JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2022, VOL. 117, NO. 539, 1183–1185: Theory and Methods Discussion
https://doi.org/10.1080/01621459.2022.2096039

Discussion of "Confidence Intervals for Nonparametric Empirical Bayes Analysis"

Marianna Pensky

Department of Mathematics, University of Central Florida, Orlando, FL

1. Introduction

We would like to start with congratulating the authors. Empirical Bayes estimation is a very old, well studied problem. However, construction of confidence intervals in empirical Bayes setting has been neglected, in spite of the fact that, in the majority of practical situations, one is interested in confidence bounds rather than point estimators.

The authors present several procedures for construction of confidence intervals, such as simultaneous confidence intervals via F-localization and AMARI confidence intervals for specific values of z. They provide general constructions of the confidence intervals and study their lengths and coverage probabilities. One of the great successes of the paper is that it offers algorithms in the case of a conditional distribution of a general form. The theoretical results are stated in asymptotic form, so that adequate coverage is guaranteed only as the number of observations tends to infinity. Subsequently, Ignatiadis and Wager examine separately the most important cases where the conditional distribution $P(z|\mu)$ belongs to the binomial, the Poisson or the Gaussian family. This investigation reveals, how much the construction of the confidence intervals and their lengths depend on the conditional distribution $P(z|\mu)$ as well as the class of prior densities $\mathcal{G}$.

This does not come as a surprise since the empirical Bayes estimation problem is an ill-posed problems. For this reason, we feel that, while construction of confidence intervals in the case of a generic conditional distribution is a very valuable undertaking, one can gain undeniable advantages by constructing confidence intervals separately for specific conditional distribution families. The latter also highlights what choices of $\mathcal{G}$ is most appropriate. Below, we consider construction of confidence intervals in the cases of the binomial, the Poisson and the Gaussian families when $h(\mu) = \mu^m$. If point or interval estimators are available for $1 \le m \le M$, then, due to linearity of $\theta_G(z)$ with respect to $h(\mu)$, the latter will allow to obtain point or interval estimators for $\theta_G(z)$ when $h(\mu) \in \mathscr{P}(M)$, where $\mathscr{P}(M)$ is the set of polynomials of degree at most M.

It is easy to see that one can construct a point estimator for $\theta_G(z) = a_G(z)/f_G(z)$ by estimating separately the top and the bottom of the fraction (see, e.g., Pensky 1997) and taking into account that $f_G(z) = \mathbb{P}(Z = z)$ in the discrete case, and $f_G(z)$ is the pdf of Z at z in the continuous case. In addition, for estimating $a_G(z)$, one can use the following statement:

Proposition 1 (Pensky 1997, 2002). Let there exists a function $\psi_z(y)$ such that, for every y, z, and μ

$$\int p(y|\mu)\psi_z(y)d\nu(y) = p(z|\mu)h(\mu), \tag{1}$$

where $\nu(y)$ is the Riemann-Stiltjes measure. Then,

$$a_G(z) = \int p(z|\mu)h(\mu)dG(\mu) = \int \psi_z(y)f_G(y)d\nu(y) = \mathbb{E}_G\left[\psi_z(Z)\right]. \tag{2}$$

Finally, combining the steps, one can obtain confidence intervals for $a_G(z) = \mathbb{E}_G\left[\psi_z(Z)\right]$ and $f_G(z)$, and combine them, using the following simple lemma for construction of the confidence interval of the fraction:

Lemma 1. Let $\widehat{a_G}(z)$ and $\widehat{f_G}(z)$ be such that $\mathbb{P}\left\{|\widehat{a_G}(z) - a_G(z)| \le |a_G(z)|\,\delta_1(z)\right\} \ge 1 - \alpha/2$ and $\mathbb{P}\left\{|\widehat{f_G}(z) - f_G(z)| \le |f_G(z)|\,\delta_2(z)\right\} \ge 1 - \alpha/2$. Let $\widehat{\theta_G}(z) = \widehat{a_G}(z)/\widehat{f_G}(z)$. If $\delta_1(z) < 1/2$, then

$$\mathbb{P}\left\{|\widehat{\theta_G}(z) - \theta_G(z)| \le 2\,|\widehat{\theta_G}(z)|\,(\delta_1(z) + \delta_2(z))\right\} \ge 1 - \alpha.$$

2. Confidence Intervals for the Binomial Family

The case when $p(z|\mu) \sim \text{Binomial}(N, \mu)$, $z = 0, \ldots, N$, is perhaps the one which brings the ill-posedness of the problem to light. Indeed, since one has only $(N + 1)$ distinct values of $f_G(z)$, infinite-dimensional classes of priors $\mathcal{G}$ will likely lead to ambiguity in the value of θ_G. One of the natural sets for $\mathcal{G}$ is the set of polynomials $\mathscr{P}(N)$ of degree at most N. Specifically, the following statement is valid.

Lemma 2. Let $h(\mu) = \mu^m$ where $m \ge 1$. Let $g(\mu) = g_1(\mu) + g_2(\mu)$ where $g_1 \in \mathscr{P}(N)$ and $g_2 \perp \mathscr{P}(N)$, orthogonal to $\mathscr{P}(N)$. Then, there exists $g_2(\mu)$ such that $\int_0^1 p(z|\mu)g_2(\mu)d\mu = 0$ for

CONTACT Marianna Pensky Marianna.Pensky@ucf.edu Department of Mathematics, University of Central Florida, Orlando, FL 32817.

$z = 0, \ldots, N$, but $\int_0^1 h(\mu)p(z|\mu)g_2(\mu)d\mu \neq 0$ for some $z \in \{0, 1, \ldots, N\}$.

Validity of the lemma follows from the fact that $p(z|\mu) \in \mathscr{P}(N)$, while $h(\mu)p(z|\mu) \in \mathscr{P}(N+m)$ for $z = 0, 1, \ldots, N$.

Similar statements can be proved for $(N+1)$-dimensional spaces other than $\mathscr{P}(N)$. The latter means that, unless one fixes an $(N+1)$-dimensional space $\mathcal{G}$, the empirical Bayes estimator $\theta_G(z)$ is not identifiable. On the other hand, as soon as such space is fixed, one can construct confidence bounds for $a_G(z)$ using standard techniques used in the finite-dimensional linear regression problems.

In conclusion, the empirical Bayes estimation problem is very ill-posed in the case of the binomial distribution and cannot be solved without imposing finite-dimensional constraints on the set $\mathcal{G}$.

3. Confidence Intervals for the Poisson Family

While the binomial case leaves one wondering whether empirical Bayes estimation problem is always so hard, the Poisson case presents at example, where the problem is relatively well-posed and can be solved without imposing assumptions on class $\mathcal{G}$.

In the Poisson case, ν is a counting measure and equation (1) with $h(\mu) = \mu^m$ yields $\psi_z(y) = I(y = z + m)$, so that $\theta_G(z) = f_G(z+m)/f_G(z)$. Hence, given n observations $Z_1, \ldots, Z_n$ on Z, the problem of construction of a confidence interval for $\theta_G(z)$ reduces to construction of confidence intervals for the binomial probability p, where $p = f_G(z+m)$ or $p = f_G(z)$. The latter is a well studied problem (see, e.g., Brown, Cai, and DasGupta 2001; Brown, Cai, and DasGupta 2002). For example, if $\kappa = q_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the normal distribution, $p = f_G(z+k)$ and $\widehat{p} = n^{-1}\sum_{i=1}^n I(Z_i = z+k)$, where $k = m$ for $\widehat{a_G}(z) = \widehat{f_G}(z+m)$ and $k = 0$ for $\widehat{f_G}(z)$, and

$$V(p, n, \kappa) = \widehat{p} - 0.5|\kappa n^{-1/2} + \sqrt{\widehat{p}(1-\widehat{p}) + 0.25\,\kappa^2 n^{-1}},$$

then, the Wilson confidence interval (Brown, Cai, and DasGupta 2001) yields

$$\frac{|\widehat{p} - p|}{p} \leq \frac{\kappa}{\sqrt{n}} \frac{V(p, n, \kappa)}{\left[\widehat{p} - \kappa\, n^{-1/2}(1 + \kappa^2 n^{-1})^{-1}\, V(p, n, \kappa)\right]} = \frac{\kappa}{\sqrt{n}} \frac{\sqrt{\widehat{p}(1-\widehat{p})}}{\widehat{p}} + O\left(\frac{1}{n}\right) \tag{3}$$

Combination of (3) and Lemma 1 leads to the following $(1-\alpha)$-confidence interval

$$|\widehat{\theta_G}(z) - \theta_G(z)| \leq \frac{2\kappa\,\widehat{\theta_G}(z)}{\sqrt{n}} \left(\frac{\sqrt{\widehat{f_G}(z+m)(1-\widehat{f_G}(z+m))}}{\widehat{f_G}(z+m)} + \frac{\sqrt{\widehat{f_G}(z)(1-\widehat{f_G}(z))}}{\widehat{f_G}(z)} \right) + O\left(\frac{1}{n}\right)$$

Note that construction of this interval does not involve any assumptions on the family of priors $\mathcal{G}$, which confirms that the case of the Poisson conditional distribution is very mildly ill-posed.

4. Construction Ideas in the Gaussian Case

In the case of the normal distribution, the ill-posedness of the problem appears as a requirement of estimating derivatives of the pdf $f_G(z)$ on the basis of observations $Z_1, \ldots, Z_n$. Denote the standard normal pdf by $\phi(x)$ and observe that, for $h(\mu) = \mu^m$, one has

$$a_G(z) = B_m(z) \equiv \int_{-\infty}^{\infty} \mu^m \phi(z - \mu) dG(\mu) \tag{4}$$

It turns out that $B_m(z)$ can be represented as a linear combination of the derivatives $f_G^{(j)}(z)$, $j = 0, \ldots, m$, of the pdf $f_G(z)$, with the coefficients being polynomials in z. Indeed, $B_0(z) = f_G(z)$ and it is well known that $B_1(z) = f_G'(z) + z f_G(z)$, so the statement is true for $m = 0$ and $m = 1$. One can show that a similar representation holds for any value of m.

Lemma 3. For any $m = 0, 1, \ldots$ and $B_m(z)$ introduced in (4), one has

$$B_m(z) = \sum_{l=0}^{m} Q_l^{(m)}(z) f_G^{(l)}(z) \tag{5}$$

Here, $Q_l^{(m)} \in \mathscr{P}(m-l)$, $l = 0, 1, \ldots, m$, are the degree $(m-l)$ polynomials in z, defined as follows: $Q_m^{(m)}(z) = 1$ and

$$Q_l^{(m)}(z) = \sum_{j=l}^{m-1} \binom{m}{j} (-1)^{m-j-1} H_{m-j}(z)\, Q_l^{(j)}(z), \qquad 0 \leq l \leq m-1 \tag{6}$$

where $H_M(x)$ is the Mth probabilistic Hermite polynomial, given by

$$H_M(x) = (-1)^M e^{\frac{x^2}{2}} \frac{d^M}{dx^M}\left(e^{-\frac{x^2}{2}}\right), \quad M = 0, 1, \ldots$$

Proof. Formula (5) can be proved by induction. Certainly, (5) holds for $m = 1$. Assume that (5) is correct for $m = 1, \ldots, M-1$. Note that (see, Abramowitz and Stegun 1964, (22.5.18)), for any $M = 0, 1, \ldots$, the Mth derivative of $f_G(z)$ can be written as

$$f_G^{(M)}(z) = (-1)^M \int_{-\infty}^{\infty} H_M(y - \mu)\phi(\mu) dG(\mu), \quad M = 0, 1, \ldots \tag{7}$$

Due to the equations (22.5.19) of Abramowitz and Stegun (1964) and (8.958.2) of Gradshteyn and Ryzhik (2007), derive

$$H_M(y - \mu) = \sum_{k=0}^{M} \binom{M}{k} (-\mu)^{M-k} H_k(y)$$

Substituting this expression into (7), obtain

$$f_G^{(M)}(z) = B_M(z) + \sum_{k=1}^{M} \binom{M}{k} (-1)^k H_k(z)\, B_{M-k}(z).$$

Rearranging the last formula and plugging in the value of $B_{M-k}(z)$ from (5), due to induction assumption, derive that

$$B_M(z) = f_G^{(M)}(z) + \sum_{k=1}^{M} \binom{M}{k} (-1)^{k-1} H_k(z) \sum_{l=0}^{M-k} Q_l^{(M-k)} f_G^{(l)}(z) \tag{8}$$

Finally, introducing $j = M - k$ and changing the order of summation in (8), obtain (6) with $m = M$, so Lemma holds for $m = M$. □

At last, we need to show that $Q_l^{(m)} \in \mathscr{P}(m-l)$ for $l \le m$. It is easy to see that the statement holds for $m = 1$. We assume that $Q_l^{(m)} \in \mathscr{P}(m-l)$ for $l \le m \le M-1$, and show that the same is true for $m = M$. Note that it immediately follows from (8) that $Q_M^{(M)}(z) = 1 \in \mathscr{P}(0)$. Moreover, since $H_{M-j}(z) \in \mathscr{P}(M-j)$ and $Q_l^{(j)}(z) \in \mathscr{P}(j-l)$, by examining (6) with $m = M$ and $l \le M-1$, we obtain that $Q_l^{(M)}(z) \in \mathscr{P}(M-j+j-l) = \mathscr{P}(M-l)$, which completes the proof.

Lemma 3 asserts that $a_G(z)$ is linear combination of $f_G(z)$ and its derivatives. Note that construction of a confidence interval for a nonparametric density function and its jth derivative is a standard problem that has been investigated previously in, e.g., Gine and Nickl (2010) and Chen (2017). In general, the choice of a kernel (or a wavelet basis if a wavelet estimator is employed) depends on the Sobolev ball $W_s(A)$, to which $f_G^{(j)}$ belongs. In a generic nonparametric setting, adapting to an unknown Sobolev space is a difficult problem. However, it is a lot easier if $f_G^{(j)}$ is given by Equation (7).

Consider the case where distribution G has a density function g. For any function q denote its Fourier transform $\mathcal{F}[q](\omega)$ at ω by $q^*(\omega)$. Then, using the fact that $\mathcal{F}[H_j * \phi](\omega) = (i\omega)^k \exp(-\omega^2/2)$ (see (7.374.6) of Gradshteyn and Ryzhik (2007)), one obtains that Fourier transform of $f_G^{(j)}$ is $(f_G^{(j)})^*(\omega) = (i\omega)^j \exp(-\omega^2/2) g^*(\omega)$, where $g^*(\omega) = \mathcal{F}[g](\omega)$. Hence,

$$f_G^{(j)} \in W_s(A) \Longleftrightarrow \int_{-\infty}^{\infty} \omega^{2j}(\omega^2+1)^s e^{-\omega^2} |g^*(\omega)|^2 d\omega \le A^2$$

The latter implies that $f_G^{(j)} \in W_s(A)$ with $A \equiv A(s,j) = \max\left\{(\omega^2+1)^{s+j}e^{-\omega^2}\right\}$ for any $s > 0$. In addition, $(f_G^{(j)})^*$ has an exponential decay, for example, $|(f_G^{(j)})^*(\omega)| \exp(\omega^2/4) \le A(j)$ where $A(j) = \max\left\{(\omega^2+1)^j \exp(-\omega^2/4)\right\}$. Adapting to the exponential decay of a density function or its derivatives requires using kernels (or wavelets) with unbounded supports. For $n \to \infty$, estimators based on those kernels (or wavelets) have better asymptotic properties. However, when n is finite, those kernels may be outperformed by kernels (or wavelets) with finite support, that are adapted to Sobolev spaces with finite s. One can try various values of s and then choose the confidence intervals with the shortest length. Methodology of Gine and Nickl (2010) allows one to obtain asymptotic convergence rates for the interval estimators for each value of s.

Unfortunately, due to our regrettable lack of knowledge of Julia language and the short time window, we could not compare the proposed approach to the confidence intervals in the paper.

5. Classes of Prior Densities

Our discussion above should have convinced the reader that construction of empirical Bayes confidence intervals depend considerably on the choice of the class of prior distributions $\mathcal{G}$. Since this class is unknown, one would like to have a broad range of choices for $\mathcal{G}$, unless identifiability issues are of a major concern, as it happens in the case of the binomial distribution. In particular, one would like to construct confidence intervals that are adaptive to unknown nonparametric classes. As one can see from our discussion, this is possible in the case of the Poisson or the Gaussian conditional distribution family. One of the properties of such nonparametric classes is that they are broad enough, so misspecification of $\mathcal{G}$ stops being an issue. On the flip side, for the same reason, those families are not well represented by a finite subset of $\mathcal{G}$. The present paper studies much more narrow choices of $\mathcal{G}$. For example, a Gaussian location mixture such as $\mathcal{LN}(0.25^2, [-4,4])$, used in simulations, can be easily approximated by a finite-dimensional family of distributions by discretizing the interval $[-4,4]$.

In addition, it is easy to see that, for $g \in \mathcal{LN}(0.25^2, [-4,4])$, one has $|g^*(\omega)| = \exp(-\omega^2/32)$, so this choice leads to $g^*(\omega)$ with the exponential decay. After close examination, one discovers that all families $\mathcal{G}$ studied in the paper (specifically, $\mathcal{LN}$ in (34), $\mathcal{SN}$ in (35), G^{Spiky} and G^{NegSpiky} in (36)) are comprised of prior densities with $|g^*(\omega)| \le \exp(-a\omega^2)$ for some $a > 0$, and, hence, represent "the best case scenario" as far as the inference for the Gaussian conditional density is concerned. For this reason, the reference to low convergence rates in Pensky (2017) is unsuitable, since Pensky (2017) only considered densities whose Fourier transforms $g^*(\omega)$ have polynomial decay as $|\omega| \to \infty$. By following calculations in Pensky (2017), one can easily observe that convergence rates improve whenever $|g^*(\omega)|$ has an exponential decay. Nevertheless, if this assumption is unfounded, it would lead to wrong conclusions.

The discussion above is by no means a criticism of the paper by Ignatiadis and Wager but rather a deliberation of the further directions in the construction of confidence intervals. We hope that the present paper will encourage the long overdue research on interval estimators and hypothesis testing in the empirical Bayes setting.

Funding

Marianna Pensky gratefully acknowledges support by the National Science Foundation through grant no. DMS-2014928.

References

Abramowitz, M., and Stegun, I. A. (1964), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th Dover printing, 10th GPO printing ed. New York: Dover. [1184]

Brown, L. D., Cai, T. T., and DasGupta, A. (2001), "Interval Estimation for a Binomial Proportion," *Statistical Science*, 16, 101–133. [1184]

——— (2002), "Confidence Intervals for a Binomial Proportion and Asymptotic Expansions," *The Annals of Statistics*, 30, 160–201. [1184]

Chen, Y.-C. (2017), "A Tutorial on Kernel Density Estimation and Recent Advances," *Biostatistics & Epidemiology*, 1, 161–187. [1185]

Gine, E., and Nickl, R. (2010), "Confidence Bands in Density Estimation," *The Annals of Statistics*, 38, 1122–1170. [1185]

Gradshteyn, I. S., and Ryzhik, I. M. (2007), *Table of Integrals, Series, and Products*, 7th ed. Amsterdam: Elsevier/Academic Press. [1184,1185]

Pensky, M. (1997), "A General Approach to Nonparametric Empirical Bayes Estimation," *Statistics*, 29, 61–80. [1183]

——— (2002), "Locally Adaptive Wavelet Empirical Bayes Estimation of a Location Parameter," *Annals of the Institute of Statistical Mathematics*, 54, 83–99. [1183]

——— (2017), "Minimax Theory of Estimation of Linear Functionals of the Deconvolution Density with or without Sparsity," *The Annals of Statistics*, 45, 1516–1541. [1185]