Empirical Bayes and fiducial effect-size estimation for small numbers of tests

Marta Padilla and David R. Bickel

March 13, 2015

Ottawa Institute of Systems Biology

Biochemistry, Microbiology, and Immunology Department

Mathematics and Statistics Department

University of Ottawa

451 Smyth Rd.

Ottawa, Ontario K1H 8M5

dbickel@uottawa.ca

Keywords: confidence distribution; empirical Bayes; fiducial inference; local false discovery rate; selection bias; small-scale inference; Type II maximum likelihood

Abstract

Estimation of an effect size or other parameter of interest (POI), such as an average of differential abundance levels of metabolites or of the differential expression levels of genes, may be improved by shrinking toward a null-hypothesis value to the extent of a probability that the null hypothesis is true. For example, the local false discovery rate (LFDR) is a null-hypothesis posterior probability that is estimable via empirical Bayes methods without specifying the hyperprior distributions needed for a hierarchical Bayesian approach.

We compared the following (estimated) null-hypothesis probabilities as degrees of shrinkage in order to improve POI interval estimates (IEs) and POI point estimates (PEs): a histogram-based estimator (HBE) of the LFDR, a binomial-based estimator of the LFDR, a maximum-likelihood estimator of the LFDR, an expected LFDR (ELFDR), and a fiducial probability (FP).

In multiple-hypothesis testing, the ELFDR yields reliable IEs while the HBE gives the best PEs. For single-hypothesis testing, the FP generates an IE that outperforms the confidence interval and generates a PE performing at least as well as the unbiased estimator.

We apply these POI estimators to the abundance levels of 20 plasma proteins in women with breast cancer.

1 Introduction

In many biostatistics applications, research aims to select and prioritize for further study genes, proteins, metabolites, DNA sites, or other biological features. The task of selection and prioritization usually either relies on volcano plots or is performed in two steps: first some features are selected based on an arbitrary statistical significance threshold, and then the features are prioritized according to a simple estimate of their parameter of interest (POI) such as a fold change between a case condition and a control condition (Montazeri et al., 2010). An alternative is to combine the statistical hypothesis testing with the estimation

of the POI of the features into a single score that would only need the specification of one threshold. Such an approach is given by a *shrinkage* estimator of the POI. As indicated by its name, a shrinkage estimator gives estimates of the POIs that are smoothly shrunk towards the null hypothesis value, which is usually 0, meaning no difference in the mean log abundance levels of proteins, log level of gene expression, etc. The shrinkage of the POI is performed by effectively adjusting it according to a significance measure such as the p-value.

Consider a multiple testing problem in which the data involves N biological features for every individual, corresponding to the observable random variables $\langle X_1, X_2, \ldots, X_N \rangle$. If there are n individuals, then the observed outcome of X_i corresponding to the ith feature is a vector $x_i \in \mathbb{R}^n$ from a distribution in the parametric family $\{P_{\theta_i,\gamma_i} : \theta_i \in \Theta, \gamma_i \in \Gamma\}$, where θ_i is a POI in a set Θ and γ_i is a nuisance parameter in a set Γ. Examples of POIs are changes in the abundance level of proteins or gene expression levels in response to some cause. The function τ transforms each observed data vector x_i into a scalar statistic denoted by t_i. Let $t_i = \tau(x_i)$ and $T_i = \tau(X_i)$ define the statistics used to test the null hypothesis that $\theta_i = \theta_0$, where $\theta_0 \in \Theta$. Usually, θ_0 refers to no effect. In turn, the p-value associated with the null hypothesis is denoted p_i and is a function of t_i.

Example 1. The usual biological problem involves the comparison of two different groups of individuals with N genes, proteins, metabolites, or other biological features, labeled *treatment* and *control* for convenience. For example, let $\langle X_1, X_2, \ldots, X_N \rangle$ and $\langle Y_1, Y_2, \ldots, Y_N \rangle$ denote the log-abundance levels of N proteins for the treatment group and for the control group, respectively. The observed data vectors in the respective groups for the ith feature are $x_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}^m$. In addition, the data are such that $X_i \sim \mathrm{N}(\xi_i, \sigma_i^2)$ and $Y_i \sim \mathrm{N}(\xi_i', \sigma_i'^2)$, where ξ_i and ξ_i' are the unknown mean of the respective normal distributions and σ_i and σ_i' are the unknown standard deviation of independent measurements of feature i in the treatment group and in the control group, respectively. The parameter of interest for feature i is the POI $\theta_i = (\xi_i - \xi_i')/\sigma$, where σ is a standard deviation based on both groups, on the control group or a pooled standard deviation (Cohen, 1988). Thus, θ_i is the differential

abundance level of the ith protein.

The null hypothesis toward which POI estimates will be shrunk is that $\theta_i = \theta_0$, the alternative hypothesis is that $\theta_i \neq \theta_0$, and $\Theta \subset \mathbb{R}$ is the parameter space. The value for the null hypothesis is known and usually $\theta_0 = 0$, thus it corresponds to no effect, that is, if $\theta_i = 0$ the ith protein has equivalent abundance level in the treatment and in the control group. On the other hand, the truth of the ith alternative hypothesis $\theta_i \neq 0$ indicates that the corresponding protein is affected by (or associated with) the treatment, disease, or other perturbation. ▲

The purpose of this paper is to study and compare the suitability of several measures of significance as degrees of shrinkage of estimates of the POI for both single comparisons and multiple comparisons. Section 2 describes the measures of significance used in this study. They consist of estimators of the local false discovery rate (LFDR) (Efron et al., 2001) and of measures of significance based on confidence intervals. The estimates of the LFDR are particularly suitable as degrees of shrinkage (Montazeri et al., 2010; Yanofsky and Bickel, 2010). The LFDR is defined as the probability of the truth of the null hypothesis and is estimated by empirical Bayes methods. It has been successfully applied in many applications involving thousand of features for which specific LFDR estimators were designed (Efron, 2004a; Efron and Tibshirani, 2007; Efron, 2007). Whereas those estimators were designed for large numbers of features, other LFDR estimators were proposed for their use in the analysis of data sets with small and medium number of features. Examples of such data sets are those involving proteins, metabolites or gene expression measured by traditional experimental techniques (Padilla and Bickel, 2012; Li et al., 2013).

Like LFDR estimators, null hypothesis probabilities based on confidence intervals can also serve to shrink POI estimates. Two such probabilities will be considered.

To estimate the POI on the basis of each of the significance measures described in Section 2, we follow the approach proposed by Bickel (2012). That approach uses the LFDR and the confidence-based probabilities of the null hypothesis as the degrees of shrinkage of the POI

estimates, as explained in Section 3. That results in shrunken interval estimates as well as shrunken point estimates of the POI (Bickel, 2012).

In Section 4, an application to biological data illustrates the differences in the results obtained by the various measures of significance. The data sets consist of the abundance levels of 20 proteins of women with breast cancer; the POI to be estimated in this case is the log fold change.

Section 5 describes the simulation studies conducted to quantify the performance of the interval and point estimators of the POI. Finally, a discussion and a summary of the findings are given in Section 6.

2 Significance measures for the degree of shrinkage

The LFDR and the confidence-based methods used to determine the degree of shrinkage in this work are briefly explained here distributed under Sections 2.1 and 2.2, respectively.

2.1 LFDR estimators

This section reviews the LFDR estimators that are compared in this work with respect to the performance of the point and interval estimates they generate. They are: Type II maximum likelihood estimator (MLE) (Section 2.1.1), histogram-based estimator (HBE) (Section 2.1.2) and binomial-based estimator (BBE1) (Section 2.1.3).

Consider a variable A_i that denotes whether the ith alternative hypothesis is true: $A_i = 1$ if $\theta_i \neq \theta_0$, and $A_i = 0$ if $\theta_i = \theta_0$. Then, $A_i = 1$ means that the ith feature is affected. For the ith feature, $g(t_i)$ denotes the marginal probability density of the test statistic. The LFDR is defined as $\psi_i = \Pr(A_i = 0|T_i = t_i) = \Pr(\theta_i = \theta_0|t_i)$ and, by the Bayes's theorem

$$\psi_i = \Pr(A_i = 0|T_i = t_i) = \Pr(\theta_i = \theta_0|t_i) = \frac{\pi_0 g_0(t_i)}{g(t_i)}, \tag{1}$$

where $\pi_0 = \Pr(\theta_i = \theta_0)$ denotes the prior probability that the ith null hypothesis is true,

g_0 is the probability density function of T_i conditional on the null hypothesis ($\theta_i = \theta_0$), which is known a priori, and g_{alt} is the probability density function of T_i conditional on the alternative hypothesis ($\theta_i \neq \theta_0$). In addition, the marginal probability density is assumed to be a mixture probability density

$$g(t_i) = \pi_0 g_0(t_i) + (1 - \pi_0) g_{\text{alt}}(t_i). \tag{2}$$

Since π_0 and $g_{\text{alt}}(t_i)$ are unknown, they have to be estimated from the data to obtain the estimated LFDR $\widehat{\psi}_i$. Then,

$$\widehat{\psi}_i = \frac{\widehat{\pi}_0 g_0(t_i)}{\widehat{\pi}_0 g_0(t_i) + (1 - \widehat{\pi}_0)\widehat{g}_{\text{alt}}(t_i)}. \tag{3}$$

where $\widehat{\pi}_0$ and $\widehat{g}_{\text{alt}}(t_i)$ denote the estimates of π_0 and $g_{\text{alt}}(t_i)$, respectively. The techniques used to estimate $\widehat{\pi}_0$ and $\widehat{g}_{\text{alt}}(t_i)$ define many of the different LFDR estimators, some of which are briefly reviewed below.

2.1.1 Maximum likelihood estimator of the LFDR

Good (1966) defined two types of *maximum likelihood estimator (MLE).* A *Type I MLE* refers to a non hierarchical model, whereas a *Type II MLE* refers to a hierarchical model. The Type I MLE estimates the ith POI using the Type I MLE defined by $\widehat{\delta}_i = \arg\sup_{\delta_{\text{alt}} \in \Delta} f_\delta(t_i)$, where $f_\delta(t_i)$ is a density function with POI δ.

The Type II MLE considered here assumes that g_0 and g_{alt} in equations (2) and (3) belong to the same known family of probability density functions $\{g_\delta : \delta \in \Delta\}$ where Δ is the parameter space and $\delta_0, \delta_{\text{alt}} \in \Delta$ are the parameters related to the null and the alternative hypothesis, respectively. The parameter δ_{alt} is unknown and $\delta_{\text{alt}} \neq \delta_0$, thus, if the ith alternative hypothesis is true $\delta_i = \delta_{\text{alt}}$, otherwise $\delta_i = \delta_0$. Likewise, two parameters have to be estimated to obtain $\widehat{\psi}_i$ in equation (3): π_0 and δ_{alt}. To estimate the LFDR by this method, the data vector x_i for the ith feature is first reduced to the absolute value of a scalar

statistic t_i thus $u_i = |t_i|$. Therefore, $g_{\delta_0}(u_i)$ and $g_{\delta_{\text{alt}}}(u_i)$ denote the probability densities of the reduced data under the null hypothesis and the alternative hypothesis, respectively. According to equations (1) to (2) and changing the notation $g_{\text{alt}} = g_{\delta_{\text{alt}}}$, $g_0 = g_{\delta_0}$ the true value of the LFDR for the ith feature is given by,

$$\psi_i = \frac{\pi_0 g_{\delta_0}(u_i)}{\pi_0 g_{\delta_0}(u_i) + (1-\pi_0) g_{\delta_{\text{alt}}}(u_i)}, \tag{4}$$

where ψ_i and the parameters δ_{alt} and π_0 are unknown, while $\delta_0 = 0$. The Type II MLE assumes that the parameters π_0 and δ_{alt} are the maximum likelihood estimates of the true parameters, then they are given by

$$\left\langle \hat{\delta}_{\text{alt}}, \hat{\pi}_0 \right\rangle = \arg \sup_{\langle \delta_{\text{alt}}, \pi_0 \rangle \in \Theta \times [0,1]} \prod_{j=1}^{N} \left(\pi_0 g_{\delta_0}(u_j) + (1-\pi_0) g_{\delta_{\text{alt}}}(u_j) \right). \tag{5}$$

To avoid finding the maximum likelihood over the wrong estimates, restrictions have to be imposed on the space where $\left\langle \widehat{\delta}_{\text{alt}}, \widehat{\pi}_0 \right\rangle$ lie, that is $\Delta \times [0,1]$, where Δ is the set of 0 and positive real numbers. Therefore, with substitution into equation (4), the estimated LFDR for the ith feature is

$$\widehat{\psi}_i = \frac{\widehat{\pi}_0 g_{\delta_0}(u_i)}{\widehat{\pi}_0 g_{\delta_0}(u_i) + (1-\hat{\pi}_0) g_{\hat{\delta}_{\text{alt}}}(u_i)}, \tag{6}$$

This method could be modified by adding components and their parameters to the mixture model. We call such $\widehat{\psi}_i$ *MLE* for simplification.

Example 2. Following Example 1, the statistic t_i is the observed two-sample t statistic for $\langle x_i, y_i \rangle$ such that $t_i = T(x_i, y_i)$ and $T_i = T(X_i, Y_i)$ is distributed as the noncentral Student t distribution g_{δ_i} with ν_i degrees of freedom (Welch, 1947) and noncentral parameter δ_i. The noncentral parameter associated with the null hypothesis is $\delta_0 = 0$ and the parameter associated with the alternative hypothesis is $\delta_{\text{alt}} > 0$. ▲

2.1.2 Histogram-based estimator of the LFDR

Efron (2004a) introduced the method that we call *histogram-based estimator (HBE)* for simultaneous inference of a large number of tests, at least N>100. Later, our simulations showed that HBE performed similarly than other estimators for a number of tests as small as $N \geq 12$ (Padilla and Bickel, 2012). HBE uses $z_i = \Phi^{-1}(p_i)$, where Φ^{-1} is the cumulative distribution function of N(0,1), and assumes that the proportion of affected cases is smaller than about 10%. The density function $h(z_i)$ is a mixture density like equation (2) and $h_0(z_i)$ is the density corresponding to the null hypothesis. The null hypothesis density $h_0(z_i)$ can be assigned theoretically to have a distribution N(0,1), or empirically (Efron, 2007). The possibility to select a theoretical or empirical null hypothesis density is an advantage of this method, but it may lead to very different results. For this reason, in this work we consider both options and denote HBE the method that uses a theoretical null for $h(z_i)$ and *HBEE* the method *HBE-empirical null,* which uses an estimated empirical null (Efron, 2007). In addition, given that we work with small-scale data sets, we have used a low value ($df = 3$) for the number of degrees of freedom for fitting the estimated density $h(z_i)$.

2.1.3 Binomial-based estimator of the LFDR

The *binomial-based estimator (BBE1)* is a conservative LFDR estimator based on the binomial distribution that have been proposed by Bickel (2013). This method is based on the nonlocal LFDR (NLFDR), which considers a rejection region for the given statistics $T_i \in \mathcal{T}$ instead than a local value. Thus, NLFDR depends on the specification of $\mathcal{T}$, which is common to all the features. The equation according to the Bayes Theorem, is

$$\Psi_i(\mathcal{T}) = \Pr(A_i = 0 | T_i \in \mathcal{T}) = \frac{E(N_0(\mathcal{T}))}{E(N_+(\mathcal{T}))} = \frac{\pi_0 \Pi_0(\mathcal{T})}{\Pi(\mathcal{T})},$$

where $N_0(\mathcal{T})$ denotes the number of false discoveries, $N_+(\mathcal{T})$ denotes the total number of discoveries (Efron, 2010), $\pi_0 = P(A_i = 0)$ is often set to 1, $\Pi_0(\mathcal{T}) = \Pr(T_i \in \mathcal{T} \mid A_i = 0)$,

and $\Pi(\mathcal{T})$=Pr($T_i \in \mathcal{T}$), which can be given by $\widehat{\Pi}(\mathcal{T}; N_+(\mathcal{T})) = N_+(\mathcal{T})/N$. The fact that $\pi_0 = 1$ is indicated in the name of the method, i.e. BBE1, to follow the nomenclature given in our previous paper (Padilla and Bickel, 2012), where we studied several LFDR estimators. If the test statistics are independent of each other, $N_+(\mathcal{T})$ follows the binomial distribution with parameters N and $\Pi(\mathcal{T})$. Thus, $\widehat{\Pi}(\mathcal{T})$ is the maximum-likelihood estimate (MLE) of $\Pi(\mathcal{T})$ and the estimator $\widehat{\Psi_i}(\mathcal{T})$ is the MLE of $\Pi_0(\mathcal{T})/\widehat{\Pi}(\mathcal{T})$.

2.2 Confidence-based probabilities of the null hypothesis

In addition to the previous LFDR estimators, we consider two posterior probabilities that are based on confidence distributions: the expected LFDR (Section 2.2.1) and the fiducial probability that the null hypothesis is true (Section 2.2.2). For a recent review of confidence distributions, see Nadarajah et al. (2015).

2.2.1 Expected LFDR

The *expected LFDR (ELFDR)* was introduced by Padilla and Bickel (2012). The method of this subsection takes into account the propagation of the uncertainty of the parameters on which the LFDR depends. Assuming the asymptotic normality of $\widehat{\psi}_i$, Padilla and Bickel (2012) derived $\ln \psi_i \sim \mathrm{N}(\ln \widehat{\psi}_i, \widehat{s}_i)$, where ψ_i is the LFDR-surrogate random variable of the lognormal (LN) confidence distribution $\mathrm{LN}(\ln \widehat{\psi}_i, \widehat{s}_i)$ and $\widehat{s}_i$ is the estimated standard error of $\ln \widehat{\psi}_i$. The expectation value $E(\psi_i)$ of the LFDR surrogate is thus,

$$E(\psi_i) = \exp\left(\ln \widehat{\psi}_i + \frac{\widehat{s}_i^2}{2}\right),$$

which is accordingly called the *expected LFDR (ELFDR)* (Padilla and Bickel, 2012).

In our simulations and application to protein data, we used $\widehat{\psi}_i = \widehat{\psi}_i^{\mathrm{HBE}}$ (the HBE of Efron (2004a) in Section 2.1.2) and the method of computing $\hat{s}_i$ implemented in the locfdr R package (Efron et al., 2011). As an expectation value of a probability that the null hypothesis

is true, the ELFDR is itself such a probability. Another is described in the next subsection.

2.2.2 Fiducial probability that the null hypothesis is true

In addition to the LFDR, other posterior probabilities of the truth of the null hypothesis can be used. Procedures based on the confidence interval of the POI can be made coherent by encoding them as a fiducial distribution that, like a Bayesian posterior, is a probability distribution on the parameter space (Bickel and Padilla, 2014). In that framework, the degree of shrinkage $\widehat{\psi}_i$ in equation (8) (Section 3.1) may be set to a *fiducial probability* (FP) that the null hypothesis is true. Bickel and Padilla (2014) specified conditions under which that probability is a two-sided p-value.

Example 3. The fiducial probability in Example 1 consists of the two-sided p-value derived from the Student t distribution g_{δ_0} with respect to the null hypothesis that there is no difference between the mean of the treatment and control groups, this is, that the noncentrality parameter is $\delta_0 = 0$. ▲

3 From a significance measure to a POI estimator

The purpose of this section is to briefly describe the method proposed by Bickel (2012). Such method uses an LFDR estimate or a confidence-based probability to build an estimator of a POI. For clarity, this will be explained in terms of prior and posterior probability density functions (Section 3.1) before defining it in terms of prior and posterior probability measures (Section 3.2).

3.1 The transformation explained in terms of density functions

An estimator of the POI based on the LFDR was firstly proposed by Efron (2008). This POI estimator uses a finite-mixture model, similarly as the model on which the LFDR estimation is based on. However, to estimate the POI by empirical Bayes methods, this

model requires an estimate of the prior distribution of the parameter conditional on the alternative hypothesis that not only it may be unreliable but it also influences the POI estimates. For this reason, Bickel (2012) proposes an improvement of the Efron (2008) model by an approach that obtains reliable confidence intervals without the estimation of such prior. In a Bayesian framework, it is considered that the ith POI θ_i has a prior probability density $p(\theta_i)$ modeled by a mixture probability density in the form of equation (2), thus $p(\theta_i) = \pi_0 p_0(\theta_i) + (1-\pi_0) p_1(\theta_i)$, where $p_1(\theta_i)$ is a given prior probability density conditional on the truth of the alternative hypothesis. Taking into account the definition of the LFDR in equation (1) , the posterior probability density of θ_i is

$$p(\theta_i|t_i) = \psi_i p_0(\theta_i|t_i) + (1-\psi_i) p_1(\theta_i|t_i), \tag{7}$$

where the prior probability density conditional on the null hypothesis $p_0(\theta_i)$ is 0 everywhere except at $\theta_i = \theta_0$, that is $p_0(\theta_i) = \delta_{\theta_0}$, the Dirac delta function at θ_0, and thus $p_0(\theta_i|t_i) = p_0(\theta_i)$. By the Bayes' theorem, $p_1(\theta_i|t_i)$ depends on $p_1(\theta_i)$ since $p_1(\theta_i|t_i) = p_1(\theta_i) f_1(t_i;\theta_i) / f_1(t_i)$, with $f_1(t_i) = \int f_1(t_i;\theta) p_1(\theta) d\theta$. In a subjective Bayesian framework $p_1(\theta_i)$ is known because it encodes the levels of belief of an agent. However, in an empirical Bayes framework, $p_1(\theta_i)$ is unknown and must be estimated from data, although it cannot be reliably estimated (Efron, 2010). Therefore, the quantities $p_1(\theta_i)$ and ψ_i have to be estimated from data to yield an estimate of $p(\theta_i|t_i)$, denoted $\check{p}(\theta_i|t_i)$. In turn, $\check{p}(\theta_i|t_i)$ gives an estimate of a credible interval or a point estimate such as $\int \theta_i \check{p}(\theta_i|t_i) d\theta_i$, the estimate of the posterior mean $\int \theta_i p(\theta_i|t_i) d\theta_i$.

The approach proposed by Bickel (2012) to derive the empirical Bayes method for point and interval estimation of the POI is analogous to the above explained method to derive $\check{p}(\theta_i|t_i)$, but without the need of estimating the prior $p_1(\theta_i)$. For this, Bickel (2012) uses the *significance function* S_{t_i}, which is a function such that $S_{t_i}(\theta_0)$ is a one sided p-value for the null hypothesis that $\theta_i = \theta_0$, for any $\theta_i \in \Theta$ (see Section 3.2). Suppose that S_{t_i} is the CDF of

the POI of the ith feature. Thus, the derivative $\partial S_{t_i}(\theta)/\partial\theta\big|_{\theta=\theta_i}$ is proportional to a probability density at θ_i. This probability density is called the *conditional confidence density* and is denoted by $p_1(\theta_i; t_i)$. A confidence density encodes the certainty levels about the intervals where θ_i lies, and from which the confidence intervals (at a given confidence level) can be extracted. In turn, let $p(\theta_i; t_i)$ denote the *marginal confidence density,* which is given by

$$p(\theta_i; t_i) = \psi_i p_0(\theta_i) + (1 - \psi_i) p_1(\theta_i; t_i) \tag{8}$$

The Bickel (2012) approach to obtaining an estimate of θ_i is based on obtaining an estimate of $p(\theta_i; t_i)$, denoted by $\widehat{p}(\theta_i; t_i)$ and given by $\widehat{p}(\theta_i; t_i) = \widehat{\psi}_i p_0(\theta_i) + \left(1 - \widehat{\psi}_i\right) p_1(\theta_i; t_i)$, where the $\widehat{\psi}_i$ is the estimated LFDR computed according to a selected method. Therefore, the Bickel (2012) approach can be illustrated by the replacement of the Bayesian posterior probability densities $p_1(\theta_i|t_i)$ and $p(\theta_i|t_i)$ in equation (7) by the confidence densities $p_1(\theta_i; t_i)$ and $p(\theta_i; t_i)$. This approach results in POI interval estimates and POI point estimates, such as $\int \theta_i \widehat{p}(\theta_i; t_i)\, d\theta_i$, the estimate of the confidence posterior mean $\int \theta_i p(\theta_i; t_i)\, d\theta_i$. Both interval and point estimates are shrunk towards the parameter value corresponding to the null hypothesis.

Any LFDR estimator or other significance measure (see Sec. 2) can be used as $\widehat{\psi}_i$ in equation (8). Examples from the literature include not only quantities explicitly presented as LFDR estimators (Efron, 2004b; Efron and Tibshirani, 2007; Efron, 2007; Padilla and Bickel, 2012) but also a fiducial probability (Bickel and Padilla, 2014). In order to determine the conditions at which it is preferable to use one estimator in place of others, in this work we study and compare the performance of several POI estimators built with different significance measures and the *unshrunken estimator* (UE), which is the trivial significance measure or degree of shrinkage that is always 0. This study is made with respect to the number of features, proportion of unaffected features and detectability levels.

3.2 The transformation defined in terms of distributions

Following the reasoning and notation given by Bickel (2012), let P^{t_i} denote the probability distribution of CDF S_{t_i}, which is the significance function defined in Section 3.1. Let ϑ_i denote the random variable of distribution P^{t_i}, that is $\vartheta_i \sim P^{t_i}$. Thus, $P^{t_i}(\vartheta_i \leq \theta) = S_{t_i}(\theta)$ for any $\theta \in \Theta$. The quantity $p_1(\theta_i; t_i)$ defined in Section 3.1 is the conditional probability density function of ϑ_i.

On the other hand, let $P^{(i)}$ denote the probability distribution that is obtained by considering the Bayesian definition of LFDR in equation (1), so that the random variable A_i is equal to 0 with probability π_0 and to 1 with probability π_1 for all $i = 1, \ldots, m$. Then, $P^{(i)}$ is a probability distribution such that, for all $\theta \in \Theta$,

$$\begin{aligned} P^{(i)}(A_i = 0) &= \psi_i \\ P^{(i)}(\vartheta_i \leq \theta | A_i = 1) &= P^{t_i}(\vartheta_i \leq \theta) = S_{t_i}(\theta) \\ P^{(i)}(\vartheta_i \leq \theta | A_i = 0) &= \delta_{\theta_0}(\vartheta_i \leq \theta) = 1_{[\theta_0, \infty)}(\theta) \end{aligned} \tag{9}$$

therefore, P^{t_i} is the conditional posterior distribution of ϑ_i while the marginal posterior CDF of ϑ_i is

$$\begin{aligned} P^{(i)}(\vartheta_i \leq \theta) &= P^{(i)}(A_i = 0) P^{(i)}(\vartheta_i \leq \theta | A_i = 0) + P^{(i)}(A_i = 1) P^{(i)}(\vartheta_i \leq \theta | A_i = 1) \\ &= \psi_i \delta_{\theta_0} + (1 - \psi_i) P^{t_i}(\vartheta_i \leq \theta) \end{aligned} \tag{10}$$

and $\widehat{P}^{(i)}$ is the estimated $P^{(i)}$, then $\widehat{P}^{(i)} = \widehat{\psi}_i \delta_{\theta_0} + \left(1 - \widehat{\psi}_i\right) P^{t_i}$. The marginal confidence density $p(\theta_i; t_i)$ defined in Section 3.1 is the marginal probability density function of ϑ_i.

In analogy with $S_{t_i}(\theta)$, the significance function of P^{t_i}, there is a *marginal significance function* $S_{(i)}$ such that $S_{(i)}(\theta) = P^{(i)}(\vartheta_i \leq \theta)$ for all $\theta \in \Theta$. Plugging the estimated LFDR or confidence-based probability values into equations (9) and (10), the estimate of $S_{(i)}(\theta)$ is

$$\widehat{S}_{(i)}(\theta) = \widehat{\psi}_i 1_{[\theta_0, \infty)}(\theta) + \left(1 - \widehat{\psi}_i\right) S_{t_i}(\theta). \tag{11}$$

The limits of the $(1-\alpha_1-\alpha_2)100\%$ confidence interval of $P^{(i)}$ are given by $\left[S_{(i)}^{-1}(\alpha_1), S_{(i)}^{-1}(1-\alpha_2)\right]$, such as those for P^{t_i} are given by $\left[S_{t_i}^{-1}(\alpha_1), S_{t_i}^{-1}(1-\alpha_2)\right]$. The formula for the inverted $S_{(i)}(\theta)$ is given below with changed notation for simplification. For any $\beta \in [0,1]$, let $\theta_1(\beta)$ denote the β-quantile of P^{t_i}, and let $\theta(\beta; t_i)$ denote the β-quantile of $P^{(i)}$. This is, $\theta_1(\beta; t_i) = \widehat{S}_{t_i}^{-1}(\beta)$ and $\theta(\beta; t_i) = \widehat{S}_{(i)}^{-1}(\beta)$. If $\beta_i = \beta/\left(1-\widehat{\psi}_i\right)$ and $\omega_i = \widehat{\psi}_i/\left(1-\widehat{\psi}_i\right)$,

$$\theta(\beta; t_i) = \begin{cases} \theta_1(\beta_i; t_i) & \text{if } \begin{cases} \beta_i \in [0,1] \text{ and} \\ \theta_1(\beta_i; t_i) < \theta_0 \end{cases} \\ \theta_1(\beta_i - \omega_i; t_i) & \text{if } \begin{cases} \beta_i - \omega_i \in [0,1] \text{ and} \\ \theta_1(\beta_i - \omega_i; t_i) > \theta_0 \end{cases} \\ \theta_0 & \text{otherwise,} \end{cases} \tag{12}$$

where the condition $\beta_i \in [0,1]$ implies that $\widehat{\psi}_i \leq 1-\beta$ and the condition $\beta_i - \omega_i \in [0,1]$ implies that $\widehat{\psi}_i \leq \beta$.

Therefore, the estimated POI of the ith feature can be given by a point estimates or a $(1-\alpha_1-\alpha_2)100\%$ confidence interval estimates, for $\alpha_1, \alpha_2 \in [0,1]$ with $\alpha_1 + \alpha_2 < 1$. The $(1-\alpha_1-\alpha_2)100\%$ confidence interval estimates of θ_i is obtained by $[\theta(\alpha_1; t_i), \theta(1-\alpha_2; t_i)]$, the *marginal confidence interval* with confidence level $1-\alpha_1-\alpha_2$, while the point estimates of θ_i is computed by the (posterior) median $\theta(1/2)$ or the (posterior) mean of the estimated $P^{(i)}$ given by the expectation value of θ_i:

$$E(\vartheta_i) = E(\vartheta_i|\theta_i = \theta_0) \times \Pr(\theta_i = \theta_0) + E(\vartheta_i|\theta_i \neq \theta_0) \times \Pr(\theta_i \neq \theta_0) = 0 + \theta_i^{\mathrm{UE}} \times (1-\widehat{\psi}_i) \tag{13}$$

since the expectation $E(\vartheta_i|\theta_i = \theta_0) = 0$, the probability $\Pr(\theta_i \neq \theta_0) = (1-\widehat{\psi}_i)$, and θ_i^{UE} is given by $\theta_i^{\mathrm{UE}} = E(\vartheta_i|\theta_i \neq \theta_0)$, an unshrunken estimate of θ_i under the alternative hypothesis. Thus, for the UE mentioned in Section 3.1, $E(\vartheta_i) = \theta_i^{\mathrm{UE}}$.

For example, the limits of the $(1-\alpha)100\%$ marginal confidence interval for the ith

feature for $\alpha \in [0, 1/2]$ are $[\theta(\alpha/2; t_i), \theta(1-\alpha/2; t_i)] = [\theta_1(\beta_i), \theta_0]$ if $\widehat{\psi}_i < \alpha/2$, and $\theta_1(\beta_i) < \theta_0$. From the above equation and from Figure 1, we can see that interval estimates like $[\theta(\alpha/2; t_i), \theta(1-\alpha/2; t_i)]$ tends to be shorter than the conditional interval estimates for an equivalent $(1-\alpha)100\%$ confidence interval $[\theta_1(\alpha/2; t_i), \theta_1(1-\alpha/2; t_i)]$, especially if $\widehat{\psi}_i$ is conservative. In addition, the null effect θ_0 is always covered. Moreover, the point estimates given by the (posterior) median is $\theta(1/2)$. Therefore, both the interval estimates and the point estimates given by the inverse of the marginal significance function $\widehat{S}^{-1}_{(i)}(\theta)$ are shrunk toward the value of the POI corresponding to the null hypothesis.

Example 4. Following Example 1, let us consider a data set where the number of features is $N = 6$ and the number of individuals in the treatment and control groups are $m = n = 5$. Feature numbers 1 and 2 are set to be affected, then, in the treatment group $X_i \sim \mathrm{N}(\xi, 1)$ with $\xi = 1.5$ and $i = 1, 2$, and remaining features are unaffected, thus $X_i \sim \mathrm{N}(0, 1)$ for $i = 3, ..., N$ and for the control group $Y_i \sim \mathrm{N}(0, 1)$ for $i = 1, ..., N$. The POI of the ith feature is the difference between the standardized means of the groups denoted by θ_i. The null hypothesis states that $\theta_i = 0$. Then, the true POI for the affected features is $\theta_i = |\xi - 0| = \theta_{\text{alt}}$ and $\theta_i = 0$ for the remaining ones $i = 3, ..., N$. The function T is a two-sample t statistic $T(X_i, Y_i)$ that follows the Student-t distribution f_{δ_i} with ν_i degrees of freedom and noncentral parameter δ_i. For every ith feature, we compute the two-sided p-value p_i to estimate the LFDR $\widehat{\psi}_i$ by BBE1 (Section 2.1.3), then we calculate the observed statistic $t_i = T(x_i, y_i)$ and the significance (p-value) function $S_{t_i}(\theta_i)$ for the equation 11. In terms of the significance function, the one sided upper-tailed p-value p_i is such that $p_i = S_{t_i}(0)$.Figure 1 shows $S_{t_i}(\theta_i)$ and $S_{(i)}(\theta_i)$ for $i = 1$ and FP ($\widehat{\psi}_i = 0.13$) and BBE1 ($\widehat{\psi}_i = 0.57$) LFDR estimation. The limits of the 95% confidence interval are represented by vertical dotted lines at points $\theta_i(0.025; t_i)$, $\theta_i(0.975; t_i)$ and also the point estimates given by the posterior median at point $\theta_i(0.5; t_i)$ in the horizontal axis, and horizontal dotted lines at points 0.025, 0.975 and 0.5 in the vertical axis. Notice the shrinkage of the interval and point estimates given by $S^{-1}_{(i)}(\theta_i)$ towards $\theta_0 = 0$, with respect to the same interval and point estimates given by $S_{t_i}(\theta_i)$, and that

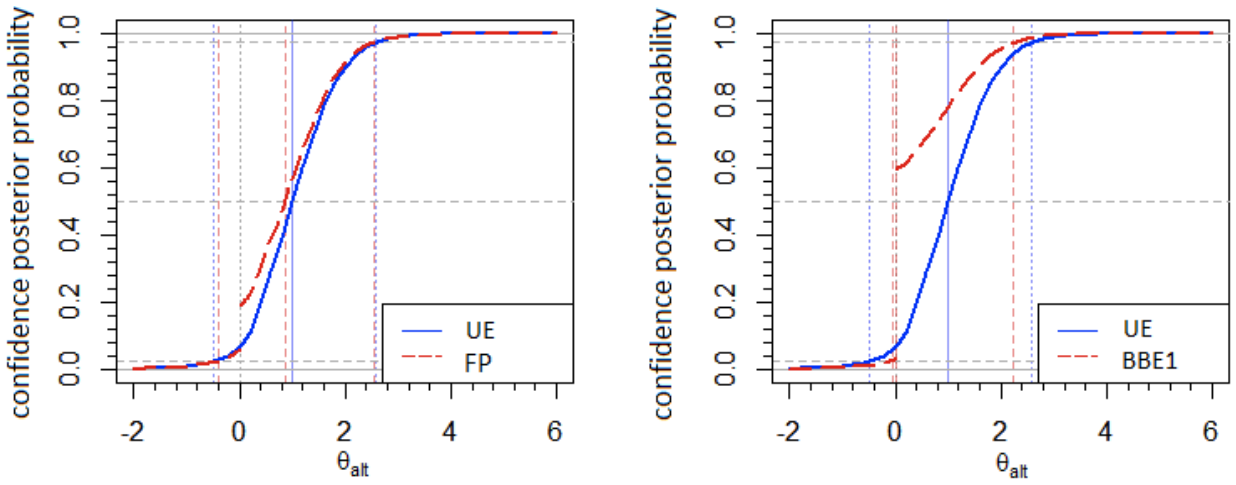

Figure 1: Example of significance function ($P(\vartheta \leq \theta_{\text{alt}})$) of two POI estimators applied on a data set consisting of 6 features, 2 of them affected, 5 control individuals and 5 individuals in the treatment group. The curves correspond to results on feature number 1 and FP (left) and BBE1 (right) as posterior probability of the truth of the null hypothesis (see Example 4). Both plots show the curve for the corresponding UE.

$S_{t_i}(\theta_i)$ corresponds to the UE (Section 2.1.1). ▲

A problem that affects the above explained construction of the $(1-\alpha_1-\alpha_2)\,100\%$ confidence interval estimates of the POI, is that it does not have the same frequentist coverage $\text{CR}(\alpha_1,\alpha_2)$, where $\text{CR}(\alpha_1,\alpha_2) = P^{(i)}(\theta_i \in [\theta(\alpha_1;T_i), \theta(1-\alpha_2;T_i)])$. In fact, in some situations the coverage (CR) can be even higher than the nominal $(1-\alpha_1-\alpha_2)\,100\%$ level. The experimental values of CR together with the width of the $(1-\alpha)\,100\%$ confidence interval and the point estimates ($\theta(1/2;t_i)$ and/or $E(\vartheta_i)$, where $E(\vartheta_i)$ is expectation value of θ_i given by equation (13)), are useful parameters to compare different estimators. From now on, the name we give to the different POI estimators is the name of the method used to get $\widehat{\psi}_i$: FP, MLE, BBE1, HBE, HBEE, ELFDR and UE.

4 Application

In this section, we show an example of the application of different shrinkage estimators on real data following the procedures explained in Example 1 and 4. The aim of this application

is to estimate the log fold change of the log-abundance levels of 20 plasma proteins in 55 women with HER2-positive breast cancer (first treatment group), 35 women with ER/PR-positive breast cancer (second treatment group) and 64 healthy women (control group), which were measured in Alex Miron's laboratory at the Dana-Farber Cancer Institute (Li, 2009). The vectors $x_i^{\text{HER2}} \in \mathbb{R}^{55}$, $x_i^{\text{ER/PR}} \in \mathbb{R}^{35}$, and $y_i \in \mathbb{R}^{64}$ denote the log-abundance (base 2 logarithm) levels of the ith protein for the HER2-positive, ER/PR-positive and healthy groups, respectively.

As in Example 1, the data were modeled as normally distributed, with the expectation values of X_i^{HER2}, $X_i^{\text{ER/PR}}$, and Y_i being ξ_i^{HER2}, $\xi_i^{\text{ER/PR}}$, and η_i, respectively. Such expectation values can be interpreted as the population means of the log-abundance of protein i. The POIs are $\theta_i^{\text{HER2}} = \left(\xi_i^{\text{HER2}} - \eta_i\right)/\sigma_i^{\text{HER2}}$ and $\theta_i^{\text{ER/PR}} = \left(\xi_i^{\text{ER/PR}} - \eta_i\right)/\sigma_i^{\text{ER/PR}}$, where σ_i^{HER2} and $\sigma_i^{\text{ER/PR}}$ are standard deviations related to each test-positive cancer group and healthy group. The significance measures described in Section 2 were used as different degrees of shrinkage to estimate the POI of the proteomics data, following the procedure explained in Section 3. In this context, the null hypothesis for each protein states that its average log-abundance level is not affected by cancer. The compared POI point estimators are transforms of: UE, FP, MLE, BBE1, HBE, HBEE and ELFDR.

The results are shown in Figure 2, which represents the estimated posterior median of the POI $\widehat{\theta}_i$ given by equation (12) evaluated at $\beta = 1/2$ using the MLE, BBE1, HBE, HBEE, FP and ELFDR versus $\widehat{\theta}_i$ using the UE. The figure show results for the HER2-positive and ER/PR-positive groups separately and indicates the identity with a dashed gray diagonal line. Notice the shrinkage effect of the point POI estimators with respect to the UE, which allows to better differentiate between the proteins affected or not by cancer by shrinking the POI of slightly affected proteins to 0. However, different POI point estimators lead to different conclusions about the individual proteins.

For example, in the HER2-positive case (Figure 2 left) the HBEE based POI point estimator identifies only one affected protein, while the FP based POI point estimator does

not identify any and the remaining POI point estimators identify several affected proteins. In turn, in the ER/PR-positive case (Figure 2 right), the HBEE based POI point estimator identifies 2 affected proteins, while again the FP based does not identify any and the remaining POI point estimators identify few affected proteins, given that most of them are shrunk to 0.

Therefore, the choice of the significance measure is crucial since some proteins would be considered as affected by cancer using one method, while considered as unaffected using other.

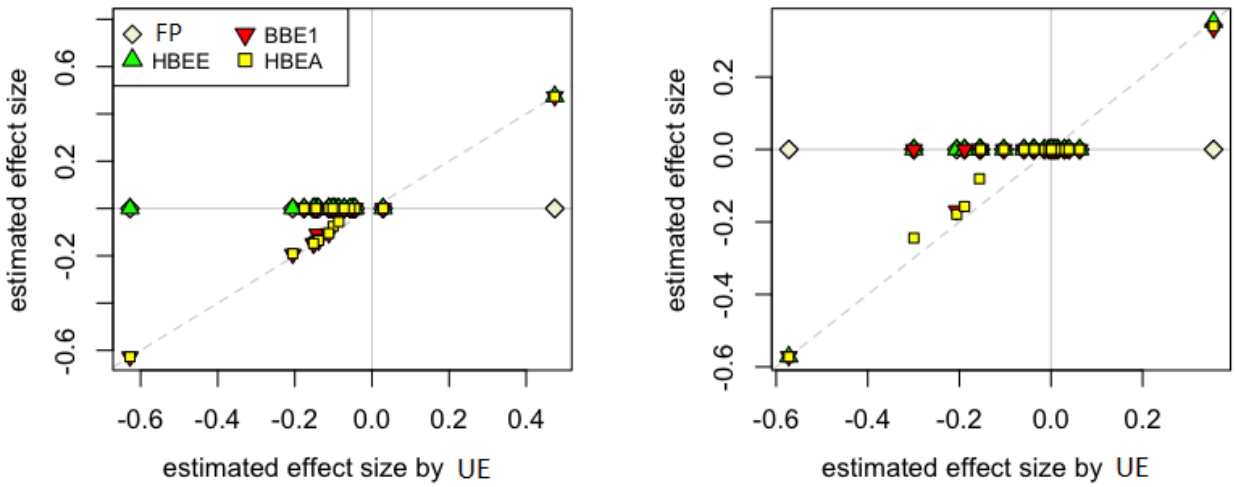

Figure 2: Illustration of shrinkage on real data sets: log2-abundance levels of 20 proteins in 2 groups of women that were positive in two different tests for breast cancer (HER2 and ER/PR) with respect to a group of healthy women. The left hand plot represents the comparison HER2-positive versus healthy women and the right hand plot represents the case ER/PR-positive versus healthy women. The plots show the posterior median of the POI $\widehat{\theta}_i$ given by equation (12) evaluated at $\beta = 1/2$ using the POI point estimators based on MLE, BBE1, HBE, HBEE, FP and ELFDR (vertical axis) versus $\widehat{\theta}_i$ using the UE (horizontal axis). Notice that the values of $\widehat{\theta}_i$ for the unaffected proteins are shrunken to 0, with respect to those obtained by the UE.

5 Simulations

In this section, we compare several POI estimators using artificial data sets that simulate the log-abundance of proteins. The different significant measures that generate the POI

Study	$\boldsymbol{N}$	$\boldsymbol{p_0}$	$\boldsymbol{\theta_{\text{alt}}}$
Fixed N	[3],6,[12,18,24,30]	$\mathbf{0, 1/N, 2/N, \ldots, (N-1)/N, 1}$	1.5,4
Fixed p_0	**1,3,6,12,18,24,30**	$5/6, 1$	1.5,4
Summary	**1,3,6,12,18,24,30**	$\geq 0, 5/6$	1.5,4

Table 1: Values of the parameters set for the simulation studies. Bold values indicate that the results are reported as a function of them. Values between brackets indicate that they do not appear in the figures.

estimators are briefly described in Section 2, while the procedure to build a POI estimator is described in Section 3. We call shrinkage POI estimators the POI estimators generated by any significant measure, to be distinguished from the UE which is also a POI estimator.

The comparison study is carried out trough three types of simulation, which show that the behavior of the POI estimators depends on specific conditions in terms of number of features N, detectability levels θ_{alt} and proportion of unaffected features p_0 as an approximation of π_0. The procedure for the estimation of the POI follows Examples 1 and 4. The performance of the different POI estimators are evaluated as POI interval estimators (IEs) and as POI point estimators (PEs).

Section 5.1 explains the general methodology involved in the simulations, including specific details related to the construction of the artificial data sets and the performance measures (Section 5.1.1) and the description of each type of simulation study (Section 5.1.2). Such general methodology is described step-by-step in Algorithms 1 and 2. The results of our studies are shown and explained in Section 5.2 and discussed in Section 6. The last simulation study is a summary in which we identify the best POI estimators based on their lowest worst-case bias and worst-case effective width of the confidence interval. Table 1 summarizes the values of the parameters set for the performed simulation studies and the reported results.

5.1 Simulations

5.1.1 General concepts

Artificial data sets Following Examples 1 and 4, each simulated data set consists of log-abundance levels of N proteins for 10 individuals equally divided in the treatment and control groups, with $\sigma_i = \sigma_i' = 1$, $\xi_i' = 0$ and $\xi_i = 0, \xi_{\text{alt}}$. Therefore, if the ith protein is affected, 5 abundance levels are drawn from $\text{N}(0, 1)$ (control group) and 5 are drawn from $\text{N}(\xi_{\text{alt}}, 1)$ (treatment group), otherwise all abundance levels for the individuals in both groups are drawn from $\text{N}(0, 1)$. Given ξ_{alt}, the true POI is $\theta_{\text{alt}} = \xi_{\text{alt}} - 0 = \xi_{\text{alt}}$.

Performance measures To evaluate the different POI estimators, we compare three parameters: coverage rate (CR), effective width (EW) of the confidence interval (CI) and root mean squared error (RMSE) of the POI point estimates. The performance of the POI estimators as IEs is assessed by CR and specially EW, while the RMSE assesses the POI estimators as PEs.

To obtain the POI interval estimates, we first selected a nominal coverage rate (NCR) and then we calculated the experimental CR of the known POI θ_{alt}. The valid IEs are those that show CR $\geq$ NCR. In these studies, we have considered two values for NCR; 95% and 0%, and a small tolerance of 5%, so that the valid estimators fulfill CR $\geq$ NCR $-$ 5. Then, for NCR=95%, the EW consists of the width of the CI (CIW) if CR $\geq$ 90% is satisfied or infinite otherwise. When NCR=0%, the CIs have 0 width, so all EWs are 0 and we have only PEs. Thus, the CR assesses the validity of the IEs and is denoted by CR^{M}, while the EW, denoted EW^{M}, accounts for the width of the valid CI. The superscript M stands for any of the significance measures that generates the POI estimators included in this study; UE, FP, MLE, BBE1, ELFDR, HBE and HBEE. In every case, the mean is computed over the B number of simulations as well as the N proteins. The computation of these performance measures is described using a pseudo-code scheme in Algorithm 2.

In order to obtain the POI point estimates we computed both the posterior median $\widehat{\theta}$

(see Example 4) and the posterior mean estimates $E(\theta)$ (by equation (13)). The error of the PEs is evaluated through the RMSE with respect to the known POI θ_{alt}. $\text{RMSE}^{\text{M}}_{\widehat{\theta}}$ and $\text{RMSE}^{\text{M}}_{E(\theta)}$ denote the RMSE of the $\widehat{\theta}$ and the $E(\theta)$ estimates, respectively (see Algorithm 2).

Algorithm 1 Pseudo-code of the simulation process.

1. For each value of the 5 input parameters (N, θ_{alt}, N_0, NCR, B)
2. For each simulation j = 1, 2, ..., B
   (a) Build 2 IID X^j and Y^j, the vectors of log-abundance levels of N proteins of $m = n = 5$ sick and healthy individuals, respectively. Since the value N_0 (number of unaffected proteins) is set, so is p_0 according to equation (14):
      i. For each i = 1, 2, ..., N_0, $X^j_i = [X^j_{i,1}, ..., X^j_{i,m}]$, X^j_i is randomly drawn from N(0, 1)
      ii. For each $i = N_0 + 1$, ..., N, $X^j_i = [X^j_{i,1}, ..., X^j_{i,m}]$, X^j_i is randomly drawn from N(ξ_{alt}, 1)
      iii. For each i = 1, 2, ..., N, $Y^j_i = [Y^j_{i,1}, ..., Y^j_{i,n}]$, Y^j_i is randomly drawn from N(0, 1)
      iv. Thus $X^j = [X^j_1, ..., X^j_{N_1}, X^j_{N_1+1}, ..., X^j_N]$ and $Y^j = [Y^j_1, ..., Y^j_N]$.
   (b) For each protein i = 1, 2, ..., N
      i. Use the true POI $\theta_i = 0, \theta_{\text{alt}}$ where $\theta_{\text{alt}} = \xi_{\text{alt}} - 0 = \xi_{\text{alt}}$.
      ii. Estimate the significance measures $\widehat{\psi}^{\text{M}}_{ij}$ by every method M = UE, FP, MLE, BBE1, ELFDR, HBE and HBEE.
      iii. For every $\widehat{\psi}^{\text{M}}_{ij}$, compute:
         A. the limits of the estimated confidence interval $\text{CI}^{\text{M}}_{ij}(\alpha)$ with α=1/40 when NCR $= 95\%$ and with $\alpha = 1/2$ when NCR $= 0\%$, where
         $$\text{CI}^{\text{M}}_{ij}(\alpha) = [\theta_{ij}(\alpha/2), \theta_{ij}(1-\alpha/2)]$$
         and $\theta_{ij}(\beta)$ is given by equation (12).
         B. the width $\text{CIW}^{\text{M}}_{ij}$ of $\text{CI}^{\text{M}}_{ij}(\alpha)$ by $\text{CIW}^{\text{M}}_{ij}(\alpha) = \theta_{ij}(1-\alpha/2) - \theta_{ij}(\alpha/2)$
         C. the posterior median of the POI $\widehat{\theta}^{\text{M}}_{ij}$ by $\widehat{\theta}^{\text{M}}_{ij} = \text{CI}^{\text{M}}_{ij}(1/2)$
         D. the posterior mean of the POI $\text{E}(\theta)^{\text{M}}_{ij}$ by equation (13)
         E. the error $\widehat{\xi}^{\text{M}}_{ij}$ of posterior median by $\widehat{\xi}^{\text{M}}_{ij} = (\hat{\theta}^{\text{M}}_{ij} - \theta_i)^2$
         F. the error $\bar{\xi}^{\text{M}}_{ij}$ of posterior mean by $\bar{\xi}^{\text{M}}_{ij} = (\text{E}(\theta)^{\text{M}}_{ij} - \theta_i)^2$

Algorithm 2 Pseudo-code of the bias approximations.

1. For each value of the 5 input parameters (N, $\theta_{\texttt{alt}}$, N_0, NCR, B)

   (a) For each significance measure M = UE, FP, MLE, BBE1, ELFDR, HBE and HBEE, compute:

   i. the mean of the coverage rate $\texttt{CR}^{\texttt{M}}$ by:

$$\texttt{CR}^{\texttt{M}} = \frac{1}{B}\sum_{j=1}^{B}\frac{1}{N}\sum_{j=1}^{N} 1_{\theta_i \in \texttt{CI}^{\texttt{M}}_{ij}}$$

   ii. the mean of the confidence interval width $\texttt{CIW}^{\texttt{M}}$:

$$\texttt{CIW}^{\texttt{M}} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{B}\sum_{j=1}^{B} \texttt{CIW}^{\texttt{M}}_{ij}$$

   iii. the effective width $\texttt{EW}^{\texttt{M}}$:

$$\texttt{EW}^{\texttt{M}} = \begin{cases} \texttt{CIW}^{\texttt{M}} & if\ \texttt{CR} \geq \texttt{NCR} \\ \infty & if\ \texttt{CR} < \texttt{NCR} \end{cases}$$

   iv. the RMSE of the posterior median $\texttt{RMSE}^{\texttt{M}}_{\widehat{\theta}}$:

$$\texttt{RMSE}^{\texttt{M}}_{\widehat{\theta}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{B}\sum_{j=1}^{B}\hat{\xi}^{\texttt{M}}_{ij}}$$

   v. the RMSE of the posterior mean $\texttt{RMSE}^{\texttt{M}}_{E(\theta)}$:

$$\texttt{RMSE}^{\texttt{M}}_{E(\theta)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{B}\sum_{j=1}^{B}\bar{\xi}^{\texttt{M}}_{ij}}$$

5.1.2 Description of simulations

Study for fixed N (dependence on p_0) In this simulation study, we compare the different POI estimators in several situations regarding the number of proteins N, the known true POI of the simulated data set θ_{alt} and the proportion of unaffected proteins p_0 given by,

$$p_0 = \frac{N_0}{N} = \frac{(N - N_1)}{N}, \tag{14}$$

where N_0 and N_1 are the number of unaffected and affected proteins, respectively. Thus, p_0 approximate the values of π_0 for fixed values of θ_{alt} and N.

To represent the two situations in which the differences between the null and alternative distributions are barely detectable and highly detectable, we considered two values for the POI of the affected proteins in the treatment group; a low value $\theta_{\text{alt}} = 1.5$ and a high value $\theta_{\text{alt}} = 4$ relative to the standard deviation $\sigma = 1$. Additionally, to study the dependance of the POI estimators on the scale of the data set, we considered several values of N, $N = 1, 3, 6, 12, 18, 24, 30$. Since there is an important influence of the p_0 in the LFDR methods, we include several values of p_0, which depend on N. Thus, for example, for $N = 6$, the possible values of the affected proteins are $N_1 = 0, 1, ..., 5, 6$, then $p_0 = {}^6\!/_6, {}^5\!/_6, ..., {}^1\!/_6, {}^0\!/_6$, respectively. Finally, every simulation was repeated $B = 200$ times. Therefore, for $N = 6$ we built $1 \times 2 \times 7 \times 2 \times 200$ data sets corresponding to the combinations of the parameters $(N, \theta_{\texttt{alt}}, N_1, \text{NCR}, B)$.

Study for fixed p_0 (dependence on N) In this study, we consider the usual situations in biostatistics by which it is deemed that most of the proteins under study are not affected. Such situation refers to high p_0 values. Therefore, we fixed θ_{alt} and a high p_0 value and we measured the estimators performance with regard to N. The selected values for p_0 are $p_0 = {}^5\!/_6 = 83.3\%$ and $p_0 = 100\%$, which represent few affected and none affected proteins, respectively. The values for θ_{alt}, N and B are the above mentioned. Thus, for $p_0 = {}^5\!/_6$ and $N = 6$, we considered $2 \times 5 \times 2 \times 200$ and for $p_0 = 100\%$ we considered $2 \times 1 \times 2 \times 200$ data sets, as many as combinations of $(N, \theta_{\texttt{alt}}, N_1, \text{NCR}, B)$. Notice that $p_0 = {}^5\!/_6$ is not applicable to $N = 1, 3$ and that $\theta_{\text{alt}} = 0$ when $p_0 = 100\%$, since there is no affected protein in this case. Results are shown in Section 5.2.

Best worst-case POI estimator Finally, in this last study, we compare the worst-case values of the performance measures in order to identify the best POI estimator for the studied cases with respect to N, which is the only known parameter. The worst-case for a given

performance measure consist of the worst value of such performance measure computed over cases with different values of p_0 and θ_{alt} for a fixed N. The values for N and θ_{alt} are the already mentioned $N = 1, 3, 6, 12, 18, 24, 30$ and $\theta_{\text{alt}} = 1.5, 4$. The worst-case value of the RMSE and the EW is the maximum of their respective values, while the worst-case value of the CR is the minimum value, all computed over the different cases. In summary, the "best" estimator is the one showing the smallest worst-case RMSE and EW. Therefore, the best worst-case estimator is the safest one.

Since sometimes it is possible to make assumptions about p_0, we consider two situations: we can assume that p_0 is high (few affected proteins) and we cannot make any assumption about p_0 (any number of affected proteins is possible). For the first situation, we consider all the studied cases with selected $p_0 \geq 5/6$, which defines the cases with most proteins unaffected given by $N_0 \geq p_0 \times N = 5/6 \times N$ (this excludes $N = 1, 3$), while the second situation corresponds to $p_0 \geq 0$, thus $N_0 = 0, 1, .., N$. Therefore, for fixed N, when $p_0 \geq 5/6$, the worst-case values for every performance measure (CR, EW and RMSE) are chosen among $2 \times N/6$ data sets and when $p_0 \geq 0$, the worst value are chosen among $2 \times (N + 1)$ data sets, as many as combinations of $(\theta_{\text{alt}}, N_0)$.

For example, for $N = 12$, the condition $p_0 \geq 5/6$ includes $N_0 \geq 10$ unaffected proteins ($N_1 < 2$ affected). Thus, the worst-case CR value is the minimum value of the CR evaluated at $p_0 = N_0/N = 10/12, 11/12, 12/12$ and at $\theta_{\text{alt}} = 1.5, 4$, this is $\min_{N_0 \in \{10,11,12\}, \theta_{\text{alt}} \in \{1.5,4\}} \text{CR}^{\text{M}}(12, \theta_{\text{alt}}, N_0)$. Similarly, the worst-case EW value is $\max_{N_0 \in \{10,11,12\}, \theta_{\text{alt}} \in \{1.5,4\}} \text{EW}^{\text{M}}(12, \theta_{\text{alt}}, N_0)$, and the worst-case posterior median and posterior mean estimates of the POI is $\max_{N_0 \in \{10,11,12\}, \theta_{\text{alt}} \in \{1.5,4\}} \text{RMSE}^{\text{M}}(12, \theta_{\text{alt}}, N_0)$ (see Algorithm 2).

5.2 Results

Here we show the obtained results for NCR=95%, since for NCR=0% the point estimates of the POIs are the same and CR and EW are 0. Results are graphically depicted in Figures 3 to 9, where the performance measures are represented as a function of p_0 (fixed N study) and

as a function of N (fixed p_0 study and summary), according to Table 1 and the description of each type of simulation in Section 5.1.2.

Fixed N study (dependence on p_0) Figures 3 and 4 display the CR, EW and RMSE of the posterior median of the POI for $N = 6$ and the considered values of $\theta_{\text{alt}} = 1.5, 4$ as a function of p_0 (equation (14)) ranging from 0% ($N_1 = N,\ N_0 = 0$) to 100% ($N_1 = 0,\ N_0 = N$).

Figure 3 shows that different IEs have different CRs; CRs for the IEs based on the FP, the MLE and the UE are almost constant and maximum in most of the cases, while remaining IEs have valid CRs (CRs$\geq$ 90%, with tolerance of 5%) for the highest values of p_0. Results are similar for $\theta_{\text{alt}} = 4$, although the regions of p_0 for which the IEs are valid are slightly extended to smaller p_0s (See also Figure 4). Therefore, IEs other than those based on the FP, the MLE and the UE can only be applied under the assumption of high p_0 (when $N = 6$). Moreover, the HBEE and the ELFDR based IEs show the worst CRs.

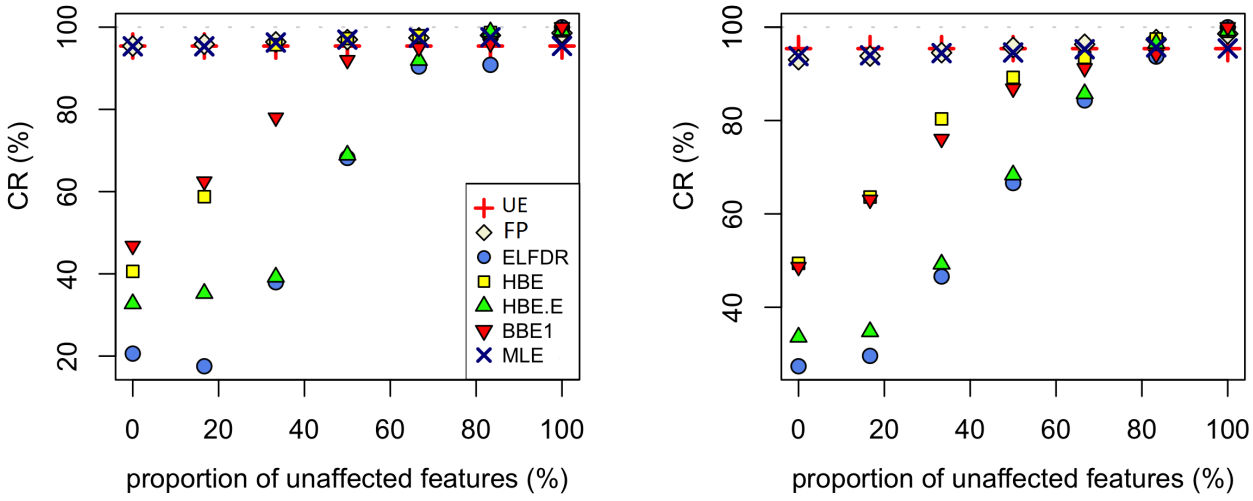

Figure 3: CR for $\theta_{\text{alt}} = 4$ (left), $\theta_{\text{alt}} = 1.5$ (right) and $N = 6$. IEs other than those based on UE, FP and MLE are valid only for high p_0. The range of valid p_0s is wider for $\theta_{\text{alt}} = 4$.

Figure 4 shows the EW as a function of p_0 for IEs with CR$\geq$ 90%, since the EW is infinite otherwise. In general, there is a decrease in the EW with the increase of p_0 for all the shrinkage IEs, while the UE shows a constant value which is also the maximum value. The best IE is the one having the smallest EW, which depends on p_0 and N (Figures 4 and

also 6). Results for $\theta_{\text{alt}} = 1.5$ are similar, although the MLE based IE has a much smoother EW than for $\theta_{\text{alt}} = 4$. In summary, for $N = 6$ and any p_0 any shrinkage IE has smaller EWs than the UE, and the overall smallest EW is given by the BBE1 based IE at $p_0 = 100\%$.

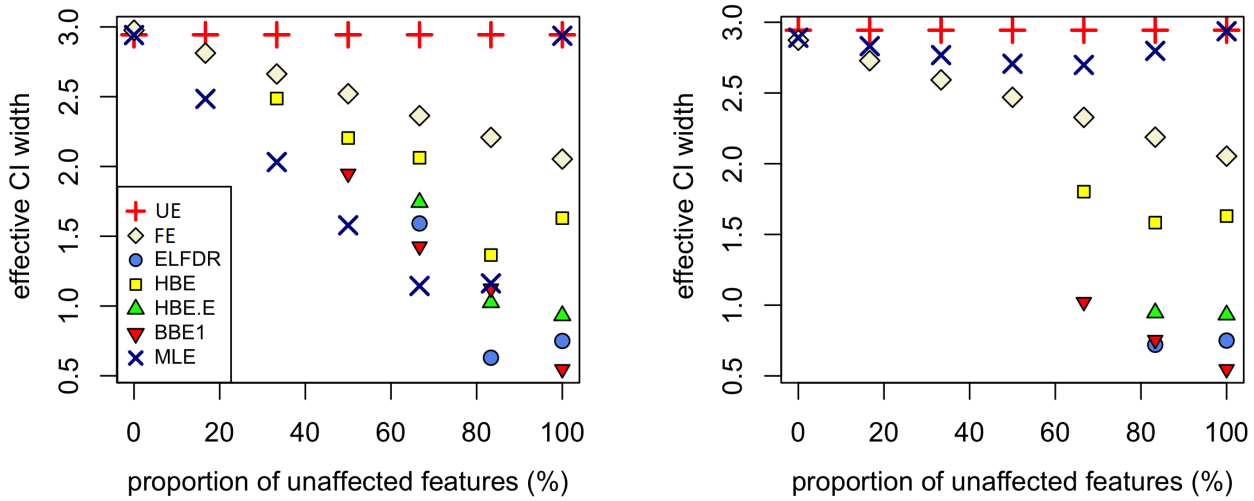

Figure 4: EW for $\theta_{\text{alt}} = 4$ (left), $\theta_{\text{alt}} = 1.5$ (right) and $N = 6$. In general, there is a decrease of EW with respect to increasing p_0. The MLE is the IE most dependent on θ_{alt}. Only results for which CR$\geq 90\%$ are displayed.

The RMSE of $\widehat{\theta}$ is displayed as a function of p_0 in Figure 5. It can be seen a general increase of the RMSE with respect to decreasing p_0 for PEs other than the ones based on the MLE and the UE. The UE and the MLE based PEs show almost constant and similar RMSEs at all p_0 values, whereas the PEs based on the ELFDR, the HBEE and the BBE1 are very sensitive to p_0. In general, the shrinkage PEs show smaller error than the UE only for high values of p_0s, the threshold value of which depends on θ_{alt}. Specifically, for $N = 6$ and $p_0 = 100\%$, the BBE1 based is clearly the best PE.

Fixed p_0 study (dependence on N) Figure 6 and 7 depicts the EW and the RMSE of the posterior median of the POI for the considered values of $\theta_{\text{alt}} = 1.5, 4$ and $p_0 = 83.3, 100\%$ as a function of N.

Figures 6 (top row) and 7 (left) show that the EWs of most IEs depend on both θ_{alt} and N, but they are always smaller than the EW of the IE based on the UE. For $p_0 = 5/6 = 83.3\%$,

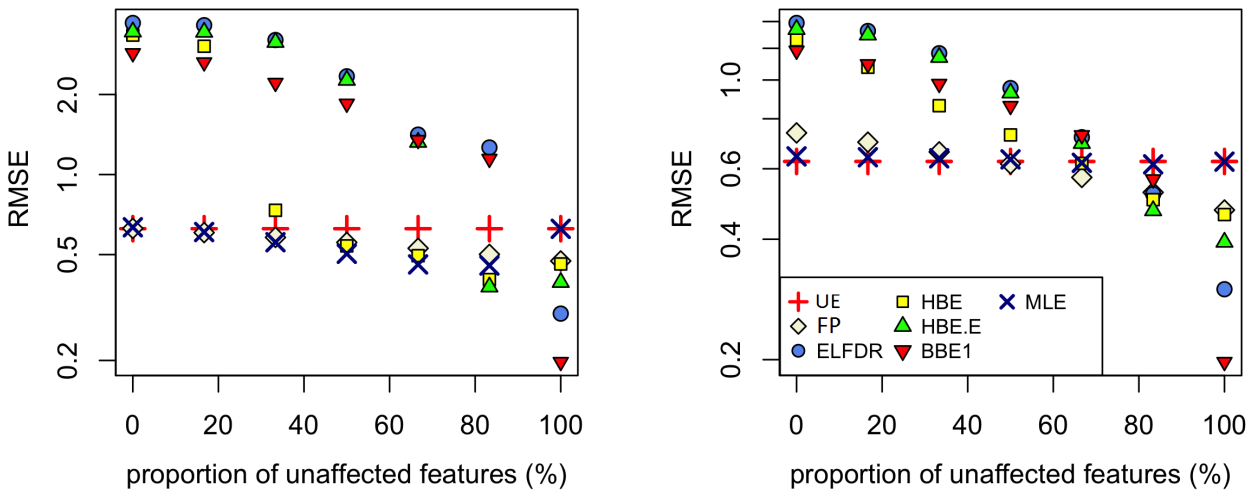

Figure 5: RMSE of the posterior median of the POI for $\theta_{\text{alt}} = 4$ (left), $\theta_{\text{alt}} = 1.5$ (right) and $N = 6$. RMSE increases with increasing N_1(decreasing p_0) for most PEs.

Figure 6 (top row) shows that the EW of the FP based IE is constant with respect to N and θ_{alt}, while the MLE based IE has very different EWs with respect to θ_{alt}. The behavior of the MLE based IE is in agreement with the results displayed in Figure 4, which highlighted its high sensitivity to θ_{alt}. The EWs of remaining shrinkage IEs show a slight dependence on N, specially when θ_{alt} is small. Such dependence on N is enhanced when $p_0 = 100\%$ (thus $\theta_{\text{alt}} = 0$), as shown in Figure 7 (left). In addition, for both $p_0 = 83.3\%$ and $p_0 = 100\%$ the IE generated by the UE has the maximum constant EW value, which does not depend on either N nor θ_{alt}. Therefore, for high p_0 values any shrinkage IE shows smaller EW than the UE disregarding N and θ_{alt}. Taking into account the results from previous simulation study for fixed N shown in Figure 4, we can conclude that any shrinkage IE overcomes the UE, specially for high p_0 values. Specifically, the ELFDR based IE is the best IE when $p_0 = 83.3\%$, while the BBE1 based is the best IE when $p_0 = 100\%$, and the EWs of both IEs are quite constant with respect to N at the considered θ_{alt} values.

Regarding the RMSE of $\widehat{\theta}$, in general the PEs also depend on N and θ_{alt}, however any shrinkage PE does not always show smaller bias. As seen in Figure 6 (bottom row) and right hand of Figure 7, more PEs become dependent on N as θ_{alt} decreases. Furthermore, when θ_{alt} is small (and p_0 is high) most of the shrinkage PEs have better RMSEs than the

UE. For example, for $p_0 = 83.3\%$ and $\theta_{\text{alt}} = 1.5$ (Figure 6, bottom row left) and also for $p_0 = 100\%$ (Figure 7, right) all the shrinkage PEs have smaller RMSE than the UE, while for $p_0 = 83.3\%$ and $\theta_{\text{alt}} = 4$ only the PEs based on the FP, the MLE and the HBE overcome the UE. This effect agrees with results from previous simulation study for fixed N seen in Figure 5. Finally, we can see that the HBE based is the best PE when $p_0 = 83.3\%$, while the BBE1 based is the best PE when $p_0 = 100\%$.

The special case $N = 1$ is depicted in Figure 7 for the case of an unaffected protein, given that $p_0 = 100\%$. In this case only the POI estimators generated by the FP and the UE can be used, and we can see that the FP based both IE and PE overcomes the UE.

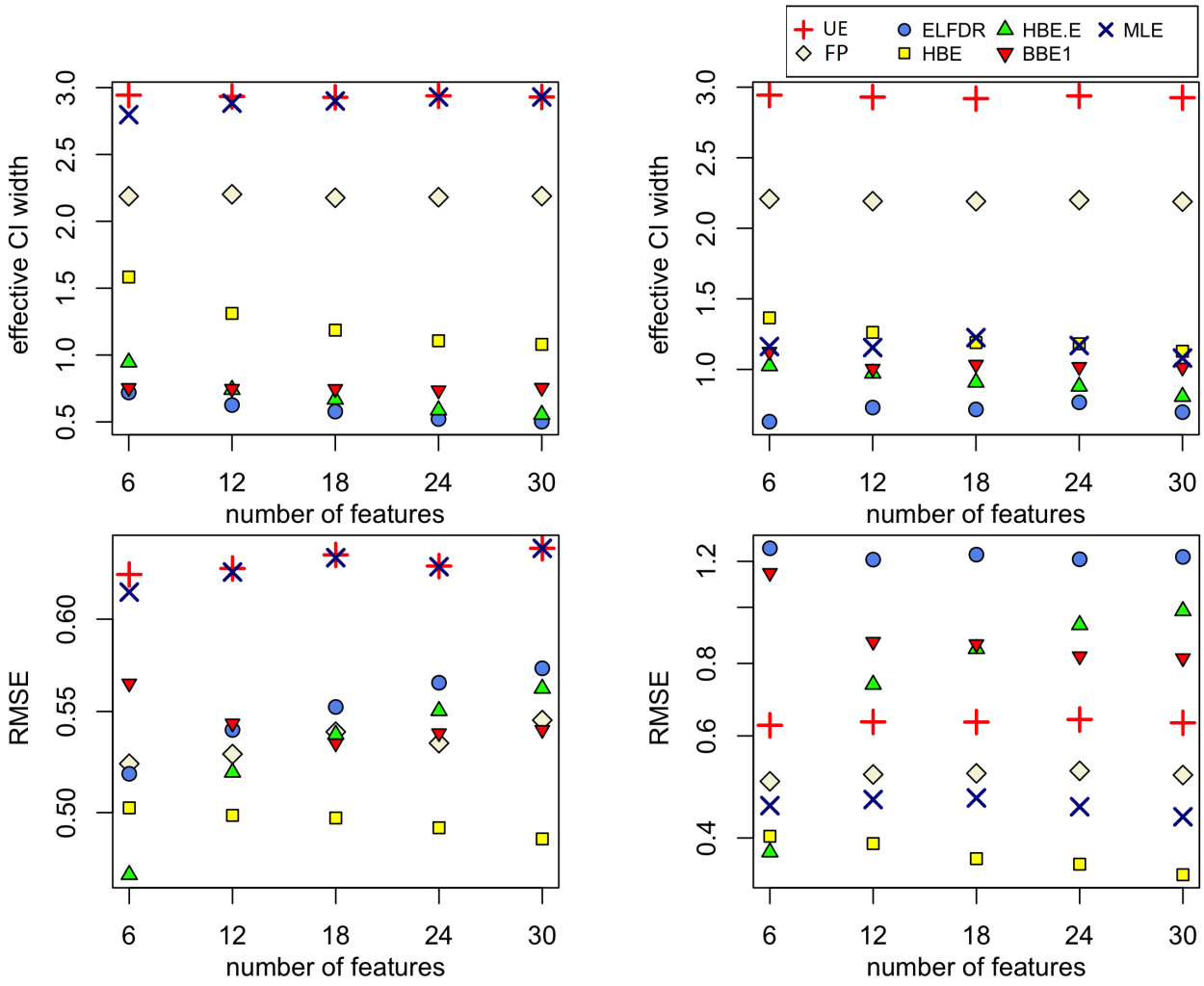

Figure 6: EW (top row) and RMSE of the posterior median of the POI (bottom row) for $\theta_{\text{alt}} = 1.5$ (left column) and $\theta_{\text{alt}} = 4$ (right column) for $p_0 = 5/6 = 83.3\%$.

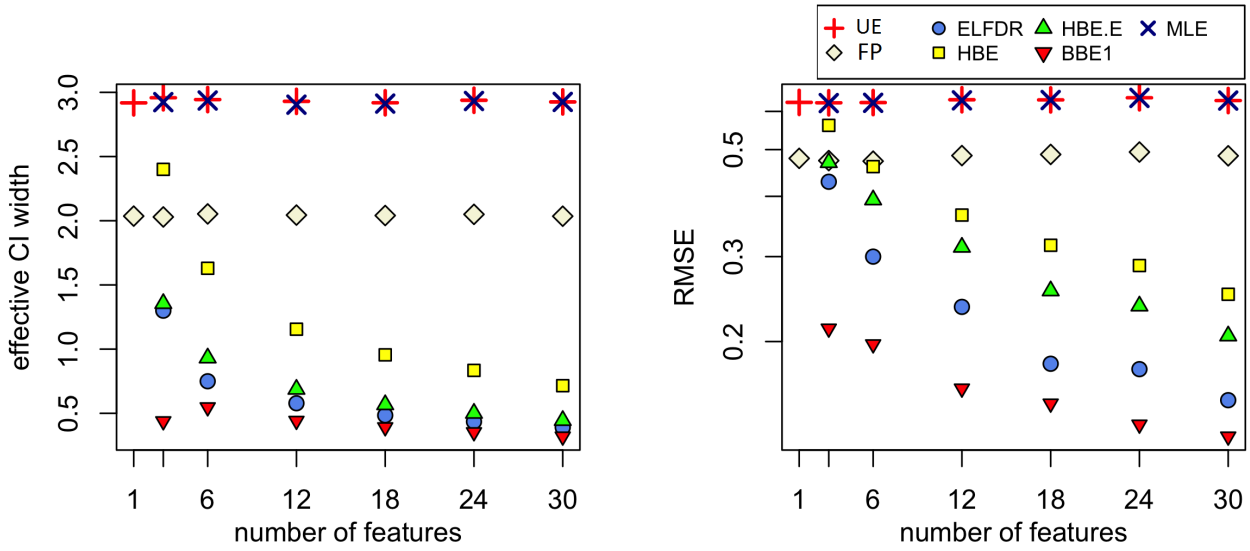

Figure 7: EW (left) and RMSE of posterior median estimates (right) for $p_0 = 1 = 100\%$.

Best worst-case POI estimator Figure 8 displays the worst-case values of EW for $p_0 \geq 0$ (left), and $p_0 \geq 5/6 = 83.3\%$ (right), while Figure 9 displays the worst-case values of the RMSE of $\widehat{\theta}$ (top row) and of $E(\theta)$ (bottom row) for $p_0 \geq 0\%$ (left column), and $p_0 \geq 5/6 = 83.3\%$ (right column). The RMSE for both point estimation approaches, $\widehat{\theta}$ and $E(\theta)$, are extremely similar for each PE (Figure 9).

Figure 8 (left) shows the worst-case EW over all the $p_0 \geq 0\%$ values as a function of N. Only the IEs based on the FP, the MLE and the UE are shown because, as seen in the fixed N simulation study in Figure 3, they are the only IEs that have CR$\geq 90\%$ at all p_0 values. In addition, the left column of Figure 9, displays the worst-case RMSE values over all the $p_0 \geq 0\%$. According to these figures, the FP generates the worst worst-case IE while the MLE and the UE generate the best PEs. Notice that the results of the simulation study for fixed N (Figure 4 and 5) show that the worst-case EW value is given by FP at $p_0 = 0\%$ and $\theta_{\text{alt}} = 4$, and that the worst-case RMSE values of all the shrinkage PEs (also at $p_0 = 0\%$) are much larger than the RMSE from the UE and MLE based PE.

In the special case $N = 1$, unlike for N>1 the FP gives the best worst-case IE, while similarly to N>1 the UE gives the best worst-case PE.

On the other hand, the right hand plot of Figure 8 and the right column of Figure 9 depict the worst-case EW and the worst-case RMSE over all the $p_0 \geq 5/6 = 83.3\%$ values as a function of N, respectively. It can be seen that, while the ELFDR gives the best IE (Figure 8, right), it also gives the worst PE (Figure 9, right column). In addition, we can see that in general the HBE gives a good IE and a very good PE for most N values, and that the UE, the MLE and the FP generate bad IEs but good PEs. The latter results agree with those from previous simulation study for fixed p_0 (bottom row of Figure 6, and right hand plot of 7), where it is shown that more shrinkage PEs have smaller RMSE than the UE when p_0 and θ_{alt} are enough large and enough small, respectively.

6 Discussion

As shown in Section 4, it is important to select the appropriate POI estimator for a given data set. Our study shows the behavior of several POI estimators as IEs and as PEs under different situations in terms of detectability level (θ_{alt}), number of proteins (N) and proportion of unaffected proteins (p_0). Although such situations are ideal, the conclusions derived from this study can be very helpful for the analysis of data sets coming from real measurements. For such data sets, the variable N is known and p_0 can sometimes be assumed to have high values, but θ_{alt} is unknown. For this reason, to identify the most suitable POI estimator for a given data set, we defined in Section 5.1.2 the worst-case and determined the best and worst values of the performance measures, which are the EW to assess IEs and RMSE to assess PEs.

Therefore, to summarize the results presented in Section 5.2, we compare the performance measures of the worst-case for every N and for 2 assumptions about p_0: p_0 is completely unknown and p_0 has high values (Figures 8 and 9).

The situation for which any assumption about p_0 can be made is displayed in Figure 8 (left) and Figure 9 (left column). According to the discussion given in Section 5.2 for

worst-case performance measures, the UE and the MLE based are the best IEs to be used for $N > 1$. However, given that the differences in the worst EW among the UE, the MLE and the FP based IEs is very small for all Ns, and given that the worst-case EW for the FP based IE was given at $p_0 = 0\%$ (Figure 4), since $p_0 = 0\%$ is very unlikely we recommend to use the FP based IE for all Ns or the MLE based IE for $N > 1$. In relation with the RMSE, it is clear that the best PE to be used is the UE (Figure 9, left column and Figure 5).

On the other hand, for the most likely situation p_0 is high, thus only some few proteins are affected. This situation is displayed for $p_0 \geq 5/6 = 83.3\%$ in Figure 8 (right) and Figure 9 (right column). From the previous discussion in Section 5.2, we conclude that the ELFDR gives the best IE while the HBE gives the best PE for most Ns.

In summary, the choice of the POI estimator should depend on the number of null hypothesis and whether it can be assumed that most of them are true. In any case, any shrinkage IE gives more accurate interval estimates of the POI than the UE based IE, while only some shrinkage PEs give more accurate point estimates of the POI when p_0 can be assumed to be high. Table 2 summarizes the conditions for the best worst-case POI estimators shown in Figures 4 to 9.

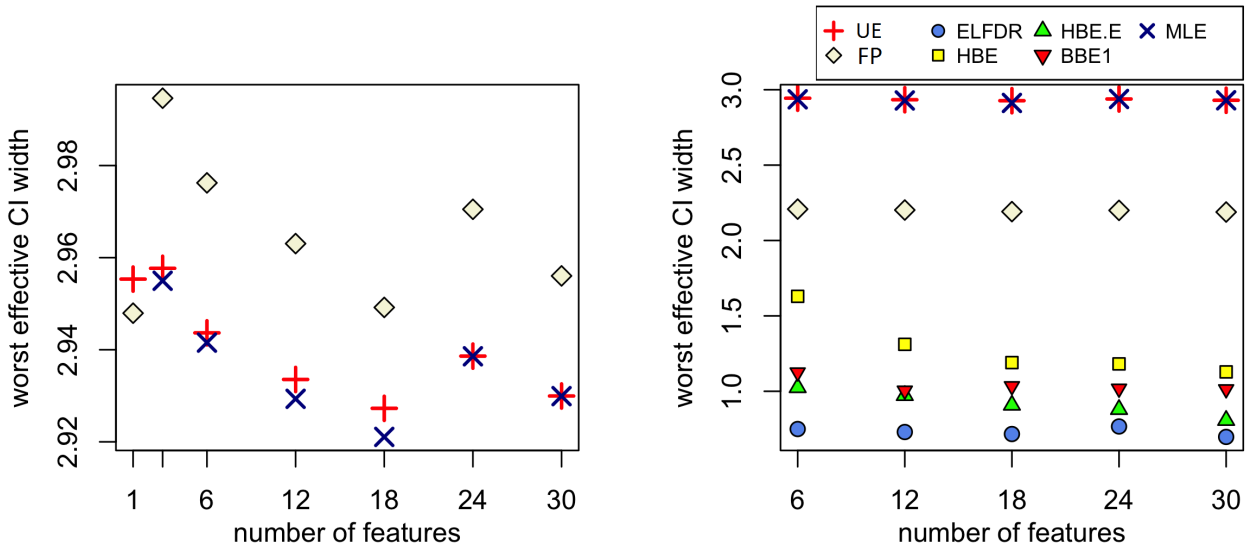

Figure 8: Worst-case mean effective width for $p_0 \geq 0\%$ (left) and $p_0 \geq 5/6 = 83.3\%$ (right).

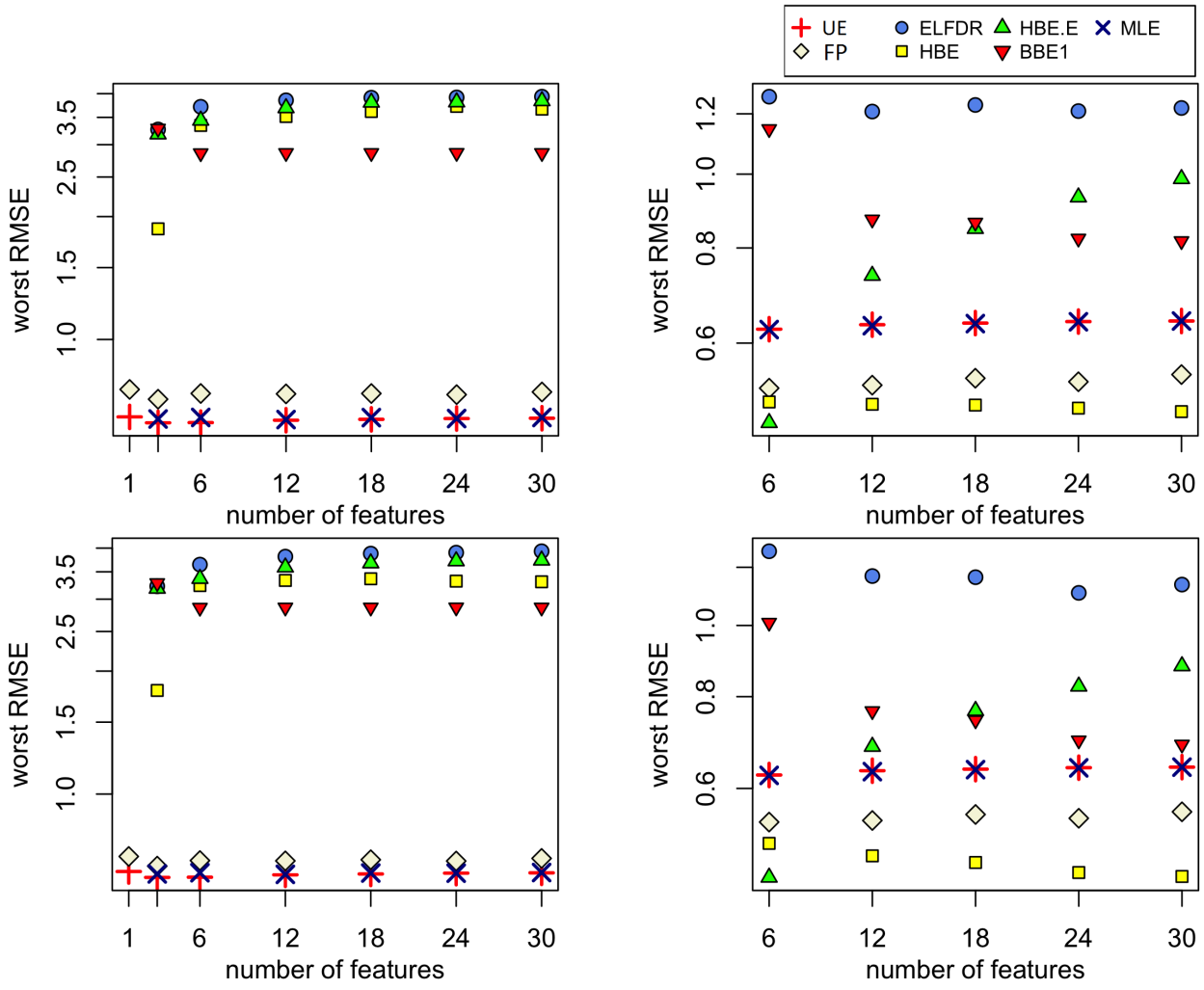

Figure 9: Worst-case RMSE of the posterior median (top row) and worst-case RMSE of the posterior mean of the POI (bottom row) for $p_0 \geq 0\%$ (left column) and $p_0 \geq 5/6 = 83.3\%$ (right column).

Acknowledgments

This research was supported by the Natural Sciences and Engineering Research Council of 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.

We used the following packages of R (R Development Core Team, 2008): `Biobase` (Gentleman et al., 2004) and `qvalue` (Dabney et al., 2011) from Bioconductor (Gentleman et al., 2004); `locfdr` (Efron et al., 2011), `fBasics` (Wuertz, 2010), and `distr` (Ruckdeschel et al.,

N	p_0	**Best worst-case shrinkage interval estimator**	**Best worst-case shrinkage point estimator**
$N = 1$	$[0, 1]$	FP	UE
$N = 1$	$[1, 1]$	FP	FP
$N \geq 3$	$[1, 1]$	BBE1	BBE1
$N = 6$	$[5/6, 1]$	ELFDR	HBEE
$N \geq 12$	$[5/6, 1]$	ELFDR	HBE
$1 < N < 24$	$[0, 1]$	MLE	UE, MLE*
$N \geq 24$	$[0, 1]$	UE, MLE*	UE, MLE*

Table 2: Best worst-case estimators. * The difference in the errors of the two estimators is negligible.

2006) from the CRAN repository.

References

Bickel, D. R., 2012. 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., 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 (4), 529–543.

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 (0), 204 – 221.

Cohen, J., 1988. Statistical power analysis for the behavioral sciences (ed 2) lawrence erlbaum. Hillsdale, NJ.

Dabney, A., Storey, J. D., with assistance from Gregory R. Warnes, 2011. qvalue: Q-value estimation for false discovery rate control. Reference Manual, R package version 1.26.0.

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

Efron, B., 2004b. Selection and estimation for large-scale simultaneous inference. Technical Report; Department of Statistics and Health, Research and Policy; Stanford University.

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

Efron, B., 2008. Microarrays, empirical Bayes and the two-groups model. Statistical Science 23, 1–22.

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

Efron, B., Tibshirani, R., 2007. On testing the significance of sets of genes. Ann Appl Stat 1, 107–129.

Efron, B., Tibshirani, R., Storey, J. D., Tusher, V., 2001. Empirical Bayes analysis of a microarray experiment. Journal of the American Statistical Association 96, 1151–1160.

Efron, B., Turnbull, B. B., Narasimhan, B., 2011. locfdr: Computes local false discovery rates. Reference Manual, R package version 1.1-7.

Gentleman, R. C., Carey, V. J., Bates, D. M., et al., 2004. Bioconductor: Open software development for computational biology and bioinformatics. Genome Biology 5, R80.

Good, I. J., 1966. How to Estimate Probabilities. IMA Journal of Applied Mathematics 2, 364–383.

Li, X., 2009. ProData. Bioconductor.org documentation for the ProData package.

Li, Z., Li, Q., Geng, L., Chen, X., Bi, K., 2013. Use of the local false discovery rate for identification of metabolic biomarkers in rat urine following genkwa flos-induced hepatotoxicity. PLoS ONE 8, 67451.

Montazeri, Z., Yanofsky, C. M., Bickel, D. R., 2010. Shrinkage estimation of effect sizes as an alternative to hypothesis testing followed by estimation in high-dimensional biology: Applications to differential gene expression. Statistical Applications in Genetics and Molecular Biology 9, 23.

Nadarajah, S., Bityukov, S., Krasnikov, N., 2015. Confidence distributions: A review. Statistical Methodology 22 (0), 23 – 46.

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.

R Development Core Team, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ruckdeschel, P., Kohl, M., Stabla, T., Camphausen, F., May 2006. S4 classes for distributions. R News 6 (2), 2–6.

Welch, B. L., 1947. The generalization of student's problem when several different population variances are involved. Biometrika 34, 28–35.

Wuertz, D., 2010. fbasics: Rmetrics - markets and basic statistics. Reference Manual, R package version 2110.79.

Yanofsky, C. M., Bickel, D. R., 2010. Validation of differential gene expression algorithms: Application comparing fold-change estimation to hypothesis testing. BMC Bioinformatics 11, art. 63.