begin{aligned} E_{\hat{\eta}\mid Y_1,\eta}&\Pr(\theta_1 \le q_{\alpha'(\hat{\eta},y_1,\alpha)}(y_1, \hat{\eta}) \mid \theta_1 \sim f(\theta_1 \mid y_1, \eta)) \\ &= E_{\hat{\eta}\mid Y_1,\eta}r(\hat{\eta}, \eta, y_1, \alpha'(\eta, y_1, \alpha)) \quad (2.10) \end{aligned}$$

is to α. Under usual conditions, since θ_1 is continuous, if $\hat{\eta}$ is a consistent estimator of η (as p tends to infinity) then (2.10) will converge to α. For fixed p, although exact evaluation of (2.10) is not possible, Theorem 2.1 (which follows) is encouraging, since it shows that in many cases (2.10) falls in an interval containing α.

Theorem 2.1. Suppose both $f(\theta_1 \mid y_1, \eta)$ and $g(\hat{\eta} \mid y_1, \eta)$ are stochastically ordered families in η for fixed y_1. Then the conditional expected "tail probability" (2.10) is bounded above by $\alpha + \max(I_1, I_2)$ and below by $\alpha + \min(I_1, I_2)$, where

$$\begin{aligned} I_1 = \int_{\hat{\eta}>\eta} &[\alpha'(\hat{\eta}, y_1, \alpha) \\ &- r(\hat{\eta}, \eta, y_1, \alpha'(\eta, y_1, \alpha))]g(\hat{\eta} \mid y_1, \eta)\, d\hat{\eta} \end{aligned}$$

and

$$I_2 = \int_{\hat{\eta}<\eta} [\alpha'(\hat{\eta}, y_1, \alpha) - r(\hat{\eta}, \eta, y_1, \alpha'(\eta, y_1, \alpha))] g(\hat{\eta} \mid y_1, \eta)\, d\hat{\eta}.$$

Proof. We prove the case where both $f(\theta_1 \mid y_1, \eta)$ and $g(\hat{\eta} \mid y_1, \eta)$ are stochastically increasing in η, with the proof for the other cases following similarly. Thus $q_\alpha(y_1, \eta) \uparrow \eta$ for fixed y_1; in fact, from (2.2), $r(\hat{\eta}, \eta, y_1, \alpha) \uparrow \hat{\eta}$ whereas $r(\hat{\eta}, \eta, y_1, \alpha) \downarrow \eta$. Since $g(\hat{\eta} \mid y_1, \eta)$ is stochastically increasing in η,

$$R(\eta, y_1, \alpha) = E_{\hat{\eta}|y_1,\eta} r(\hat{\eta}, \eta, y_1, \alpha) \uparrow \eta \quad (2.11)$$

(e.g., see Lehmann 1986, lemma 2, chap. 3). Also, the mild regularity condition of Lemma 2.1 insures that $R(\eta, y_1, \alpha) \uparrow \alpha$. Next, let $\eta_1 < \eta_2$, and consider for a specified α_0, $\alpha'(\eta_1, y_1, \alpha_0)$ and $\alpha'(\eta_1, y_1, \alpha_0)$ arising from $R(\eta_1, y_1, \alpha') = \alpha_0$ and $R(\eta_2, y_1, \alpha') = \alpha_0$, respectively. By (2.11), $R(\eta_1, y_1, \alpha)$ lies below $R(\eta_2, y_1, \alpha)$, whence $\alpha'(\eta_1, y_1, \alpha) > \alpha'(\eta_2, y_1, \alpha)$; that is, $\alpha'(\eta, y_1, \alpha) \downarrow \eta$. Thus if $\eta \le \hat{\eta}$,

$$r(\hat{\eta}, \hat{\eta}, y_1, \alpha'(\hat{\eta}, y_1, \alpha)) \le r(\hat{\eta}, \eta, y_1, \alpha'(\hat{\eta}, y_1, \alpha)) \le r(\hat{\eta}, \eta, y_1, \alpha'(\eta, y_1, \alpha)). \quad (2.12)$$

In addition, the inequalities in (2.12) are reversed if $\hat{\eta} \le \eta$. The left side of (2.12) equals $\alpha'(\hat{\eta}, y_1, \alpha)$ and thus is decreasing in $\hat{\eta}$; the right side of (2.12) is increasing in $\hat{\eta}$; however, we cannot conclude monotonicity for $r(\hat{\eta}, \eta, y_1, \alpha'(\hat{\eta}, y_1, \alpha))$. Figure 2 offers a generic view of the situation. Finally, since $E_{\hat{\eta}|y_1,\eta} r(\hat{\eta}, \eta, y_1, \alpha'(\eta, y_1, \alpha)) = R(\eta, y_1, \alpha'(\eta, y_1, \alpha)) = \alpha$ by definition, the bounds in the theorem follow.

Remark 2. From Figure 2 we see that $I_1 \times I_2 < 0$, whence (2.10) falls in an interval containing α.

Remark 3. The fact that Y_1 (or a function of Y_1) enters directly into the posterior (hence into all of our subsequent expressions) makes qualitative examination of conditional coverage of our bias-correction method given Y_1 straightforward. Analytic examination of conditional coverage given other characteristics of the data does not seem promising, except in cases where a pivotal exists, as in Remark 1.

If η is of dimension k then calculation of R requires a k-dimensional numerical integration and the solution of (2.5) requires a root-finding algorithm. A possible alternative to the numerical integration is to utilize the approach of Cox (1975), who suggested expansion of $r(\hat{\eta}, \eta, y_1, \alpha)$ in $\hat{\eta}$ about η; that is, $r(\hat{\eta}, \eta, y_1, \alpha) \approx r(\eta, \eta, y_1, \alpha) + (\hat{\eta} - \eta)^T \nabla_r(\eta) + 1/2(\hat{\eta} - \eta)^T H_r(\eta)(\hat{\eta} - \eta)$, where $(\nabla_r(\eta))_i = (\partial r/\partial \hat{\eta}_i) \mid \eta$ and $(H_r(\eta))_{ij} = (\partial^2 r/\partial \hat{\eta}_i\, \partial \hat{\eta}_j) \mid_\eta$, whence

$$R(\eta, y_1, \alpha) \approx \alpha + E_{\hat{\eta}|Y_1,\eta}(\hat{\eta} - \eta)^T \nabla_r(\eta) + 1/2\, \mathrm{tr}[H_r(\eta) \times E_{\hat{\eta}|Y_1,\eta}(\hat{\eta} - \eta)(\hat{\eta} - \eta)^T]. \quad (2.13)$$

Denoting the right side of (2.13) by $R'(\eta, y_1, \alpha)$, analogous to (2.5) we may solve $R'(\hat{\eta}, y_1, \alpha') = \alpha$ for α'. Note that even if $g(\hat{\eta} \mid y_1, \eta)$ is a standard distribution so that $E_{\hat{\eta}|y_1,\eta}(\hat{\eta})$ and $\Sigma_{\hat{\eta}|y_1,\eta}$ are readily available, (2.13) still requires the evaluation of $2k + \binom{k}{2}$ numerical derivatives.

3. THE MARGINAL POSTERIOR APPROACH

In the parametric EB setting several authors have attempted to account for the variation in estimating the hyperparameter η by introducing a hyperprior distribution on η. Corresponding quantiles of the resulting *marginal posterior* are used in place of those of the estimated posterior. As a mixture of posteriors, this marginal posterior typically has more spread than the estimated posterior, so intervals longer than the naive ones result. This section is intended to illuminate this marginal-posterior approach.

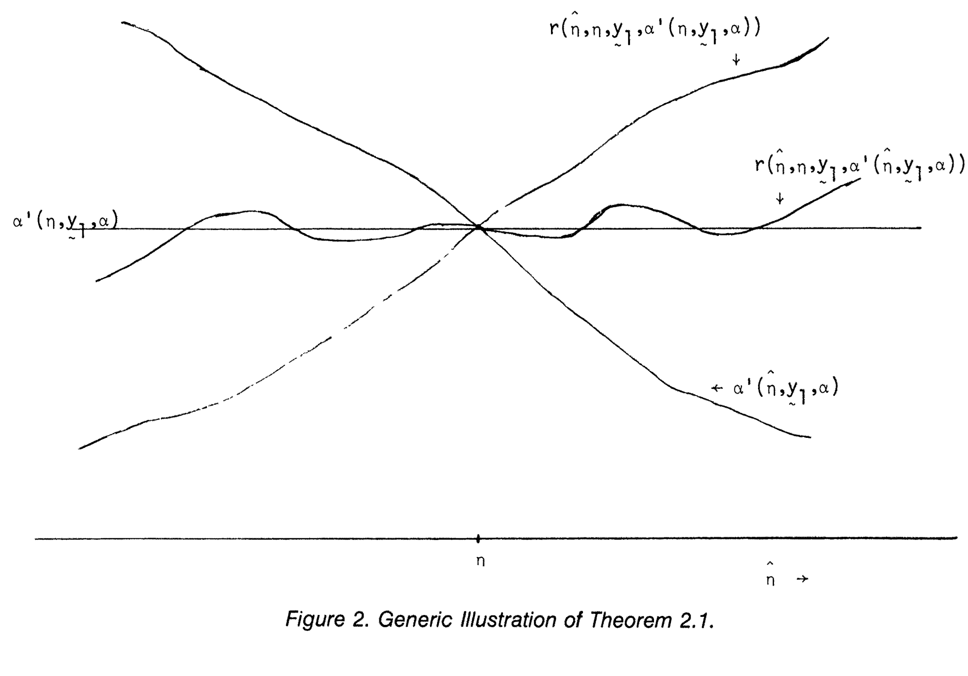

Figure 2. Generic Illustration of Theorem 2.1.

To formalize the setup we again confine ourselves to the exchangeable case using the notation of Section 1. Suppose $\hat{\eta}(\mathbf{Y})$ is an estimator of η that is sufficient for the marginal family $m(\mathbf{Y} \mid \eta)$ and has density $\rho(\hat{\eta} \mid \eta)$ with respect to Lebesgue measure. Let $\tau(\eta)$ be a continuous hyperprior on η, which induces the conditional distribution $h(\eta \mid \hat{\eta}) \propto \rho(\hat{\eta} \mid \eta) \times \tau(\eta)$, which in turn induces the marginal posterior for θ_i:

$$l_h(\theta_i \mid y_i, \hat{\eta}) = \int f(\theta_i \mid y_i, \eta) h(\eta \mid \hat{\eta})\, d\eta. \tag{3.1}$$

We use a subscript h to indicate which mixing distribution was used with the posterior. The naive intervals (1.3) and (1.4) would be replaced with corresponding lower and upper points of l_h. Hence coverage in the sense of (1.5) or (1.6) will vary with the specification of τ or, equivalently, h. This pure Bayesian approach is less targeted at achieving specified EB coverage than that of Section 2. For example, there is no obvious relationship between using a vague hyperprior and achieving nominal EB coverage through the resulting (3.1). In fact, Laird and Louis (1987) were empirically successful in the normal/normal problem (Ex. 2.3) with known prior mean and unknown prior variance using l_ρ [i.e., mixing with respect to $\rho(\eta \mid \hat{\eta})$, the sampling density with η and $\hat{\eta}$ exchanged]. The key issue (a non-Bayesian one) concerns the existence and nature of an h that will be successful in achieving nominal EB coverage. (We defer a rough discussion of this issue until the end of the section.) For instance, if the naive EB confidence interval is too long (as in Case 1 in Sec. 4) this approach seems doomed to failure; we need to *correct,* not lengthen.

When ρ is available in closed form the numerical integration in (3.1) can be carried out directly (Deely and Lindley 1981; Rubin 1982). Morris (1987) suggested approximating l_h using the member of the posterior family f whose first two moments agree with those of l_h. Laird and Louis (1987) suggested approximating (3.1) by the use of a Type III parametric bootstrap. That is, given $\hat{\eta}$, draw θ_i^* from $\pi(\theta \mid \hat{\eta})$. Then draw Y_i^* from $f(y \mid \theta_i^*)$, and finally calculate $\eta^* = \hat{\eta}(\mathbf{Y}^*)$. Repeating this process N times, we obtain η_j^* $(j = 1, \ldots, N)$ distributed as $\rho(\cdot \mid \hat{\eta})$. The discrete mixture distribution

$$\sum_{j=1}^{N} f(\theta_i \mid y_i, \eta_j^*)/N \tag{3.2}$$

is taken as the estimator of (3.1), and quantiles of (3.2), obtained by a root finder, are used instead of those of (3.1).

Note that (3.2) is an unbiased estimator of l_ρ and converges almost surely to l_ρ as $N \to \infty$, leading to criticism of its use in the comments following the Laird and Louis article. But if the objective is EB coverage, l_ρ [or an estimate of it, such as (3.2)] may be as good as l_h. An important point is that since ρ (hence l_ρ) changes as $\hat{\eta}$ changes, the performance of the Laird and Louis approach can be quite sensitive to the choice of $\hat{\eta}$ [see Ex. 3.2 (which follows) and Table 1 (Sec. 4)]. The empirical success of (3.2) suggests that for the examples of which it has been applied, with a good choice of $\hat{\eta}$, $\rho(\cdot \mid \hat{\eta})$ is a good choice of h. For any 1–1 onto transformation of η given $\hat{\eta}$, $s_{\hat{\eta}}(\eta)$, having density ψ, the Type III parametric bootstrap enables estimation of l_ψ by $\sum_{j=1}^N f(\theta_i \mid y_i, s_{\hat{\eta}}(\eta_j^*))/N$ analogous to (3.2). There may exist a choice of $s_{\hat{\eta}}(\cdot)$ such that ψ "matches" h; that is, $l_\psi = l_h$. This extension is attractive in that, like (3.2), it does not require that ρ be given in closed form.

If ρ *is* available in closed form then for any τ the Type III parametric bootstrap provides an importance-sampling Monte Carlo integration (Geweke 1988; Hammersley and Handscomb 1964) of (3.1) of the form

$$\sum_{j=1}^{N} f(\theta_i \mid y_i, \eta_j^*) k_{\hat{\eta}}(\eta_j^*) \Big/ \sum_{j=1}^{N} k_{\hat{\eta}}(\eta_j^*), \tag{3.3}$$

where $k_{\hat{\eta}}(\cdot) = \rho(\hat{\eta} \mid \cdot)\tau(\cdot)/\rho(\cdot \mid \hat{\eta})$. Note that the standardizing constant for h is not required. Implementation of the marginal-posterior approach for a specified τ in the absence of a closed form for ρ is unclear. We consider earlier examples in this context.

Example 3.1. Consider the normal/normal example (Ex. 2.3). Assume μ unknown but B known. The sampling distribution $\rho(\hat{\mu} \mid \mu)$ is $N(\mu, 1/(Bp))$. For a flat hyperprior τ, $h(\mu \mid \hat{\mu})$ is $N(\hat{\mu}, 1/(Bp))$. Hence $\psi = h$ for $s_{\hat{\mu}}(\mu) = \mu$ (Laird and Louis 1987, theorem 1). If we assume B unknown as well, theorem 2 of Laird and Louis shows that no choice of $s_{\hat{\mu}}$ will produce $\psi = h$.

Example 3.2. Consider again the exponential/inverse gamma example (Ex. 2.1). Recall that the sampling distribution $\rho(\hat{\eta} \mid \eta)$ is $\mathrm{IG}(p, 1/(\eta p))$. Then the hyperprior associated with l_ρ is neither simple nor natural. Under the flat hyperprior $\tau_1(\eta) = 1$, $\eta > 0$, $h_1(\eta \mid \hat{\eta})$ is gamma$(p + 1, \hat{\eta}/p)$, and there is no obvious choice of $s_{\hat{\eta}}$ having distribution h_1, but we can use (3.3) to "match" (3.1). Under the hyperprior $\tau_2(\eta) = \eta^{-1}$, $\eta > 0$, $h_2(\eta \mid \hat{\eta})$ is gamma$(p, \hat{\eta}/p)$, and $s_{\hat{\eta}}(\eta) = \hat{\eta}^2/\eta$ does have density exactly h_2. Pepple (1988) placed a flat hyperprior on $1/\eta$ but then approximated the resulting marginal posterior by a gamma distribution whose first two moments agree with those of the exact l_h.

We return to the question of when l_h may be expected to give approximate nominal EB coverage. For any marginal posterior [such as those in (3.1)–(3.3)], let $C_\alpha^{(i)}(Y_i, \hat{\eta})$ be a $1 - \alpha$ posterior (Bayes) credible set for θ_i; that is, $P_{\theta_i \mid y_i, \eta}(\theta_i \in C_\alpha^{(i)}(Y_i, \hat{\eta})) = 1 - \alpha$. Let $I(\theta_i, Y_i, \hat{\eta})$ equal 1 if $(\theta_i, Y_i, \hat{\eta})$ are such that $\theta_i \in C_\alpha^{(i)}(Y_i, \hat{\eta})$, and let it equal 0 otherwise. Then provided the distribution of $\eta \mid y_i$ is proper, $E_{\eta \mid y_i}[P_\eta(\theta_i \in C_\alpha^{(i)}(Y_i, \hat{\eta}) \mid y_i)] = E_{\theta_i, \hat{\eta} \mid y_i} I(\theta_i, y_i, \hat{\eta}) = E_{\hat{\eta} \mid y_i} P_{\theta_i \mid y_i, \hat{\eta}}(\theta_i \in C_\alpha^{(i)}(Y_i, \hat{\eta})) = E_{\hat{\eta} \mid y_i}(1 - \alpha) = 1 - \alpha$.

Thus for any l_h such that the distribution of $\eta \mid y_i$ is proper, on average (over $\eta \mid y_i$) $C_\alpha^{(i)}(Y_i, \hat{\eta})$ meets (1.6); it provides conditional EB coverage given Y_i. A *good* l_h, however, requires that $P_\eta(\theta_i \in C_\alpha^{(i)}(Y_i, \hat{\eta}) \mid y_i) \approx 1 - \alpha$ for *each* η. To address this more demanding issue, consider the following rough argument [motivated by Laird (1988, personal communication)], which provides insight in the case where a pivotal exists. Dropping Y_i in (3.1) and replacing θ_i by ξ_i, let $q_\alpha^{(h)}(\eta)$ denote the αth quantile of $l_h(\xi_i \mid \eta)$, and let $q_\alpha(\eta)$ denote the αth quantile of the true distribution of $\xi_i \mid \eta$ [obtained from $f(\theta_i \mid y_i, \eta)$]. Defining $r^{(h)}(\hat{\eta}, \eta, \alpha) = P_{\xi_i|\eta}(\xi_i \le q_\alpha^{(h)}(\hat{\eta}))$, we show when the expectation of $r^{(h)}$ over $\hat{\eta} \mid \eta$ will fall in an interval containing α. Since mixing by h will typically "spread out" the posterior (hence the distribution of $\xi_i \mid \eta$), we assume that for α small (near 0), $q_\alpha^{(h)}(\eta) < q_\alpha(\eta)$ whereas for α large (near 1), $q_\alpha^{(h)}(\eta) > q_\alpha(\eta)$. Suppose additionally that h is such that for α small $r^{(h)}$ is approximately convex in $\hat{\eta}$ whereas for α large $r^{(h)}$ is approximately concave in $\hat{\eta}$. (We argue when this might be the case subsequently.) Finally, let $\hat{\eta}$ be unbiased for η. Then for α small,

$$\begin{aligned} r^{(h)}(\eta, \eta, \alpha) &\le E_{\hat{\eta}|\eta} r^{(h)}(\hat{\eta}, \eta, \alpha) \\ &\le E_{\hat{\eta}|\eta} r(\hat{\eta}, \eta, \alpha) \equiv R(\eta, \alpha), \end{aligned} \quad (3.4)$$

where r and R are as in (2.2) and (2.3) with y_1 deleted because of the pivotal. But also,

$$r^{(h)}(\eta, \eta, \alpha) \le r(\eta, \eta, \alpha) = \alpha \le R(\eta, \alpha). \quad (3.5)$$

The last inequality in (3.5) usually holds because, when α is small, α' such that $\alpha = R(\eta, \alpha')$ is usually less than α and $R(\eta, \alpha) \uparrow \alpha$ by Lemma 2.1. Together, (3.4) and (3.5) suggest that for α small $E_{\hat{\eta}|\eta} r^{(h)}$ will be close to α. For α large our assumptions reverse the inequalities in (3.4) and (3.5) and thus a similar conclusion holds.

To return to the question of the convexity or concavity of $r^{(h)}$, suppose the distribution of $\xi_i \mid \eta$ is unimodal. Then the cdf of $\xi_i \mid \eta$ will be an increasing convex (concave) function below (above) the mode. Hence if h is such that $q_\alpha^{(h)}(\eta)$ is approximately convex (concave) in η for α small (large), then $r^{(h)}$ will be approximately convex (concave) in $\hat{\eta}$ for α small (large). We recall that under families stochastically ordered in η, $q_\alpha^{(h)}(\eta)$ will be monotone in η. Using the definition of $q_\alpha^{(h)}(\eta)$, implicit differentiation enables an expression for its second derivative. We omit details.

4. SIMULATED COVERAGE PROBABILITIES AND INTERVAL LENGTHS

In this section we present the results of simulation studies comparing the methods discussed in the previous two sections. We first offer results for the bias-corrected naive (BCN) method, then some limited results for the marginal posterior method. Finally, we give the unconditional EB coverages for both methods in a unifying example.

Case 1

First, we illustrate the BCN method's ability to achieve conditional EB coverage regardless of the length of the naive intervals using the normal/normal problem of Example 2.3. We assume (as do Laird and Louis in their numerical work) that the prior mean μ is known and equal to 0 w.l.o.g., but that the prior variance τ^2 is unknown. To implement bias correction given Y_1 we use $\hat{B} = p/(p + \sum_{i=1}^{p} Y_i^2)$ (Raghunathan 1987). This estimator of B is smooth with distribution having support (0, 1), unlike the MLE, MVUE, or truncated versions of them proposed by Morris (1983b) and Laird and Louis (1987). We then obtain $\alpha'(\hat{B}, y_1, \alpha)$, and compare intervals based on this bias correction with the naive EB interval (1.3) and the classical frequentist interval [simply $Y_1 \pm \Phi^{-1}(\alpha)$ in this case]. We took $B = .5$, $p = 10$, and the nominal $\gamma = .90$, and used 5,000 replications.

Figure 3 shows the resulting simulated coverage probability of these three intervals for θ_1 conditional on y_1. The

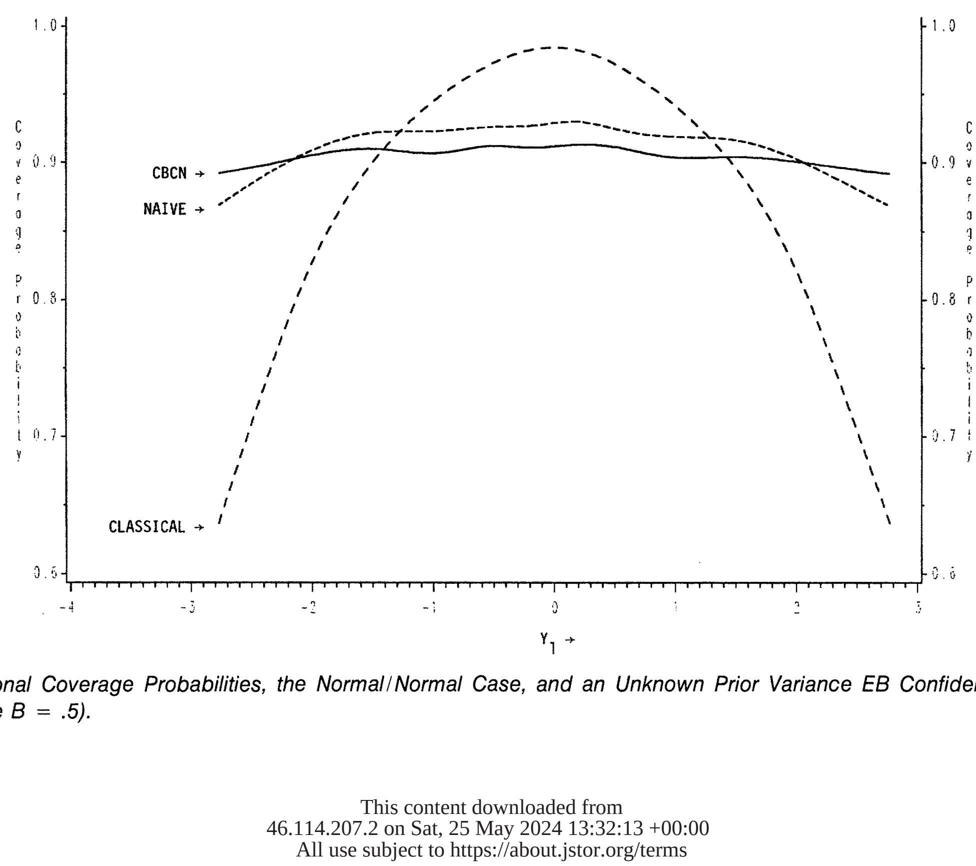

Figure 3. Conditional Coverage Probabilities, the Normal/Normal Case, and an Unknown Prior Variance EB Confidence Interval for θ_1 at Nominal γ = .9 (true B = .5).

points plotted range from the .025 to the .975 percentile points of Y_1's unconditional distribution, which in this case is $N(0, 2)$. Note that the classical method's conditional behavior is conservative for central Y_1's but very poor in the tails. The unusual aspect of this example is the pattern of lengths and conditional coverage of the naive EB intervals—too short and below the nominal level in the tails of Y_1's distribution, too long and well above the nominal level in the middle. This is a result of the bias in our estimator $\hat{B}$. The conditional BCN (CBCN) method gives intervals that flatten out this pattern over Y_1's distribution. In addition, the simulated CBCN intervals were uniformly shorter than the inappropriately centered naive ones. They also had nearly constant average lengths, ranging from about 82% as long as classical in the tails of Y_1's distribution to about 75% as long as classical in the middle of the distribution. Of course, the fact that the CBCN method achieves conditional EB coverage over Y_1's distribution implies unconditional EB coverage overall.

Case 2

As a second example of the BCN method, consider the regression problem introduced in Example 2.4. For illustrative purposes we consider simple linear regression with $\theta_i = (\alpha_i, \beta_i)^T$, assuming $p = 5$ simultaneous regressions, each having only $n_i = 5$ observations. For convenience we take the X_{ij} equally spaced, centered and scaled for each i. Let both the model variance σ^2 and the prior variance τ^2 be known and equal to 1 w.l.o.g. Since in this case a pivotal exists and α' is a function only of α, unconditional bias correction (UBCN) produces exactly unconditional EB coverage, (1.5). An exact CBCN given Y_i may also be implemented, since by Remark 1 and Example 2.4, if we choose an independent, unbiased estimate of μ_θ, we can again find α' as a function only of α.

Since our design makes the slope β_1 and the intercept α_1 independent in the posterior family, we may obtain bias-corrected intervals for them separately. Taking the true values of the hyperparameters to be $\mu_\alpha = 0$ and $\mu_\beta = 1$, and again using 5,000 replications, Figure 4 shows simulated coverage probabilities conditional on $\hat{\beta}_1 = \sum_{j=1}^5 X_{1j}Y_{1j}$ for the classical, naive EB, UBCN, and CBCN intervals for β_1. The points plotted cover ± 3 standard deviations of the unconditional distribution of $\hat{\beta}_1$, which is $N(1, 2)$. Again note the very poor conditional behavior of the classical method, and the poor conditional and unconditional behavior of the naive method. Of course the UBCN method guarantees nominal unconditional behavior, but also exhibits good conditional behavior in this case. The CBCN method's behavior is perfect as advertised, its curve being completely flat at $\gamma = .90$ to the accuracy of the simulation (standard error $\approx .004$). The fact that we assumed all variances known means that all the methods have a constant interval length. In this example the lengths are classical, 3.29; naive EB, 2.33; UBCN, 2.55; and CBCN, 2.60. We can similarly exactly bias correct a simultaneous EB confidence rectangle for (α_1, β_1), conditional on both $\hat{\alpha}_1 = \sum_{j=1}^5 Y_{1j}/5$ and $\hat{\beta}_1$, and thus unconditionally.

Case 3

To shed light on the question raised in Section 3 of a good choice of marginal posterior, we return to the exponential/inverse gamma case of Example 3.2. We compare the sensitivity of the achieved EB coverage probabilities to the choice of $\hat{\eta}$ using the Laird and Louis bootstrap, the τ_1 (flat hyperprior) matching bootstrap, and the τ_2 matching bootstrap methods. From the discussion in Example 2.1 if $\hat{\eta}_c = c/\sum_{i=1}^p \log(Y_i + 1)$, appropriate choices for c include p (MLE), $p - 1$ (UMVUE), and $p + 1$ (best invariant under suitable squared error loss). Choice of $\hat{\eta}_c$ affects the scale parameter of the sampling

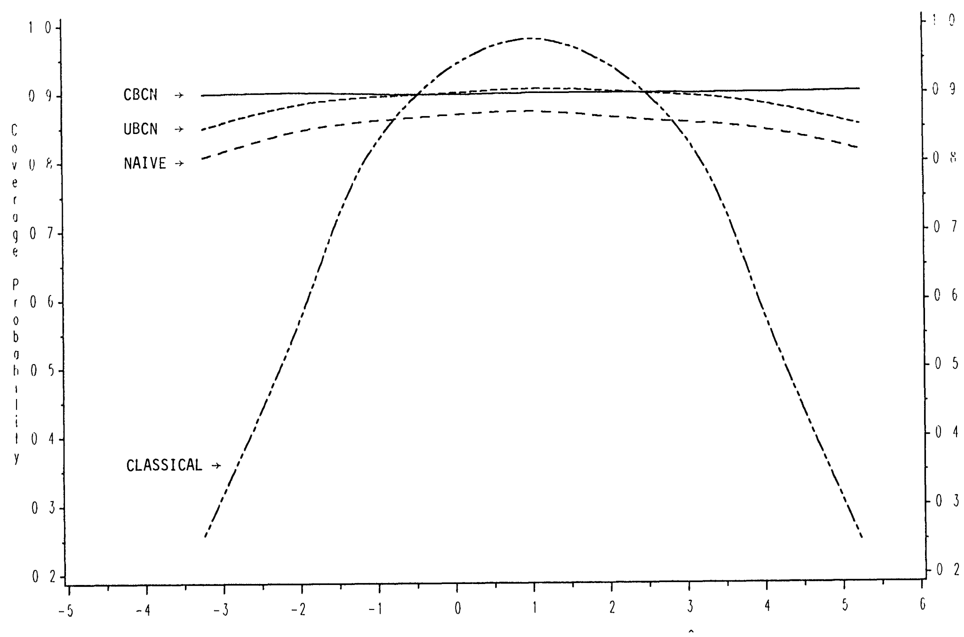

Figure 4. Conditional Coverage Probabilities and the Simple Linear Regression Case Individual EB Confidence Interval for β_1 at Nominal γ = .9 (true μ_β = 1).

density for drawing the bootstrap $\hat{\eta}^*$'s. We ran a simulation of 3,000 replications ($N = 400$ bootstrap observations per replication) with $\eta = 2$, $p = 5$, and nominal $\gamma = .95$ to compare these three methods over the three choices of $\hat{\eta}$. The results are summarized in Table 1, which shows achieved EB coverage probability, with interval length in parentheses. The Laird and Louis bootstrap is extremely sensitive to choice of $\hat{\eta}$; the τ_1 matching bootstrap is stable but fails to achieve nominal coverage probability; the τ_2 matching bootstrap is stable with respect to choice of $\hat{\eta}$ and achieves nominal coverage.

Case 4

Finally, we compare all the methods discussed in the context of the exponential/inverse gamma problems of Examples 2.1 and 3.2. For fixed η and p, we generated θ_i's iid as IG(η, 1), and then generated the Y_i's independently as $\exp(\theta_i)$, $i = 1, \ldots, p$. Each simulation is again based on 3,000 replications; for the methods requiring a bootstrap, we again used $N = 400$ bootstrap trials per replication.

Table 2 shows lower endpoint, upper endpoint, interval length, and unconditional EB coverage probability (all averaged over both i and the replications) for the classical, naive EB, UBCN, Laird and Louis bootstrap, τ_1 matching bootstrap, and τ_2 matching bootstrap methods for $p = 5$; true $\eta = 2$, 5; and nominal individual coverage probabilities $\gamma = .90$ and .95. The bias-corrected method is affected by the choice of $\hat{\eta}$ in three places: the computation of the R function (2.3) (we need the distribution of $\hat{\eta} \mid \eta$), solving (2.5), and the estimated posterior distribution. In our simulation, for the naive and bias-corrected naive methods we show results obtained using the marginal UMVUE. Results (not shown) obtained using the marginal MLE gave longer (i.e., too conservative) bias-corrected intervals (extending farther to the right), but shorter naive intervals. For the three bootstrap methods, we also used the UMVUE for $\hat{\eta}$, since from Table 1 this is the best choice for the Laird and Louis method, the only bootstrap sensitive to this choice. Recall also that unbiasedness is assumed in our rough argument at the end of Section 3.

Several points can be made from Table 2. As expected, the classical intervals faithfully achieve the desired coverages, but are quite long compared to the better EB intervals. The naive EB intervals fail to achieve nominal coverage and are very poor for large η with our small p. The bias-corrected naive intervals, on the other hand, achieve the desired nominal coverage to the accuracy of the table (the coverage probabilities have a standard error of about .005). The Laird and Louis and τ_2 matching bootstrap intervals generally achieve the desired coverage, yet the latter are substantially shorter. The intervals based on matching the flat hyperprior τ_1 are shifted to the left of those based on τ_2 and generally fail to achieve the desired coverage probability; apparently this hyperprior is putting too much weight on large values of η.

Table 1. Comparison of Marginal Posterior Methods: $\gamma = .95$

Estimator of η	Laird and Louis	τ_1 matching	τ_2 matching
UMVUE	.954 (7.50)	.930 (4.51)	.951 (5.66)
MLE	.931 (4.50)	.928 (4.33)	.950 (5.61)
Best invariant	.865 (2.76)	.926 (4.14)	.948 (5.40)

Table 2. Comparison of Bias-Corrected and Marginal Posterior Methods: $p = 5$

Interval method	Average lower endpoint	Average upper endpoint	Average interval length	Average unconditional coverage probability
$\eta = 2, \gamma = .90$				
Classical	.335	19.5	19.2	.901
Naive EB	.355	3.87	3.51	.839
Bias-corrected	.331	4.74	4.41	.897
Laird and Louis	.339	5.15	4.81	.904
τ_1 matching	.287	3.23	2.95	.868
τ_2 matching	.311	4.00	3.69	.894
$\eta = 2, \gamma = .95$				
Classical	.268	39.1	38.8	.952
Naive EB	.306	5.53	5.22	.900
Bias-corrected	.285	7.84	7.55	.952
Laird and Louis	.283	7.79	7.50	.954
τ_1 matching	.246	4.46	4.51	.930
τ_2 matching	.265	5.93	5.66	.951
$\eta = 5, \gamma = .90$				
Classical	.084	4.89	4.81	.899
Naive EB	.134	.690	.556	.771
Bias-corrected	.116	1.03	.914	.902
Laird and Louis	.114	1.04	.928	.899
τ_1 matching	.092	.620	.528	.863
τ_2 matching	.102	.810	.708	.901
$\eta = 5, \gamma = .95$				
Classical	.068	9.87	9.81	.948
Naive EB	.120	.859	.739	.846
Bias-corrected	.103	1.67	1.57	.956
Laird and Louis	.096	1.41	1.31	.951
τ_1 matching	.081	.816	.735	.918
τ_2 matching	.089	1.10	1.01	.947

5. CONCLUSION

In this article we have developed a general method to conditionally correct the bias in naive empirical Bayes confidence intervals. We have also attempted to clarify and expand on the idea of using bootstrap observations to accomplish a marginal posterior Bayes solution. We conclude that the bias-correction method is attractive because of its general applicability, straightforward implementation, and direct attack on the deficiencies of the naive EB interval. The marginal posterior approach can also be quite successful, although the choice of a good mixing distribution h (equivalently, a good hyperprior τ) is critical and might require preliminary investigation. Furthermore, implementation of this approach for a given τ in the absence of a closed form for ρ, the sampling density of $\hat{\eta}$, is not clear.

[*Received February 1988. Revised June 1989.*]

REFERENCES

Abramowitz, M., and Stegun, I. (1967), *Handbook of Mathematical Functions With Formulas, Graphs and Mathematical Tables* (Applied Mathematics Series, No. 55), Washington, DC: National Bureau of Standards.

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

Carlin, B. P., and Gelfand, A. E. (1989), "A Sample Reuse Method for Accurate Parametric Empirical Bayes Confidence Intervals," Technical Report 89-1, University of Connecticut, Dept. of Statistics.

Cox, D. R. (1975), "Prediction Intervals and Empirical Bayes Confidence Intervals," in *Perspectives in Probability and Statistics,* ed. J. Gani, New York: Academic Press, pp. 47–55.

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

Deely, J. J., and Lindley, D. V. (1981), "Bayes Empirical Bayes," *Journal of the American Statistical Association,* 76, 833–841.

Efron, B. (1987), Comment on "Empirical Bayes Confidence Intervals Based on Bootstrap Samples," by N. M. Laird and T. A. Louis, *Journal of the American Statistical Association,* 82, 754.

Geweke, J. (1988), "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica,* 57, 1317–1340.

Hammersley, J. M., and Handscomb, D. C. (1964), *Monte Carlo Methods,* London: Methuen.

James, W., and Stein, C. (1961), "Estimation With Quadratic Loss," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1), Berkeley: University of California Press, pp. 361–379.

Johnson, N. L., and Kotz, S. (1970), *Distributions in Statistics: Continuous Univariate Distributions—2,* New York: John Wiley.

Laird, N. M., and Louis, T. A. (1987), "Empirical Bayes Confidence Intervals Based on Bootstrap Samples," *Journal of the American Statistical Association,* 82, 739–750.

Lawless, J. F. (1981), "Mean Squared Error Properties of Generalized Ridge Estimators," *Journal of the American Statistical Association,* 76, 462–466.

Lehmann, E. L. (1986), *Testing Statistical Hypotheses* (2nd ed.), New York: John Wiley.

Morris, C. N. (1983a), "Parametric Empirical Bayes Inference: Theory and Applications," *Journal of the American Statistical Association,* 78, 47–59.

——— (1983b), "Parametric Empirical Bayes Confidence Intervals," in *Scientific Inference, Data Analysis, and Robustness,* New York: Academic Press, pp. 25–50.

——— (1987), "Determining the Accuracy of Bayesian Empirical Bayes Estimates in the Familiar Exponential Families," Technical Report 46, University of Texas, Center for Statistical Sciences.

Pepple, P. (1988), "Simultaneous Estimation of Exponential Means Using Bayesian Methods," *Communications in Statistics,* 17, 357–374.

Raghunathan, T. E. (1987), "Bootstrap Bayes," technical report, University of Washington, Dept. of Biostatistics.

Robbins, H. (1983), "Some Thoughts on Empirical Bayes Estimation," *The Annals of Statistics,* 11, 713–723.

Rubin, D. B. (1982), "Estimation in Parallel Randomized Experiments," *Journal of Educational Statistics,* 6, 377–401.

——— (1984), "Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician," *The Annals of Statistics,* 12, 1151–1172.