

Optimal Estimation of False Discovery Rates

Brit B. Turnbull
Stanford University

November 20, 2007

Abstract

In large-scale multiple testing – performed, for example, in studies involving genomics or proteomics – it is useful to estimate the false discovery rate (FDR), which can be interpreted as the probability that a null hypothesis is true. Estimation of FDR involves specification of a null distribution, which Efron (2007) estimates by maximum likelihood under the assumption that test statistics corresponding to false null hypotheses are sufficiently extreme. We introduce a modification to the method that produces consistent FDR estimates even when the assumption is violated. The new method is applied to data from a microarray study comparing HIV-positive to HIV-negative subjects and from the Beta-Blocker Evaluation of Survival Trial (BEST) Proteomics Study, which aims to find biomarkers related to survival time of heart disease patients with SELDI-TOF mass spectrometry.

1 Introduction

Large-scale multiple testing involves testing hundreds or thousands or more of null hypotheses simultaneously. Dudoit, Shaffer, and Boldrick (2003) review

several approaches to the problem, including controlling the false discovery rate (FDR), introduced by Benjamini and Hochberg (1995), and estimating FDR, as in, for example, Tusher, Tibshirani, and Chu (2001).

Estimation of FDR, described in detail in Section 2, relies heavily on the null distribution to which observed test statistics are compared. The “theoretical null distribution” – the marginal distribution of each test statistic under the null hypothesis and possibly other assumptions – can be inappropriate and can lead to biased FDR estimates (Efron 2004). To avoid the problems with the theoretical null distribution, Efron (2004) proposes replacing it with the “empirical null distribution,” calculated using the observed test statistics.

Proposed methods for estimating the empirical null distribution involve Poisson and linear regression (Efron 2004), characteristic functions and Fourier analysis (Jin and Cai 2007), and maximum likelihood in an exponential family (Efron 2007). This paper introduces a modification to the maximum likelihood method, the default in the R function `locfdr`.

Let N be the number of tests. Denote the null hypotheses by H_1, \dots, H_N , and the corresponding observed test statistics by z_1, \dots, z_N . We assume that the test statistics have been transformed if necessary so that, for $i = 1, \dots, N$,

$$H_i : z_i \sim N(0, 1). \tag{1}$$

For example, if t_i is the observed test statistic, with theoretical or permutation null cumulative distribution function (cdf) F ,

$$z_i = \Phi^{-1}(F(t_i)), \tag{2}$$

where Φ is the standard normal cdf.

We follow the “two-groups model” (Efron 2008): if H_i is true, we call z_i a null test statistic, and, if H_i is false, we call z_i a non-null test statistic.

The maximum likelihood method estimates the null distribution from the observed test statistics that fall in a fixed interval around zero, by assuming all such z_i are null. If some non-null test statistics fall in the interval, the FDR estimates are asymptotically biased. We introduce a modification that, by varying the interval size with the number of tests, produces consistent FDR estimates, even when some non-null test statistics are near zero.

We present two motivating examples: the human immunodeficiency virus (HIV) microarray dataset and the Beta-blocker Evaluation of Survival Trial (BEST) proteomics dataset.

The HIV microarray data set, featured in Efron (2007), is the expression levels of $N = 7680$ genes in each of 8 subjects, 4 HIV positive and 4 HIV negative. For the HIV data set, for $i = 1, \dots, 7680$, H_i is the statement that the expression of gene i is unrelated to HIV infection, (1), (2), with F the cdf of the T distribution with 6 degrees of freedom (left panel of Figure 1).

The BEST proteomics data set, described in detail in Turnbull (2006), consists of data from 598 subjects, all with heart disease. The variable of interest is a survival outcome – time to event and censoring indicator – and the features are $N = 355$ peak intensities in spectra output from runs of a surface-enhanced laser desorption ionization time-of-flight (SELDI-TOF) mass spectrometer on samples of the subjects’ serum. For the BEST data set, for $i = 1, \dots, 355$, H_i is the statement that the intensity of peak i is unrelated to survival time, (1) with z_i the Cox score statistic corresponding to peak i (right panel of Figure 1).

Section 2 defines FDR, Section 3 reviews the maximum likelihood method for estimating the FDR numerator, Section 4 calculates the asymptotic bias and approximate variance of the estimate, Section 5 presents the proposed method, and Section 6 provides a discussion.

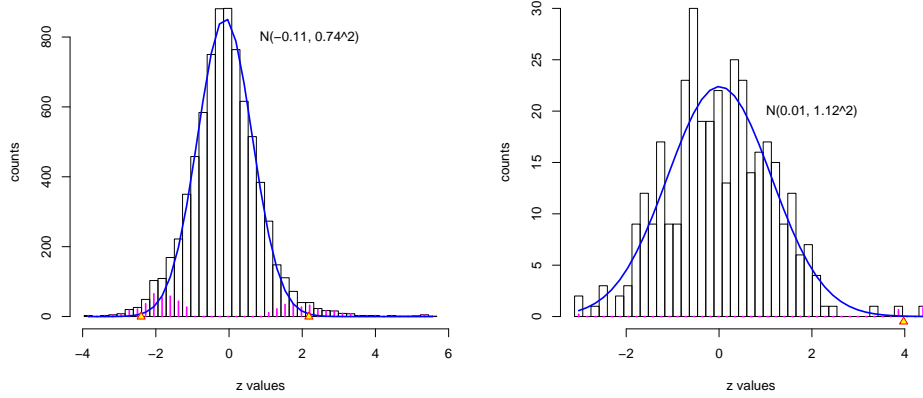


Figure 1: Histograms of z values. Left panel: transformed t statistics from the HIV microarray study. The center of the histogram follows a scaled $N(-0.11, 0.74^2)$ distribution (blue curve). Right panel: Cox scores from the BEST proteomics study. The center of the histogram follows a scaled $N(0.01, 1.12^2)$ distribution (blue curve). Bars represent the estimated scaled distribution of the non-null z values. Triangles mark the interval outside of which the fdr is estimated at less than 0.2.

2 False Discovery Rates

Here we discuss FDR in the Bayesian framework of Efron et. al (2001). Let p_0 represent the prior probability that a null hypothesis is true and $p_1 = 1 - p_0$ the prior probability that it is false. Let f_0 be the probability density function (pdf) of a null test statistic and f_1 the pdf of a non-null test statistic. Define the null subdensity

$$f_0^+(z) = p_0 f_0(z)$$

and the non-null subdensity

$$f_1^+(z) = p_1 f_1(z).$$

The z_i comprise a sample from the mixture density

$$f(z) = f_0^+(z) + f_1^+(z). \quad (3)$$

The local false discovery rate (fdr) is the ratio of the null subdensity to the mixture density,

$$\text{fdr}(z) = f_0^+(z)/f(z). \quad (4)$$

The tail area false discovery rate (Fdr) is defined similarly, but in terms of cdf's instead of pdf's. Everything that follows can be applied to Fdr as well as fdr.

We estimate fdr by estimating f_0^+ and f individually, then plugging into (4),

$$\widehat{\text{fdr}}(z) = \widehat{f_0^+}(z)/\widehat{f}(z). \quad (5)$$

The mixture density f can be estimated from z_1, \dots, z_N using any density estimation method, for example Poisson regression (Efron 2004). This paper focuses on estimation of the null subdensity, f_0^+ .

Statement (1) says that the theoretical null distribution is standard normal, but f_0 might not be (Efron 2004). The HIV data conform instead to a $N(0.11, 0.75^2)$ distribution, and the BEST data are roughly $N(0.01, 1.12^2)$ (Figure 1).

We assume that f_0 is a normal density, but with unknown mean δ_0 and unknown standard deviation σ_0 . Estimation of f_0^+ thus requires estimation of the parameters $(p_0, \delta_0, \sigma_0)$.

3 The Maximum Likelihood Method

The maximum likelihood method estimates f_0^+ by fitting a scaled and truncated normal density to the subset of observed test statistics within an in-

terval around zero. The model relies on the

$$\text{Strong zero assumption: } f_1(z) = 0 \text{ for } z \in [-t, t], \quad t \text{ fixed.} \quad (6)$$

An asymmetric interval $[t_1, t_2]$ containing 0 can also be used; for notational simplicity, we assume symmetry in what follows.

Define the truncated normal distribution as the probability distribution with density proportional to the normal distribution in $[-t, t]$ and zero elsewhere,

$$f_0^t(z) = f_0(z)/H_0(\delta_0, \sigma_0), \quad z \in [-t, t], \quad (7)$$

where

$$H_0(\delta_0, \sigma_0) = \int_{-t}^t f_0 = \Phi\left(\frac{t - \delta_0}{\sigma_0}\right) - \Phi\left(\frac{-t - \delta_0}{\sigma_0}\right). \quad (8)$$

Let

$$\theta_t \equiv \int_{-t}^t f_0^+ = p_0 H_0(\delta_0, \sigma_0), \quad (9)$$

and note that

$$\theta_t \stackrel{\text{SZA}}{=} \int_{-t}^t f = \text{Prob}\{z_i \in [-t, t]\}, \quad (10)$$

where $\stackrel{\text{SZA}}{=}$ represents equality under the strong zero assumption.

Ignoring the values of the test statistics outside $[-t, t]$ means reducing the data to

$$N_t = \#\{z_i \in [-t, t]\}, \quad (11)$$

$$\mathbf{z}_t = \{z_i \in [-t, t]\}. \quad (12)$$

By (10), under the strong zero assumption,

$$N_t \sim \text{Bin}(N, \theta_t)$$

and the elements of \mathbf{z}_t are *iid* with distribution the truncated normal (7).

	t	$\hat{\delta}_0$	$\hat{\sigma}_0$	\hat{p}_0	$\#\{i : \widehat{\text{fdr}}(z_i) < 0.2\}$	smallest $\widehat{\text{fdr}}$'s	
HIV	1	-0.13	0.65	0.84	427	$< 10^{-13}$	$< 10^{-12}$
	2	-0.12	0.79	0.96	121	$< 10^{-8}$	$< 10^{-8}$
	3	-0.13	0.86	0.99	66	$< 10^{-7}$	$< 10^{-6}$
BEST	1	-0.12	1.40	1	1	0.13	0.99
	2	0.01	1.13	1	1	0.01	0.36
	3	-0.02	1.08	0.99	2	0.01	0.18

Table 1: Empirical Null and FDR estimates depend on the threshold t .

Therefore, the likelihood under the strong zero assumption is

$$L_t(N, \mathbf{z}_t | \delta_0, \sigma_0, \theta_t) = \binom{N}{N_t} \theta_t^{N_t} (1 - \theta_t)^{N - N_t} \prod_{\mathcal{I}_t} f_0^t(z_i), \quad (13)$$

where $\mathcal{I}_t = \{i : z_i \in [-t, t]\}$.

Maximizing (13) yields

$$\hat{\theta}_t = N_t/N \quad (14)$$

and numerically calculated values for $\hat{\delta}_0$ and $\hat{\sigma}_0$. Definition (9) gives $\hat{p}_0 = \hat{\theta}_t / H_0(\hat{\delta}_0, \hat{\sigma}_0)$, completing the estimation of the null subdensity f_0^+ .

Null subdensity estimates, and therefore FDR estimates, can depend heavily on the choice of t (Table 1).

Furthermore, FDR estimates resulting from a fixed value of t are asymptotically biased if the strong zero assumption is violated. In what follows, we calculate the asymptotic bias and approximate variance of the null subdensity estimate. We then show that using an optimal value of t produces consistent FDR estimates under a weak assumption.

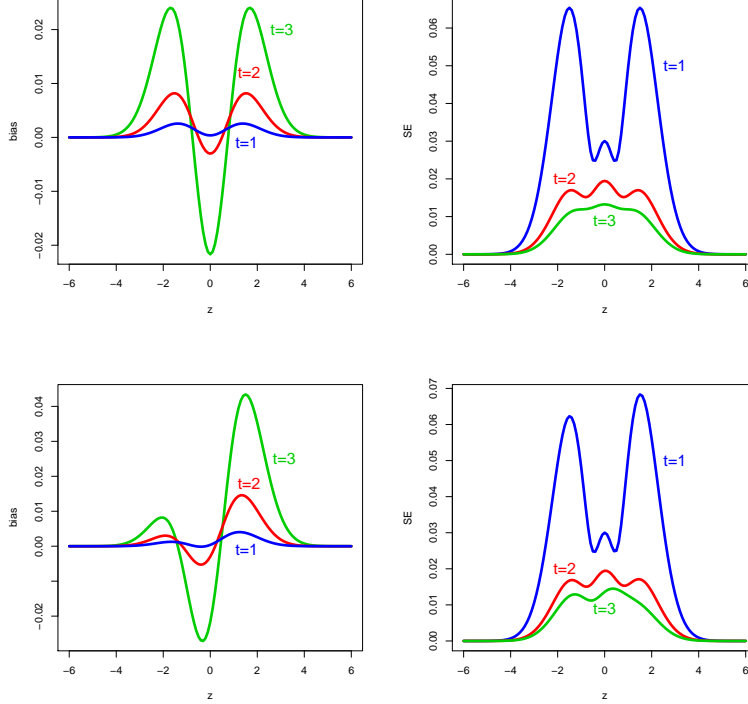


Figure 2: Bias and Standard Error of the Null Subdensity Estimate $\widehat{f}_0^+(z)$, by Threshold t . Top panels: bias and standard error for Example Model 1. Bottom panels: bias and standard error for Example Model 2.

4 Bias and Variance of the Null Subdensity Estimate

Choosing t is a bias-variance trade-off (Figure 2). Increasing t allows further violation of the strong zero assumption, leading to larger bias. Decreasing t decreases N_t , thus increasing variance. In this section, we calculate the asymptotic bias and approximate variance of $\widehat{f}_0^+(z)$, so that in Section 5 we can determine the optimal value of t .

The calculations that follow assume nothing about f_1 , so they hold for any

two-group mixture model (3) in which f_0 is a normal density. For illustration, we use the example models with $p_0 = 0.9$, f_0 the standard normal pdf, and

Example Model 1 : $f_1 \sim 50 - 50$ mix of $N(-3, 1)$ and $N(3, 1)$;

Example Model 2 : $f_1 \sim N(3, 1)$.

4.1 Bias

Since the standard error will be of order $1/\sqrt{N}$, we can ignore the diminishing bias, which, in the maximum likelihood estimation, is of order $1/N$. Instead we focus on the asymptotic bias, which is due to violation of the strong zero assumption and does not diminish with increasing N . That is, we will calculate $\widetilde{f}_0^+(z) - f_0^+(z)$, where $\widetilde{f}_0^+(z) = \lim_{N \rightarrow \infty} \widehat{f}_0^+$ (left panels of Figure 2). Because f_0^+ is a continuous function of $(p_0, \delta_0, \sigma_0)$, \widetilde{f}_0^+ is easily calculated from the limits $(\tilde{p}_0, \tilde{\delta}_0, \tilde{\sigma}_0)$ of the estimates $(\hat{p}_0, \hat{\delta}_0, \hat{\sigma}_0)$.

By (14),

$$\tilde{\theta}_t = \int_{-t}^t f,$$

and, instead of the truncated normal (7), the true model for \mathbf{z}_t is the truncated mixture,

$$f^t(z) = f(z)/\tilde{\theta}_t. \quad (15)$$

The limit $(\tilde{\delta}_0, \tilde{\sigma}_0)$ can be found using the fact that, for a unimodal function, the limit of the argmax is the argmax of the limit. This implies that $(\tilde{\delta}_0, \tilde{\sigma}_0)$ maximizes the strong-zero-assumption-based likelihood (13).

$$\begin{aligned} (\hat{\delta}_0, \hat{\sigma}_0) &= \operatorname{argmax}_{\delta_0, \sigma_0} \frac{1}{N_t} \sum_{\mathcal{I}_t} \log f_0^t(z_i) \\ N \rightarrow \infty \quad \downarrow \\ (\tilde{\delta}_0, \tilde{\sigma}_0) &= \operatorname{argmax}_{\delta_0, \sigma_0} E_{f^t} \log f_0^t(Z), \end{aligned}$$

where E_{f^t} indicates that the expectation is taken under the truncated mix-

ture density (15). Let

$$\mathcal{L}(\delta_0, \sigma_0) = E_{f^t} \log f_0^t(Z).$$

We find $(\tilde{\delta}_0, \tilde{\sigma}_0)$ using the Newton-Raphson algorithm, which requires \mathcal{L} 's gradient and Hessian, provided in the following two theorems.

Generalizing (8), for $p = 1, \dots, 4$, let

$$H_p = H_p(\delta_0, \sigma_0) = \int_{-t}^t z^p f_0(z) d(z),$$

and let

$$E_p = H_p / \int_{-t}^t f(z) dz,$$

where f is the true mixture density (3).

Theorem 4.1. *The gradient of \mathcal{L} with respect to (δ_0, σ_0) is*

$$\nabla \mathcal{L}(\delta_0, \sigma_0) = \frac{1}{\sigma_0^2} \begin{pmatrix} E_1 - \delta_0 \\ (E_2 - 2\delta_0 E_1 + \delta_0^2) / \sigma_0 \end{pmatrix} - \frac{1}{\sigma_0 H_0} \begin{pmatrix} H_1 \\ H_2 \end{pmatrix}. \quad (16)$$

Theorem 4.2. *The Hessian of \mathcal{L} with respect to (δ_0, σ_0) is*

$$\begin{aligned} \nabla^2 \mathcal{L}(\delta_0, \sigma_0) = & \frac{1}{\sigma_0^2} \left\{ \frac{1}{H_0^2} \begin{pmatrix} H_1^2 & H_1 H_2 \\ H_1 H_2 & H_2^2 \end{pmatrix} \right. \\ & - \frac{1}{H_0} \begin{pmatrix} H_2 & H_3 - 2H_1 \\ H_3 - 2H_1 & H_4 - 3H_2 \end{pmatrix} \\ & \left. + \frac{1}{\sigma_0} \begin{pmatrix} 0 & 2(\delta_0 - E_1) \\ 2(\delta_0 - E_1) & -3(E_2 - 2\delta_0 E_1 + \delta_0^2) / \sigma_0 \end{pmatrix} \right\}. \end{aligned}$$

All theorems are proved in the appendix.

A positively biased $\hat{\sigma}_0$ means a flattened estimate of f_0^+ : underestimating

near 0 and overestimating in the tails (top left panel of Figure 2). A positively biased $\hat{\delta}_0$ means a right-shifted estimate of f_0^+ : overestimating for positive z and underestimating for negative z (bottom left panel of Figure 2). A positively biased \hat{p}_0 creates bias in $\widehat{f_0^+}$ in proportion to f_0 .

4.2 Variance

The variance of the null subdensity estimate is approximated via four applications of the delta method.

Define the matrix

$$V = \frac{1}{N_t} \begin{pmatrix} E_2 - E_1^2 & E_3 - E_1 E_2 \\ E_3 - E_1 E_2 & E_4 - E_2^2 \end{pmatrix}. \quad (17)$$

Lemma 4.3. *The covariance matrix of $(\hat{\delta}_0, \hat{\sigma}_0)$ can be approximated by*

$$\text{Cov}(\hat{\delta}_0, \hat{\sigma}_0) \approx \mathcal{V}/N_t,$$

where

$$\mathcal{V} = \frac{\tilde{\sigma}_0^4}{N_t(V_{11}V_{22} - V_{12}^2)} \begin{pmatrix} V_{22} + 4\tilde{\delta}_0(\tilde{\delta}_0 V_{11} - V_{12}) & \tilde{\sigma}_0(2\tilde{\delta}_0 V_{11} - V_{12}) \\ \tilde{\sigma}_0(2\tilde{\delta}_0 V_{11} - V_{12}) & \tilde{\sigma}_0^2 V_{11} \end{pmatrix}, \quad (18)$$

and V_{ij} is the (i, j) element of V in (17).

Let $\tilde{H}_p = H_p(\tilde{\delta}_0, \tilde{\sigma}_0)$, and define the vector

$$\tilde{\mathbf{h}} = \frac{1}{\tilde{H}_0} \begin{pmatrix} \tilde{H}_1 \\ \tilde{H}_2 - \tilde{H}_0 \end{pmatrix}. \quad (19)$$

Lemma 4.4. *The covariance matrix of $(\hat{p}_0, \hat{\delta}_0, \hat{\sigma}_0)$ can be approximated by*

$$\text{Cov}(\hat{p}_0, \hat{\delta}_0, \hat{\sigma}_0) \approx \frac{1}{N} \begin{pmatrix} \frac{\tilde{\theta}_t(1-\tilde{\theta}_t)}{\tilde{H}_0^2} + \frac{\tilde{\theta}_t}{\tilde{\sigma}_0^2 \tilde{H}_0^2} \tilde{\mathbf{h}}' \boldsymbol{\nu} \tilde{\mathbf{h}} & -\frac{1}{\tilde{\sigma}_0 \tilde{H}_0} \boldsymbol{\nu} \tilde{\mathbf{h}}' \\ -\frac{1}{\tilde{\sigma}_0 \tilde{H}_0} \boldsymbol{\nu} \tilde{\mathbf{h}} & \boldsymbol{\nu} / \tilde{\theta}_t \end{pmatrix}. \quad (20)$$

Finally, letting

$$z_{\delta\sigma} = (z - \delta) / \sigma,$$

define

$$\tilde{\mathbf{z}}_0 = \frac{\tilde{p}_0}{\tilde{\sigma}_0} \begin{pmatrix} z_{\tilde{\delta}_0 \tilde{\sigma}_0} \\ z_{\tilde{\delta}_0 \tilde{\sigma}_0}^2 - 1 \end{pmatrix}.$$

Theorem 4.5. *For a given value z , the variance of $\widehat{f}_0^+(z)$ can be approximated by*

$$\text{Var}(\widehat{f}_0^+(z)) \approx \frac{\varphi^2(z_{\tilde{\delta}_0 \tilde{\sigma}_0})}{N} \begin{pmatrix} \frac{\tilde{\theta}_t(1-\tilde{\theta}_t)}{\tilde{H}_0^2} + \frac{\tilde{\theta}_t}{\tilde{\sigma}_0^2 \tilde{H}_0^2} \tilde{\mathbf{h}}' \boldsymbol{\nu} \tilde{\mathbf{h}} - \frac{2}{\tilde{\sigma}_0 \tilde{H}_0} \tilde{\mathbf{h}}' \boldsymbol{\nu} \tilde{\mathbf{z}}_0 + \frac{1}{\tilde{\theta}_t} \tilde{\mathbf{z}}_0' \boldsymbol{\nu} \tilde{\mathbf{z}}_0 \end{pmatrix}, \quad (21)$$

where φ represents the standard normal pdf.

The standard error of the null subdensity, as a function of z , is bat-shaped, trimodal with the relative heights of the modes varying with t (right panels of Figure 2). Since the standard error decreases with increased N but the asymptotic bias stays constant, it is optimal to decrease t with increasing N .

5 The Optimal Threshold

To optimize t in estimating f_0^+ , we use as a loss function the mean squared error (MSE) of \widehat{f}_0^+ and minimize the risk

$$R_N(t) = \int_{-\infty}^{\infty} \text{MSE} \left[\widehat{f}_0^+(z) \right] w(z) dz, \quad (22)$$

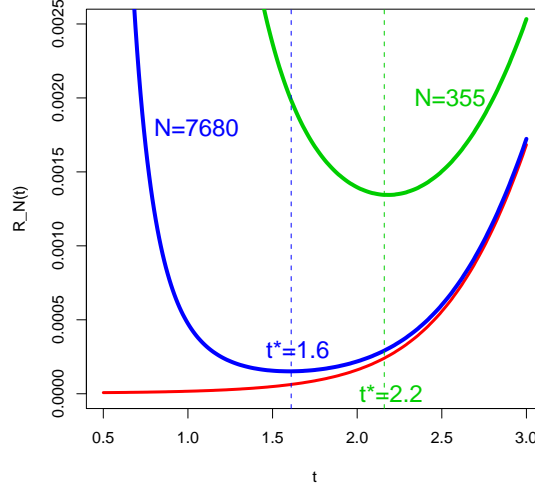


Figure 3: $R_N(t)$ vs. t , for Example Model 1. Red curve: Squared asymptotic bias of $\widehat{f}_0^+(z)$, integrated over z , which does not depend on N . Green curve: $R_N(t)$ (22) for $N = 355$ as in the BEST study. Blue curve: $R_N(t)$ for $N = 7680$, as in the HIV study.

where the dependence on t comes in the calculation of $\widehat{f}_0^+(z)$ (Section 3), and the dependence on N comes from $Var(\widehat{f}_0^+(z))$ (Section 4.2). One can choose the weighting function w to reflect the importance of good estimation of FDR at different values of z . For example, typically one is most interested in FDR estimates in the tails of f , so w may be greater for z further from 0. For simplicity, in what follows, we set $w(z) = 1$ for all z . Thus, $R_N(t)$ is the sum of the squared bias of $\widehat{f}_0^+(z)$, integrated over z , and the variance of $\widehat{f}_0^+(z)$, integrated over z (Figure 3).

The proposed method of FDR estimation is to estimate $f_0^+(z)$ as in Section 3 with t set to

$$t^*(N) = \operatorname{argmin}_t R_N(t).$$

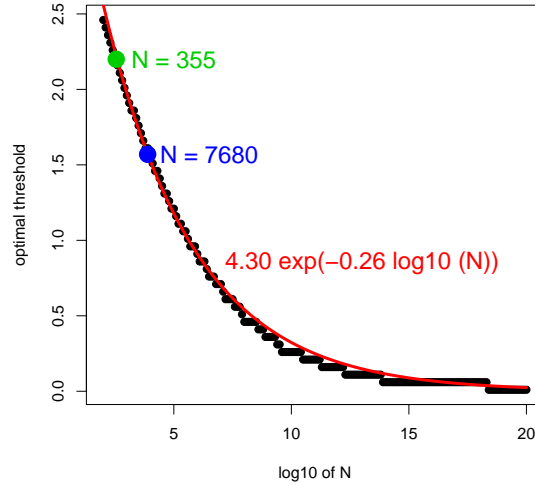


Figure 4: Optimal threshold $t^*(N)$ vs. $\log_{10}N$, for Example Model 1. Black points: $t^*(N)$, calculated as in Figure 3. Red curve: approximation to $t^*(N)$ for Example Model 1. Green point refers to $N = 355$, as in the BEST dataset. Blue point refers to $N = 7680$, as in the HIV dataset.

Using Example Model 1, this gives $t^*(355) = 2.2$ for the BEST data, $t^*(7680) = 1.6$ for the HIV data, and, in general,

$$t^*(N) \approx 4.3e^{-0.26 \log_{10} N} \quad (23)$$

(Figure 4). The R function `locfdr` uses approximation (23).

	iteration	t^*	$\hat{\delta}_0$	$\hat{\sigma}_0$	\hat{p}_0	$\#\{i : \widehat{\text{fdr}}(z_i) < 0.2\}$	smallest $\widehat{\text{fdr}}$'s	
HIV	1	1.6	-0.12	0.75	0.93	159	$< 10^{-9}$	$< 10^{-9}$
	2	1.3	-0.11	0.74	0.92	176	$< 10^{-10}$	$< 10^{-9}$
BEST	1	2.2	0.00	1.12	1	1	0.01	0.30
	2	∞	0.01	1.12	1	1	0.01	0.31

Table 2: Optimal values t^* and estimates

6 Discussion

The null subdensity estimate $\widehat{f}_0^{+*}(z)$ that results from using $t = t^*(N)$ is consistent for $f_0^+(z)$ under the

$$\text{Weak zero assumption: } \exists t > 0 : f_1(z) = 0 \text{ for } z \in [-t, t]. \quad (24)$$

When the strong zero assumption is violated, $\widehat{f}_0^{+*}(z)$ of Section 3 (with fixed t) is asymptotically biased. To achieve consistency, the proposed method induces dependence of the bias on N by using $t = t^*(N)$.

The dependence of t^* on N comes about because the choice of t is a bias-variance trade-off (Section 4). For small N , it is worth taking on more bias in exchange for the corresponding decrease in variance, while, for large N , it is not (Figure 3).

As N approaches infinity, $t^*(N)$ approaches the largest t for which the zero assumption holds in truth. Consequently, under the weak zero assumption, as N approaches infinity, the bias of $\widehat{f}_0^{+*}(z)$ vanishes (Figure 5).

Because the model (3) is unknown, we iterate between finding $t^*(N)$ and estimating the components of (3). For example, one may begin with Example Model 1, which gives approximation (23), then use the resulting estimated model to obtain a refined $t^*(N)$. One such iteration is enough in practice (Table 2).

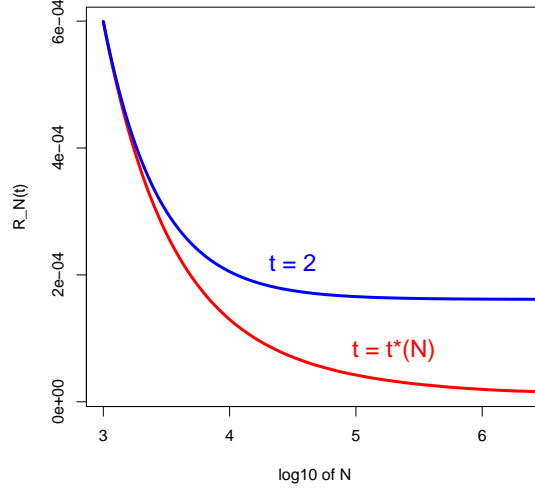


Figure 5: $R_N(t)$ vs. $\log_{10} N$, for Example Model 1. Blue curve: $t = 2$. Red curve: $t = t^*(N)$.

A Proofs

A.1 Proof of Theorem 4.1

Lemma A.1. *For $p = 0, 1, \dots$, the gradient of $H_p(\delta_0, \sigma_0)$ is*

$$\nabla H_p(\delta_0, \sigma_0) = \frac{1}{\sigma_0} \begin{pmatrix} H_{p+1} - pH_{p-1} \\ H_{p+2} - (p+1)H_p \end{pmatrix}, \quad (25)$$

with $H_{-1} \equiv 0$.

Find \mathcal{L} by taking the expectation of (13) with respect to the truncated mixture (15), and, treating E_1 and E_2 as constants, differentiate with respect to (δ_0, σ_0) , using Lemma A.1.

A.2 Proof of Theorem 4.2

Differentiate (16) with respect to (δ_0, σ_0) , using Lemma A.1.

A.3 Proof of Lemma 4.3

Lemma A.2. (Efron 2007) *The truncated normal distribution (7) is a two-parameter exponential family with natural parameters*

$$\beta = \begin{pmatrix} \delta_0/\sigma_0^2 \\ -1/2\sigma_0^2 \end{pmatrix}, \quad (26)$$

corresponding sufficient statistics

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} \sum_{\mathcal{I}_t} z_i/N_t \\ \sum_{\mathcal{I}_t} z_i^2/N_t \end{pmatrix}, \quad (27)$$

and cumulant generating function

$$\psi(\delta_0, \sigma_0) = \frac{1}{2} \left(\frac{\delta_0}{\sigma_0} \right)^2 + \frac{1}{2} \log(2\pi\sigma_0^2) + \log H_0(\delta_0, \sigma_0). \quad (28)$$

Lemma A.3. (Efron 2007) *For fixed N_t , the covariance matrix of $\hat{\beta}$, the estimate of the natural parameter vector β (26) of the truncated normal exponential family, can be approximated by*

$$\text{Cov}(\hat{\beta}) \approx V^{-1}/N_t.$$

Lemma A.4. *The Jacobian of (δ_0, σ_0) with respect to β is*

$$J = \frac{d(\delta_0, \sigma_0)}{d\beta} = \sigma_0^2 \begin{pmatrix} 1 & 2\delta_0 \\ 0 & \sigma_0 \end{pmatrix}. \quad (29)$$

By the delta method,

$$\text{Cov}(\hat{\beta}) \approx \left(\frac{d\hat{\beta}}{dY} \right) \text{Cov}(Y) \left(\frac{d\hat{\beta}}{dY} \right)' = V^{-1} \left(\frac{V}{N_t} \right) V^{-1} = V^{-1}/N_t.$$

Again by the delta method,

$$\text{Cov}(\hat{\delta}_0, \hat{\sigma}_0) \approx \left(\frac{d(\delta_0, \sigma_0)}{d\beta} \right) \text{Cov}(\hat{\beta}) \left(\frac{d(\delta_0, \sigma_0)}{d\beta} \right)', \quad (30)$$

where the derivatives are evaluated at $E\hat{\beta}$, so

$$\text{Cov}(\hat{\delta}_0, \hat{\sigma}_0) \approx \tilde{J}V^{-1}\tilde{J}'/N_t,$$

where \tilde{J} is (29) evaluated at $(\tilde{\delta}_0, \tilde{\sigma}_0)$. Then,

$$\begin{aligned} \tilde{J}V^{-1}\tilde{J}' &= \frac{\tilde{\sigma}_0^4}{V_{11}V_{22} - V_{12}^2} \begin{pmatrix} 1 & 2\tilde{\delta}_0 \\ 0 & \tilde{\sigma}_0 \end{pmatrix} \begin{pmatrix} V_{22} & -V_{12} \\ -V_{12} & V_{11} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2\tilde{\delta}_0 & \tilde{\sigma}_0 \end{pmatrix} \\ &= \frac{\tilde{\sigma}_0^4}{V_{11}V_{22} - V_{12}^2} \begin{pmatrix} V_{22} + 4\tilde{\delta}_0(\tilde{\delta}_0 V_{11} - V_{12}) & \tilde{\sigma}_0(2\tilde{\delta}_0 V_{11} - V_{12}) \\ \tilde{\sigma}_0(2\tilde{\delta}_0 V_{11} - V_{12}) & \tilde{\sigma}_0^2 V_{11} \end{pmatrix}. \end{aligned}$$

A.4 Proof of Lemma 4.4

Lemma A.5. *The covariance matrix of $(\hat{\theta}_t, \hat{\delta}_0, \hat{\sigma}_0)$ can be approximated by*

$$\text{Cov}(\hat{\theta}_t, \hat{\delta}_0, \hat{\sigma}_0) \approx \begin{pmatrix} \tilde{\theta}_t(1 - \tilde{\theta}_t)/N & \mathbf{0}' \\ \mathbf{0} & \mathcal{V}/N_t \end{pmatrix}, \quad (31)$$

where $\mathbf{0}$ is the column vector of two zeroes.

By the delta method,

$$\text{Cov}(\hat{p}_0, \hat{\delta}_0, \hat{\sigma}_0) \approx \left(\frac{d(p_0, \delta_0, \sigma_0)}{d(\theta_t, \delta_0, \sigma_0)} \right) \text{Cov}(\hat{\theta}_t, \hat{\delta}_0, \hat{\sigma}_0) \left(\frac{d(p_0, \delta_0, \sigma_0)}{d(\theta_t, \delta_0, \sigma_0)} \right)', \quad (32)$$

where the derivatives are evaluated at $(\tilde{\theta}_t, \tilde{\delta}_0, \tilde{\sigma}_0)$.

From (9), $p_0 = \theta_t/H_0$. Differentiating,

$$\frac{d(p_0, \delta_0, \sigma_0)}{d(\theta_t, \delta_0, \sigma_0)} = \begin{pmatrix} 1/H_0 & \mathbf{0}' \\ -\frac{\pi_0}{\sigma_0 H_0} \mathbf{h} & I \end{pmatrix}. \quad (33)$$

Plugging (33), evaluated at $(\tilde{\theta}_t, \tilde{\delta}_0, \tilde{\sigma}_0)$, and (31) into (32) yields

$$C \approx \begin{pmatrix} 1/\tilde{H}_0 & -\frac{\tilde{\theta}_t}{\tilde{\sigma}_0 \tilde{H}_0} \tilde{\mathbf{h}}' \\ \mathbf{0} & I \end{pmatrix} \begin{pmatrix} \tilde{\theta}_t(1 - \tilde{\theta}_t)/N & \mathbf{0}' \\ \mathbf{0} & \mathcal{V}/N_t \end{pmatrix} \begin{pmatrix} 1/\tilde{H}_0 & \mathbf{0}' \\ -\frac{\tilde{\theta}_t}{\tilde{\sigma}_0 \tilde{H}_0} \tilde{\mathbf{h}} & I \end{pmatrix}, \quad (34)$$

where $\mathbf{0}$ is the column vector of two zeroes and I is the 2×2 identity matrix. Multiplying out (34) yields (20).

A.5 Proof of Theorem 4.5

By the delta method,

$$\text{Var}(\widehat{f_0^+}(z)) \approx \left(\frac{\partial f_0^+(z)}{\partial(p_0, \delta_0, \sigma_0)} \right) \text{Cov}(\hat{p}_0, \hat{\delta}_0, \hat{\sigma}_0) \left(\frac{\partial f_0^+(x)}{\partial(p_0, \delta_0, \sigma_0)} \right)', \quad (35)$$

where the derivatives are evaluated at $(\tilde{p}_0, \tilde{\delta}_0, \tilde{\sigma}_0)$. Differentiating,

$$\frac{\partial f_0^+(z)}{\partial(p_0, \delta_0, \sigma_0)} = \varphi_{\tilde{\delta}_0 \tilde{\sigma}_0}(z) \begin{pmatrix} 1 \\ \tilde{\mathbf{z}}_0 \end{pmatrix}. \quad (36)$$

Plugging (36) and (20) into (35) and multiplying out yields (21).

B Acknowledgments

The author thanks B. Efron for the HIV data, P.W. Lavori and J.F. Plehn for the BEST data, K. Strimmer for feedback on `locfdr`, and B. Efron, P.W.

Lavori, N. Zhang and S. Chatterjee for helpful comments.

References

- [1] Benjamini Y and Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57**:289–300.
- [2] Dudoit S, Shaffer JP, Boldrick JC (2003) Multiple Hypothesis Testing in Microarray Experiments. *Statist. Sci.* **18**(1):71–103.
- [3] Efron B (2004) Large-scale Simultaneous Hypothesis Testing: the choice of a null hypothesis. *JASA* **99**(465):96–104.
- [4] Efron B (2007) Size, Power, and False Discovery Rates. *Ann. Statist.* **35**(4):1351–1377.
- [5] Efron B (2008) Microarrays, Empirical Bayes, and the Two-Groups Model. To appear, *Statist. Sci.*, with discussion.
- [6] Efron B, Tibshirani R, Storey JD, Tusher V (2001) Empirical Bayes Analysis of a Microarray Experiment. *JASA* **96**:1151-1160.
- [7] Efron B, Turnbull BB, Narasimhan B (2007) `locfdr` R package. cran.r-project.org
- [8] Jin J, Cai TT (2007) Estimating the Null and the Proportion of Non-null Effects in Large-Scale Multiple Comparisons. *JASA* **102**(478):495–506.
- [9] Turnbull BB (2006) BEST Proteomics Data. www.stanford.edu/people/brit.turnbull/BESTproteomics.pdf

- [10] Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *PNAS USA* **98**:5116–5121.