Statistical Science
2019, Vol. 34, No. 2, 219–223
https://doi.org/10.1214/19-STS708
Main article: https://doi.org/10.1214/18-STS674

Comment: Empirical Bayes Interval Estimation

Wenhua Jiang

Abstract. This is a contribution to the discussion of the enlightening paper by Professor Efron. We focus on empirical Bayes interval estimation. We discuss the oracle interval estimation rules, the empirical Bayes estimation of the oracle rule and the computation. Some numerical results are reported.

Key words and phrases: Empirical Bayes, interval estimation, oracle rule, generalized MLE.

We congratulate Professor Efron on the enlightening article and thank him for illuminating the oracle and finite Bayes faces of empirical Bayes inference. Our comments below, focused on empirical Bayes interval estimation, is very much inspired by his discussion on the topic.

An interesting aspect of the interval estimation problem is its lack of a purely posterior-based oracle rule in the compound framework. Suppose $X_i|\theta_i \sim f(x|\theta_i)\nu(dx)$ and for simplicity confine our discussion of oracle rules to making inference about θ_i by decision rules as a function of X_i only, that is, rules of the form $[a(X_i), b(X_i)]$ as we are interested in interval estimation. The objective of compound interval estimation could then be written as

$$\text{minimize } n^{-1}\sum_{i=1}^{n}\mathbb{E}\{b(X_i)-a(X_i)\}_+ \tag{1}$$
$$\text{subject to } n^{-1}\sum_{i=1}^{n}\mathbb{P}\{a(X_i)\le\theta_i\le b(X_i)\} = 1-\alpha$$

with a prespecified $\alpha \in (0,1)$. Robbins (1951), Hannan and Robbins (1955) and Robbins (1956) argued that for any loss function $L(t(X_i),\theta_i)$ and decision rule $t(x)$, the compound risk is identical to the Bayes risk for the empirical prior G_n, defined by $G_n(A) = n^{-1}\sum_{i=1}^{n} I_{\{\theta_i\in A\}}$ for any Borel set A. This fundamental theorem of empirical Bayes can be written as

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}L\big(t(X_i),\theta_i\big) = \iint L\big(t(x),\theta\big)f(x|\theta)\nu(dx)G_n(d\theta),$$

which directly provides the oracle Bayes rule

$$t_{G_n}(x) = \arg\min_t \int L(t,\theta)G_n(d\theta|x),$$

where $G_n(d\theta|x) = f(x|\theta)G_n(d\theta)/\int f(x|\theta)G_n(d\theta)$ is the posterior distribution. For the interval estimation problem (1), this leads to the oracle Bayes rule

$$\begin{aligned}&\big[a_{G_n,\lambda}(x), b_{G_n,\lambda}(x)\big] \\ &\quad = \arg\min_{[a(x),b(x)]}\int \big[\{b(x)-a(x)\}_+ \\ &\qquad - \lambda I_{\{a(x)\le\theta\le b(x)\}}\big]G_n(d\theta|x)\end{aligned} \tag{2}$$
with a λ satisfying
$$\iint I_{\{a_{G_n,\lambda}(x)\le\theta\le b_{G_n,\lambda}(x)\}} \times f(x|\theta)G_n(d\theta)\nu(dx) = 1-\alpha.$$

However, the constraint in (2) still involves an integration with respect to the joint measure.

Wenhua Jiang is Young Investigator, School of Mathematical Sciences, Fudan University and Shanghai Center for Mathematical Sciences, 220 Handan Road, Shanghai 200433, China (e-mail: wenhuajiang@yahoo.com).

In the finite Bayes setting, where θ_i are i.i.d. variables from a (nonempirical) prior G_n, one may solve instead of (2) the HPD (highest posterior density) credible interval problem

$$\begin{aligned}&[a_{G_n}(x), b_{G_n}(x)]\\ &\quad = \underset{[a(x),b(x)]}{\arg\min} \int \{b(x)-a(x)\}_+ G_n(d\theta|x) \\ &\quad \text{subject to} \int I_{\{a(x)\le\theta\le b(x)\}} G_n(d\theta|x) = 1-\alpha\end{aligned} \tag{3}$$

based on the posterior alone. Indeed, the credible interval (3) provides a feasible solution to the optimization problem (1), but the solution is unfortunately suboptimal in general. Nevertheless, due to the optimality of (3) under the more stringent posterior coverage constraint, we may still use its performance as a benchmark to gauge the performance of empirical Bayes interval estimators under the two criteria in (1). From this point of view, the Bayes credible interval has a frequentist face. We shall call (3) the oracle Bayes rule as it is expected to outperform empirical Bayes methods due to its dependence on the oracular knowledge of G_n and its optimality in the sense of (3).

We shall discuss now empirical Bayes methods, that is, interval estimators based on estimated priors $\hat{G}_n$. For simplicity, we confine this discussion to the normal case where $X_i|\theta_i \sim N(\theta_i, 1)$. Efron's paper offers interesting insights into Morris' (1983) approach with normal prior, the broader g-modeling methods (Efron, 2014, 2016) and an ingenious bootstrap correction (Laird and Louis, 1987). We shall add the general maximum likelihood empirical Bayes (GMLEB, Jiang and Zhang, 2009) to this mix in our simulation study.

We note that the empirical Bayes interval estimation problem does not admit simple f-modeling in the nonparametric setting in general, at least for one-sided credible intervals, as $G_n(d\theta)$ can be recovered by differentiating $1-\alpha = \int_{-\infty}^{t} G_n(d\theta|x)$.

An interesting aspect of the Gaussian prior $G_n \sim N(0, A)$ is that the oracle Bayes rule (3) is also an optimal solution of the compound problem in (1). This can be seen as follows. When $G_n \sim N(0, A)$, both the interval length and posterior coverage probability in (3) are not dependent on x, so that the credible interval $[a_{G_n}(x), b_{G_n}(x)]$ is identical to the oracle solution $[a_{G_n,\lambda}(x), b_{G_n,\lambda}(x)]$ in (2) for a fixed $\lambda = \lambda_\alpha$ not depending on x. Thus, when $G_n \approx N(0, A)$ and $\hat{G}_n \approx N(0, A)$, the plug-in empirical Bayes credible interval $[a_{\hat{G}_n}(x), b_{\hat{G}_n}(x)]$ is an approximately optimal solution of the compound problem (1). However, when $G_n \approx N(0, A)$ fails to hold, parametric estimates of G_n may not provide sufficient coverage probability.

Let $f_G(x) = \int f(x|\theta)G(d\theta)$ denote the mixture of the density $f(x|\theta)$ with prior G. In the GMLEB approach, we simply replace the empirical prior in the oracle rule with the generalized maximum likelihood estimator (GMLE, Kiefer and Wolfowitz, 1956),

$$\hat{G}_n = \arg\max_G \prod_{i=1}^{n} f_G(X_i).$$

This can be viewed as nonparametric g-modeling as no constraint is imposed on G in the above maximization. Corresponding to the oracle rule (3), the resulting HPD credible interval is

$$\begin{aligned}&[a_{\hat{G}_n}(x), b_{\hat{G}_n}(x)]\\ &\quad = \underset{[a(x),b(x)]}{\arg\min} \int \{b(x)-a(x)\}_+ \hat{G}_n(d\theta|x) \\ &\quad \text{subject to} \int I_{\{a(x)\le\theta\le b(x)\}} \hat{G}_n(d\theta|x) = 1-\alpha.\end{aligned} \tag{4}$$

The computation of the credible interval (4) is a delicate matter. The GMLE $\hat{G}_n$ is usually implemented by the EM algorithm: putting a large number of equally spaced grids in the range of observations and updating the weights in iterations. Koenker and Mizera (2014) proposed a convex optimization approach to computing the GMLE, which reduces the computational effort by several orders of magnitude. The efficiency of their algorithm (R package REBayes) has been demonstrated in Koenker and Mizera (2014) and later research (e.g., Jiang and Zhang, 2016). In Figure 1, $n = 1000$ parameters θ_i were drawn from $N(0, 1)$ distribution. The GMLE seems to overfit as the solution calculated by the REBayes package is supported on three points (red dashed). The EM algorithm with 100 iterations yields a smooth density $\hat{g}(\theta)$ (green dotted). This seems to suggest early stopping of the EM or a penalized EM with the objective function

$$-\sum_{i=1}^{n} \log f_G(X_i) + \lambda \int (g''(x))^2\, dx \tag{5}$$

as a smooth version of the GMLE. An early success story of such smooth EM is its application in tomography (Vardi, Shepp and Kaufman, 1985). In

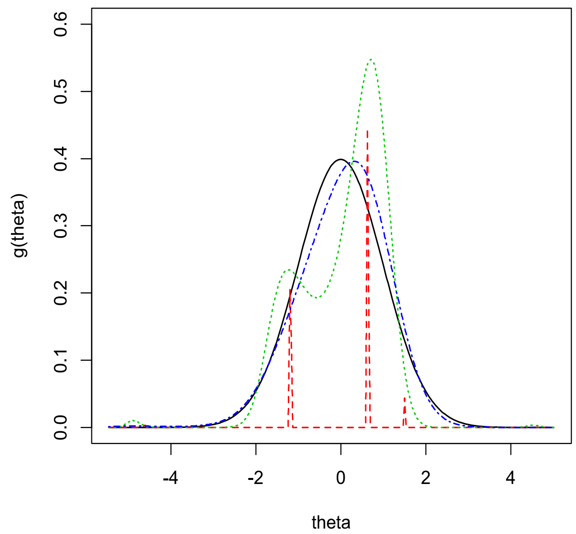

FIG. 1. *Black solid line represents the density of prior* $N(0,1)$. *Red dashed is the estimated prior density* $\hat{g}(\theta)$ *by the REBayes package*; *green dotted curve from the* 100 *EM-iterations*; *blue dotdash curve from the minimum of penalized log-likelihood* (5) *among the first* 100 *EM-iterations*.

Figure 1, the blue curve represents the $\hat{g}(\theta)$ from the minimum of penalized log-likelihood (5) among the first 100 EM-iterations. It is much closer to the true $N(0,1)$ prior density, that is, it controls overfitting. Actually, the minimum appeared at a pretty early stage.

SOME NUMERICAL RESULTS

Here we evaluate the coverage probability of several empirical Bayes credible intervals. Two omnibus measures of accuracy: the average coverage probability and average length,

$$\text{(6)}\qquad \frac{1}{n}\sum_{i=1}^{n} I\{a(X_i) \le \theta_i \le b(X_i)\}, \qquad \frac{1}{n}\sum_{i=1}^{n}\{b(X_i) - a(X_i)\}_+$$

are reported. The classical $100(1-\alpha)\%$ confidence interval is $x_i \pm z_{\alpha/2}$, where $z_{\alpha/2}$ is the upper $(\alpha/2)$-point of the standard normal distribution. The James–Stein-based credible interval is

$$\text{(7)}\qquad \hat{\theta}_i^{\text{JS}} \pm z_{\alpha/2}\{1 - ((k-1)/k)\hat{B}\}^{1/2},$$

where $\hat{\theta}_i^{\text{JS}} = (1-\hat{B})x_i + \hat{B}\bar{x}$, $\hat{B} = ((k-3)/(k-1))/(1+\hat{A}_+)$, $\hat{A} = S/(k-1) - 1$ and $S = \sum_{i=1}^{n}(x_i - \bar{x})^2$. Morris' (1983) correction to (7) gives a wider interval $\hat{\theta}_i^{\text{JS}} \pm z_{\alpha/2}s_i$ where $s_i^2 = 1 - ((k-1)/k)\hat{B} + (2/(k-3))\hat{B}^2(x_i - \bar{x})^2$. For the GMLEB and the g-modeling, we report their HPD intervals. As a benchmark, the oracle interval $[a_{G_n}(X_i), b_{G_n}(X_i)]$ in (3) is included, where $G_n(u) = n^{-1}\sum_{i=1}^{n} I\{\theta_i \le u\}$ is the empirical distribution of the unknown means. Throughout the simulation we set $\lambda = 20$ for the GMLEB. We used R package deconvolveR with default settings to compute the g-modeling intervals.

We first use a lognormal-normal example. We drew $n = 25, 100, 400$ and 1600 values of θ_i from a lognormal$(0,1)$ distribution and n corresponding observations $X_i \overset{\text{ind}}{\sim} N(\theta_i, 1)$. Each interval uses a nom-

TABLE 1
Coverage probabilities and lengths of empirical Bayes credible intervals with nominal coverage level of 95%. *The true prior is* lognormal(0, 1)

	25		100		400		1600	
n	C	L	C	L	C	L	C	L
Classical	0.949	3.920	0.948	3.920	0.951	3.920	0.950	3.920
James-Stein	0.936	3.250	0.950	3.422	0.953	3.535	0.953	3.544
James-Stein corrected	0.944	3.314	0.951	3.433	0.954	3.537	0.953	3.544
G-modeling	0.941	3.145	0.954	3.153	0.961	3.298	0.957	3.415
GMLEB	0.940	3.069	0.955	3.082	0.961	2.991	0.963	2.894
Oracle	0.961	2.227	0.950	2.495	0.951	2.545	0.951	2.566

TABLE 2
Coverage probabilities and lengths of the corrected g-modeling and GMLEB credible intervals by the bootstrap algorithm

n	15		25	
	C	L	C	L
G-modeling	0.933	3.125	0.941	3.145
G-modeling corrected	0.951	3.507	0.960	3.521
GMLEB	0.927	2.978	0.940	3.069
GMLEB corrected	0.953	3.357	0.959	3.385

inal coverage level of 95%. We report the average coverage probability and the average interval length based on 100 replications. The results are summarized in Table 1. The classical confidence intervals are guaranteed to attain the nominal level. The widths of them are all 3.920 under 95% level. The uncorrected James–Stein intervals attain the nominal level except $n = 25$. Morris' correction works well for small n. Both James–Stein intervals are more compact than the classical intervals. Because the lognormal(0, 1) distribution is quite asymmetry, it is expected that there is still room to improve upon the interval length. This is achieved by the g-modeling and the GMLEB credible intervals. Especially, the GMLEB provides much more compact intervals although the level of coverage is slightly conservative. For small sample sizes, the g-modeling and the GMLEB intervals could be corrected by the bootstrap procedure as in Section 6 of Efron's paper. In Table 2, we display the corrections for $n = 15$ and 25 based on $B = 500$ bootstrap replications. It is clear that the bootstrap correction improves the coverage.

A more difficult case is as follows. The prior distribution is a mixture: 90% of a degenerate distribution at zero and 10% of a uniform distribution over $[-3, 3]$. The g-modeling can accommodate the possible big atom at $\theta = 0$ by including in model matrix Q a column $\boldsymbol{e}_0 = (0, \dots, 1, \dots, 0)$, with the 1 at $\theta = 0$. The GMLEB can also take into consideration of the sparsity. We initialized the weight at zero by a Fourier method as in Jiang and Zhang (2009) and put the remaining weights uniformly. In Table 3, the g-modeling method gives more precise estimates than the James–Stein in terms of interval size, while the performance of the GMLEB is stronger, both benefiting from a good starting mass at 0.

ACKNOWLEDGEMENT

The author would like to express his sincere thankfulness to Professor Cun-Hui Zhang to pointing out the oracle Bayes rule in (2).

TABLE 3
Coverage probabilities and lengths of empirical Bayes credible intervals with nominal coverage level of 95%, $\theta_i \sim 0.9\delta_0 + 0.1\,\mathrm{Unif}[-3, 3]$, $i = 1, \dots, n$

n	200		400		800		1600	
	C	L	C	L	C	L	C	L
Classical	0.948	3.920	0.949	3.920	0.950	3.920	0.950	3.920
James–Stein	0.943	1.810	0.944	1.856	0.942	1.851	0.941	1.859
James–Stein corrected	0.944	1.846	0.945	1.873	0.943	1.859	0.941	1.863
G-modeling	0.968	1.608	0.967	1.529	0.967	1.483	0.965	1.495
GMLEB	0.946	1.092	0.951	1.104	0.955	1.050	0.956	1.017
Oracle	0.963	1.019	0.962	1.036	0.963	1.032	0.962	1.048

REFERENCES

EFRON, B. (2014). Two modeling strategies for empirical Bayes estimation. *Statist*. *Sci*. **29** 285–301. MR3264543

EFRON, B. (2016). Empirical Bayes deconvolution estimates. *Biometrika* **103** 1–20. MR3465818

HANNAN, J. F. and ROBBINS, H. (1955). Asymptotic solutions of the compound decision problem for two completely specified distributions. *Ann*. *Math*. *Stat*. **26** 37–51. MR0067444

JIANG, W. and ZHANG, C.-H. (2009). General maximum likelihood empirical Bayes estimation of normal means. *Ann*. *Statist*. **37** 1647–1684. MR2533467

JIANG, W. and ZHANG, C.-H. (2016). Generalized likelihood ratio test for normal mixtures. *Statist*. *Sinica* **26** 955–978. MR3559938

KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann*. *Math*. *Stat*. **27** 887–906. MR0086464

KOENKER, R. and MIZERA, I. (2014). Convex optimization, shape constraints, compound decisions, and empirical Bayes rules. *J*. *Amer*. *Statist*. *Assoc*. **109** 674–685. MR3223742

LAIRD, N. M. and LOUIS, T. A. (1987). Empirical Bayes confidence intervals based on bootstrap samples. *J*. *Amer*. *Statist*. *Assoc*. **82** 739–757. MR0909979

MORRIS, C. N. (1983). Parametric empirical Bayes inference: Theory and applications. *J*. *Amer*. *Statist*. *Assoc*. **78** 47–65. MR0696849

ROBBINS, H. (1951). Asymptotically subminimax solutions of compound decision problems. In *Proc*. *Second Berkeley Symp*. *Math*. *Statist*. *Probab*. **1** 131–148. Univ. California Press, Berkeley, CA. MR0044803

ROBBINS, H. (1956). An empirical Bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955 **1** 157–163. Univ. California Press, Berkeley and Los Angeles. MR0084919

VARDI, Y., SHEPP, L. A. and KAUFMAN, L. (1985). A statistical model for positron emission tomography. *J*. *Amer*. *Statist*. *Assoc*. **80** 8–20. MR0786595