A fiducial continuum from confidence sets to empirical Bayes set estimates as the number of comparisons increases

David R. Bickel

December 31, 2014

Ottawa Institute of Systems Biology
Department of Biochemistry, Microbiology, and Immunology
University of Ottawa
451 Smyth Road
Ottawa, Ontario, K1H 8M5

+01 (613) 562-5800, ext. 8670
dbickel@uottawa.ca

Abstract

Methods based on fiducial inference to hierarchical models that are formulated for set estimation propagate uncertainty to varying degrees. Some of the set estimates reduce to the usual confidence sets for a sufficiently low number of comparisons.

Keywords: coherent fiducial distribution; confidence distribution; empirical Bayes; fiducial inference; foundations of statistics; hierarchical model; interpretation of probability; local false discovery rate

1 Introduction

Many frequentists strive to approach each statistical problem with appropriate tools without feeling constrained by any decision theory. For example, several statisticians apply the usual confidence intervals for low numbers of comparisons but find empirical Bayes methods useful for high numbers of comparisons (e.g., Efron, 2010).

Two problems confronting this eclectic approach to statistics result from its lack of a unifying theoretical foundation. First, there is typically no continuity between a p-value reported as a level of evidence for a hypothesis in the absence of the information needed to know or estimate relevant prior, and a posterior probability reported as a level for a hypothesis in the presence of such information. Second, the empirical Bayes methods recommended do not propagate the uncertainty due to estimating the prior.

Both problems are addressed by applying a coherent form of fiducial inference to hierarchical models (§2), yielding empirical Bayes set estimates that reflect uncertainty in estimating the prior (§3). The maximum likelihood version provides continuity from single comparisons to large numbers of comparisons, generalizing the above uses of empirical Bayes methods and of the p-values as inverse confidence sets (§3.1.2). The methods are illustrated by an application to a well-known data set (§4).

2 Preliminary parameter distributions

2.1 Confidence-based distributions

2.1.1 Confidence distributions

Let $\mathfrak{H}$ denote a measurable set of subsets of the space Θ of the interest parameter θ, which, with a nuisance parameter γ in a set Γ, determine $P_{\theta,\gamma}$, the probability distribution of an observable random n-vector X. A *procedure of nested confidence sets* for θ is any function $\mathcal{K} = \mathcal{K}_{\bullet}(\bullet, \bullet)$ that transforms an n-vector x in a sample space $\mathcal{X}$, a shape vector $s \in \mathcal{S}$ (Polansky, 2007, p. 4), and a *confidence level* $1 - \alpha$ to some set $\mathcal{K}_x(1-\alpha; s) \in \mathfrak{H}$ such that $1-\alpha_1 \leq 1-\alpha_2$ implies $\mathcal{K}_x(1-\alpha_1; s) \subseteq \mathcal{K}_x(1-\alpha_2; s)$ and such that the coverage probability of the *nested confidence set* $\mathcal{K}_X(1-\alpha; s)$ is equal to the confidence level:

$$P_{\theta,\gamma}(\theta \in \mathcal{K}_X(1-\alpha; s)) = 1 - \alpha \tag{1}$$

for all $\alpha \in [0,1]$ and $s \in \mathcal{S}$.

For any $x \in \mathcal{X}$, a *confidence distribution* is an observed (data-dependent) additive measure K_x on $(\Theta, \mathfrak{H})$ that satisfies $0 < K_x(\Theta) \leq 1$ and

$$K_x(\Theta_1) \in \{1 - \alpha : \alpha \in [0,1], s \in \mathcal{S}, \mathcal{K}_x(1-\alpha; s) = \Theta_1\} \tag{2}$$

for all $\Theta_1 \in \mathfrak{H}_x(\mathcal{K})$, where $\mathfrak{H}_x(\mathcal{K}) = \{\mathcal{K}_x(1-\alpha; s) : \alpha \in [0,1], s \in \mathcal{S}\}$ (Bickel and Padilla, 2014). For a scalar θ, $\mathfrak{H}_x(\mathcal{K}) = \mathfrak{H}$ is typically a Borel field (Bickel, 2012c), in which case a confidence distribution transforms the composite hypothesis that $\theta \in \Theta_1$ given any $\Theta_1 \in \mathfrak{H}$ to a number between 0 and 1 that is equal to the coverage probability $1-\alpha$ of the observed confidence set $\mathcal{K}_x(1-\alpha; s) = \Theta_1$ for any $s \in \mathcal{S}$. For an example with a vector θ, Bayesian posterior distributions with asymptotically probability matching priors (Datta and Mukerjee, 2004) are approximate confidence distributions. Equation (2) generalizes previous definitions of the confidence distribution, many of which appear in the literature cited by Nadarajah et al. (2015).

2.1.2 Coherent fiducial distributions

Confidence distributions lead to fiducial distributions that are genuine probability measures according to the framework introduced by Bickel and Padilla (2014), summarized as follows. A confidence distribution that is also a probability measure is called a *basic fiducial distribution*. It is called "basic" because it provides the foundation for a *coherent fiducial distribution* Π_x, which is any probability measure on $(\Theta, \mathfrak{H})$ that is a basic fiducial distribution, that is conditionally equal to one or more coherent fiducial distributions, or that is the probability measure of a transformation of a random variable distributed as another coherent fiducial distribution. Since a coherent fiducial distribution is not necessarily a confidence distribution, not all coherent fiducial distributions are basic.

The concept of the coherent fiducial distribution provides the ability to manipulate each basic fiducial distribution resulting from the given confidence set method as if it were a Bayesian posterior distribution resulting from a given prior distribution. Both coherent fiducial distributions and Bayesian posterior distributions are special cases of *certainty distributions*, each of which is a data-dependent probability measure on a measurable space corresponding to a parameter and the sets of parameters constituting the domain of the measure. Taking actions that minimize expected loss with respect to a coherent fiducial distribution or other certainty distribution avoids sure loss according to the classical theories of subjective probability (Bickel, 2012a; Bickel and Padilla, 2014). This avoidance of sure loss contrasts with the "noncoherence" inherent in a wide class of fiducial distributions defined according on the basis of confidence distributions according to the formulation of Wilkinson (1977).

2.1.3 Certainty density functions

For the observation $x \in \mathcal{X}$, let the probability density function c_x denote a Radon-Nikodym derivative of a certainty distribution C_x on $(\Theta, \mathfrak{H})$. If C_x is a coherent fiducial distribution or a Bayesian posterior distribution, $c_x(\theta)$ is called a *coherent fiducial density* or a *posterior density*, respectively, for all $\theta \in \Theta$.

2.2 Ideal hierarchical distributions

Many recent empirical Bayes estimators are based on the following hierarchical two-component mixture model of a data set that has been reduced to a single scalar statistic for each of N null hypotheses (Efron, 2010). Examples of such statistics include test statistics, p-values, and, as in Efron (2004), probit transformations of p-values. Let $\tau : \mathcal{X} \to \mathbb{R}$ denote a $\mathcal{X}$-measurable map, let f_0 and f_1 are probability density functions (PDFs) corresponding to the ith null and alternative hypotheses, respectively, and let $\pi_0 \in [0,1]$ and $\pi_1 = 1 - \pi_0$. For all $i \in \{1, \ldots, N\}$, let A_i denote a Bernoulli random variable such that $P(A_i = a) = \pi_a$ for $a = 0, 1$, and $x_i \in \mathcal{X}$, the observed statistic $t_i = \tau(x_i)$ associated with the null hypothesis that $\theta_i = \theta_0$ is assumed to be a realization of the random statistic T_i of probability density f_0 with probability π_0 and of probability density f_1 with probability π_1. Thus, each T_i has a marginal probability density function f such that $f(t) = \pi_0 f_0(t) + \pi_1 f_1(t)$ for all $t \in \mathcal{T}$. The corresponding N-tuples are $\mathbf{t} = (t_1, \ldots, t_N)$ and $\mathbf{T} = (T_1, \ldots, T_N)$.

The *local false discovery rate* (LFDR) for the ith statistic is defined as the Bayesian posterior probability that the ith null hypothesis is true (Efron, 2010):

$$\psi_i = P(A_i = 0 | T_i = t_i) = \frac{\pi_0 f_0(t_i)}{f(t_i)}. \tag{3}$$

For each $i = 1, \ldots, N$, Bickel (2012b) used the posterior probabilities of hypotheses' truths to define the random variable ϑ_i to be equal to some $\theta_0 \in \Theta$ if the ith null hypothesis is true or to be distributed with density function c_{x_i} otherwise:

$$\vartheta_i \sim c_{[i]} = P(A_i = 0|T_i = t_i)\,\delta_0 + P(A_i = 1|T_i = t_i)\,c_{x_i} = \psi_i\delta_0 + (1-\psi_i)\,c_{x_i}, \tag{4}$$

where δ_0 is the Dirac delta function with 100% probability mass at θ_0. Thus, $\int_{\mathcal{H}} c_{[i]}(\theta)\,d\theta = \psi_i 1_{\mathcal{H}}(\theta) + (1-\psi_i)\int_{\mathcal{H}} c_{x_i}(\theta)\,d\theta$ for any $\mathcal{H} \in \mathfrak{H}$, where $1_{\mathcal{H}}(\theta)$ indicates whether $\theta \in \mathcal{H}$.

Since the distribution of ϑ_i may be constructed from purely inferential distributions such as confidence distributions or default priors, its uncertainty does not necessarily represent physical variability in the system of study, as would be the case if the data were generated from ϑ_i. Rather, it is based on the principle that if $c_{[i]}$ would be the certainty distribution from which inferences or other actions would be derived were it known that $A_i = 0$, then the conditional certainty distribution determining actions must be equal to $c_{[i]}$ conditional on $A_i = 0$. This would be ensured by minimizing expected loss with respect to $c_{[i]}$, called the *ideal hierarchical density function*, in the same way that posterior expected loss is minimized with respect to Bayesian posterior densities. The PDF $c_{[i]}$ is "ideal" in that it is unknown because π_0 and f_1 are unknown. Alternatively, $c_{[i]}$ may be treated as a Bayesian posterior density function for the construction of set estimates such as fixed-tail interval estimates.

3 Empirical Bayes distributions and set estimates

3.1 Plug-in hierarchical distributions and set estimates

3.1.1 General plug-in concepts

The unknown probability π_0 and the unknown alternative hypothesis PDF f_1 may be estimated by $\widehat{\pi}_0$ and $\widehat{f}_1$, respectively. By substituting $\widehat{\pi}_0$ and $\widehat{f}_1$ into equation (3), this yields $\widehat{\psi}_i$ as the point estimate of ψ_i, the LFDR. The ideal hierarchical density function $c_{[i]}$ is then replaced by the *plug-in hierarchical density function*

$$\widehat{c}_{[i]} = \widehat{\psi}_i\delta_0 + \left(1-\widehat{\psi}_i\right)c_{x_i},$$

defined by substituting $\widehat{\psi}_i$ for ψ_i in (4). Bickel (2012b) made that substitution under the special case of a "theoretical null" histogram-based estimator of the LFDR (Efron, 2007). Sections 3.1.2 and 3.1.3 give other special cases.

A $(1-\alpha)$ *plug-in hierarchical set estimate* is a set $\widehat{C}_{[i]}(1-\alpha)$ such that $\int_{\widehat{C}_{[i]}(1-\alpha)} \widehat{c}_{[i]}(\theta)\,d\theta = 1-\alpha$ for any $\alpha \in [0,1]$. A example of such a set estimate for scalar parameters of interest is the $(1-\alpha)$ *fixed-tail interval* $\left[\widehat{\theta}_i(\beta_1), \widehat{\theta}_i(\beta_2)\right]$, defined such that $0 \le \beta_1 \le \beta_2 \le 1$, $\beta_2 - \beta_1 = 1-\alpha$, and $\beta = \int_{-\infty}^{\widehat{\theta}_i(\beta)} \widehat{c}_{[i]}(\theta)\,d\theta$ for all $\beta \in [0,1]$. In other words, $\widehat{\theta}_i$ is the quantile function of the random variable $\widehat{\vartheta}_i \sim \widehat{c}_{[i]}$ (Bickel, 2012b). The interval estimate is called *equal-tailed* if $\beta_1 = 1-\beta_2 = \alpha/2$.

The $(1-\alpha)$ *highest density sets* defined such that

$$\theta_1 \notin \widehat{C}_{[i]}(1-\alpha), \theta_2 \in \widehat{C}_{[i]}(1-\alpha) \implies \widehat{c}_{[i]}(\theta_1) < \widehat{c}_{[i]}(\theta_2)$$

apply to vector interest parameters. However, they lead to $\theta_0 \in \widehat{C}_{[i]}(1-\alpha)$ with probability 1 for all $i = 1, \dots, N$ unless $\alpha = 1$, $\pi_0 = 0$, or there is a $\theta \in \Theta$ such that $c_{x_i}(\theta) > 0$. Thus, a highest density set is not necessarily connected. A separation between the portion of the set estimate at θ_0 and a portion corresponding to f_1 may be advantageous (Efron, 2010, §11.4).

3.1.2 Hierarchical distributions based on maximum likelihood estimation

A simple class of parametric models of dimension $D \ge 2$ assumes there are parameter values $\xi_0, \dots, \xi_{D-1}$ in some set Ξ and a family $\{g(\bullet;\xi) : \xi \in \Xi\}$ of probability density functions on $\mathbb{R}$ such that $f_j = g(\bullet;\xi_j)$, with ξ_0 known and ξ_j unknown for $j = 1, \dots, D-1$. There are also unknown parameters $\pi_0, \dots, \pi_{D-1} \in [0,1]$ such that $\sum_{j=0}^{D-1} \pi_j = 1$. The *maximum likelihood estimate* (MLE) of $((\pi_0, \dots, \pi_{D-1}), (\xi_0, \dots, \xi_{D-1}))$ is

$$\left((\widehat{\pi}_0, \dots, \widehat{\pi}_{D-1}), \left(\widehat{\xi}_1, \dots, \widehat{\xi}_{D-1}\right)\right) = \arg \sup_{\pi_0,\dots,\pi_{D-1}\in[0,1],\xi_1,\dots,\xi_{D-1}\in\Xi:\sum_{j=0}^{D-1}\pi_j=1} L\left((\pi_0, \dots, \pi_{D-1}), (\xi_1, \dots, \xi_{D-1}); \mathbf{t}\right);$$

$$L\left((\pi_0, \dots, \pi_{D-1}), (\xi_1, \dots, \xi_{D-1}); \mathbf{t}\right) = \prod_{i=1}^{N} \sum_{j=0}^{D-1} \pi_j g(t_i;\xi_j). \tag{5}$$

On the basis of equation (3),

$$\widehat{\psi}_i^{\mathrm{MLE}} = \frac{\widehat{\pi}_0 g(t_i;\xi_0)}{\sum_{j=0}^{D-1} \widehat{\pi}_j g\left(t_i;\widehat{\xi}_j\right)} \tag{6}$$

is considered the MLE of ψ_i. Using three different families of densities, Padilla and Bickel (2012), Yang et al. (2013a), and Yang et al. (2013b) evaluated its performance relative to other point estimators of the LFDR for $D = 2$. This two-component MLE had been extended to arbitrary numbers of components (Pawitan et al., 2005; Muralidharan, 2010; Bickel, 2014).

Using any LFDR MLE $\widehat{\psi}_i^{\text{MLE}}$ as $\widehat{\psi}_i$ in Section 3.1.1 yields an *MLE hierarchical density function* and the corresponding $(1-\alpha)$ *MLE hierarchical set estimate*s. The following theorems indicate conditions under which the proposed methods degenerate to standard eclectic practice.

Theorem 1. *Consider the class of models given by Section 3.1.2 and the family* $\{g(\bullet;\xi) : \xi \in \Xi\}$ *consisting entirely of continuous density functions, where* Ξ *is a continuous vector space. If* $N = 1$*, then the MLE hierarchical set estimates are the same as the nested confidence sets with probability 1.*

Proof. By the assumptions that the density functions and Ξ are continuous, $\widehat{\pi}_0 = 0$ with probability 1. Thus, $\widehat{\psi}_i^{\text{MLE}} = 0$ with probability 1. □

3.1.3 Hierarchical distributions based on a hyperparameter confidence limit

Let $\widehat{\pi}_0(1-\varepsilon;\bullet)$ denote any function such that $\widehat{\pi}_0(1-\varepsilon;\mathbf{T})$ is a random variable satisfying

$$P_{\pi_0,\xi_0,\xi_1}\left(\widehat{\pi}_0(1-\varepsilon;\mathbf{T}) \le \pi_0\right) = \varepsilon \tag{7}$$

for all $\varepsilon, \pi_0 \in [0,1]$ and $\xi_0, \xi_1 \in \Xi$. It follows that $[0, \widehat{\pi}_0(1-\varepsilon;\mathbf{t})]$ is an observed $(1-\varepsilon)$ confidence interval of π_0. This, in analogy with equation (6), suggests

$$\widehat{\psi}_i(1-\varepsilon) = \frac{\widehat{\pi}_0(1-\varepsilon;\mathbf{t})\, g(t_i;\xi_0)}{\widehat{\pi}_0(1-\varepsilon;\mathbf{t})\, g(t_i;\xi_0) + (1-\widehat{\pi}_0(1-\varepsilon;\mathbf{t}))\, g\left(t_i;\widehat{\xi}_1\right)} \tag{8}$$

with $\varepsilon < 1/2$ as a conservative estimate of the LFDR of the ith test. Substituting $\widehat{\psi}_i(1-\varepsilon)$ for $\widehat{\psi}_i$ in Section 3.1.1 defines a $(1-\varepsilon)$*-limit hierarchical density function* and the corresponding $(1-\alpha)$ $(1-\varepsilon)$*-limit hierarchical set estimates.*

Approximate confidence intervals for π_0 may be computed using estimates of the variance of π_0 estimators bootstrap estimates of the variance of π_0 estimators. Such variance estimates have been determined by bootstrapping (Lai, 2007) and by a step-down procedure (Owen, 2005). The method presented in the following example instead relies on first-order likelihood theory.

Example 1. Under the model of Section 3.1.2 with ξ_1 known ($\Xi = \{\xi_1\}$), the likelihood root statistic is

$$r(\pi_0; \mathbf{t}) = \text{sign}(\widehat{\pi}_0 - \pi_0)\sqrt{2\left(\ln L(\widehat{\pi}_0, \xi_1; \mathbf{t}) - \ln L(\pi_0, \xi_1; \mathbf{t})\right)}$$

where $L(\pi_0, \xi_1; \mathbf{t}) = \prod_{i=1}^{N} (\pi_0 g(t_i; \xi_0) + (1 - \pi_0) g(t_i; \xi_1))$, simplifying equation (5) for the $D = 2$ case. Since $r(\pi_0; \mathbf{T})$ is asymptotically $\mathrm{N}(0, 1)$ under the usual regularity conditions, the resulting interval $\{\pi_0 \in [0, 1] : 1 - \Phi(r(\pi_0; \mathbf{t})$ is an approximate $(1 - \varepsilon)$ confidence interval of π_0, where Φ is the distribution function of $\mathrm{N}(0, 1)$. Thus, for use in equation (8), let

$$\widehat{\pi}_0(1 - \varepsilon; \mathbf{t}) = \sup\{\pi_0 \in [0, 1] : 1 - \Phi(r(\pi_0; \mathbf{t})) \geq \varepsilon\} = r^{-1}\left(\Phi^{-1}(1 - \varepsilon); \mathbf{t}\right).$$

3.1.4 Hierarchical distributions based on an expected LFDR

For simplicity, consider the mixture model of Section 3.1.2 but with ξ_1 known, and assume $\widehat{\pi}_0(1 - \varepsilon; \mathbf{T})$ is a continuous random variable. The probability distribution $K_{\mathbf{t}}$ defined by

$$K_{\mathbf{t}}(\varpi_0 \leq \widehat{\pi}_0(1 - \varepsilon; \mathbf{t})) = P_{\pi_0, \xi_0, \xi_1}(\widehat{\pi}_0(1 - \varepsilon; \mathbf{T}) \geq \pi_0),$$

where $\varpi_0 \sim K_{\mathbf{t}}$, is a coherent fiducial distribution for π_0 (§2.1.2). By equation (7), $K_{\mathbf{t}}(\varpi_0 \leq \widehat{\pi}_0(1 - \varepsilon; \mathbf{t})) = 1 - \varepsilon$, that is, $\widehat{\pi}_0(1 - \varepsilon; \mathbf{t})$ is the $(1 - \varepsilon)$ quantile of the random variable ϖ_0. Thus, equation (8) implies that $\widehat{\psi}_i(1 - \varepsilon)$ is the $(1 - \varepsilon)$ quantile of the LFDR random variable $\Psi_i(\varpi_0) = \varpi_0 f_0(t_i) / (\varpi_0 f_0(t_i) + (1 - \varpi_0) f_1(t_i))$, defined by substituting ϖ_0 for π_0 in equation (3). Since $K_{\mathbf{t}}$ is defined for the observed value $\mathbf{t}$, the distribution of $\Psi_i(\varpi_0)$ is defined with $\mathbf{t}$ fixed.

The expected value of $\Psi_i(\varpi_0)$ with respect to $K_{\mathbf{t}}$ is $E_{\mathbf{t}}(\Psi_i(\varpi_0)) = \int \Psi_i(\pi_0)\, dK_{\mathbf{t}}(\pi_0)$. This *expected LFDR* (ELFDR) is a fiducial-Bayes posterior probability that the ith null hypothesis is true. Substituting $E_{\mathbf{t}}(\Psi_i(\varpi_0))$ for $\widehat{\psi}_i$ in Section 3.1.1 defines an *ELFDR hierarchical density function* and the corresponding $(1 - \alpha)$ *ELFDR hierarchical set estimate.* Under asymptotic normality, the ELDFR may be approximated by a simple function of a point estimate of the LFDR and an estimate of the logarithm of the LFDR (Padilla and Bickel, 2012).

3.2 Propagated hierarchical distributions and set estimates

In this section, the confidence intervals of Section 3.1.3 and the coherent fiducial distribution of Section will be used to construct a coherent fiducial distribution of the LFDR for the propagation of error according to the framework of integrating with respect to fiducial distributions that is described in (Bickel and Padilla, 2014). In contrast with previous uses of confidence distributions for the propagation of the uncertainty of false discovery rates (Padilla and Bickel, 2012; Bickel, 2013) and the plug-in methods of Section 3.1, the uncertainty will be propagated according to the laws of probability all the way to the interval estimates.

Putting the coherent fiducial distributions c_{x_i} and $K_{\mathbf{t}}$ together, the random variable corresponding to the parameter of interest is not the ϑ_i of equation (4) but rather is $\vartheta_i\left(\varpi_0\right)$, which is distributed according to the *propagated hierarchical density function* defined by

$$\Psi_i\left(\varpi_0\right)\delta_0+\left(1-\Psi_i\left(\varpi_0\right)\right)c_{x_i},$$

the mixture probability density function with components δ_0 and c_{x_i} and with random mixing proportions $\Psi_i\left(\varpi_0\right)$ and $\left(1-\Psi_i\left(\varpi_0\right)\right)$. As above, the observation $\mathbf{t}$ is fixed, with the randomness coming through ϖ_0 and through the mixture in terms of the posterior probability $\Psi_i\left(\varpi_0\right)$ that the ith null hypothesis is true.

In analogy with Section 3.1.1, let $\widetilde{\theta}_i$ denote the quantile function of the random variable $\vartheta_i\left(\varpi_0\right)$, yielding $\widetilde{\theta}_i\left(1-\epsilon\right)$ as the $\left(1-\epsilon\right)$ quantile of $\vartheta_i\left(\varpi_0\right)$ for any $\epsilon\in[0,1]$. The $\left(1-\alpha\right)$ *propagated hierarchical set estimate* defined as $\left[\widetilde{\theta}_i\left(\beta_1\right),\widetilde{\theta}_i\left(\beta_2\right)\right]$ contains $\vartheta_i\left(\varpi_0\right)$ with fiducial-Bayes posterior probability $1-\alpha=\beta_2-\beta_1$. Since $\varpi_0=\widehat{\pi}_0\left(U;\mathbf{t}\right)$ for the random variable U uniform between 0 and 1, realizations of $\Psi_i\left(\varpi_0\right)$ and then of $\vartheta_i\left(\varpi_0\right)$ may be generated repeatedly by Monte Carlo, with the resulting β_1 and β_2 empirical quantiles serving as numeric approximations of $\widetilde{\theta}_i\left(\beta_1\right)$ and $\widetilde{\theta}_i\left(\beta_2\right)$, respectively.

4 Application

A study was conducted across 8 exam sites to study the effect of a training program on student test scores. Rubin (1981) reported the mean and variance of the SAT score differences between students participating in

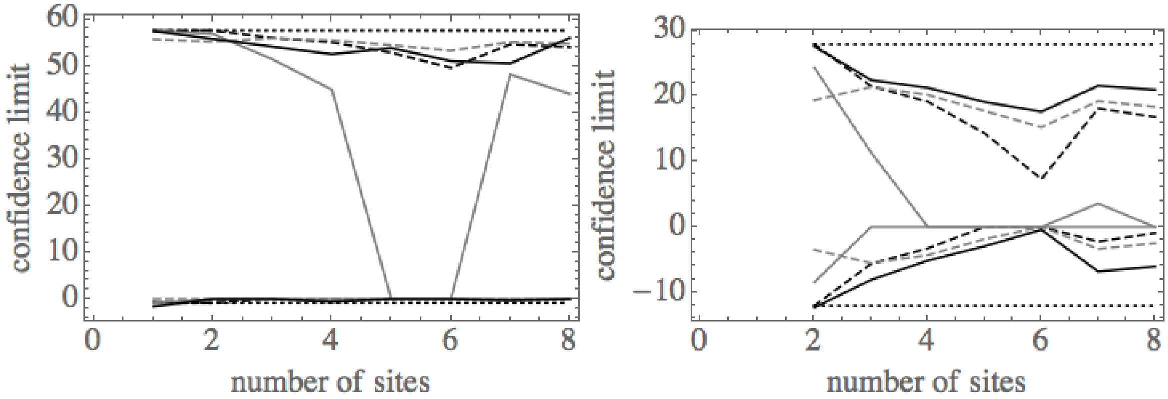

Figure 1: Upper and lower equal-tail 95% confidence limits of the expected test score difference for Site A (left panel) and Site B (right panel) according to nested confidence sets (dotted black; §2.1.1), MLE hierarchical set estimates (dashed black; §3.1.2, Example 1), 75%-limit hierarchical set estimates (solid gray; §3.1.3), ELFDR hierarchical set estimates (dashed gray; §3.1.4), and propagated hierarchical set estimates (solid black; §3.2, via 1000 Monte Carlo draws) versus N, the number of test sites in the data set considered.

the program and non-participants for each exam site. The standardized mean difference for the ith exam site is assumed to have been drawn from $\mathrm{N}(\theta_i, 1)$ with θ_i unknown. To determine the effect of N, the number of sites, on the interval estimates, 8 data sets were generated: Site A alone ($N = 1$), Sites A-B ($N = 2$), ..., Sites A-G ($N = 7$), and Sites A-H ($N = 8$).

The hierarchical model of Section 3.1.2 uses $f_0 = g(\bullet; 0)$ and $f_1 = g(\bullet; 2)$, the normal probability density functions of unit variance and means 0 and 2, respectively. (For simplicity of exposition, $\xi_1 = 2$ is assumed as if it were known.)

Equal-tail 95% interval estimates for the various methods considered are displayed as Figure 1. The three hierarchical set estimates appear similar to the nested confidence interval when the numbers of site in the data set is lower. At one extreme, the nested confidence intervals have no shrinkage. At the other extreme, the 75%-limit hierarchical set estimates exhibit excessive shrinkage toward 0, the null hypothesis value. The MLE hierarchical set estimates, the ELFDR hierarchical set estimates, and the propagated hierarchical set estimates strike a happy medium, with the latter two methods exhibiting less sensitivity to the number of sites.

Acknowledgments

This research was partially supported by the Natural Sciences and Engineering Research Council of Canada, by Agriculture and Agri-Food Canada, by the Canada Foundation for Innovation, by the Ministry of Research and Innovation of Ontario, and by the Faculty of Medicine of the University of Ottawa.

References

Bickel, D. R., 2012a. Coherent frequentism: A decision theory based on confidence sets. Communications in Statistics - Theory and Methods 41, 1478–1496.

Bickel, D. R., 2012b. Empirical Bayes interval estimates that are conditionally equal to unadjusted confidence intervals or to default prior credibility intervals. Statistical Applications in Genetics and Molecular Biology 11 (3), art. 7.

Bickel, D. R., 2012c. A frequentist framework of inductive reasoning. Sankhya A 74, 141–169.

Bickel, D. R., 2013. Simple estimators of false discovery rates given as few as one or two p-values without strong parametric assumptions. Statistical Applications in Genetics and Molecular Biology 12, 529–543.

Bickel, D. R., 2014. Small-scale inference: Empirical Bayes and confidence methods for as few as a single comparison. International Statistical Review 82, 457–476.

Bickel, D. R., Padilla, M., 2014. A prior-free framework of coherent inference and its derivation of simple shrinkage estimators. Journal of Statistical Planning and Inference 145, 204–221.

Datta, G. S., Mukerjee, R., 2004. Probability Matching Priors: Higher Order Asymptotics. Springer, New York.

Efron, B., 2004. Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. Journal of the American Statistical Association 99, 96–104.

Efron, B., 2007. Size, power and false discovery rates. Annals of Statistics 35, 1351–1377.

Efron, B., 2010. Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction. Cambridge University Press, Cambridge.

Lai, Y., 2007. A moment-based method for estimating the proportion of true null hypotheses and its application to microarray gene expression data. Biostatistics 8, 744–755.

Muralidharan, O., 2010. An empirical Bayes mixture method for effect size and false discovery rate estimation. Annals of Applied Statistics 4, 422–438.

Nadarajah, S., Bityukov, S., Krasnikov, N., 2015. Confidence distributions: A review. Statistical Methodology 22, 23–46.

Owen, A., 2005. Variance of the number of false discoveries. Journal of the Royal Statistical Society. Series B: Statistical Methodology 67, 411–426.

Padilla, M., Bickel, D. R., 2012. Estimators of the local false discovery rate designed for small numbers of tests. Statistical Applications in Genetics and Molecular Biology 11 (5), art. 4.

Pawitan, Y., Murthy, K., Michiels, S., Ploner, A., 2005. Bias in the estimation of false discovery rate in microarray studies. Bioinformatics 21, 3865–3872.

Polansky, A. M., 2007. Observed Confidence Levels: Theory and Application. Chapman and Hall, New York.

Rubin, D. B., 1981. Estimation in parallel randomized experiments. Journal of Educational Statistics 6, pp. 377–401.

Wilkinson, G. N., 1977. On resolving the controversy in statistical inference (with discussion). Journal of the Royal Statistical Society B 39, 119–171.

Yang, Y., Aghababazadeh, F. A., Bickel, D. R., 2013a. Parametric estimation of the local false discovery rate for identifying genetic associations. IEEE/ACM Transactions on Computational Biology and Bioinformatics 10, 98–108.

Yang, Z., Li, Z., Bickel, D. R., 2013b. Empirical Bayes estimation of posterior probabilities of enrichment: A comparative study of five estimators of the local false discovery rate. BMC Bioinformatics 14, art. 87.