Statistica Sinica **20** (2010), 911-926

GENERALIZED THRESHOLDING ESTIMATORS FOR HIGH-DIMENSIONAL LOCATION PARAMETERS

Min Zhang[1], Dabao Zhang[1] and Martin T. Wells[2]

[1]*Purdue University and* [2]*Cornell University*

Abstract: Analyzing high-throughput genomic, proteomic, and metabolomic data usually involves estimating high-dimensional location parameters. Thresholding estimators can significantly improve such estimation when many parameters are zero, i.e., parameters are sparse. Several such estimators have been constructed to be adaptive to parameter sparsity. However, they assume that the underlying parameter spaces are symmetric. Since many applications present asymmetry parameter spaces, we introduce a class of generalized thresholding estimators. A construction of these estimators is developed using a Bayes approach, where an important constraint on the hyperparameters is identified. A generalized empirical Bayes implementation is presented for estimating high-dimensional yet sparse normal means. This implementation provides generalized thresholding estimators which are adaptive to both sparsity and asymmetry of high-dimensional parameters.

Key words and phrases: Asymmetric parameter space, Bayes construction, empirical Bayes, sparse parameter space, thresholding.

1. Introduction

Availability of high-throughput biotechniques poses a fundamental but challenging issue for estimating high-dimensional yet sparse parameters. For example, spotted microarrays with a two-color competitive hybridization process provide a genome-wide comparison of gene expression levels under different conditions (Schena, Shalon, Davis and Brown (1995); Brown and Botstein (1999)). Such genomic studies are usually provided with the logarithmic fold changes of thousands of genes. In proteomic (and metabolomic) study, mass spectrometry has been widely used to quantitatively profile proteins (and metabolites), where abundances of a large number of underlying compounds are recorded (Feng, Liu, Lou and Liu (2008)). With appropriate preprocessing and normalization, these genome-, proteome-, and metabolome-wide measures are taken to estimate the underlying true parameters, and identify the limited number of non-zero components.

In general, we consider estimating a high-dimensional location parameter $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_p)$ from the noisy data $\mathbf{Y}_p = (y_1, y_2, \cdots, y_p)$, where

$$y_i - \mu_i \overset{i.i.d.}{\sim} \varphi(\cdot), \quad \varphi(\cdot) \text{ is a symmetric log-concave density function.} \tag{1.1}$$

In the case that $\varphi(\cdot)$ has zero mean and standardized variance, the "oracle" estimator $\hat{\mu}_i = y_i 1_{\{|\mu_i|>1\}}$ yields the risk $E[\sum_{i=1}^p (\hat{\mu}_i - \mu_i)^2] = \sum_{i=1}^p \min(1, \mu_i^2)$, where $1_A(x)$ equals to one if $x \in A$, and zero otherwise. When many components in $\boldsymbol{\mu}$ are smaller than one, the "oracle" estimator can dominate the original one. Motivated by such observation, Donoho and Johnstone (1994) proposed to estimate the i-th parameter component μ_i with a *hard threshold* $\delta_h(y_i, \tau) = y_i 1_{(-\infty,-\tau)\cup(\tau,\infty)}(y_i)$, and a *soft threshold* $\delta_s(y_i, \tau) = sign(y_i)(|y_i| - \tau)1_{(-\infty,-\tau)\cup(\tau,\infty)}(y_i)$, where τ is a thresholding parameter.

When a large number of the components in $\boldsymbol{\mu}$ are zero, both hard and soft thresholds can dramatically reduce the risk with properly chosen τ. However, it is challenging to construct this thresholding parameter. On the basis of Stein's unbiased risk estimator (SURE) in estimating a multivariate normal mean (Stein (1981)), Donoho and Johnstone (1995) proposed the SURE method which performs well when non-zero parameters are of small sizes. However, it does not perform as well when non-zero parameters are of large sizes. Abramovich, Benjamini, Donoho and Johnstone (2006) proposed a method based on Benjamini and Hochberg's (1995) idea of controlling the false discovery rate (FDR) in multiple tests, and Fan and Li (2001) proposed a penalized least squares estimator with the smoothly clipped absolute deviation (SCAD) penalty. Both estimators are adaptive to sparse parameter space but also rely on properly pre-specified parameters.

Johnstone and Silverman (2004) proposed an empirical Bayes method that does not require pre-specification of the thresholding parameter. Its data-driven threshold is adaptive to sparsity of the high-dimensional parameters. However, their thresholding estimator inherently assumes that the underlying parameter space is symmetric, and the probabilities to observe negative and positive values are equal, many data settings violate this assumption. For example, van de Peppel, Kemmeren, van Bakel, Radonjic, van Leenen and Holstege (2003) presented a microarray dataset with only 4,936 negative values out of a total of 16,734 endogenous genes. That is, the probability of observing a negative value is estimated to be 0.295 with standard error 0.0035. In extremal cases, all parameters are non-negative as observed in removing background noise for mass spectrometry data (shown in Section 4.2).

In Section 2 we introduce a class of generalized thresholding estimators that extend the traditional ones. The construction of the estimators employs a Bayes approach, where an important constraint on the hyperparameters is identified. In Section 3, we extend the empirical Bayes (EB) implementation of Johnstone and Silverman (2005) to a generalized empirical Bayes (GEB) implementation for estimating high-dimensional, asymmetric, and sparse normal means. Analyses of two datasets are presented in Section 4 to compare the EB and GEB

estimators. In Section 5, a simulation study is conducted to further compare the GEB estimator with other estimator, including those based on SURE, FDR and SCAD.

2. Generalized Thresholding Estimators

2.1. Definitions

Existing thresholding estimators are antisymmetric, which assumes that the parameter space of interest is symmetric. Here we introduce a class of generalized thresholding estimators that include the traditional thresholding estimators as special cases. As demonstrated in later sections, appropriately constructed generalized thresholding estimators can be adaptive to both sparsity and asymmetry of the underlying parameter spaces.

Definition 2.1. For $\tau_- \leq 0$ and $\tau_+ \geq 0$, $\delta(y, \tau_-, \tau_+)$ is a generalized thresholding estimator if (i) $\delta(y, \tau_-, \tau_+)$ is increasing in $y \in \mathbb{R}$; (ii) $-|y| \leq \delta(y, \tau_-, \tau_+) \leq |y|, \forall y \in \mathbb{R}$; (iii) $\delta(y, \tau) = \delta(y, -\tau, \tau)$ is antisymmetric for any $\tau \geq 0$; (iv) $\delta(y, \tau_-, \tau_+) = 0$ if and only if $\tau_- \leq y \leq \tau_+$.

Here τ_- and τ_+ are the lower and upper thresholds, respectively, which were bounded by the universal threshold $\sqrt{2 \log p}$ in Donoho and Johnstone (1994). When $\tau_- = -\tau_+$, $\delta(y, \tau_-, \tau_+)$ reduces to a thresholding estimator that works well for symmetric data. If, on the other hand, the non-zero parameters can only be positive (or negative), it is preferable to construct a generalized thresholding estimator $\delta(y, -\infty, \tau_+)$ (or $\delta(y, \tau_-, \infty)$). In practice, testing the symmetry of $\mathbf{Y}_p$ could be done before such a generalized thresholding estimator is employed.

Corresponding to the hard and soft thresholding estimators, we can define, with the thresholds (τ_-, τ_+), two types of generalized thresholding estimators: *generalized hard threshold* refers to estimating the i-th parameter μ_i with

$$\hat{\mu}_i = \delta_{\text{hard}}(y_i, \tau_-, \tau_+) = y_i 1_{(-\infty, \tau_-) \cup (\tau_+, \infty)}(y_i);$$

generalized soft threshold refers to estimating the i-th parameter μ_i with

$$\hat{\mu}_i = \delta_{\text{soft}}(y_i, \tau_-, \tau_+) = (y_i - \tau_-) 1_{(-\infty, \tau_-)}(y_i) + (y_i - \tau_+) 1_{(\tau_+, \infty)}(y_i).$$

2.2. A Bayes construction

Here we construct generalized thresholding estimators for $\boldsymbol{\mu}$ in model (1.1) using a Bayes approach. Let $\gamma_+(\mu) = 2\gamma(\mu) 1_{[0,\infty)}(\mu)$ and $\gamma_-(\mu) = 2\gamma(\mu) 1_{(-\infty, 0]}(\mu)$, where $\gamma(\cdot)$ is a unimodal and symmetric distribution function. Consider a Bayes estimator of $\boldsymbol{\mu}$ by assuming that the components of $\boldsymbol{\mu}$ have the prior

$$\mu_i \overset{i.i.d.}{\sim} (1 - w_- - w_+)\delta_0(\mu) + w_- \gamma_-(\mu) + w_+ \gamma_+(\mu), \tag{2.1}$$

where $\delta_0(\cdot)$ is Dirac's delta function. Here w_- and w_+ are the weights for the negative and positive parts with density distributions $\gamma_-(\mu)$ and $\gamma_+(\mu)$, respectively.

The posterior distribution of the parameter μ_i consists of a positive part, a negative part, and a mass at zero, which make it possible to derive a generalized thresholding estimator. Indeed, with properly prespecified w_- and w_+, such estimators can be constructed with the posterior median of μ_i

$$\hat{\mu}(y_i; w_-, w_+) = \text{median}(\mu_i | y_i; w_-, w_+). \tag{2.2}$$

The performance of the estimator $\hat{\mu}(y_i; w_-, w_+)$ depends on the choices of the hyperparameters (w_-, w_+), the density distribution of the noise, and (γ_-, γ_+) specified in the prior (2.1). Here (w_-, w_+) describe not only the sparsity but also the asymmetry of the parameter $\boldsymbol{\mu}$. Intuitively, an optimal (w_-, w_+) can be elicited by maximizing the marginal likelihood function. As shown later, the estimator $\hat{\mu}(y_i; w_-, w_+)$ is a generalized thresholding estimator when (w_-, w_+) lies in a region determined by the constant

$$a = \frac{\varphi(0)}{g_+(0) + \varphi(0)} \in (0, 1), \tag{2.3}$$

where $g_+(0) = \int_0^\infty \varphi(\mu)\gamma_+(\mu)d\mu$. This constant specifies the flatness of the priors $\gamma_+(\cdot)$ and $\gamma_-(\cdot)$ relative to the density of the noise.

Theorem 2.1. *With the simplex*

$$\mathcal{S}(a) = \{(w_-, w_+) \in [0,1]^2 : (2a-1)w_- + w_+ \le a, w_- + (2a-1)w_+ \le a\}, \tag{2.4}$$

$\hat{\mu}(y; w_-, w_+)$ *is a generalized thresholding estimator if and only if* $(w_-, w_+) \in \mathcal{S}(a)$.

As mixing weights, (w_-, w_+) can be any point in $\mathcal{S}(1)$ which corresponds to the grey area in Figure 2.1. Theorem 2.1 says that, in order for the Bayes estimator $\hat{\mu}(y; w_-, w_+)$ to have the same sign as the observed data y, (w_-, w_+) needs to be chosen from the shaded area, i.e., the intersection area under the two solid lines that defines $\mathcal{S}(a)$. However, the estimator developed by Johnstone and Silverman (2004) essentially requires $(w_-, w_+) \in \mathcal{S}(0)$, i.e., the lower part of the dashed line lies completely in the shaded area (see Figure 2.1). The proposed Bayes estimator gains more flexibility by offering a much larger set of admissible values for (w_-, w_+).

When $\gamma_+(0) \to 0$ and $\gamma_-(0) \to 0$, a essentially goes to one, which puts less constraint for the above Bayes estimator to be a generalized thresholding estimator. However, as shown in the following theorem, the tails of the priors cannot be too heavy in order for this estimator to have the bounded shrinkage property, the proof of which follows Johnstone and Silverman (2004).

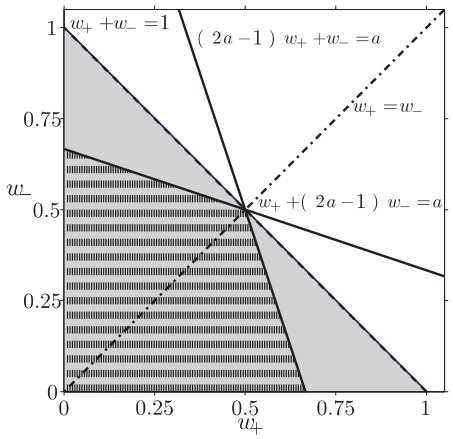

Figure 2.1. Conditions on (w_-, w_+) to construct generalized thresholding estimators. The proposed Bayes estimator $\hat{\mu}(y; w_-, w_+)$ is a generalized thresholding estimator when $(2a-1)w_- + w_+ \le a$ and $w_- + (2a-1)w_+ \le a$; the estimator of Johnstone and Silverman (2004) has $w_- = w_+ \in [0, 0.5]$.

Theorem 2.2. *Assume* (i) *there exists* $\rho > 0$ *such that* $\varphi(y)\exp\{\rho y\}$ *is decreasing for sufficiently large* y, *and* (ii) *there exist* $\Lambda > 0$ *and* $M > 0$ *such that*

$$\sup_{u>M} \left| \frac{d}{du} \log \gamma_+(u) \right| \le \Lambda < \rho. \tag{2.5}$$

Then for $(w_-, w_+) \in \mathcal{S}(a)$, *there exists a constant* c *such that, for all* w_-, w_+, *and* y, *the generalized thresholding estimator* $\hat{\mu}(y; w_-, w_+)$ *has the bounded shrinkage property*

$$\begin{cases} y - \hat{\mu}(y; w_-, w_+) \le \tau_+(w_-, w_+) + c & \text{for } y \ge 0 \\ \hat{\mu}(y; w_-, w_+) - y \le -\tau_-(w_-, w_+) + c & \text{for } y \le 0, \end{cases}$$

where the thresholds $\tau_-(w_-, w_+)$ *and* $\tau_+(w_-, w_+)$ *are determined by* (w_-, w_+).

3. Estimating High-Dimensional Normal Means

When $\varphi(\cdot)$ in model (1.1) is the standard normal density, i.e., $\varphi(\cdot) = \phi(\cdot)$, ρ in (2.5) can be chosen arbitrarily large, so the two assumptions in Theorem 2.2 essentially place no extra constraint on $\gamma_+(\cdot)$ (or on $\gamma_-(\cdot)$). Johnstone and Silverman (2004) developed an empirical Bayes thresholding estimator (EB hereafter) based on a quasi-Cauchy prior for $\boldsymbol{\mu}$, and compared this estimator with others in the literature. The EB estimator showed superior performance. Here we adopt

this quasi-Cauchy prior to construct a generalized thresholding estimator (GEB hereafter).

Specifically, we can construct the generalized thresholding estimators with a quasi-Cauchy prior, i.e., taking

$$\begin{cases} \gamma_+(\mu|\theta_+) = 2(\frac{1}{\theta_+}-1)^{-1/2}\phi(\frac{\mu}{1/\theta_+-1})1_{[0,\infty)}(\mu), & \theta_+ \sim \text{Beta}(0.5,1), \\ \gamma_-(\mu|\theta_-) = 2(\frac{1}{\theta_-}-1)^{-1/2}\phi(\frac{\mu}{1/\theta_--1})1_{(-\infty,0]}(\mu), & \theta_- \sim \text{Beta}(0.5,1), \end{cases} \tag{3.1}$$

or, equivalently,

$$\begin{cases} \gamma_+(\mu) = \sqrt{\frac{2}{\pi}}\left(1-\frac{\mu(1-\Phi(\mu))}{\phi(\mu)}\right) & 1_{[0,\infty)}(\mu), \\ \gamma_-(\mu) = \sqrt{\frac{2}{\pi}}\left(1+\frac{\mu\Phi(\mu)}{\phi(\mu)}\right) & 1_{(-\infty,0]}(\mu), \end{cases}$$

which has tails similar to those of Cauchy densities, i.e., much heavier than Gaussian distribution as desired.

Assuming that $\varphi(\cdot) = \phi(\cdot)$ in model (1.1), we then have

$$\begin{cases} g_-(y_i) = \int_{-\infty}^0 \varphi(y_i-\mu)\gamma_-(\mu)d\mu = \frac{2\Phi(y_i)-\sqrt{2\pi}\phi(y_i)-2y_i\phi(y_i)}{y_i^2\sqrt{2\pi}}, \\ g_+(y_i) = \int_0^\infty \varphi(y_i-\mu)\gamma_+(\mu)d\mu = \frac{2[1-\Phi(y_i)]-\sqrt{2\pi}\phi(y_i)+2y_i\phi(y_i)}{y_i^2\sqrt{2\pi}}. \end{cases}$$

Since $\phi(0) = 1/\sqrt{2\pi}$ and $g_+(0) = \lim_{y\downarrow 0} g_+(y) = 1/\sqrt{8\pi}$, $a = 2/3$ from (2.3), and $\mathcal{S}(a)$ is given by

$$\begin{cases} w_+ + 3w_- \le 2 \\ 3w_+ + w_- \le 2. \end{cases}$$

Maximizing the marginal distribution of $\mathbf{Y}_p$ for $(w_-, w_+) \in \mathcal{S}(2/3)$, we then construct a generalized thresholding estimator with the posterior median, which is essentially an empirical Bayes estimator.

4. Data Analyses Examples

4.1. Microarray data analysis

A common issue in genomic study with microarray data is to identify genes differentially expressed under different conditions. For example, van de Peppel et al. (2003) designed a microarray experiment to examine, in comparison to non-heat-shock cells, the heat-shock response of primarily cultured human umbilical vein endothelial cells (HUVECs). The datasets are available from ArrayExpress (`http://www.ebi.ac.uk/aerep/`) with accession number E-UMCU-2. For illustration, we only use the dataset collected three hours after heat shock (van de Peppel et al. (2003)), in which the differential expression levels of a total of 16,734

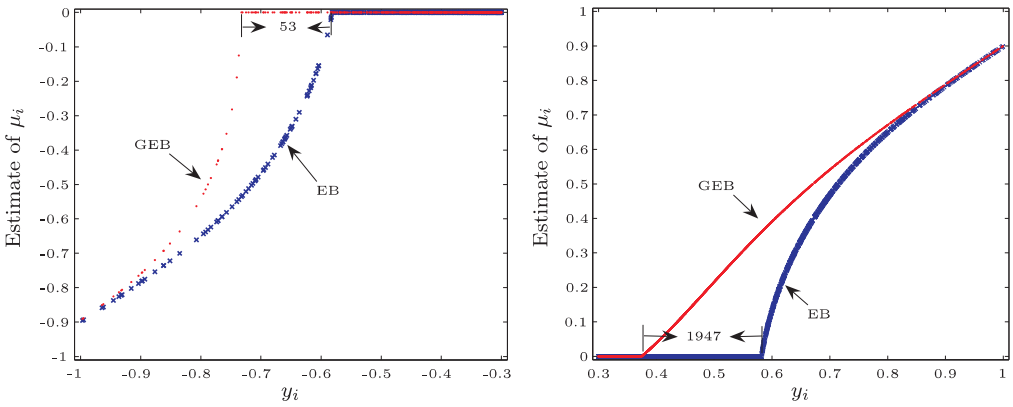

Figure 4.2. Identified up-regulated genes (left) and down-regulated genes (right) using the EB and GEB estimators, respectively.

endogenous genes are normalized using the 960 external control genes (Zhang, Zhang and Wells (2006)).

Figure 4.2 shows the results from the EB and GEB estimators, respectively. The estimated μ_i for any gene with $y_i \in (-0.3, 0.3)$ is zero, and therefore is not shown here. A positive (or negative) μ_i means that the i-th gene is up-regulated (or down-regulated), and a zero μ_i means that it is not differentially expressed. With the GEB estimator, 3,275 genes were identified to be up-regulated and 139 genes were identified to be down-regulated. In contrast, the EB estimator identified 53 more genes to be down-regulated and 1,947 less genes to be up-regulated, as the symmetry assumption forces the upper and lower thresholds to be the same (shown in Figure 4.2). While for small and large scale y_i, whether positive or negative, the two estimates coincide with each other, their performances on medium scale y_i are quite different due to the asymmetric data.

4.2. Mass spectrometry data analysis

Mass spectrometry (MS) plays an important role in discovering clinically relevant peptides/proteins and eventually understanding biological cancer processes. After preprocessing, experimental MS data present many peaks residing on certain mass-to-charge ratios (m/z), that are due to either chemical background noise or peptide fragments. A critical step in identifying proteins using experimental MS data is to remove the peaks caused by chemical background noise while keeping those corresponding to peptide fragments in order to match them with proteins in databases.

Shown in the top panel of Figure 4.3 is an experimental MS dataset from Keller, Purvine, Nesvizhskii, Stolyar, Goodlett and Koler (2002). The log-intensity values, after preprocessing and normalization, are shown in the cental panel of Figure 4.3, in which each point corresponds to a peak observed in the top panel. The EB estimator identified 14 peaks with estimated hyperparameters $\hat{w}_- = \hat{w}_+ = 0.0639$; the GEB estimator identified 33 peaks with the estimated hyperparameters $\hat{w}_- = 0$ and $\hat{w}_+ = 0.2767$ (see Figure 4.3). All peaks identified here observed positive log-transformed intensities. Apparently, forcing $w_- = w_+$ in the EB estimator has significantly reduced the number of identified peaks, which affects the efficiency of searching for proteins in databases.

5. A Simulation Study

We conducted a simulation study to compare the GEB estimator with the EB estimator and other approaches based on SURE, FDR and SCAD. Assuming $\varphi(\cdot) = \phi(\cdot)$ in model (1.1), we simulated 1,000 datasets in each setting of $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)$ with $p = 1,000$, and used each method to estimate all parameters. The risk $R(q) = \sum_{i=1}^{p} E[|\hat{\mu}_i - \mu_i|^q]$ was computed for different values of $q \in (0, 2]$, and the number of false positives (NFP) and the number of false negatives (NFN) were summarized for comparison.

Let $k_+ = \#\{i : \mu_i > 0\}$ and $k_- = \#\{i : \mu_i < 0\}$. We uniformly took all positive parameters with values at μ_+, and all negative parameters with values at μ_-. With sparse non-zero parameters, we considered three cases of asymmetric parameter spaces: (i) $|\mu_-| = \mu_+$ but $k_- \neq k_+$ (see Table 5.1); (ii) $|k_-| = k_+$ but $\mu_- \neq \mu_+$ (see Table 5.2); and (iii) $|\mu_-| \neq \mu_+$ and $k_- \neq k_+$ (see Table 5.2). With $\alpha = 3.7$ and λ equal to the universal threshold (Fan and Li (2001)), the SCAD estimator reported the largest risk $R(2)$ in most of the parameter settings. When instead setting λ equal to the threshold suggested by the SURE approach, it performed similarly to the SURE approach (results are not shown). We report $R(2)$, NFP, and NFN for other estimators in Table 5.1 and Table 5.2 where, in each case, the largest values of $R(2)$, NFP and NFN are in bold type.

In terms of the risk $R(2)$, the comparisons between EB, SURE and FDR agree with the observations of Johnstone and Silverman (2004). Thus SURE is competitive to EB when most of the signals are small, but it may perform poorly otherwise; an appropriately chosen false discovery rate q can make FDR outperform others, but an inappropriate choice can result in very large risk $R(2)$ of FDR. FDR can control the NFPs very well but may report large NFNs for an inappropriate choice of q, while SURE tends to report a larger number of false positives, and thus a smaller number of false negatives than others.

In general, GEB performed competitively with EB in terms of $R(q)$, and it was more stable than EB and the other methods. In cases with many small

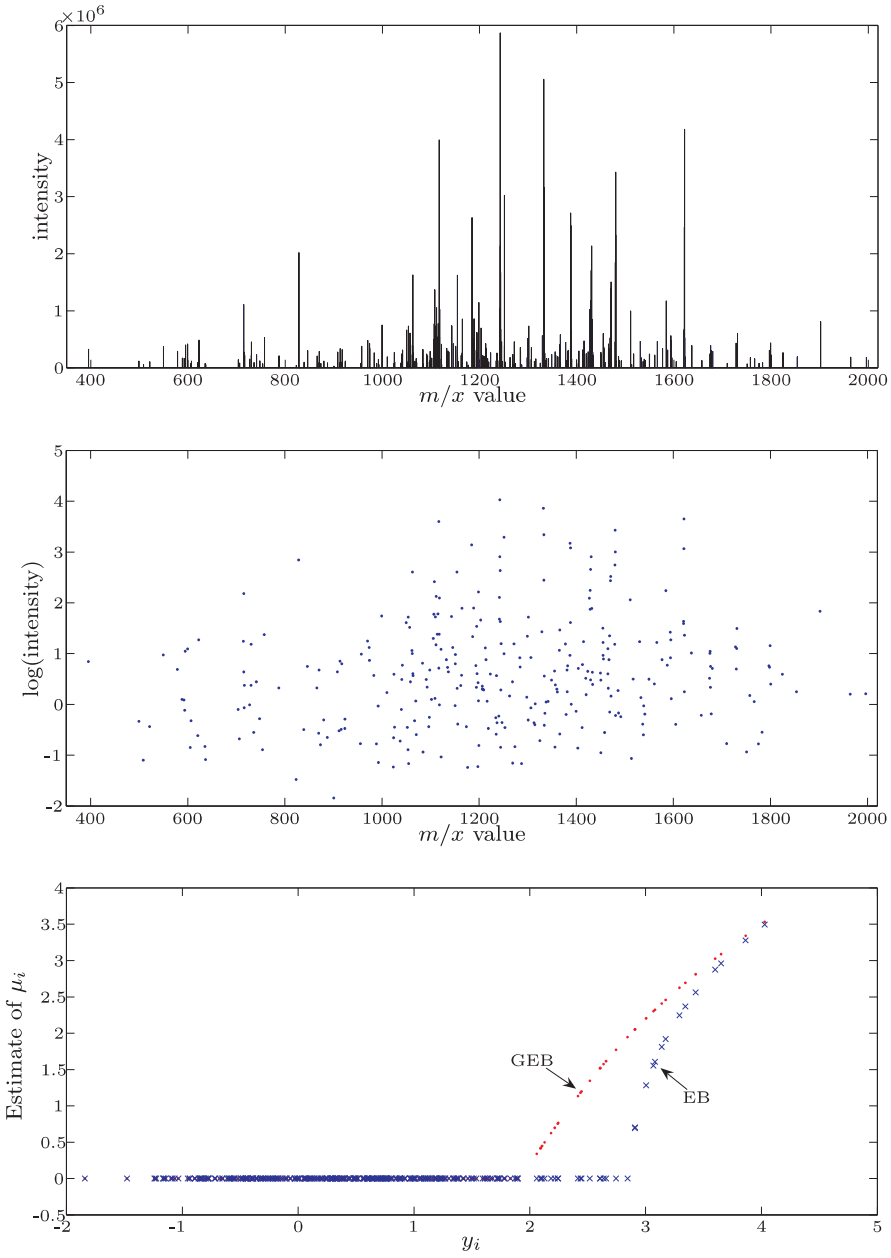

Figure 4.3. An experimental MS dataset (top), normalized peaks (central) and identified peaks (bottom).

signals, GEB gained more over EB in terms of the risk $R(q)$ for large q. Note that this gain was smaller when q decreased in $(0, 2]$, and essentially disappeared when

Table 5.1. Comparing the GEB estimator with others when $|\mu_-| = \mu_+$. $R(2)$, NFP, and NFN are recorded for each estimator with the largest values in bold type. The corresponding standard errors are bracketed.

k_-/k_+	μ_-/μ_+	Criterion	GEB	EB	SURE	FDR ($q = 0.01$)	FDR ($q = 0.1$)	FDR ($q = 0.4$)
0/5	$-3/3$	$R(2)$	37.13 (0.30)	39.29 (0.25)	63.70 (1.42)	79.78 (1.17)	**160.36** (5.88)	115.20 (6.72)
		NFP	0.20 (0.02)	0.10 (0.01)	**103** (3.31)	6.55 (0.16)	30.32 (1.44)	31.93 (3.26)
		NFN	3.84 (0.04)	**4.15** (0.03)	0.86 (0.04)	2.45 (0.04)	2.44 (0.05)	2.28 (0.04)
	$-4/4$	$R(2)$	30.98 (0.55)	36.50 (0.58)	**66.24** (1.39)	51.04 (0.75)	41.00 (1.69)	53.34 (0.98)
		NFP	0.51 (0.03)	0.26 (0.02)	**107** (3.25)	1.11 (0.10)	2.05 (0.39)	3.86 (0.11)
		NFN	1.53 (0.04)	1.97 (0.04)	0.14 (0.01)	**2.42** (0.04)	1.52 (0.03)	0.72 (0.03)
	$-5/5$	$R(2)$	18.05 (0.43)	18.46 (0.52)	**66.51** (1.38)	25.40 (0.70)	19.94 (0.54)	48.19 (1.03)
		NFP	2.73 (0.13)	0.38 (0.02)	**107** (3.24)	0.06 (0.01)	0.68 (0.03)	4.21 (0.11)
		NFN	0.19 (0.01)	0.41 (0.02)	0.01 (0.00)	**0.86** (0.03)	0.28 (0.02)	0.08 (0.01)
	$-7/7$	$R(2)$	25.86 (0.46)	8.35 (0.21)	**66.52** (1.38)	6.18 (0.21)	14.45 (0.42)	47.13 (1.02)
		NFP	14.15 (0.45)	0.42 (0.02)	**107** (3.24)	0.06 (0.01)	0.72 (0.03)	4.27 (0.11)
		NFN	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	**0.01** (0.00)	0.00 (0.00)	0.00 (0.00)
0/50	$-3/3$	$R(2)$	219 (1.01)	267 (1.09)	215 (0.87)	**386** (0.98)	276 (1.14)	295 (1.56)
		NFP	4.02 (0.12)	2.57 (0.06)	**211** (2.19)	0.16 (0.02)	2.79 (0.06)	25.07 (0.28)
		NFN	18.42 (0.13)	25.22 (0.14)	2.21 (0.06)	**40.35** (0.14)	24.59 (0.14)	11.09 (0.11)
	$-4/4$	$R(2)$	142 (1.03)	174 (1.25)	221 (0.89)	**302** (1.86)	172 (1.32)	262 (1.67)
		NFP	6.46 (0.10)	5.24 (0.08)	**219** (2.00)	0.34 (0.02)	4.71 (0.08)	30.71 (0.29)
		NFN	3.19 (0.06)	5.57 (0.08)	0.16 (0.01)	**16.98** (0.12)	5.87 (0.08)	1.62 (0.04)
	$-5/5$	$R(2)$	92.40 (0.69)	103 (0.83)	221 (0.90)	125 (1.45)	112 (0.99)	**250** (1.67)
		NFP	7.55 (0.10)	6.44 (0.09)	**220** (1.99)	0.46 (0.02)	5.27 (0.08)	31.65 (0.28)
		NFN	0.23 (0.01)	0.52 (0.02)	0.01 (0.00)	**3.29** (0.06)	0.65 (0.03)	0.11 (0.01)
	$-7/7$	$R(2)$	73.87 (0.48)	76.90 (0.51)	221 (0.90)	57.22 (0.49)	100.50 (0.81)	**249** (1.67)
		NFP	7.77 (0.10)	6.77 (0.09)	**220** (1.99)	0.51 (0.02)	5.34 (0.08)	31.73 (0.28)
		NFN	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	**0.01** (0.00)	0.00 (0.00)	0.00 (0.00)
5/50	$-3/3$	$R(2)$	255 (1.06)	286 (1.14)	228 (0.87)	**422** (1.03)	296 (1.20)	313 (1.57)
		NFP	3.97 (0.09)	3.00 (0.06)	**219** (2.04)	0.15 (0.02)	3.13 (0.06)	27.52 (0.03)
		NFN	22.18 (0.14)	26.66 (0.15)	2.23 (0.06)	**43.98** (0.14)	26.26 (0.15)	11.53 (0.11)
	$-4/4$	$R(2)$	170 (1.17)	185 (1.26)	233 (0.89)	**323** (1.96)	183 (1.35)	279 (1.70)
		NFP	6.96 (0.10)	6.01 (0.09)	**228** (1.85)	0.38 (0.02)	5.20 (0.08)	33.48 (0.30)
		NFN	4.55 (0.07)	5.64 (0.08)	0.15 (0.01)	**18.10** (0.13)	6.11 (0.08)	1.64 (0.04)
	$-5/5$	$R(2)$	107 (0.81)	111 (0.84)	234 (0.89)	135 (1.50)	121 (1.01)	**267** (1.69)
		NFP	8.34 (0.11)	7.43 (0.10)	**228** (1.84)	0.53 (0.02)	5.76 (0.08)	34.48 (0.30)
		NFN	0.45 (0.02)	0.53 (0.02)	0.01 (0.00)	**3.46** (0.06)	0.07 (0.03)	0.10 (0.01)
	$-7/7$	$R(2)$	82.66 (0.52)	84.50 (0.54)	234 (0.89)	62.86 (0.50)	109 (0.84)	**266** (1.69)
		NFP	8.69 (0.11)	7.79 (0.10)	**228** (1.84)	0.57 (0.02)	5.81 (0.08)	3.45 (0.30)
		NFN	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	**0.01** (0.00)	0.00 (0.00)	0.00 (0.00)
50/450	$-3/3$	$R(2)$	972 (1.66)	924 (1.40)	830 (1.36)	**2565** (3.91)	1156 (2.73)	923 (1.78)
		NFP	259 (0.47)	**500** (0.00)	323 (0.85)	1.24 (0.04)	22.20 (0.16)	121 (0.39)
		NFN	10.58 (0.12)	0.00 (0.00)	2.59 (0.05)	**256** (0.47)	81.09 (0.31)	16.81 (0.13)
	$-4/4$	$R(2)$	865 (1.67)	899 (1.53)	835 (1.38)	**1339** (3.94)	743 (1.99)	868 (1.61)
		NFP	263 (0.40)	**500** (0.00)	326 (0.83)	2.22 (0.05)	25.95 (0.17)	125 (0.39)
		NFN	1.04 (0.03)	0.00 (0.00)	0.10 (0.01)	**62.42** (0.25)	9.97 (0.10)	1.07 (0.03)
	$-5/5$	$R(2)$	760 (1.48)	830 (1.47)	835 (1.38)	656 (2.32)	653 (1.49)	**862** (1.59)
		NFP	265 (0.33)	**500** (0.00)	326 (0.83)	2.47 (0.05)	26.43 (0.17)	125 (0.39)
		NFN	0.05 (0.01)	0.00 (0.00)	0.00 (0.00)	**7.15** (0.09)	0.57 (0.02)	0.03 (0.01)
	$-7/7$	$R(2)$	667 (1.28)	744 (1.31)	835 (1.39)	524 (1.15)	645 (1.41)	**862** (1.59)
		NFP	265 (0.36)	**500** (0.00)	326 (0.83)	2.51 (0.05)	26.45 (0.17)	12.53 (0.39)
		NFN	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	**0.01** (0.00)	0.00 (0.00)	0.00 (0.00)

q approached zero (see Figure 5.4.b). GEB also showed gains over EB for large signals, see Table 5.1, Table 5.2 and Figure 5.4.c. The case with $|\mu_-| = \mu_+ = 7$ and $(k_-, k_+) = (0, 5)$ was exceptional, with the risk $R(2)$ of GEB much larger than that of EB due to the much larger number of NFP from GEB. In most cases, the GEB estimator reported slightly larger NFP than EB. However, when the

Table 5.2. Comparing the GEB estimator with others when $|\mu_-| \neq \mu_+$. $R(2)$, NFP, and NFN are recorded for each estimator with the largest values in bold type. The corresponding standard errors are bracketed.

k_-/k_+	μ_-/μ_+	Criterion	GEB	EB	SURE	FDR ($q = 0.01$)	FDR ($q = 0.1$)	FDR ($q = 0.4$)
5/45	−5/3	$R(2)$	219 (1.15)	246 (1.05)	215 (0.88)	**358** (1.06)	256 (1.12)	289 (1.58)
		NFP	3.74 (0.07)	2.95 (0.06)	**212** (2.18)	0.15 (0.01)	3.11 (0.06)	25.82 (0.28)
		NFN	17.71 (0.13)	21.99 (0.13)	1.96 (0.06)	**35.85** (0.12)	21.62 (0.13)	9.79 (0.10)
	−7/3	$R(2)$	212 (1.01)	241 (1.00)	215 (0.88)	**343** (0.84)	253 (1.07)	289 (1.58)
		NFP	4.35 (0.11)	2.99 (0.06)	**212** (2.18)	0.15 (0.01)	3.12 (0.06)	2.58 (0.28)
		NFN	17.29 (0.13)	21.78 (0.13)	1.96 (0.06)	**35.09** (0.11)	21.48 (0.13)	9.77 (0.10)
	−7/5	$R(2)$	92.62 (0.70)	99.83 (0.79)	221 (0.90)	118 (1.36)	111 (0.98)	**250** (1.67)
		NFP	7.25 (0.10)	6.47 (0.09)	**220** (1.99)	0.47 (0.02)	5.28 (0.08)	31.68 (0.28)
		NFN	0.24 (0.01)	0.47 (0.02)	0.01 (0.00)	**2.93** (0.05)	0.58 (0.02)	0.09 (0.01)
45/5	−5/3	$R(2)$	121 (0.73)	116 (0.85)	220 (0.90)	149 (1.43)	125 (0.98)	**254** (1.68)
		NFP	6.79 (0.09)	6.07 (0.09)	**220** (2.01)	0.44 (0.02)	5.07 (0.08)	31.07 (0.29)
		NFN	3.94 (0.04)	2.46 (0.04)	0.18 (0.01)	**6.50** (0.06)	2.64 (0.04)	1.03 (0.03)
	−7/3	$R(2)$	103 (0.54)	91.80 (0.56)	220 (0.90)	84.98 (0.52)	115 (0.83)	**253** (1.67)
		NFP	7.01 (0.10)	6.35 (0.09)	**220** (2.00)	0.47 (0.02)	5.14 (0.08)	31.11 (0.29)
		NFN	**3.68** (0.04)	1.93 (0.03)	0.18 (0.01)	3.45 (0.03)	2.05 (0.03)	0.95 (0.03)
	−7/5	$R(2)$	81.62 (0.60)	79.15 (0.54)	221 (0.90)	63.11 (0.61)	101 (0.83)	**249** (1.67)
		NFP	7.61 (0.10)	6.75 (0.91)	**220** (1.99)	0.51 (0.02)	5.33 (0.01)	31.72 (0.28)
		NFN	0.21 (0.01)	0.04 (0.01)	0.00 (0.00)	**0.30** (0.02)	0.06 (0.01)	0.01 (0.00)
50/450	−5/3	$R(2)$	888 (1.51)	913 (1.42)	830 (1.39)	**2324** (3.57)	1094 (2.52)	916 (1.78)
		NFP	263 (0.37)	**500** (0.00)	323 (0.85)	1.4 (0.04)	22.68 (0.16)	122 (0.39)
		NFN	0.67 (0.03)	0.00 (0.00)	2.35 (0.05)	**225** (0.41)	71.54 (0.27)	15.00 (0.13)
	−7/3	$R(2)$	875 (1.46)	905 (1.40)	830 (1.39)	**2299** (3.44)	1093 (2.51)	916 (1.78)
		NFP	263 (0.34)	**500** (0.00)	323 (0.85)	1.41 (0.04)	22.69 (0.16)	122 (0.39)
		NFN	0.60 (0.02)	0.00 (0.00)	2.35 (0.05)	**223** (0.41)	71.45 (0.27)	15.0 (0.13)
	−7/5	$R(2)$	747 (1.50)	820 (1.49)	835 (1.42)	640 (2.20)	652 (1.50)	**862** (1.59)
		NFP	265 (0.34)	**500** (0.00)	326 (0.83)	2.48 (0.05)	26.44 (0.17)	125 (0.39)
		NFN	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	**6.32** (0.08)	0.50 (0.02)	0.03 (0.01)
450/50	−5/3	$R(2)$	835 (1.62)	839 (1.46)	835 (1.38)	815 (2.41)	696 (1.60)	**868** (1.61)
		NFP	260 (0.47)	**500** (0.00)	326 (0.83)	2.37 (0.05)	26.05 (0.17)	125 (0.39)
		NFN	9.29 (0.11)	0.00 (0.00)	0.25 (0.02)	**28.15** (0.14)	7.72 (0.08)	1.64 (0.04)
	−7/3	$R(2)$	754 (1.46)	762 (1.32)	835 (1.38)	691 (1.41)	689 (1.54)	**868** (1.61)
		NFP	260 (0.46)	**500** (0.00)	326 (8.30)	2.40 (0.05)	26.07 (0.17)	125 (0.39)
		NFN	9.26 (0.11)	0.00 (0.00)	0.25 (0.02)	**21.39** (0.11)	7.21 (0.08)	1.62 (0.04)
	−7/5	$R(2)$	679 (1.32)	752 (1.33)	835 (1.39)	537 (1.30)	646 (1.42)	**862** (1.59)
		NFP	265 (0.33)	**500** (0.00)	326 (0.83)	2.50 (0.05)	26.45 (0.17)	125 (0.39)
		NFN	0.05 (0.01)	0.00 (0.00)	0.00 (0.00)	**0.70** (0.03)	0.06 (0.01)	0.01 (0.00)
50/50	−5/3	$R(2)$	299 (1.18)	287 (1.12)	330 (0.91)	**437** (1.63)	310 (1.30)	424 (1.80)
		NFP	12.71 (0.14)	13.94 (0.14)	25.19 (1.27)	0.62 (0.03)	8.15 (0.10)	**53.46** (0.36)
		NFN	16.42 (0.13)	14.44 (0.11)	1.41 (0.04)	**35.67** (0.13)	17.85 (0.12)	6.67 (0.08)
	−7/3	$R(2)$	281 (1.07)	268 (0.97)	330 (0.91)	377 (1.02)	302 (1.21)	**423** (1.80)
		NFP	14.03 (0.14)	14.29 (0.14)	**252** (1.27)	0.64 (0.03)	8.20 (0.10)	53.50 (0.36)
		NFN	16.14 (0.13)	14.00 (0.11)	1.41 (0.04)	**32.66** (0.12)	17.40 (0.11)	6.61 (0.08)
	−7/5	$R(2)$	169 (0.81)	169 (0.81)	333 (0.94)	156 (1.28)	187 (1.08)	**395** (1.80)
		NFP	20.54 (0.18)	20.34 (0.18)	**255** (1.22)	0.95 (0.03)	9.90 (0.11)	57.11 (0.36)
		NFN	0.16 (0.01)	0.15 (0.01)	0.00 (0.00)	**2.19** (0.05)	0.34 (0.02)	0.05 (0.01)

signals were relatively dense, i.e., $(k_-, k_+) = (450, 50)$ or $(50, 450)$, GEB reported much smaller NFPs than EB which essentially reported each one as a signal.

Acknowledgement

The support of the Purdue Research Foundation and NSF Grant 06-12031 are gratefully acknowledged.

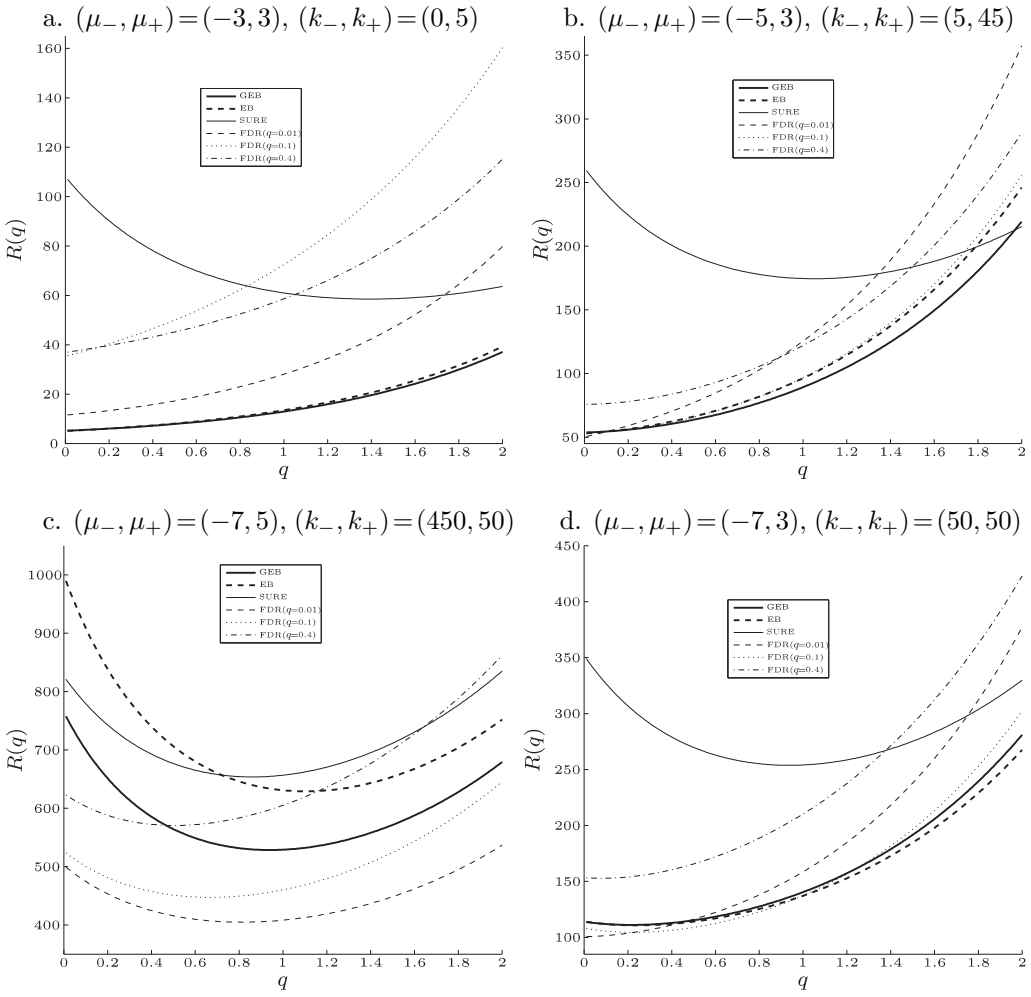

Figure 5.4. Comparing GEB with others in terms of the risks $R(q)$.

Appendix. Proof of Theorem 2.1

We first show some preliminary results on related functions, and then proceed to prove Theorem 2.1 by integrating these results.

Lemma A.3. *If $\varphi(\cdot)$ is symmetric and log-concave, and $\gamma(\cdot) = [\gamma_-(\cdot) + \gamma_+(\cdot)]/2$ is unimodal and symmetric, then* (i) $g_+(y) = g_-(-y)$, $y \in \mathbb{R}$; (ii) $g_+(y)/\varphi(y)$ *is increasing on* $\mathbb{R}$; (iii) $g_-(y)/\varphi(y)$ *is decreasing on* $\mathbb{R}$.

Proof.

(i) This statement follows directly.

(ii) Since $\varphi(\cdot)$ is log-concave, it is PF_2. That is, for all $v \geq 0$ and $y_1 \leq y_2$, $\varphi(y_1)\varphi(y_2-v) \geq \varphi(y_2)\varphi(y_1-v)$, which implies that $\varphi(y-v)/\varphi(y)$ is increasing in y for all $v \geq 0$. Hence, $g_+(y)/\varphi(y)$ is increasing on $y \in \mathbb{R}$ since

$$\frac{g_+(y)}{\varphi(y)} = \int_0^\infty \frac{\varphi(y-v)}{\varphi(y)} \gamma_+(v)dv.$$

(iii) Similarly, for all $u \leq 0$ and $y_1 \leq y_2$, we have $\varphi(y_1-u)\varphi(y_2) \geq \varphi(y_1)\varphi(y_2-u)$, which implies that $\varphi(y-u)/\varphi(y)$ is decreasing in y for all $u \leq 0$. The conclusion follows the fact that

$$\frac{g_-(y)}{\varphi(y)} = \int_{-\infty}^0 \frac{\varphi(y-u)}{\varphi(y)} \gamma_-(u)du.$$

Let $\tilde{w}_-$ and $\tilde{w}_+$ denote the posterior probabilities of μ being positive and negative, respectively. Then it follows that

$$\tilde{w}_+(y; w_-, w_+) = \frac{w_+}{w_+ + (1-w_- - w_+)/(g_+(y)/\varphi(y)) + w_- g_-(y)/\varphi(y)/(g_+(y)/\varphi(y))},$$
$$\tilde{w}_-(y; w_-, w_+) = \frac{w_-}{w_- + (1-w_- - w_+)/(g_-(y)/\varphi(y)) + w_+ g_+(y)/\varphi(y)/(g_-(y)/\varphi(y))}.$$

Therefore, Lemma A.3 leads to the following.

Lemma A.4. *Assume $(w_-, w_+) \in \mathcal{S}(1)$. With the same $\varphi(\cdot)$ and $\gamma(\cdot)$ as in Lemma* A.3, (i) *$\tilde{w}_+(y; w_-, w_+)$ is increasing in $y \in \mathbb{R}$;* (ii) *$\tilde{w}_-(y; w_-, w_+)$ is decreasing in $y \in \mathbb{R}$.*

Now we use Lemma A.4 to identify the conditions for the proposed estimator to have the same sign as the observed data.

Proposition A.5. *With the same $\varphi(\cdot)$ and $\gamma(\cdot)$ as in Lemma* A.3*, the Bayes estimator $\hat{\mu}(y; w_-, w_+)$ satisfies:* (i) *$\hat{\mu}(y; w_-, w_+) \geq 0$, $\forall y \geq 0$, if and only if $w_- + (2a-1)w_+ \leq a$;* (ii) *$\hat{\mu}(y; w_-, w_+) \leq 0$, $\forall y \leq 0$, if and only if $(2a-1)w_- + w_+ \leq a$.*

Proof. Note that $g_+(0) = g_-(0)$. The proposition follows from Lemma A.4 and

(i) $\hat{\mu}(y; w_-, w_+) \geq 0$, $\forall y \geq 0$, if and only if $\tilde{w}_-(0; w_-, w_+) \leq 1/2$;

(ii) $\hat{\mu}(y; w_-, w_+) \leq 0$, $\forall y \leq 0$, if and only if $\tilde{w}_+(0; w_-, w_+) \leq 1/2$.

Proposition A.6. *Assume $(w_-, w_+) \in \mathcal{S}(a)$. With the same $\varphi(\cdot)$ and $\gamma(\cdot)$ as in Lemma* A.3*, we have $|\hat{\mu}(y; w_-, w_+)| \leq |y|$, $\forall y \in \mathbb{R}$.*

Proof. With Proposition A.5, it suffices to prove that, for any $y>0$, $\hat{\mu}(y; w_-, w_+) \leq y$ or, equivalently, $p(\mu > y|Y = y; w_-, w_+) \leq 1/2$, $\forall y > 0$. Note that, for $y > 0$,

$$p(\mu > y|Y = y; w_-, w_+) = \frac{w_+ \int_y^\infty \varphi(y-\mu)\gamma_+(\mu)d\mu}{(1-w_- - w_+)\varphi(y) + w_+ g_+(y) + w_- g_-(y)}.$$

By Lemma A.3, we have, for $y > 0$,

$$\frac{g_-(y)}{\varphi(y)} \leq \frac{g_-(0)}{\varphi(0)} \Rightarrow \varphi(y) \geq \frac{\varphi(0)}{g_-(0)} g_-(y).$$

Therefore,

$$\begin{aligned} p(\mu>y|Y=y; w_-, w_+) &\leq \frac{w_+ \int_y^\infty \varphi(y-\mu)\gamma_+(\mu)d\mu}{(1-w_- - w_+)(\varphi(0)g_-(y)/g_-(0)) + w_+ g_+(y) + w_- g_-(y)} \\ &= \frac{\int_y^\infty \varphi(y-\mu)\gamma_+(\mu)d\mu}{g_+(y) + g_-(y) + (g_-(y)/((1-a)w_+))\big[a - w_+ - (2a-1)w_-\big]} \\ &\leq \frac{\int_y^\infty \varphi(y-\mu)\gamma_+(\mu)d\mu}{g_+(y) + g_-(y)}, \end{aligned}$$

where the last inequality holds because $w_+ + (2a-1)w_- \leq a \leq 1$.

We conclude the proof of the proposition by proving that,

$$\frac{\int_y^\infty \varphi(y-\mu)\gamma_+(\mu)d\mu}{g_+(y) + g_-(y)} \leq \frac{1}{2}, \forall y > 0. \tag{A.1}$$

Since $\gamma(\mu) = (1/2)[\gamma_+(\mu) + \gamma_-(\mu)]$ is unimodal and symmetric, for $y > 0$, $t \geq 0$,

$$\begin{aligned} &\gamma_+(y-t) + \gamma_-(y-t) \geq \gamma_+(y+t) + \gamma_-(y+t) \\ \Rightarrow &\int_0^\infty \varphi(t)[\gamma_+(y-t) + \gamma_-(y-t)]dt \geq \int_0^\infty \varphi(t)[\gamma_+(y+t) + \gamma_-(y+t)]dt. \end{aligned}$$

Note that

$$\begin{aligned} \int_0^\infty \varphi(t)[\gamma_+(y-t) + \gamma_-(y-t)]dt &= \int_{-\infty}^y \varphi(y-u)[\gamma_+(u) + \gamma_-(u)]du, \\ \int_0^\infty \varphi(t)[\gamma_-(y+t) + \gamma_+(y+t)]dt &= \int_y^\infty \varphi(y-u)[\gamma_+(u) + \gamma_-(u)]du, \\ g_+(y) + g_-(y) &= \int_{-\infty}^\infty \varphi(y-u)[\gamma_+(u) + \gamma_-(u)]du. \end{aligned}$$

Therefore,

$$\begin{aligned} g_+(y) + g_-(y) &\geq 2\int_y^\infty \varphi(y-u)[\gamma_+(u) + \gamma_-(u)]du \\ &= 2\int_y^\infty \varphi(y-u)\gamma_+(u)du. \end{aligned}$$

Hence, we have the inequality (A.1).

With the proof by Johnstone and Silverman (2004, p.1619), we can easily establish the following proposition.

Proposition A.7. *If $\varphi(\cdot)$ is symmetric and log-concave, then $\hat{\mu}(y; w_-, w_+)$ is increasing in $y \in \mathbb{R}$.*

Theorem 2.1 can be proved with the following proposition, which follows directly from Propositions A.5 and A.7.

Proposition A.8. *Assume $(w_-, w_+) \in \mathcal{S}(a)$. With the same $\varphi(\cdot)$ and $\gamma(\cdot)$ as in Lemma* A.3*, there exist $\tau_+(w_-, w_+) \geq 0$ and $\tau_-(w_-, w_+) \leq 0$ such that $\hat{\mu}(y; w_-, w_+) = 0$ if and only if $\tau_-(w_-, w_+) \leq y \leq \tau_+(w_-, w_+)$. Furthermore, if $w_+ = w_-$, then $\tau_+(w_-, w_+) = -\tau_-(w_-, w_+)$ and $\hat{\mu}(-y; w_-, w_+) = -\hat{\mu}(y; w_-, w_+)$.*

References

Abramovich, F., Benjamini, Y., Donoho, D. L. and Johnstone, I. M. (2006). Adapting to unknown sparsity by controlling the false discovery rate. *Ann. Statist.* **34**, 584-653.

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.

Brown, P. O. and Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. *Nature Genet.* **21**, 33-37.

Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425-455.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90**, 1200-1224.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96**, 1348-1360.

Feng, X., Liu, X., Lou, Q. and Liu, B.-F. (2008). Mass spectrometry in systems biology: an overview. *Mass Spectrometry Review* **27**, 635-660.

Johnstone, I. M. and Silverman, B. W. (2004). Needles and straw in haystacks: empirical Bayes estimates of possibly sparse sequence. *Ann. Statist.* **32**, 1594-1649.

Johnstone, I. M. and Silverman, B. W. (2005). EbayesThresh: R programs for empirical Bayes thresholding. *J. Statist. Software* **12**, 1-38.

Keller, A., Purvine, S., Nesvizhskii, A. I., Stolyar, S., Goodlett, D. R. and Koler, E. (2002). Experimental protein mixture for validating mass spectral analysis. *Omics* **6**, 207-212.

Schena, M., Shalon, D., Davis, R. W. and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467-470.

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9**, 1135-1151.

van de Peppel, J., Kemmeren, P., van Bakel, H., Radonjic, M., van Leenen, D. and Holstege, F. C. (2003). Monitoring global messenger RNA changes in externally controlled microarray experiments. *EMBO Reports* **4**, 387-393.

Zhang, D., Zhang, M. and Wells, M. T. (2006). Multiplicative background correction for spotted microarrays to improve reproducibility. *Genetical Research* **97**, 195-206.

Department of Statistics, Purdue University, 150 N. University Street, West Lafayette, IN 47907-2067, U.S.A.

E-mail: minzhang@stat.purdue.edu

Department of Statistics, Purdue University, 150 N. University Street, West Lafayette, IN 47907-2067, U.S.A.

E-mail: zhangdb@stat.purdue.edu

Department of Statistical Science, Cornell University, 301 Malott Hall, Ithaca, NY 14853, U.S.A.

E-mail: mtw1@cornell.edu

(Received August 2008; accepted February 2009)