AD-A216 198

A SAMPLE REUSE METHOD FOR ACCURATE PARAMETRIC EMPIRICAL BAYES CONFIDENCE INTERVALS

BY

BRADLEY P. CARLIN and ALAN E. GELFAND

TECHNICAL REPORT NO. 423
NOVEMBER 7, 1989

PREPARED UNDER CONTRACT
N00014-89-J-1627 (NR-042-267)
FOR THE OFFICE OF NAVAL RESEARCH

Reproduction in Whole or in Part is Permitted
for any purpose of the United States Government

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

DTIC
ELECTE
DEC 29 1989
S B D

89 12 29 051

(4)

A SAMPLE REUSE METHOD FOR ACCURATE PARAMETRIC EMPIRICAL BAYES CONFIDENCE INTERVALS

BY

BRADLEY P. CARLIN and ALAN E. GELFAND

TECHNICAL REPORT NO. 423
NOVEMBER 7, 1989

Prepared Under Contract
N00014-89-J-1627 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director

Reproduction in Whole or in Part is Permitted
for any purpose of the United States Government

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

DTIC
ELECTE
DEC 29 1989
S B D

A SAMPLE REUSE METHOD FOR ACCURATE PARAMETRIC EMPIRICAL BAYES CONFIDENCE INTERVALS

Bradley P. Carlin and Alan E. Gelfand

ABSTRACT

Parametric empirical Bayes methods of point estimation for a vector of unknown parameters date to the landmark paper of James and Stein (1961). The usual approach is to use the mean of the estimated posterior distribution of each parameter, where the estimation of the prior parameters ("hyperparameters") is accomplished through the marginal distribution of the data. While point estimates computed this way usually perform well, interval estimates based on the estimated posterior (called "naive" EB intervals) are not. They fail to account for the variability in the estimation of the hyperparameters, generally resulting in sub-nominal coverage probability in the "EB" sense defined in Morris (1983a).

In this paper we extend the work of Carlin and Gelfand (1989), who proposed a conditional bias correction method for developing EB intervals which corrects the deficiencies in the naive intervals. We show how bias correction can be implemented in general via a Type III parametric boostrap procedure, a sample reuse method first employed by Laird and Louis (1987). Theoretical and simulation results indicate that intervals which are accurate with respect to nominal coverage ensue. We give two specific applications (to binomial test data and Poisson failure rate data) where we compute simultaneous point and bias corrected interval estimates.

KEY WORDS: Confidence interval; parametric empirical Bayes; parametric bootstrap; bias correction; conditional calibration.

1. INTRODUCTION

In this paper we consider the problem of multiparameter interval estimation in the parametric empirical Bayes (PEB) framework. We consider the familiar exchangeable model, where at the first stage, given θ_i the data vectors $\underset{\sim}{Y}_i$ are independently distributed as $f_i(\underset{\sim}{y}_i|\theta_i)$, $i = 1, \ldots, p$. At the second stage, the θ_i are supposed i.i.d. with distribution $\pi(\theta|\eta)$ over Θ, where η indexes the family π. The marginal distribution of $\underset{\sim}{Y}_i$ is $m_i(\underset{\sim}{y}_i|\eta) = \int f_i(\underset{\sim}{y}_i|\theta_i)\,\pi(\theta_i|\eta)d\theta_i$, and the conditional independence structure of our model implies that marginally the $\underset{\sim}{Y}_i$ are independent. Thus the joint marginal distribution of all the data is $m(\mathrm{y}|\eta) = \prod_{i=1}^{p} m_i(\underset{\sim}{y}_i|\eta)$. The posterior distribution of θ_i depends on the data only through $\underset{\sim}{y}_i$ and is denoted by $f_i(\theta_i|\underset{\sim}{y}_i, \eta)$, though in the sequel we suppress the subscript on f to simplify the notation.

If η were known then point or interval estimates for θ_i would be computed via the posterior $f(\theta_i|\underset{\sim}{y}_i, \eta)$. Generally, however, η is unknown. A pure Bayesian approach would place a third stage prior (also known as a "hyperprior") $\tau(\eta)$ on η and then base inference about θ_i on the "marginal posterior,"

$$l_h(\theta_i|\mathrm{y}) = \int f(\theta_i|\underset{\sim}{y}_i, \eta)h(\eta|\mathrm{y})d\eta \tag{1.1}$$

where $h(\eta|\mathrm{y}) \propto m(\mathrm{y}|\eta)\tau(\eta)$. The hyperprior $\tau(\eta)$ is often given vague specification. Usually computation of (1.1) is an arduous task. In recent years substantial effort has been devoted to developing methodology to calculate the distribution (1.1) and its characteristics. (see for example Naylor and Smith 1982, Tierney and Kadane 1986, Smith et. al. 1985, Smith et. al. 1987, and Gelfand and Smith 1989).

An alternative is the PEB approach, which treats η as a fixed unknown, and replaces the integration over η in (1.1) with estimation of η (usually via maximum likelihood) from the marginal distribution $m(\mathrm{y}|\eta)$, obtaining $\hat{\eta} = \hat{\eta}(\mathrm{y})$. Inference is based upon the "estimated posterior,"

DTIC COPY INSPECTED 1
or
n
/
y Codes
and/or
ial
A-1

$$f(\theta_i|\underline{y}_i, \hat{\eta}). \tag{1.2}$$

Note that (1.2) may be considered an approximation to (1.1) of order $O(p^{-1})$ in the sense that, under mild regularity conditions,

$$E(g(\theta_i)|y) = E(g(\theta_i)|\underline{y}_i, \hat{\eta})[1 + O(p^{-1})] \tag{1.3}$$

(Kass and Steffey, 1988). Expression (1.3) formalizes the fact that PEB estimates are approximately fully Bayesian posterior means.

There is also a substantial amount of literature which demonstrates that, as an estimator of $g(\theta_i)$, $E(g(\theta_i)|\underline{y}_i, \hat{\eta})$ often performs well in a decision theoretic sense (see Morris, 1983a for a summary). Unfortunately, interval estimation of θ_i through (1.2) has been less successful; such "naive" confidence intervals based on appropriate percentiles of the estimated posterior (either highest posterior density or "equal tail" intervals) are generally too short, and hence fail to attain the nominal coverage probability. The explanation for this problem from a PEB point of view is that we are ignoring the variability in $\hat{\eta}$; from a Bayesian point of view, that we are ignoring the posterior uncertainty about η. More precisely

$$Var(\theta_i|y) = E_{\eta|y}[Var(\theta_i|\underline{y}_i, \eta)] + Var_{\eta|y}[E(\theta_i|\underline{y}_i, \eta)], \tag{1.4}$$

and so the variance estimate based on (1.2), $Var(\theta_i|\underline{y}_i, \hat{\eta})$, will, according to (1.3), only approximate the first term in (1.4). Morris (1983b, 1987) develops improved approximations to (1.4) in special cases while Kass and Steffey (1988) give a general first order approximation.

We propose a more direct attack on the interval estimation problem. We first formalize our objective, nominal conditional or unconditional EB coverage. We then describe how to correct the bias in the naive interval estimate to meet the objective. Most

importantly, we show that for many interesting problems implementation of the bias correction can be accomplished via a sample reuse method. In particular we employ the Type III parametric bootstrap introduced by Laird and Louis (1987), applying it to two discrete exponential family data sets.

In the EB framework (Morris 1983a,b, Hill 1988) a statistical model is appealingly specified as a collection of *joint* probability distributions indexed by some parameter,

$$\mathcal{P} = \{p_\eta(y, \theta)\,,\ \eta \in \mathrm{H}\}. \tag{1.5}$$

A member of this family is expressible in the sampling form as $p_\eta(y, \theta) = f(y \mid \theta)\, \pi(\theta \mid \eta)$, or in the inferential form as $p_\eta(y, \theta) = f(\theta \mid y, \eta)\, m(y \mid \eta)$. Performance of an inference procedure is evaluated over the variability inherent in *both* θ and the data. Thus an unconditional EB confidence set of size $1 - \alpha$ for $g(\theta)$ is a subset $\iota_\alpha(\mathrm{Y})$ of Θ such that

$$\inf_{\mathcal{P}} P_\eta\{g(\theta) \in \iota_\alpha(\mathrm{Y})\} \geq 1 - \alpha. \tag{1.6}$$

Equation (1.6) has been criticized as being a weak statement. First, we would likely prefer $P_\eta\{g(\theta) \in \iota_\alpha(\mathrm{Y})\} \approx 1 - \alpha$ over all distributions in $\mathcal{P}$. Second, we would likely prefer a probability statement which offers conditional calibration given some appropriate data summary (Rubin, 1984). For instance a pure Bayesian interval would be based upon $\ell(\theta_r \mid \mathrm{y})$, and so is conditionally calibrated given all the data. We propose modifying (1.6) to a conditional statement by integrating instead over the distribution $p(\theta_r, \mathrm{y} \mid b(\mathrm{y}), \eta)$ for a suitable statistic $b(\mathrm{Y})$. That is, $\iota_\alpha(\mathrm{Y})$ is a conditional $1 - \alpha$ EB confidence set for $g(\theta)$ given $b(\mathrm{Y})$ if, for each $b(\mathrm{y}) = b$ and η,

$$P_\eta\{g(\theta) \in \iota_\alpha(\mathrm{Y}) \mid b(\mathrm{y}) = b\} \approx 1 - \alpha. \tag{1.7}$$

In Section 2, we review the "bias correcting" approach (given in Carlin and Gelfand, 1989) for obtaining intervals achieving this sort of coverage probability using various choices of $b(\mathrm{Y})$. We then settle on $b(\mathrm{Y}) = \underline{Y}_r$ as a natural choice, and elaborate on sam-

ple reuse methods to implement the bias correction. Section 3 discusses the mechanics of hyperparameter estimation and EB point and interval estimation of θ_i in the context of two discrete models, the binomial and the Poisson. For illustration, the analysis of two data sets is given in Section 4. Finally in Section 5 we summarize and comment on some remaining problems.

2. PEB MODELLING AND BIAS CORRECTION

Adopting the EB framework we define the concepts of sufficiency and ancillarity (see Hill, 1988) for the model (1.5).

Definition 1. A statistic T is *sufficient* for $g(\theta)$ in (1.5) if and only if

- T is sufficient for $g(\theta)$ in the posterior family, i.e., $f(g(\theta)\,|\,t,\eta) = f(g(\theta)\,|\,y,\eta)$ for all $\eta \in H$, and,
- T is sufficient for η in the marginal family, i.e., $m(y\,|\,t,\eta) = m(y\,|\,t)$ for all $\eta \in H$.

For example, under the exchangeable structure of Section 1, if $g(\theta) = \theta_i$, then usually $T = (\underline{Y}_i, \hat{\eta})$.

Using this definition it is straightforward to show that T carries all the information about η in the model with respect to inference about $g(\theta)$, i.e. we may replace Y by T in (1.6) and (1.7). In particular for θ_i we need only study intervals based on $\underline{Y}_i$ and $\hat{\eta}$.

Definition 2. A statistic A is *ancillary* for $g(\theta)$ in (1.5) if and only if $T = (\hat{\eta}, a)$ is minimally sufficient for $g(\theta)$, where $\hat{\eta}$ is an estimate of η, and $p(a\,|\,\eta) = p(a)$ for all $\eta \in H$.

Thus if A is ancillary, its distribution with respect to the marginal family is free of η. In this definition, minimal sufficiency is with respect to (1.5) using Definition 1.

In the sequel we take $g(\theta) = \theta$, and let $[U \mid V]$ be a generic notation for the distribution of U given V. To strengthen the statement (1.6) suppose we replace the integration over $[\theta_i, \underline{Y}, \hat{\eta} \mid \eta]$ in (1.6) with an integration over $[\theta_i, \underline{Y}, \hat{\eta} \mid b(Y), \eta]$. If we take $b(Y) = Y$ then we are in fact integrating over $[\theta_i \mid \underline{Y}, \eta]$, the posterior distribution of θ_i. If we choose $b(Y)$ EB-ancillary, then typically $(b, \hat{\eta}, \underline{y})$ are such that fixing any two of them determines the third. This means that usually $[\theta_i, \underline{Y}, \hat{\eta} \mid b, \eta]$ $= [\theta_i, \underline{Y} \mid \hat{\eta}, b, \eta] \cdot [\hat{\eta} \mid b, \eta] = [\theta_i \mid \underline{Y}, \eta] \cdot [\hat{\eta} \mid \eta]$, where we have used Basu's well-known theorem in the last step. Thus if we can find an EB-ancillary statistic, we can develop intervals with EB-coverage conditional on the ancillary merely by integrating first over the full posterior and then over the sampling distribution of our estimator $\hat{\eta}$. However in the discrete cases we study, and in fact rather generally, exact ancillary statistics are not available. We have not investigated the notion of approximate ancillaries. If we instead take $b(Y) = \underline{Y}$, then our integration is over $[\theta_i, \hat{\eta} \mid \underline{Y}, \eta] = [\theta_i \mid \underline{Y}, \eta] \cdot [\hat{\eta} \mid \underline{Y}, \eta]$. This choice of b is appealing since $\underline{Y}$ is sufficient for θ_i in the posterior family. Additionally, this conditioning is straightforward to implement.

If we denote the αth quantile of the posterior $f(\theta_i \mid \underline{y}_i, \eta)$ by $q_\alpha(\underline{y}_i, \eta)$, then the so-called equal tail "naive" EB interval can be written as

$$(q_{\alpha/2}(\underline{y}_i, \hat{\eta}), q_{1-\alpha/2}(\underline{y}_i, \hat{\eta})) \tag{2.1}$$

As has already been noted, intervals of this type fail to satisfy (1.6), and their coverage conditional on $\underline{Y}$ is also poor. Most of the work in the area of EB confidence intervals has focused on "lengthening" these intervals by using the marginal posterior (1.1), either exactly or approximately. Deely and Lindley (1981) and Rubin (1982) perform the numerical integration and compute l_A directly. Morris (1987) advocates use of the member of $f(\theta_i \mid \underline{y}_i, \eta)$ whose first two moments agree with the first two (estimated) moments of l_A. Laird and Louis (1987) suggest the use of a parametric bootstrap sampling method

(which we discuss in some detail below) to approximate (1.1) where h is taken to be $p(\hat{\eta}|\eta)$, the sampling density of $\hat{\eta}$, with the arguments interchanged.

In our view, the weakness of these approaches is that they fail to address the salient issue, the existence and nature of an h which will be successful in achieving nominal EB coverage. Put simply, they are all concerned with "matching" l_α in (1.1), without specifying how to choose h to begin with. This is understandable, as the whole approach is not directly aimed at attaining a specified level of EB coverage (either conditional or unconditional). Furthermore, while the lengthening of the naive intervals created by the "mixing" in (1.1) is usually desirable, if our estimator $\hat{\eta}$ is badly biased, the naive intervals can actually turn out to be too *long*. The real issue is how to correct the bias in (2.1). A direct approach which would be applied to each tail separately is as follows:

Suppose we define

$$r(\hat{\eta}, \eta, \underline{y}_i, \alpha) = P\{\theta_i \leq q_\alpha(\underline{y}_i, \hat{\eta}) \,|\, \theta_i \sim f(\theta_i|\underline{y}_i, \eta)\} \tag{2.2}$$

and

$$R(\eta, \underline{y}_i, \alpha) = E_{\hat{\eta}|\underline{y}_i, \eta}\{r(\hat{\eta}, \eta, \underline{y}_i, \alpha)\}. \tag{2.3}$$

If we then solve

$$R(\eta, \underline{y}_i, \alpha') = \alpha \tag{2.4}$$

for $\alpha' = \alpha'(\eta, \underline{y}_i, \alpha)$, we conditionally "correct the bias" in using $\hat{\eta}$. Of course (2.4) is not solvable as it stands since η is unknown. Using $\hat{\eta}$, we propose to obtain a bootstrap estimate of the left hand side of (2.4) and thus solve instead for $\alpha'(\hat{\eta}, \underline{y}_i, \alpha)$. We correct both the lower and upper percentiles of our naive interval, obtaining α_L' and α_U', and then take

$$(q_{\alpha_1}(\underset{\sim}{y}_i, \hat{\eta}), q_{\alpha_2}(\underset{\sim}{y}_i, \hat{\eta})) \tag{2.5}$$

as our bias corrected interval for θ_i.

Carlin and Gelfand (1989) showed that this bias corrected confidence interval is unique provided $\partial r/\partial \alpha$ exists. They further showed that if the distributions $f(\theta_i|\underset{\sim}{y}_i, \eta)$ and $g(\hat{\eta}|\underset{\sim}{y}_i, \eta)$ are stochastically ordered in η, then conditional bias correction "works" with respect to (1.7), i.e. $E_{\hat{\eta}|\underset{\sim}{y}_i, \eta} r(\hat{\eta}, \eta, \underset{\sim}{y}_i, \alpha'(\hat{\eta}, \underset{\sim}{y}_i, \alpha))$, the expected conditional bias corrected tail probability, falls in an interval containing α (for some η, it will be slightly more than α; for others, slightly less). They also provide simulation support in several elementary cases. Note that while we describe bias correction of the naive interval (2.1) we could just as well have chosen to bias correct intervals resulting from (1.1). Such correction could be used to possibly obtain shorter intervals achieving desired coverage levels.

Let us briefly investigate the mechanics of bias correction. If $\pi(\theta|\eta)$ is chosen as the standard conjugate prior for $f(\underset{\sim}{y}_i|\theta_i)$ then of course the posterior $f(\theta_i|\underset{\sim}{y}_i, \eta)$ belongs to the same standard family, and r in (2.2) takes the relatively simple form

$$r(\hat{\eta}, \eta, \underset{\sim}{y}_i, \alpha) = F_{\underset{\sim}{y}_i, \eta}[F^{-1}_{\underset{\sim}{y}_i, \eta}(\alpha)] \tag{2.6}$$

where F is the posterior c.d.f. To compute R in (2.3) necessitates integrating over the distribution $g(\hat{\eta}|\underset{\sim}{y}_i, \eta)$. If g is available in closed form this will require a numerical integration (either Monte Carlo or other quadrature method). Then we compute the marginal MLE $\hat{\eta}$ and solve equation (2.4) at $\eta = \hat{\eta}$ for α' via regula falsi. This replacement of an unknown population parameter by its estimate from the marginal empirical c.d.f. is referred to as a "parametric bootstrap" by Laird and Louis (1987).

Examples where this procedure can be carried out efficiently and easily are discussed in Carlin and Gelfand (1989). However, in many cases (in particular the discrete settings

in Section 3) a closed form for $g(\hat{\eta}|\underline{y}, \eta)$ is unavailable. In fact, in several elementary conjugate situations, the marginal MLE $\hat{\eta}$ itself has no closed form and can only be computed numerically. In such cases we estimate $R(\eta, \underline{y}, \alpha)$ through the use of a "Type III parametric bootstrap," a sample reuse method introduced by Laird and Louis (1987) and modified here for the case of EB coverage conditional on $\underline{Y}$.

Let us first review the unconditional version. To estimate expectations under the sampling density of $\hat{\eta}$, $p(\hat{\eta}|\eta)$, we obtain $\hat{\eta}^*$ from the distribution $p(\cdot|\hat{\eta})$ as follows. Draw $\theta_1^*, \ldots, \theta_p^*$ i.i.d. from $\pi(\theta|\hat{\eta})$, then draw $\underline{Y}_i^*$ independently from $f(\underline{y}|\theta_i^*)$, $i = 1, \ldots, p$, and finally compute $\hat{\eta}^*$ from the "pseudodata" $\{\underline{Y}_i^*\}$ in the same way that $\hat{\eta}$ was computed from the data $\{\underline{Y}_i\}$. Concisely, we have

$$\hat{\eta} \rightarrow \{\theta_i^*\} \rightarrow \{\underline{Y}_i^*\} \rightarrow \hat{\eta}^* \tag{2.7}$$

For correction conditional on $\underline{Y}_i$, we modify the Laird and Louis procedure in order to draw observations from g rather than p by changing (2.7) to

$$\hat{\eta} \rightarrow \{\theta_k^*, k \neq i\} \rightarrow \{\underline{Y}_k^*, k \neq i\} \rightarrow \hat{\eta}^* = \hat{\eta}^*(\underline{y}_i, \{\underline{Y}_k^*, k \neq i\}) \tag{2.8}$$

Repeating this process N times we obtain $\hat{\eta}_j^* \overset{\text{iid}}{\sim} g(\cdot|\underline{y}_i, \hat{\eta})$, $j = 1, \ldots, N$, and our Type III parametric bootstrap estimate of $R(\eta, \underline{y}_i, \alpha')$ becomes

$$\sum_{j=1}^{N} r(\hat{\eta}_j^*, \hat{\eta}, \underline{y}_i, \alpha')/N. \tag{2.9}$$

We equate (2.9) to α, and solve for α' by regula falsi as above. Note that since only the $\underline{Y}_i^*$ are needed to calculate $\hat{\eta}^*$, if $m(\underline{y}|\eta)$ can be sampled from directly we can omit the generation of the θ_i^* (see Example 3.2).

The Type III parametric bootstrap can obviously be modified to implement bias correction conditional on $b(Y)$ other than $\underline{Y}_i$. However the theoretical results below (2.5) are only established given $\underline{Y}_i$.

3. SPECIFIC DISCRETE MODELS

We now apply the preceding methodology to perhaps the two most common discrete models.

Example 3.1: Binomial-beta. Suppose the data vectors $\underline{Y}$ are simply vectors of zeroes and ones, each element corresponding to a success or failure on the jth independent Bernoulli trial in the ith population, $j = 1, \ldots, n_i$, $i = 1, \ldots, p$. At the first stage of the hierarchy, we assume the probability of success on any given trial to be the same for all trials within the ith population, but possibly different across populations. By sufficiency we reduce to $X_i = \sum_{j=1}^{n_i} Y_{ij}$ and so $X_i | \theta_i \overset{ind}{\sim} \text{Bin}(n_i, \theta_i)$. Under the conjugate prior $\theta_i | \eta \overset{ind}{\sim} \text{Beta}(a, b)$, $i = 1, \ldots, p$, where $\eta = (a, b)$, $a, b > 0$, the posterior distribution is $f(\theta_i | x_i, a, b) = \text{Beta}(a + x_i, b + n_i - x_i)$, and the marginal distribution is $m(x_i | a, b) = \binom{n_i}{x_i} B(a + x_i, b + n_i - x_i) / B(a, b)$, where $B(\cdot, \cdot)$ represents the beta function. The marginal likelihood,

$$L(a, b) = m(\underline{x} | a, b) = \prod_{i=1}^{p} \binom{n_i}{x_i} B(a + x_i, b + n_i - x_i) / B(a, b), \tag{3.1}$$

is maximized via numerical methods to produce the marginal MLE, $\hat{\eta} = (\hat{a}, \hat{b})$. The naive EB confidence interval (2.1) is computed as $(\Upsilon^{-1}_{(\hat{a} + x_i, \hat{b} + n_i - x_i)}(\alpha/2), \Upsilon^{-1}_{(\hat{a} + x_i, \hat{b} + n_i - x_i)}(1 - \alpha/2))$, where $\Upsilon_{(c,d)}$ is the c.d.f. of a beta distribution with parameter c and d.

To implement the bias correction we note that (2.2) becomes

$$r(\hat{\eta}, \eta, x_i, \alpha) = \Upsilon_{(a + x_i, b + n_i - x_i)}(\Upsilon^{-1}_{(\hat{a} + x_i, \hat{b} + n_i - x_i)}(\alpha)) \tag{3.2}$$

which is available numerically using a scientific subroutine library. Using the Type III parametric bootstrap procedure (2.9) becomes

$$\sum_{j=1}^{N} \Upsilon_{(\hat{a} + x_i, \hat{b} + n_i - x_i)}(\Upsilon^{-1}_{(a_j^* + x_i, b_j^* + n_i - x_i)}(\alpha')) / N \tag{3.3}$$

which we equate to α and solve for α'. Setting α to the desired nominal level (e.g., $\alpha = (.05, .95)$ for 90% conditional EB coverage), we solve for the corresponding (α_L', α_U'), and compute our interval for θ_i using (2.5). Note that we bias correct each θ_i-interval separately, $i = 1, \ldots, p$, but of course the correction in each case depends on the data through $(\hat{a}, \hat{b})$. If we desire intervals corrected only for unconditional EB coverage, our bootstrap equation becomes

$$\sum_{j=1}^{N} \Upsilon_{(\hat{a}+x_i^*, \hat{b}+n_i-x_i^*)}(\Upsilon^{-1}_{(\hat{a}_j^*+x_{ij}^*, \hat{b}_j^*+n_i-x_{ij}^*)}(\alpha'))/N = \alpha, \tag{3.4}$$

Note that (3.4) differs from (3.3) only in replacing the given value x_i by the bootstrapped x_{ij}^*'s. Now we are "averaging over X_i," as in (1.6), rather than conditioning on X_i. We must still bias correct each θ_i-interval separately, unless we have a balanced experiment (all n_i equal). In this case the x_i's are marginally exchangeable, hence so are the x_i^*'s, and so we need only solve (3.4) once for (α_L', α_U') before using (2.5).

Example 3.2: Poisson-gamma. Suppose we have data of the form (x_i, t_i), $i = 1, \ldots, p$, where the x_i are observed counts during the time interval $(0, t_i)$. For example, the data might be calls arriving at p different switchboards in the same county. We assume $X_i | \theta_i \overset{\text{ind}}{\sim}$ Poisson $(\theta_i t_i)$, $i = 1, \ldots, p$, where the t_i are known "time exposures." Under the conjugate prior, $\theta_i | \eta \overset{\text{iid}}{\sim}$ Gamma (a, b) (again $\eta = (a, b)$, $a, b > 0$), the posterior distribution $\theta_i | x_i$ is Gamma $(a + x_i, (t_i + 1/b)^{-1})$, and the marginal distribution of X_i is Negative Binomial, i.e.,

$$m(x_i | a, b) = \binom{x_i+a-1}{a-1}\left(\frac{t_i}{t_i + 1/b}\right)^{x_i}\left(\frac{1/b}{t_i + 1/b}\right)^{a}. \tag{3.5}$$

If $(\hat{a}, \hat{b})$ maximize the marginal likelihood ($\hat{a}$ and $\hat{b}$ are not available in closed form; see Section 4), then the naive EB confidence interval for θ_i is

$$\left(D^{-1}_{2(\hat{a}+x_i)}(\alpha/2)/[2(t_i+1/\hat{b})],\ D^{-1}_{2(\hat{a}+x_i)}(1-\alpha/2)/[2(t_i+1/\hat{b})]\right), \tag{3.6}$$

where D_k denotes the c.d.f. of a chi-square distribution with k (not necessarily integer) degrees of freedom.

In this case (2.2) becomes

$$D_{2(a+x_i)}\{[(t_i+1/b)\,/\,(t_i+1/\hat{b})]\ D^{-1}_{2(\hat{a}+x_i)}(\alpha)\}. \tag{3.7}$$

For EB correction conditional on $X_i = x_i$, analogous to (3.3), the Type III parametric bootstrap (2.9) becomes

$$\sum_{j=1}^{N} D_{2(\hat{a}+x_i)}\{[(t_i+1/\hat{b})\,/\,(t_i+1/b_j^*)]\ D^{-1}_{2(a_j^*+x_i)}(\alpha')\}\ /N \tag{3.8}$$

which we equate to α and solve for α'. For unconditional correction analogous to (3.4) we have the equation

$$\sum_{j=1}^{N} D_{2(\hat{a}+x_{ij}^*)}\{[(t_i+1/\hat{b})\,/\,(t_i+1/b_j^*)]\ D^{-1}_{2(a_j^*+x_{ij}^*)}(\alpha')\}\ /N = \alpha \tag{3.9}$$

The remark after (2.9) reminds us that in this case we would generate negative binomial X_{ij}^*'s directly. In addition, if $t_i = t$ for all i, we need only solve (3.9) once for (α_L', α_U') before using (2.5).

In this example if we take the gamma scale hyperparameter b to be known (say $b = 1$ w.l.o.g.), and if we assume $t_i = t$ (= 1 w.l.o.g.), then the marginal family (3.5) is Negative Binomial $(a, 1/2)$. The method of moments (instead of maximum likelihood) estimator of a is $\hat{a} = \overline{X} = \sum_{i=1}^{p} X_i/p$ and the distribution of $\hat{a}\,|\,x_i, a$ follows from writing $\hat{a} = (W + X_i)/p$ where $W = \sum_{k \neq i} X_k \sim$ Negative Binomial $(a(p-1), 1/2)$. Thus we can integrate (2.4) directly, avoiding the Type III resampling algorithm.

4. DATA EXAMPLES

We illustrate the implementation of the bias corrected naive EB confidence interval approach with two discrete data sets. Results of several simulation studies evaluating coverage probabilities and interval lengths are discussed in Carlin (1989).

The data set in Table 1 comes from Burton and Turvey (1988), and gives the results of a psychological evaluation of six independent subjects. Each subject was given three hollow wooden balls, two of which contained a small pyramid and the third either a hemisphere, a block, a cylinder, or a cone. On each of the four resulting trials, the subject had to guess which ball was not like the others simply by shaking, turning, etc. Table 1 gives the results of this (balanced) experiment, where X_i is the number of questions answered *incorrectly* by subject i, and r_i is the raw failure rate, which of course is also the usual classical point estimate of θ_i, the true failure rate for subject i.

(Insert Table 1 about here)

In terms of modelling this experiment, our binomial-beta model seems natural. At the first stage (given θ_i), a subject's four responses could reasonably be assumed to be i.i.d. Bernoulli trials, and the beta provides a broad choice for the second stage distribution of the $\{\theta_i\}$. In fact, since our prior belief is that the questions are relatively easy, the simpler family Beta(1, b), $b > 1$ seems adequate. Finally this data set benefits substantially from empirical Bayes modelling, since the small amount of information on each subject severely limits frequentist inference.

(Insert Table 2 about here)

The results of our analysis are given in Table 2. Since we are in a balanced (i.i.d.) case, the results for subjects 1 and 2 (who both answered two questions incorrectly) are identical, as are those for subjects 4, 5 and 6 (all of whom made no mistakes). Note the familiar shrinking of the classical point estimates toward the overall proportion of

questions answered incorrectly (5/24 = .208) by the EB point estimator. Classical confidence intervals (based on Fisher's exact test statistic for θ_i) and naive EB intervals are shown, along with versions that are bias corrected unconditionally (via (3.4)) and conditionally (via (3.3)) for nominal coverage $\gamma = .90$. We chose $N = 1000$ bootstrap reps in solving these two equations, using FORTRAN augmented by the IMSL subroutine library. One can see that both types of bias corrected intervals have lengths that are between those of the classical and naive EB methods. However, since nominal conditional coverage implies nominal unconditional coverage, neither bias corrected interval can be uniformly shorter. In particular we see that the conditional intervals are shorter when $X = 2$, but that the unconditional intervals are shorter when $X = 0$.

The data presented in Table 3 record numbers of pump failures, X_i, observed in thousands of hours, t_i, for $p = 10$ different systems of a certain nuclear power plant. The observations are listed in increasing order of raw failure rate $r_i = X_i/t_i$, which again is the classical point estimate of the true failure rate θ_i for the ith system. This data originally appeared in Worledge, Stringham, and McClymont (1982), and was subjected to an empirical Bayes analysis by Gaver and O'Muircheartaigh (1987).

(Insert Table 3 about here)

Our approach is that of Example 3.2 above, using the conjugate Gamma(a, b) prior for θ_i. Gaver and O'Muircheartaigh also explore this approach, but after computing the estimated posterior (also gamma, of course) in the usual way, they obtain an approximate EB confidence interval for θ_i by assuming that the posterior distribution of $\varepsilon_i = \log(\theta_i)$ is approximately normal. Thus their EB point estimate for θ_i is $\exp(\hat{\mu}_i)$, and their naive 90% EB confidence interval for θ_i is given by

$$\left(\exp(\hat{\mu}_i - 1.645\hat{\sigma}_i),\ \exp(\hat{\mu}_i + 1.645\hat{\sigma}_i)\right), \tag{4.1}$$

where $\hat{\mu}_i$ and $\hat{\sigma}_i$ are the mean and standard deviation, respectively, of the (log-gamma) estimated posterior for ε_i. (Actually, the authors' concern about conjugate priors' overshrinkage of outliers leads them to prefer a heavier-tailed log-Student's t prior on ε_i. However, they conclude that this assumption does not greatly affect the results. To make a fair comparison, we will compare our intervals only with their gamma-based intervals.)

Solving (3.8) and (3.9) involves computing N MLE vectors (a_j^*, b_j^*), one for each bootstrapped pseudodata sample. In order to expedite the maximization Gaver and O'Muircheartaigh suggest using as starting values crude moments estimators obtained by equating the first two sample moments of the crude rates, $\bar{r}$ and s_r^2, to the corresponding moments in the marginal family, namely, $E(r_i) = E(X_i)/t_i = ab$ and $Var(r_i) = Var(X_i)/t_i^2 = ab/t_i + ab^2$. This results in

$$\hat{a}_{\mathrm{MOM}} = \bar{r}^2/[s_r^2 - \bar{r}\overline{t^{-1}}] \tag{4.2}$$

and

$$\hat{b}_{\mathrm{MOM}} = [s_r^2 - \bar{r}\overline{t^{-1}}]/\bar{r} \tag{4.3}$$

where $\overline{t^{-1}} = \sum_{i=1}^{p} t_i^{-1}/p$. (These estimators do not exist if $s_r^2 \leq \bar{r}\overline{t^{-1}}$, whence any other reasonable values, possibly the "parent" values $(\hat{a}, \hat{b})$, can be used.) Our algorithm begins with Newton-Raphson. If it diverges, we use a local grid search to find a better (higher likelihood) place to restart. We iterate grid search and Newton-Raphson up to twenty times, finally giving up and regenerating new pseudodata if the algorithm still fails. We were able to keep the "failure rate" under 1% using this algorithm.

Gaver and O'Muircheartaigh felt that the intervals (4.1) were likely to be too short due to hyperparameter estimation. Their ad hoc remedy was to compute approximate joint 95% confidence regions for their computed values of $(\hat{a} = .8223, \hat{b} = .7943)$ using a chi-squared likelihood ratio technique. They then searched over all $(\hat{a}, \hat{b})$ in this re-

gion, taking the largest value of $\exp(\hat{\mu}_i + 1.645\hat{\sigma}_i)$ obtained as their "corrected" upper confidence limit for θ_i. (A similar procedure could have been undertaken for the lower limit, but was not since the authors' interest was, understandably, only in how *large* the failure rates could be.) While this process certainly does increase the upper confidence limit, it is not clear what confidence level to attach to it.

(Insert Table 4 about here)

Table 4 gives the results for observations 1, 5, 6, and 10 using the classical interval, $(1/2 \cdot D^{-1}_{2t_i}(\alpha/2),\ 1/2 \cdot D^{-1}_{2(t_i+1)}(1-\alpha/2))$, applied to the raw rates, the naive and corrected Gaver and O'Muircheartaigh (G & O) methods, the naive EB interval (2.1), and the conditional and unconditional bias corrected methods (3.8) and (3.9) above. Bias corrected percentile (α') values used in these methods are shown in parentheses below the corresponding interval endpoint.

Some conclusions are: while the classical point estimates seem all right, the corresponding interval estimates are very wide. The naive G & O point estimates have been uniformly shrunk closer to zero than the regular EB point estimates, while the corresponding upper confidence limits are uniformly *further* from zero. The naive G & O interval estimates are not necessarily too short; in the case of the smallest observation, the naive G & O upper limit is already larger than either of the bias corrected upper limits! The "corrected" G & O upper limit is always much larger than that of any of the other EB methods, reflecting the substantial conservatism embodied in this procedure. Note that more bias correcting (more extreme values of α', hence longer confidence intervals) is present for cases having shorter history (smaller t_i) -- observation #5, for example. This jibes with our intuition about EB point estimation: that cases with less information have more uncertainty associated with them and possibly exhibit more extreme shrinkage patterns. It also appears that conditioning on the value of a more highly variable X_i (like observation #5) results in a longer interval than would have been obtained had only unconditional coverage been sought.

5. CONCLUSION

In this paper we have described a sample reuse method (the Type III parametric bootstrap) to correct the bias in naive empirical Bayes confidence intervals when tractable distribution theory is unavailable. Through data examples we have shown that this method is easily implemented yielding intervals which retain much of the intuitive appeal associated with PEB and shrinkage estimation. Future effort is directed at more general applications. For example, the conjugate prior assumption may be dropped, using numerical methods to evaluate (2.6) in the absence of a convenient form for the posterior. Additionally we might choose to bias correct intervals based on the marginal posterior (1.1), rather than the naive intervals based on the estimated posterior (1.2). Beginning with a richer class of intervals could lead to shorter corrected intervals achieving nominal coverage. Finally, efforts to unite the PEB and hierarchical Bayesian literature, as in Section 1 of this paper as well as Kass and Steffey (1988), continue to be important.

REFERENCES

1. Burton, G., and Turvey, M. (1988), "Haptic Perception of Shape," manuscript in process, Department of Psychology, University of Connecticut.

2. Carlin, B. P. (1989), "Accurate Empirical Bayes Interval Estimation," unpublished Ph.D. dissertation, University of Connecticut.

3. Carlin, B. P., and Gelfand, A. E. (1989), "Approaches for Empirical Bayes Confidence Intervals," to appear *Journal of the American Statistical Association.*

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

5. Gaver, D. P., and O'Muircheartaigh, I. G. (1987), "Robust Empirical Bayes Analyses of Event Rates," *Technometrics*, 29, 1-15.

6. Gelfand, A. E., and Smith, A. F. M. (1989), "Sampling Based Approaches to Calculating Marginal Densities," technical report, Department of Mathematics, University of Nottingham.

7. Hill, J. R. (1988), "A General Framework for Model-Based Statistics," unpublished manuscript, EDS Research, Albuquerque, New Mexico.

8. 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.

9. Kass, R. E., and Steffey, D. (1988), "Approximate Bayesian Inference in Conditionally Independent Hierarchical Models (Parametric Empirical Bayes Models)," Technical Report #386, Department of Statistics, Carnegie-Mellon University.

10. 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.

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

12. —— (1983b), "Parametric Empirical Bayes Confidence Intervals," in *Scientific Inference, Data Analysis, and Robustness*, New York: Academic Press, pp. 25-50.

13. —— (1987), "Determining the Accuracy of Bayesian Empirical Bayes Estimates in the Familiar Exponential Families," Technical Report 46, University of Texas, Center for Statistical Sciences.

14. Naylor, J. C., and Smith, A. F. M. (1982), "Applications of a Method for the Efficient Computation of Posterior Distributions," *Applied Statistics*, 31, 214-225.

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

16. —— (1984), "Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician," *The Annals of Statistics*, 12, 751-772.

17. Smith, A. F. M., Skene, A. M., Shaw, J. E. H., Naylor, J. C., and Dransfield, M. (1985), "The Implementation of the Bayesian Paradigm," *Communications in Statistics, Theory and Methods*, 14, 1079-1102.

18. Smith, A. F. M., Skene, A. M., Shaw, J. E. H., and Naylor, J. C. (1987), "Progress with Numerical and Graphical Methods for Bayesian Statistics," *Statistician*, 36, 75-82.

19. Tierney, L., and Kadane, J. B. (1986), "Accurate Approximations for Posterior Moments and Marginal Densities," *Journal of the American Statistical Association*, 81, 82-86.

20. Worledge, D. H., Stringham, R. S., and McClymont, A. S. (1982), "PWR Power Plant Reliability Data," Interim Report NP-2592, Electric Power Research Institute.

TABLE 1. Pychological Test Data

Subject	n_i	X_i	$r_i = \hat{\theta}_i^{Classical}$
1	4	2	.50
2	4	2	.50
3	4	1	.25
4	4	0	.00
5	4	0	.00
6	4	0	.00

Source: Burton and Turvey (1988)

TABLE 2. Psychological Test Data Analysis

Subject		90% Classical CI			EB	90% Naive EB CI		
i	r_i	Lower	Upper	Length	Pt Est	Lower	Upper	Length
1,2	0.500	0.098	0.902	0.805	0.336	0.112	0.604	0.491
3	0.250	0.013	0.751	0.739	0.224	0.047	0.474	0.427
4,5,6	0.000	0.000	0.527	0.527	0.112	0.006	0.315	0.308

Subject	90% Unconditional BCN CI			90% Conditional BCN CI		
i	Lower	Upper	Length	Lower	Upper	Length
1,2	0.108	0.690	0.582	0.100	0.627	0.528
3	0.044	0.570	0.526	0.047	0.576	0.529
4,5,6	0.006	0.415	0.409	0.008	0.455	0.447

Note: Hyperparameter estimate (MLE): $\hat{b} = 3.9296$.

TABLE 3. Pump Failure Data

System$_i$	X_i	t_i	$r_i = \hat{\theta}_i^{Classical}$
1	5	94.320	.053
2	1	15.720	.064
3	5	62.880	.080
4	14	125.760	.111
5	3	5.240	.573
6	19	31.440	.604
7	1	1.048	.954
8	1	1.048	.954
9	4	2.096	1.910
10	22	10.480	2.099

Source: Electric Power Research Institute Report
(Worledge, Stringham, and McClymont 1982)

TABLE 4. Pump Failure Data Analysis

	Pt Estimate	90% Confidence Interval		Interval Length
Observation: 1				
Classical	.053	.000	3.098	3.098
"Naive" G & O	.056	.027	.114	.087
"Corrected" G & O	"	---	.125	---
Naive EB	.061	.026	.107	.081
Unconditional BCN	"	.026 (.0472)	.109 (.9533)	.083
Conditional BCN	"	.025 (.0453)	.108 (.9507)	.082
Observation: 5				
Classical	.573	.004	4.032	4.028
"Naive" G & O	.512	.207	1.258	1.051
"Corrected" G & O	"	---	1.553	---
Naive EB	.588	.195	1.154	.959
Unconditional BCN	"	.158 (.0268)	1.271 (.9705)	1.114
Conditional BCN	"	.163 (.0298)	1.284 (.9721)	1.121
Observation: 6				
Classical	.604	.006	4.087	4.081
"Naive" G & O	.577	.395	.861	.466
"Corrected" G & O	"	---	.905	---
Naive EB	.606	.401	.846	.445
Unconditional BCN	"	.394 (.0437)	.863 (.9594)	.469
Conditional BCN	"	.398 (.0471)	.863 (.9594)	.465
Observation: 10				
Classical	2.099	.396	6.444	6.048
"Naive" G & O	1.896	1.343	2.691	1.348
"Corrected" G & O	"	---	2.945	---
Naive EB	1.944	1.327	2.658	1.332
Unconditional BCN	"	1.230 (.0250)	2.735 (.9638)	1.505
Conditional BCN	"	1.311 (.0452)	2.674 (.9532)	1.363

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
4. 3

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
A Sample Reuse Method For Accurate Parametric Empirical Bayes Confidence Intervals

5. TYPE OF REPORT & PERIOD COVERED
TECHNICAL REPORT

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Bradley P. Carlin and Alan E. Gelfand

8. CONTRACT OR GRANT NUMBER(s)
N00014-89-J-1627

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Department of Statistics
Stanford University
Stanford, CA 94305

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
NR-042-267

11. CONTROLLING OFFICE NAME AND ADDRESS
Office of Naval Research
Statistics & Probability Program Code 1111

12. REPORT DATE
November 7, 1989

13. NUMBER OF PAGES
24

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Confidence interval; parametric empirical Bayes; parametric bootstrap; bias correction; conditional calibration.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
PLEASE SEE FOLLOWING PAGE.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TECHNICAL REPORT NO. 423

20. ABSTRACT

Parametric empirical Bayes methods of point estimation for a vector of unknown parameters date to the landmark paper of James and Stein (1961). The usual approach is to use the mean of the estimated posterior distribution of each parameter, where the estimation of the prior parameters ("hyperparameters") is accomplished through the marginal distribution of the data. While point estimates computed this way usually perform well, interval estimates based on the estimated posterior (called "naive" EB intervals) are not. They fail to account for the variability in the estimation of the hyperparameters, generally resulting in sub-nominal coverage probability in the "EB" sense defined in Morris (1983a).

In this paper we extend the work of Carlin and Gelfand (1989), who proposed a conditional bias correction method for developing EB intervals which corrects the deficiencies in the naive intervals. We show how bias correction can be implemented in general via a Type III parametric bootstrap procedure, a sample reuse method first employed by Laird and Louis (1987). Theoretical and simulation results indicate that intervals which are accurate with respect to nominal coverage ensue. We give two specific applications (to binomial test data and Poisson failure rate data) where we compute simultaneous point and bias corrected interval estimates.

(Sfw)