The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1242–1257
https://doi.org/10.1214/20-AOAS1341

A SEMIPARAMETRIC MIXTURE METHOD FOR LOCAL FALSE DISCOVERY RATE ESTIMATION FROM MULTIPLE STUDIES

BY SEOK-OH JEONG[1], DONGSEOK CHOI[2] AND WONCHEOL JANG[3]

[1]*Department of Statistics. Hankuk University of Foreign Studies, seokohj@hufs.ac.kr*
[2]*OHSU-PSU School of Public Health, Oregon Health & Science University, choid@ohsu.edu*
[3]*Department of Statistics, Seoul National University, wcjang@snu.ac.kr*

Antineutrophil cytoplasmic antibody associated vasculitis (AAV) is extremely heterogeneous in clinical presentation and involves multiple organ systems. While the clinical presentation of AAV is diverse, we hypothesized that all AAV share common pathways and tested the hypothesis based on three different microarray studies of peripheral leukocytes, sinus and orbital inflammation disease. For the hypothesis testing we developed a two-component semiparametric mixture model to estimate the local false discovery rates from the p-values of three studies. The two pillars of the proposed approach are Efron's empirical null principle and log-concave density estimation for the alternative distribution. Our method outperforms other existing methods, in particular when the proportion of null is not that high. It is robust against the misspecification of alternative distribution. A unique feature of our method is that it can be extended to compute the local false discovery rates by combining multiple lists of p-values.

1. Introduction. Antineutrophil cytoplasmic antibody associated vasculitis (AAV) is extremely heterogeneous in clinical presentation and involves multiple organ systems, including ranges from life threatening pulmonary hemorrhage to limited diseases of skin, nerves, orbit or eye (Kallenberg (2014), Macfarlane et al. (1983)). While the clinical presentation of AAV is diverse, all may share common pathways. We wanted to test whether we could identify common pathways of AAV from three different studies. To test the hypothesis, gene expression data were collected from three published studies—a study of peripheral leukocytes (Alcorta et al. (2007)), sinus brushings (Grayson et al. (2015)) and orbital inflammatory disease (Rosenbaum et al. (2015)). While all studies employed microarray technology, they used Affymetix HU133 A and B, Affymetrix Human Gene 1.0 ST and Affymetrix Human Genome U133 Plus 2.0, respectively. Therefore, it was not straightforward to analyze the raw expression data of three studies together due to the different number of probes of each array, different comparisons and control groups etc. From each study the associated pathways were identified based on the set of differentially upregulated genes. The aim was to identify common pathways with the combined local false discovery rate (md-fdr), which will be defined in Section 2.3, less than 0.1 as shown in Figure 1.

Since the early 21st century, microarray technology and next generation sequencing technology have revolutionized genomic researches by enabling simultaneous interrogating of tens of thousands genes. A quick search on PubMed.gov with a keyword *microarray* returns more than 80,000 items since 2000 (Coordinators (2017)). Many authors have deposited their microarray data to the gene expression omnibus (GEO) database, and anyone can easily download gene expression data from multiple studies for a secondary data analysis (Edgar, Domrachev and Lash (2002)). While one can perform a meta-analysis of raw data from multiple studies, there will be several major challenges in combining data from different platforms/technologies, such as normalizing or adjusting batch effects. In addition, the full-scale

Received September 2019; revised March 2020.
Key words and phrases. False discovery rate, log concave, microarray, mixture model, next generation sequencing data.

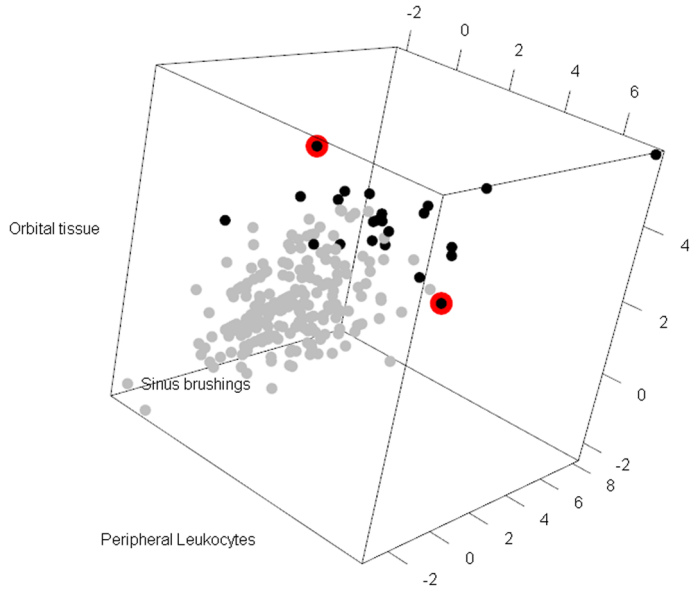

FIG. 1. *Probit-transformed p-values of common pathways from three AAV microarray studies. Black points represents* md-fdr < 0.1. *The red circled points are osteoclast dierentiation* (*K*) *and cell surface interactions at the vascular wall* (*R*). *For more details, see Section* 4.

analysis can require considerable resources and time. If one only wishes to do a quick screening analysis from the multiple lists of significantly differentially expressed genes or pathways, a method to combine the multiple lists of p-values into one can be extremely handy. Figure 1 shows the probit-transformed p-values of common pathways identified from the three studies in our motivated application; the black points represent the combined local false discovery rates less than 0.1. In this paper we propose a semiparametric mixture model that provides a unified way to compute the local false discovery rate for both single list of p-values and multiple lists of p-values.

Suppose that we observe N cases, each with its own p-value, p_i, for $i = 1, 2, \ldots, N$. Let z_i, $i = 1, 2, \ldots, N$, be the probit-transformed p-values. Each case can be considered as being from either null or alternative, with the prior probability $p_0 = \Pr\{\text{null}\}$ or $p_1 = \Pr\{\text{alternative}\} = 1 - p_0$. Hence, z-values have the following mixture density:

$$f(z) = p_0 f_0 + (1 - p_0) f_1(z), \tag{1.1}$$

where $f_0(z)$ and $f_1(z)$ are null and alternative densities.

The above mixture model appears in three contexts: (i) in multiple testing problems (fMRI, microrarray), probit-transformed p-values under H_0 follow the standard normal distribution while the marginal probability density function of the probit-transformed p-values associated with H_1 is unknown (Efron (2008)); (ii) in variable/basis selection, a mixture prior is used to achieve sparsity (Johnstone and Silverman (2005)); (iii) in prediction, Fisher's discriminant function can be regularized using a connection with the local false discovery rate theory (Efron (2009)).

In multiple testing, adjusting for multiplicity is of great interest. To do so, we may consider controlling for either the local false discovery rate, $\mathsf{fdr}(z) = \Pr(\text{null} \mid Z = z)$ or the False Discovery Rate, $\mathsf{FDR}(z) = \Pr(\text{null} \mid Z > z)$. In this paper we focus on estimating $\mathsf{fdr}(z)$ that can be viewed as the posterior probability of a case being from the null given z. Under the mixture model (1.1) it is straightforward to show

$$\mathsf{fdr}(z) = p_0 f_0(z)/f(z).$$

It is natural to assume that f_0 follows the standard normal distribution. However, Efron (2008) suggested that theoretical null distribution $N(0, 1)$ may not be suitable for f_0 and proposed to estimate f_0 with the *empirical null distribution* $N(\mu, \sigma^2)$ where μ and σ^2 are to be estimated from data. He used the *zero assumption* to estimate p_0 and f_0 but estimated the marginal density f separately with *Lindsey's method*. As a result, his estimates may not follow the original mixture structure since the estimates are given from separate procedures. In other words, the mixture model only used to define fdr but not for making inference of fdr. This can sometimes lead to problematic fdr estimates, as shown in our simulations in Section 3.

In this paper we propose a semiparametric mixture model that estimates p_0, f_0 and f_1 simultaneously with alternative being a log-concave density. The main advantages of this approach are threefold. First, it is reasonable to assume the alternative distribution belongs to a log-concave family, as long as probit transformed p-values are considered. We will discuss the robustness of this assumption in detail. Second, log-concave densities can be estimated nonparametrically without a smoothing parameter. Finally, our method can compute the local false discovery rate from multiple lists of p-values while Efron's method cannot be extended similarly. For example, we observe 3-tuple of p-values (p_{1i}, p_{2i}, p_{3i}) for $i = 1, \ldots, N$ in the three AAV microarray studies. Our goal is to estimate

$$\text{fdr}(\boldsymbol{z}_i) = \Pr(\text{null} \mid \mathbf{Z} = \boldsymbol{z}_i),$$

where $\boldsymbol{z}_i = (z_{1i}, z_{2i}, z_{3i}) = (\Phi^{-1}(1 - p_{1i}), \Phi^{-1}(1 - p_{2i}), \Phi^{-1}(1 - p_{3i}))$ and Φ is the cumulative distribution function of $N(0, 1)$ so that Φ^{-1} defines the probit transform.

Section 2 introduces our semiparametric mixture model. In Section 3 we present numerical studies to show the robustness and performance of our method. Section 4 presents the analysis results of three AAV microarray studies. Section 4 shows the robustness of our method. Section 5 concludes this paper.

2. Semiparametric mixture model.

2.1. *The proposed model.* In this section we propose a semiparametric mixture model for f.

$$f(z) = p_0 \phi_{\mu,\sigma^2}(z) + (1 - p_0) f_1(z), \tag{2.1}$$

where ϕ_{μ,σ^2} denotes the density of $N(\mu, \sigma^2)$ and f_1 is a log-concave density function for the alternative distribution.

Let $h(t)$ denote the pdf of the alternative distribution of p-values. Define $Z = \Phi^{-1}(1 - p)$ to be a probit transformed p-value. We usually assume that the p-values follow a uniform $[0, 1]$ under H_0. For the behavior of p-value under H_1, see Sellke, Bayarri and Berger (2001) and Hung et al. (1997). While we do not assume a specific class of family for alternative distribution of p-values, we claim that it is reasonable to assume *the alternative distribution of Z, probit transformed p-value* belongs to a log-concave distribution family. We will discuss the robustness of our assumption later in Section 3.

Statistical properties of the class of log-concave densities are well studied in Walther (2002, 2009). Walther (2002) also showed the existence of the nonparametric MLE of a univariate log-concave density that can be computed via an efficient algorithm, such as an active set algorithm and an iterative convex minorant algorithm (Dümbgen and Rufibach (2011)).

In general, semiparametric mixture models are not identifiable without additional assumptions on the alternative density. Assuming that the alternative distribution belongs to the location-shift family, Bordes, Delmas and Vandekerkhove (2006) showed the identifiability of the semiparametric mixture model under mild regularity conditions. Genovese and

Wasserman (2004) also addressed identifiability of mixture models for p-value distribution under the assumption that the alternative is pure, that is, $\operatorname{ess\,inf}_t h(t) = 0$. The recent work of Hunter, Wang and Hettmansperger (2007) and Balabdaoui and Doss (2018) address the identifiability issues in log-concave mixture models. Note that Balabdaoui and Doss (2018) have an extra assumption about symmetry, but our method assumes the null distribution is a normal.

Our semiparametric mixture model in (2.1) may also suffer from the nonidentifiability issue. In order to avoid this issue, we assume that the support of alternative distribution for Z is given by (a, ∞) for some a. Note that (p_0, μ, σ^2) can be determined by at least *three* distinct points arbitrarily chosen from the interval $(-\infty, a]$. In other words, the null components are estimable using data points in the interval $(-\infty, a]$, since the data points are certainly from the null distribution. Hence, the identifiability of our model is guaranteed as long as we can choose a such that there exist at least three points in $(-\infty, a]$. This argument is similar to Efron's with *zero assumption* for the support of f_1.

Recently, Hu, Wu and Yao (2016) proposed the MLE for log-concave mixture models and showed the existence and consistency of the MLE under fairly general condition. Since a normal distribution is also log concave, our semiparametric mixture model can be considered as a special case of a finite (two) mixture of log-concave distributions.

We closely follow Hu, Wu and Yao (2016) to show the consistency of our estimator. The key idea is to prevent the likelihood from being unbounded by constraining the parameter space. Hence, it suffices to check whether our semiparametric mixture model satisfies $f \in \mathcal{F}_\eta$ for some $\eta \in (0, 1]$, where $\mathcal{F}_\eta$ is defined as follows:

$$\mathcal{F}_\eta = \{f : f(x) = p\exp(\varphi_1(x)) + (1-p)\exp(\varphi_2(x)),\ p \in (0,1),\ \boldsymbol{\varphi} \in \Phi_\eta\},$$

where

$$\boldsymbol{\varphi} = (\varphi_1(x), \varphi_2(x)) = \left(-\frac{(x-\mu)^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2), \log f_1(x)\right)$$

and Φ_η is a constrained subspace

$$\Phi_\eta = \{\boldsymbol{\varphi} = (\varphi_1, \varphi_2) | \varphi_j \text{ is concave}, |\mathcal{S}(\boldsymbol{\varphi})| \geq \eta > 0\},$$

with

$$\begin{aligned}\mathcal{S}(\boldsymbol{\varphi}) &= \frac{\min\{\max_x \varphi_1(x), \max_x \varphi_2(x)\}}{\max\{\max_x \varphi_1(x), \max_x \varphi_2(x)\}} \\ &= \frac{\min\{-\frac{1}{2}\log(2\pi\sigma^2), \max_x \log f_1(x)\}}{\max\{-\frac{1}{2}\log(2\pi\sigma^2), \max_x \log f_1(x)\}}.\end{aligned}$$

The condition

$$|\mathcal{S}(\boldsymbol{\varphi})| \geq \eta > 0$$

is satisfied as long as the mode of alternative density does not increase to ∞, and it is not so much a restrictive one for any alternative density. Therefore, the log-likelihood for our method is bounded, and the estimated mixture density is consistent by Theorems 1 and 2 in Hu, Wu and Yao (2016). Furthermore, the two estimated components and the null probability are also consistent since our model is identifiable.

2.2. *EM algorithm.* It is natural to consider an EM-type algorithm to fit the proposed semiparametric model. To do so, we first define the likelihood function

$$\ell(\Delta, \mu, \sigma^2, f_1 | z_1, \ldots, z_N) = \sum_{i=1}^{N} \{\Delta_i \log \phi_{\mu,\sigma^2}(z_i) + (1-\Delta_i) \log f_1(z_i)\} + \sum_{i=1}^{N} \{\Delta_i \log p_0 + (1-\Delta_i) \log(1-p_0)\},$$

where $\Delta = (\Delta_1, \ldots, \Delta_N)^T$ is a latent variable vector indicating the group membership.

Following Chang and Walther (2007), we first run the EM algorithm for a Gaussian mixture to get initial estimates of (p_0, μ, σ^2). Then, the EM algorithm for fitting the proposed semiparametric model is given below:

1. Initialization: Set $k = 0$ and run the EM algorithm for a Gaussian mixture to get initial values $(p_0^{(0)}, \mu^{(0)}, \sigma^{2(0)})$, $f_1^{(0)}(z_i)$ for $i = 1, \ldots, N$. Put $\gamma_i^{(0)} = 0$, $i = 1, \ldots, N$.
2. E-step: Compute the posterior probability: for $i = 1, \ldots, N$,

$$\gamma_i^{(k+1)} = \mathsf{E}\left(\Delta_i | \mu^{(k)}, \sigma^{2(k)}, z_1, \ldots, z_N\right) = \frac{p_0^{(k)} \phi_{\mu^{(k)},\sigma^{2(k)}}(z_i)}{p_0^{(k)} \phi_{\mu^{(k)},\sigma^{2(k)}}(z_i) + (1-p_0^{(k)}) f_1^{(k)}(z_i)}.$$

3. M-step: Compute the log-concave estimates $f_1^{(k+1)}(z_i)$ based on z_i with weights $1 - \gamma_i^{(k+1)}$ for $i = 1, \ldots, N$. And put

$$\mu^{(k+1)} = \frac{\sum_{i=1}^N \gamma_i^{(k+1)} z_i}{\sum_{i=1}^N \gamma_i^{(k+1)}},$$

$$\sigma^{2(k+1)} = \frac{\sum_{i=1}^N \gamma_i^{(k+1)} (z_i - \mu^{(k+1)})^2}{\sum_{i=1}^N \gamma_i^{(k+1)}},$$

$$p_0^{(k+1)} = \frac{1}{N} \sum_{i=1}^{N} \gamma_i^{(k+1)}.$$

4. Replication: If $\max_{i=1,\ldots,N} |\gamma_i^{(k+1)} - \gamma_i^{(k)}| < \mathsf{TOL}$, then output $p_0^{(k+1)}$, $\mu^{(k+1)}$, $\sigma^{2(k+1)}$, $f_1^{(k+1)}(z_i)$'s, $\gamma_i^{(k+1)}$'s and STOP. Otherwise, set $k = k + 1$ and go to the E-step.

In practice, we used $\mathsf{TOL} = 5 \times 10^{-6}$.

2.3. *Extension to multiple studies.* Suppose that one forms multivariate statistics $\mathbf{Z} = (Z_1, \ldots, Z_d)$ where each component is collected from a different study and provides unique information for common scientific hypotheses. For example, from each study of three AAV studies the lists of test statistics and corresponding p-values are available. To the best of our knowledge, there is no method available to combine such information. It is straightforward to extend the concept of fdr to such multivariate cases using the proposed method

$$\mathsf{fdr}(\mathbf{Z}) = p_0 \frac{f_0(\mathbf{Z})}{f(\mathbf{Z})},$$

which we call hereafter as md-fdr. However, applying Efron's method to multivariate is not straightforward. For example, it is not easy to extend Lindsey's method to estimate the marginal distribution even for two-dimension.

Up to our knowledge, Ploner et al. (2006) was the first attempt of md-fdr modeling. They proposed a method to combining a common test statistic, say t-statistic, for assessing differential expression in microarray studies with its standard error information. They estimated f_0 with discrete smoothing of binomial data after binning the data. However, their method still requires a smoothing parameter and has an issue with boundary bias. Furthermore, their method is only applicable to combining t-test statistics with standard error estimates for each gene. On the other hand, our method is straightforward to extend to multivariate cases and is not involved in choosing a smoothing parameter.

Cule, Samworth and Stewart (2010) extended Dümbgen and Rufibach (2009) to multivariate settings and implemented multivariate log-concave estimation in R package LogConcDEAD (Cule, Gramacy and Samworth (2009)). We use LogConcDEAD to implement our semiparametric mixture model in both single study and multiple studies. The R function and examples are available in the Supplemental Material (Jeong, Choi and Jang (2020)).

3. Simulation studies. In this section we investigated the performance of our semiparametric approach with some simulation studies where we can compare the results with Efron's method. We conducted $M = 500$ Monte Carlo experiments with $N = 1000$ and considered performance measures used for classifiers in machine learning applications: false positive rate (FPR) and sensitivity with the threshold set to be $\mathsf{fdr} \le 0.2$ which is equivalent to the Bayes factor $f_1(z)/f_0(z) \ge 4p_0/(1-p_0)$. This is a very strict level compared with classical testing practice. Note that the multiple testing can be thought as an unsupervised learning where the nullity of each observation is not known. Since the true nullity of each data point is known during the simulation studies, it is appropriate to use performance measures for classifiers with two classes to investigate the performance of a multiple testing procedure. Also, in order to evaluate the accuracy of the estimated fdr, we computed the root-mean-squared error (RMSE) which is given by

$$\mathrm{RMSE} = \frac{1}{M}\sum_{r=1}^{M}\sqrt{\sum_{i:\mathsf{fdr}(z_i^{(r)})\le 0.5}\{\widehat{\mathsf{fdr}}(z_i^{(r)}) - \mathsf{fdr}(z_i^{(r)})\}^2 \Big/ \sum_{i=1}^{N} I\big(\mathsf{fdr}(z_i^{(r)}) \le 0.5\big)},$$

where $z_1^{(r)}, z_2^{(r)}, \ldots, z_N^{(r)}$ is the rth random sample generated during the Monte Carlo simulation. Note that we only took account of z_i's with $\mathrm{fdr}(z_i^{(r)}) \le 0.5$ in the definition of the RMSE, because the accuracy of fdr estimation is required mainly for the region where the $\mathsf{fdr}(z)$ is small. For example, the behavior of fdr estimates for z-values with $\mathsf{fdr}(z) \approx 0.2$ is of concern as is in our study.

3.1. *Simulations of single study.* We considered six scenarios for simulations; see Table 1. For the first three scenarios we assumed normal distributions for both null and alternative distributions with different p_0. It does not conform the assumption on the support of the alternative density when considering a normal distribution for the alternative component. However, the normal distributions are extremely thin tailed, and we may practically reckon that they are supported on a bounded interval given by the range of ± 3 standard deviation around its mean; we consider that the support assumption is valid. The next three scenarios are the same as the first three, except using gamma distributions as an alternative. Gamma distributions are log concave and are supported on $(0, \infty)$, so that there is no issue on the identifiability.

Figures 2 and 3 show that our proposed approach stably yields reasonable results for a wide range of p_0 while Efron's method tends to collapse when p_0 is relatively small. During our simulation study, R package locfdr was used for implementing Efron's method (Efron,

TABLE 1
Simulation scenarios for single study

Scenario	p_0	f_0	f_1	Scenario	p_0	f_0	f_1
1	0.95	$N(0,1)$	$N(3.5,0.5^2)$	4	0.95	$N(0,1)$	gamma$(12,0.25)$
2	0.90	$N(0,1)$	$N(3.5,0.5^2)$	5	0.90	$N(0,1)$	gamma$(12,0.25)$
3	0.80	$N(0,1)$	$N(3.5,0.5^2)$	6	0.80	$N(0,1)$	gamma$(12,0.25)$

Turnbull and Narasimhan (2015)). Note that one of the key assumptions of Efron's method is the null probability is large, say $p_0 \geq 0.90$; see (6.11) in Efron (2010). In the case that p_0 is not large enough, Efron's method for estimating p_0 and the null distribution tends to break down and give an unreasonable estimate for p_0, for example, an estimate greater than one, by the naive (with no constraint on the range of p_0 estimate) default fit to $\log f$ over the central portion of the z-values. Figure 2(a)–(d) summarize the results from the simulation setting with $p_0 = 0.95$ while (e)–(h) are from the setting with $p_0 = 0.90$; and (i)–(l) are from the setting with $p_0 = 0.80$. In the settings with $p_0 = 0.95$ and $p_0 = 0.90$, both Efron's and our method worked reasonably well. However, when $p_0 = 0.80$, Efron's method tends to overestimate p_0 and RMSE becomes substantially large. Furthermore, their sensitivity is closer to 0 in most cases while our method still works well. Figure 3 shows similar patterns for gamma alternative distributions.

Our log-concavity assumption is indeed robust in testing problems. To check this, we simulated p-values from Beta$(0.3, 1)$ following Sellke, Bayarri and Berger (2001). Note that this distribution does not satisfy log-concavity. But we verified that the distribution of probit transformed p-values gets quite close to a log-concave one. Figure 4 shows comparisons between the empirical distributions of the simulated p-values p_i's from Beta$(0.3, 1)$ and that of random numbers generated from the log-concave fit of p_i's. Clearly, Figure 4(a) presents a distributional discrepancy between the simulated p-values and a random sample generated from the fitted log-concave density, and the Kolmogorov–Smirnov (KS) test confirmed the discrepancy with p-value $< 10^{-4}$. However, after probit transformation, one can barely see the difference between the two empirical distributions in Figure 4(b) with KS statistics p-value $= 0.8693$. We repeated the above procedure 500 times, and Figure 5 summarizes the results. While most of the KS test p-values before the probit transformation are small, after the probit transformation only 0.2% of the KS tests have p-values less than or equal to 0.05.

We conducted the similar numerical studies using noncentral t, noncentral χ^2 and noncentral F as alternative distribution of test statistics. We found the results were similar to the beta distribution case; see the Supplementary Material (Jeong, Choi and Jang (2020)).

It is possible that the distribution of probit-transformed p-values from the alternative component may not be unimodal. Suppose the alternative is a two-component mixture distribution normal distribution given by

$$f_1 = 0.6 \cdot N\big(2.5, 0.25^2\big) + 0.4 \cdot N\big(3.5, 0.5^2\big),$$

which clearly does not belong to a log-concave family. Figure 6 demonstrates that our method performed reasonably well with a two-component alternative density in a simulation. Our method overestimates the alternative density (underestimating fdr) and provides deflated fdr estimates around the valley between the two components in the alternative distribution. However, this area should be declared as alternative, so underestimating fdr should not be an issue. Indeed, the monotonicity of fdr estimator can be considered as a desirable property in a two-component model from the classification point of view, and using the log-concave estimators enforces the monotonicity into fdr estimator. See Figure 6.

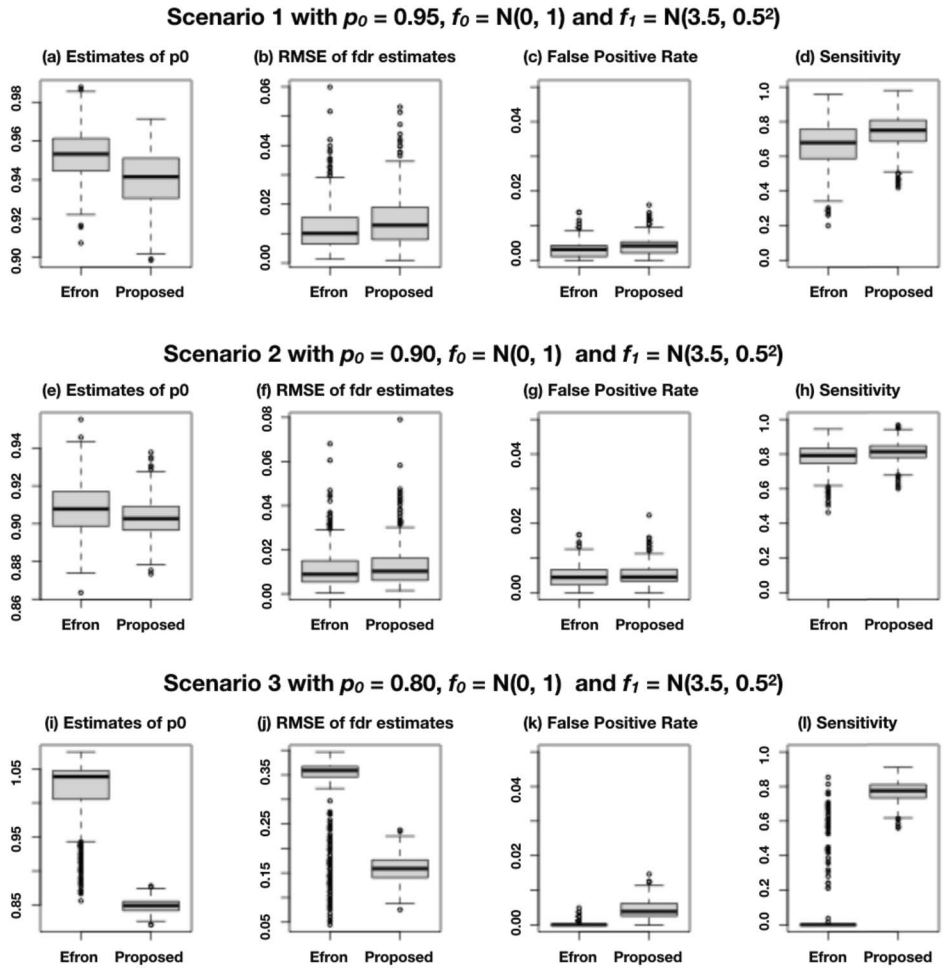

FIG. 2. *Comparison of the performances of the proposed method and Efron's method for Scenarios* 1–3 *with the normal alternative distribution in Table* 1. *Boxplots are drawn with the results from* $M = 500$ *Monte Carlo experiments. Panels* (*a*)–(*d*) *summarize the results from the simulation setting with* $p_0 = 0.95$; (*e*)–(*h*) *are from the setting with* $p_0 = 0.90$, *and* (*i*)–(*l*) *are from the setting with* $p_0 = 0.80$.

The further simulation results presented in Figure 7 suggest that the proposed method is robust even if the alternative distribution violates the log-concavity assumption. It still outperforms Efron's again, although his method does not need assumptions on the shape of alternative distribution. As long as p_0 is moderately large, the potential risk of misspecification for alternative distribution is likely to be forgiven, since it affects little to the accuracy of f estimates and, consequently, to that of fdr estimates.

In general, we are more sensitive to fdr estimation results around the decision boundary. In other words, we are more interested in whether our estimate for alternative fits reasonably well around the decision boundary area. We want to put an emphasis on that the key is to estimate the fdr around the boundary between null and alternative distributions, and our fdr estimator around this area is robust even when log-concavity assumption is violated. The risk of misspecification for the alternative distribution can be forgiven, since it affects little to the fdr estimates as long as p_0 is close to 1 in testing problems. Note that the proposed method captures and utilizes the "local" feature of data distribution only because it is affected little by the global shape of data distribution. Indeed, the fdr estimate is not required to fit well all over the support of the marginal distribution and has only to be accurate on a neighborhood near the threshold.

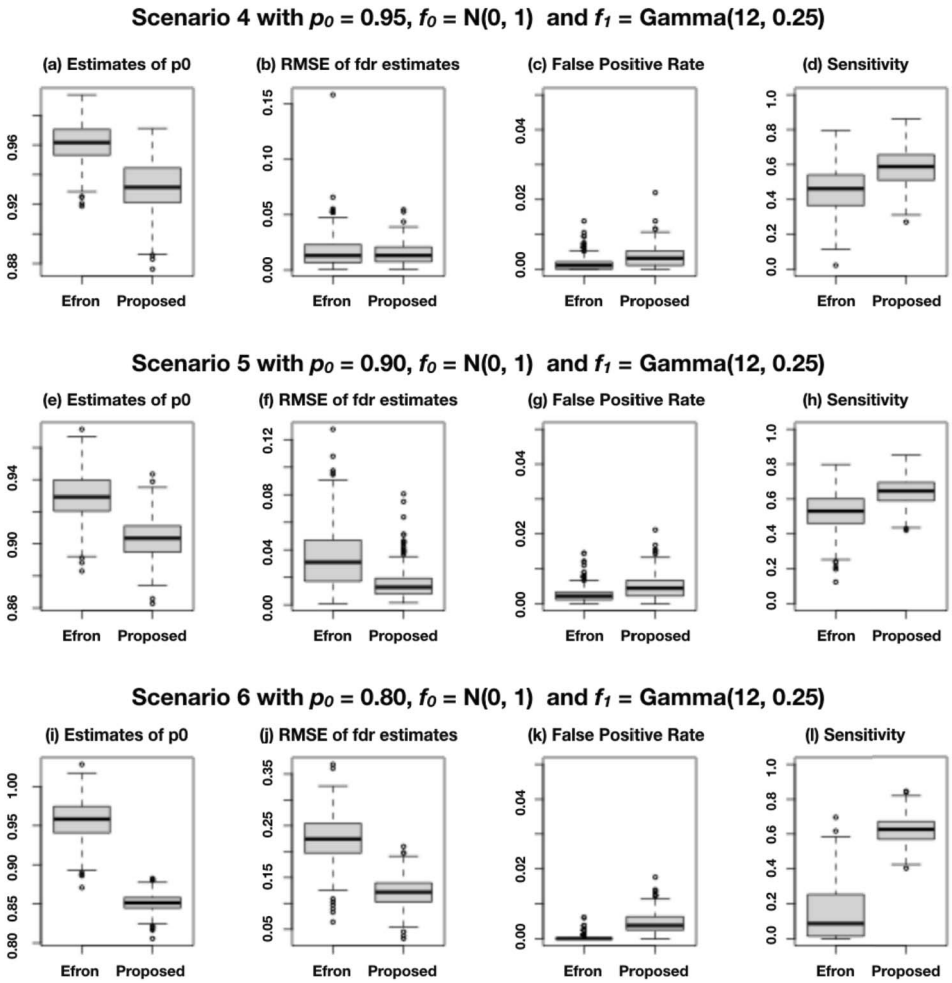

FIG. 3. *Comparison of the performances of the proposed method and Efron's method for Scenarios* 4–6 *with gamma alternative distribution in Table* 1. *The boxplots are drawn with the results from* $M = 500$ *Monte Carlo experiments. Panels* (*a*)–(*d*) *summarize the results from the simulation setting with* $p_0 = 0.95$; (*e*)–(*h*) *are from the setting with* $p_0 = 0.90$, *and* (*i*)–(*l*) *are from the setting with* $p_0 = 0.80$.

3.2. *Simulations of multiple studies*. The simulation scenarios are similar to those in the previous subsection. In each scenario the distributions of null and alternative are the same as the null and alternative distributions of the corresponding scenario in the single study setup. For the dependency of z's, we postulated two copula structures: the Gaussian copula for simulations with normal marginal distributions and Frank copula for simulations with gamma marginals; see Yan (2007) for a detailed explanation on implementing copula models with R. By doing so, we want to find whether there is an advantage of md-fdr compared to fdr from a single study (1d-fdr).

Figure 8 illustrates the empirical distributions of sensitivities from 500 Monte Carole experiments for each scenario. All the figures commonly show that md-fdr outperforms 1d-fdr. Comparing the boxplots for md-fdr with those from separate 1d-fdr, we confirm that the empirical distributions of sensitivities with md-fdr are much preferable to those in the corresponding 1d-fdr. Combining information with md-fdr could be better since it endows the test procedure with an extra flexibility in deciding the boundary of the rejection region. This is also explained by a similar phenomenon to the classification problem where the error rate may decrease when more variables are used.

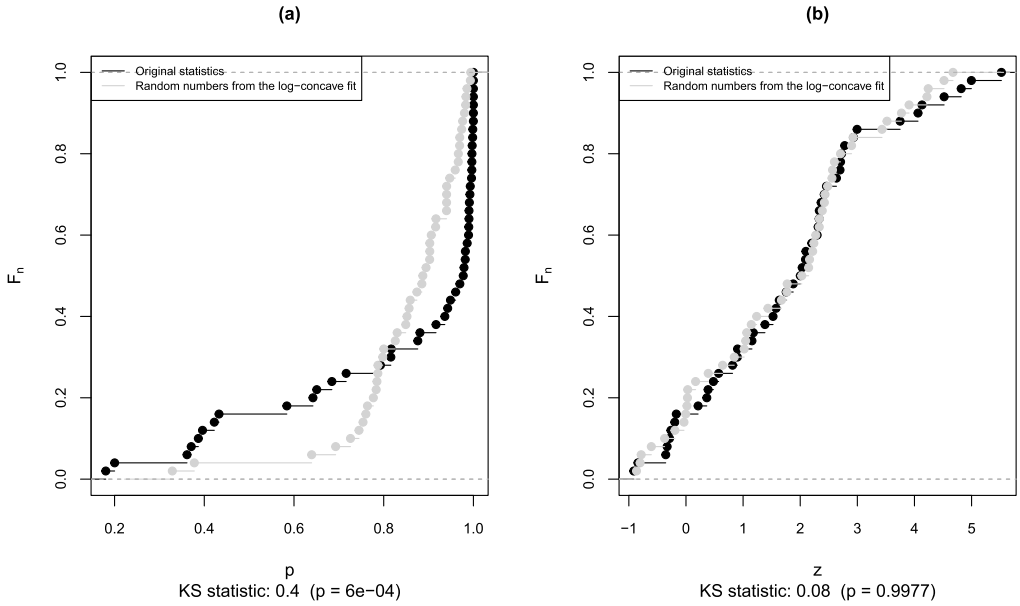

FIG. 4. *A simulation result for* Beta(0.3, 1) *which violates the log-concavity. Panel* (*a*) *is for comparing the empirical distribution of* 50 *simulated* p_i *from* Beta(0.3, 1) *and that of* $\tilde{p}_i$, $i = 1, 2, \ldots, 50$ *randomly drawn from the log-concave fit of* p_i*'s. The Kolmogorov–Smirnov test statistic for comparing these two empirical distributions was* 0.46 *with p-value* $< 10^{-4}$. *Panel* (*b*) *is for comparing the empirical distribution of the probit-transformed p-values* $z_i = \Phi^{-1}(1 - p_i)$ *and that of* $\tilde{z}_i$*'s drawn from the log-concave fit of* z_i*'s. The KS statistic was* 0.12 *with p-value* $= 0.8693$.

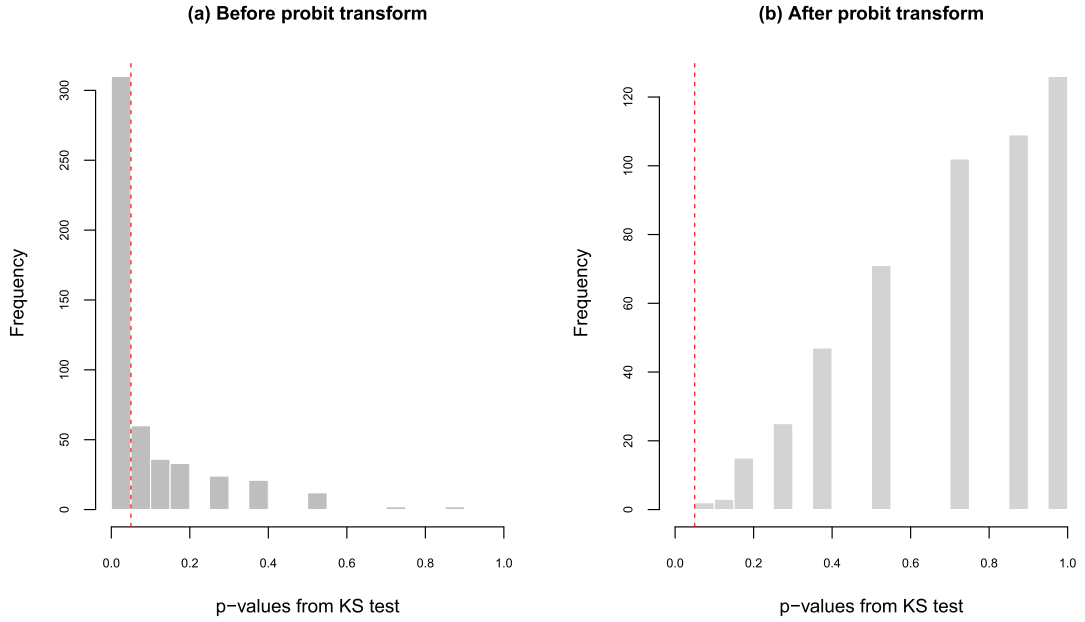

FIG. 5. *From* 500 *Monte Carlo experiments with the same setup as in Figure* 1, 500 *p-values of Kolmogorov–Smirnov test were computed for* p_i*'s and* z_i*'s. Panel* (*a*) *is the histogram for the p-values of Kolmogorov–Smirnov test which compares the empirical distribution of* p_i*'s and that of* $\tilde{p}_i$*'s randomly drawn from the log-concave fit of* p_i*'s. Among* 500 *p-values for KS test,* 57.2% *were less than or equal to* 0.05. *Panel* (*b*) *is the histogram for the p-values of KS test for* $z_i = \Phi^{-1}(1 - p_i)$*'s and* $\tilde{z}_i$*'s drawn from the log-concave fit of* z_i*'s. Only* 0.2% (*one case out of* 500 *experiments*) *of p-values were less than or equal to* 0.05.

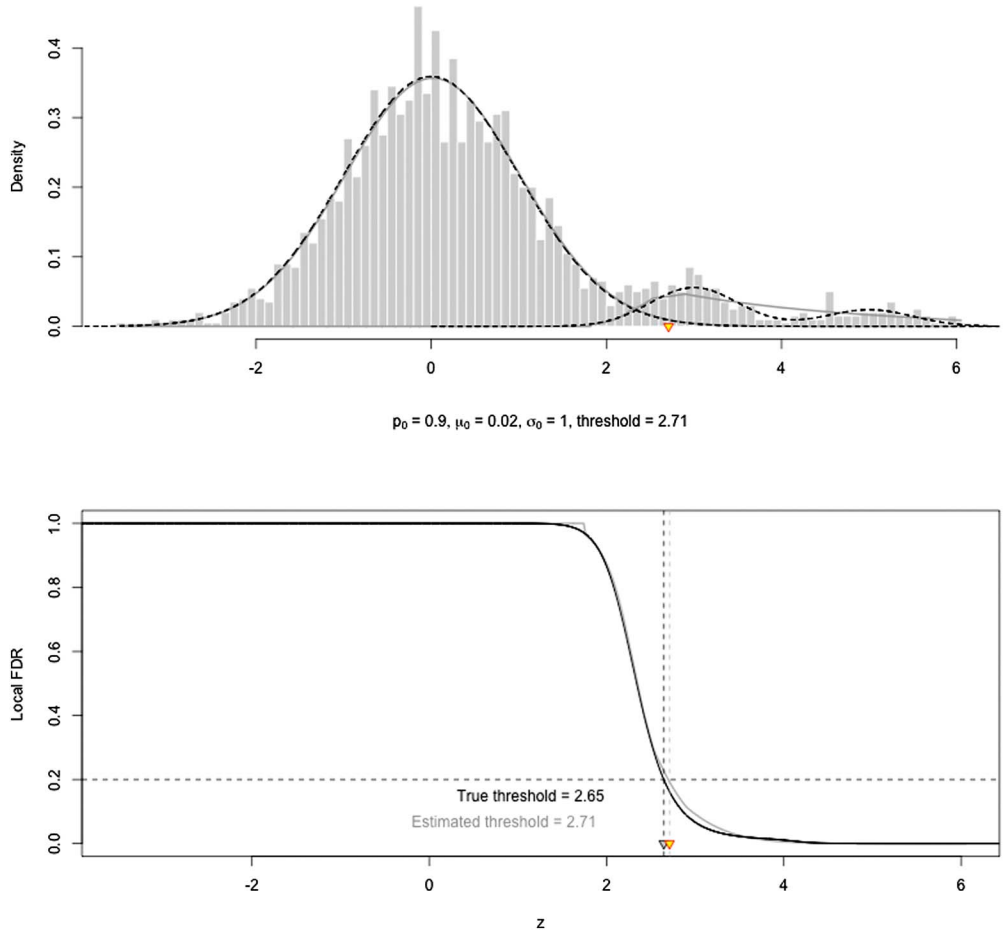

FIG. 6. *A simulation result with a bimodal alternative distribution*: $f_1 = 0.6 \cdot N(2.5, 0.25^2) + 0.4 \cdot N(3.5, 0.5^2)$. *The dotted lines in the upper panel is the true density functions used for simulating data. The black solid curve in the lower panel is the true* fdr, *and the gray curve is the* fdr *estimate.*

4. Common pathways across multiple studies. We wanted to test the hypothesis that common pathways exist in various AAV. Three published studies provided the list of significantly upregulated genes in: (1) peripheral leukocytes (Alcorta et al. (2007)), (2) sinus brushings (Grayson et al. (2015)) and (3) orbital inflammatory disease (Rosenbaum et al. (2015)) compared to healthy controls. Reactome pathway database was used to get the relevant pathways for upregulated genes from each study (Fabregat et al. (2016)) with their associated p-values. By applying our model to each list of p-values separately, there were 26, 32 and 26 pathways with 1d-fdr < 0.1 for study (1), (2) and (3), respectively. When we fitted our model to estimate md-fdr of three lists of p-values, there were 23 pathways with md-fdr < 0.1. Figure 1 shows the probit-transformed p-values of common pathways from the three studies, where black points represent md-fdr < 0.1. Table 2 presents the pathways with md-fdr < 0.1.

Many of the pathways in Table 2 support current knowledge and theories about AAV. For example, neutrophil degranulation pathway was the most significant pathway with 1d-fdr of each study and md-fdr. The involvement of neutrophil in these diseases was previously reported (Soderberg and Segelmark (2016)). Toll-like receptors cascades pathway had also small 1d-fdr in all three studies as well as md-fdr. This pathway supports the hypothesis that the pathogenesis of these diseases involves infections and innate immunity. In addition,

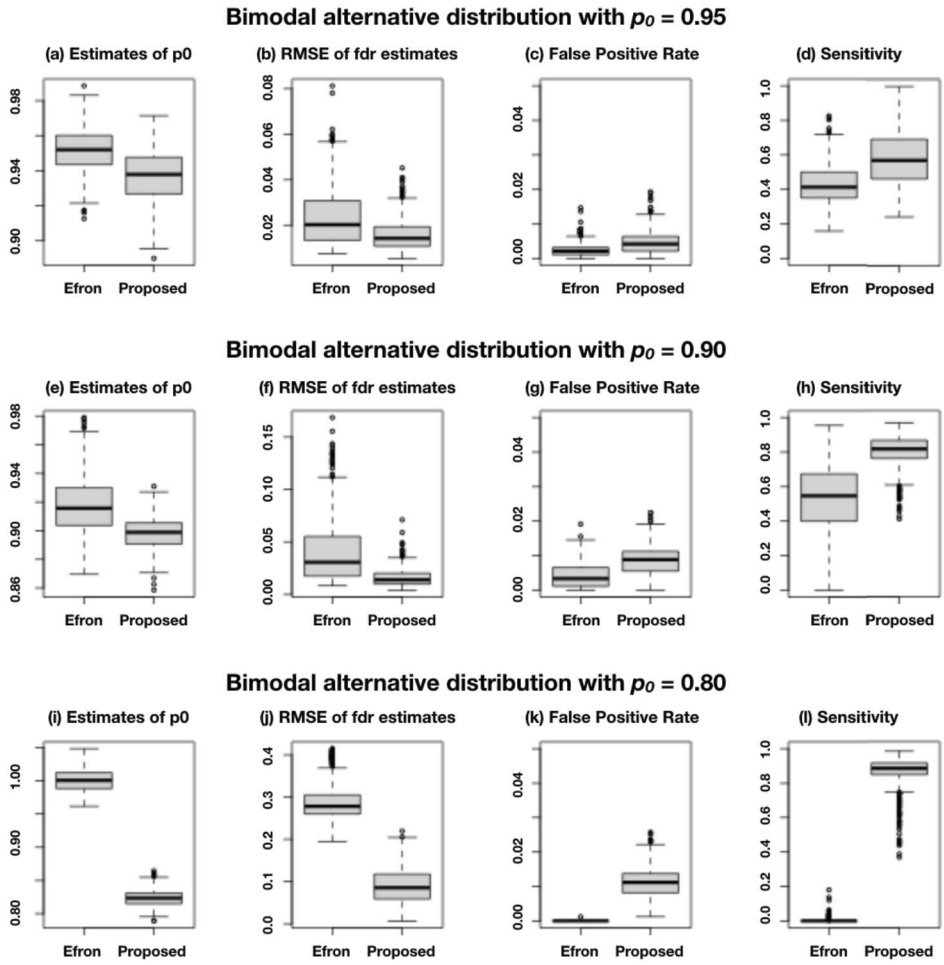

FIG. 7. *Comparison of the performances of the proposed method and Efron's method for the scenarios with a bimodal alternative distribution which violates the log-concavity assumption for alternative distributions. Boxplots are drawn with the results from* $M = 500$ *Monte Carlo experiments. Panels* (*a*)–(*d*) *summarize the results from the simulation setting with* $p_0 = 0.95$; (*e*)–(*h*) *are from the setting with* $p_0 = 0.90$, *and* (*i*)–(*l*) *are from the setting with* $p_0 = 0.80$.

our methods could identify more pathways with md-fdr< 0.1, but not necessarily all the 1d-fdr< 0.1. For examples, osteoclast differential pathway had the 1d-fdr< 0.1 in the sinus-brushings study only, while cell-surface interactions at the vascular wall pathway had the 1d-fdr< 0.1 in the sinus-brushings and the orbital-tissue studies. The former is known to be important in the pathogenesis of AAV, and the latter is highly relevant in vasculitis. Two large points in Figure 1 represent these two pathways.

Finally, potential novel common pathways were identified such as cell surface interactions at the vascular wall, amb2 integrin signaling, platelet pathways, etc., which should be investigated in future studies. The more comprehensive clinical interpretation can be found elsewhere (Friedman et al. (2019)).

5. Discussion. In this paper we were able to combine three lists of p-values of pathways from three different studies about AAVs by using the proposed semiparametric mixture model of normal and log-concave densities. Our method identified more pathways with md-fdr < 0.1 but not necessarily all the 1d-fdr < 0.1. For example, the osteoclast differential pathway had the 1d-fdr < 0.1 in only one study, but the md-fdr is significantly small. The pathogenesis

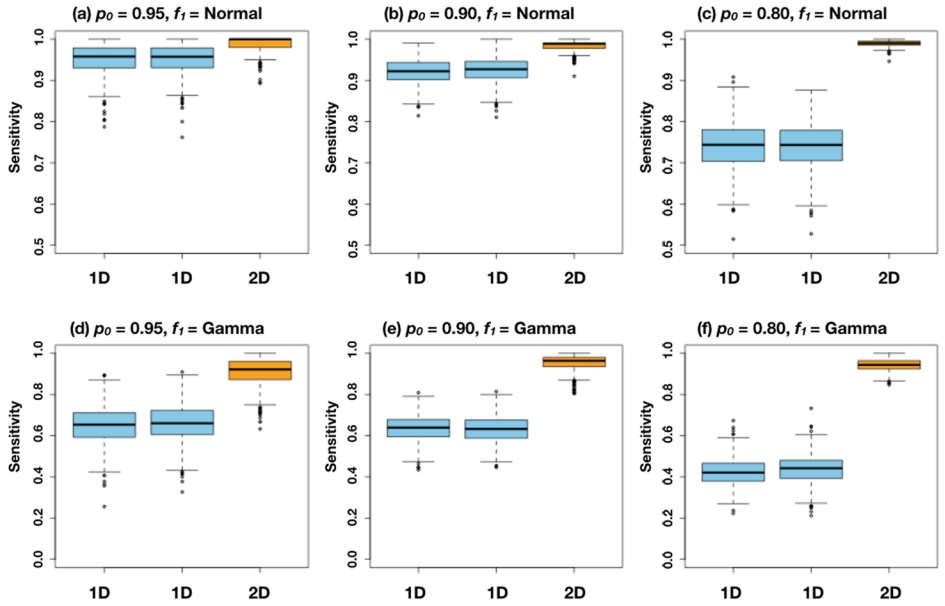

FIG. 8. *Simulation results from bivariate (two studies) setups. Boxplots are drawn with the results from* $M = 500$ *Monte Carlo experiments. Panels (a)–(c) summarize the simulation results from the setting with the normal + normal mixture marginals coupled with elliptical copula; (d)–(f) are from the setting with the normal + gamma mixture coupled with Frank copula. In each panel the first two boxplots are drawn with the results from fitting one-dimensional models separately for two components, and the last boxplot is from fitting bivariate model.*

of AAV agrees with our findings. To the best of our knowledge, there is no existing method that can compute combined FDR or fdr for multiple lists of p-values from several microarray studies.

We presented an EM-type algorithm to implement the proposed estimators for both single high-dimensional and multiple high-dimensional testing results. Our method can estimate fdr and the proportion of the null simultaneously and fit the alternative when necessary. Our method is easy to use because it does not require smoothing parameter selection. Our simulation studies showed that the proposed method outperforms other existing method in both single study and multiple studies. We presented an application which demonstrates the unique feature of the proposed semiparametric mixture model and, especially, the advantage of using md-fdr over 1d-fdr.

Recently, Wilson (2019) proposed the harmonic mean p-value (HMP) to combine dependent tests. HMP also has the Bayesian properties (Held (2019)). HMP appears to outperform other procedures that control the false discovery. It would be interesting if one extends the HMP to multiple studies with our method.

Acknowledgments. We are also grateful to Dr. Marcia Friedman at Oregon Health & Science University for providing us the motivating data used in Section 4. We also thank the Editor/Associate Editor and two referees for their constructive suggestions which enabled us to improve this manuscript significantly.

Supported by NIH Grants EY026572, EY020249 and EY010572. None of the funding organizations had a role in the design or conduct of this research. Woncheol Jang's research was supported by the National Research Foundation of Korea (NRF) grant and a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI) funded by the Korea government (MSIT) and the Ministry of Health & Welfare, Republic of Korea (No. 2017R1A2B2012816, HI19C0378). Seok-Oh Jeong's work

TABLE 2
Raw p values and estimated fdr*s for pathways with the* md-fdr < 0.10

Pathway	Peripheral Leukocytes	Sinus brushings	Orbital tissue	md-fdr
Neutrophil degranulation	$3.14 \cdot 10^{-13}$	$1.11 \cdot 10^{-16}$	$3.28 \cdot 10^{-9}$	$1.05 \cdot 10^{-12}$
Osteoclast differentiation	$1.26 \cdot 10^{-2}$	$6.19 \cdot 10^{-12}$	$3.91 \cdot 10^{-1}$	$3.75 \cdot 10^{-5}$
Cell surface interactions at the vascular wall	$8.87 \cdot 10^{-1}$	$4.49 \cdot 10^{-4}$	$5.86 \cdot 10^{-5}$	$4.19 \cdot 10^{-4}$
Signaling by Interleukins	$8.91 \cdot 10^{-5}$	$8.24 \cdot 10^{-8}$	$1.88 \cdot 10^{-5}$	$6.13 \cdot 10^{-4}$
Phagosome	$4.77 \cdot 10^{-1}$	$2.88 \cdot 10^{-8}$	$2.13 \cdot 10^{-2}$	$3.17 \cdot 10^{-3}$
Leishmaniasis	$4.12 \cdot 10^{-2}$	$1.22 \cdot 10^{-9}$	$1.72 \cdot 10^{-1}$	$3.54 \cdot 10^{-3}$
Urokinase-type plasminogen activator and uPAR-mediated signaling	$3.54 \cdot 10^{-1}$	$9.81 \cdot 10^{-8}$	$9.39 \cdot 10^{-3}$	$7.77 \cdot 10^{-3}$
Antimicrobial peptides	$7.69 \cdot 10^{-6}$	$1.31 \cdot 10^{-3}$	$1.14 \cdot 10^{-3}$	$8.23 \cdot 10^{-3}$
Malaria	$4.00 \cdot 10^{-1}$	$3.77 \cdot 10^{-6}$	$1.69 \cdot 10^{-3}$	$1.16 \cdot 10^{-2}$
Tuberculosis	$3.96 \cdot 10^{-2}$	$3.79 \cdot 10^{-8}$	$2.52 \cdot 10^{-3}$	$1.91 \cdot 10^{-2}$
IL4-mediated signaling events	$5.47 \cdot 10^{-4}$	$1.07 \cdot 10^{-1}$	$5.04 \cdot 10^{-4}$	$2.07 \cdot 10^{-2}$
Signaling by the B Cell Receptor	$7.74 \cdot 10^{-1}$	$8.67 \cdot 10^{-1}$	$6.68 \cdot 10^{-3}$	$2.18 \cdot 10^{-2}$
Extracellular matrix organization	$4.96 \cdot 10^{-1}$	$2.51 \cdot 10^{-4}$	$1.27 \cdot 10^{-3}$	$2.40 \cdot 10^{-2}$
Toll-like receptors cascades	$1.85 \cdot 10^{-5}$	$2.78 \cdot 10^{-4}$	$3.39 \cdot 10^{-3}$	$2.55 \cdot 10^{-2}$
Measles	$7.59 \cdot 10^{-1}$	$1.23 \cdot 10^{-2}$	$2.84 \cdot 10^{-3}$	$3.48 \cdot 10^{-2}$
Cytokine-cytokine receptor interaction	$6.66 \cdot 10^{-3}$	$7.78 \cdot 10^{-6}$	$4.08 \cdot 10^{-4}$	$3.55 \cdot 10^{-2}$
Response to elevated platelet cytosolic Ca2+	$6.70 \cdot 10^{-1}$	$4.50 \cdot 10^{-6}$	$9.52 \cdot 10^{-2}$	$3.82 \cdot 10^{-2}$
Chemokine signaling pathway	$3.08 \cdot 10^{-1}$	$1.97 \cdot 10^{-6}$	$1.27 \cdot 10^{-2}$	$4.12 \cdot 10^{-2}$
Complement and coagulation cascades	$1.99 \cdot 10^{-1}$	$2.70 \cdot 10^{-7}$	$4.78 \cdot 10^{-2}$	$4.44 \cdot 10^{-2}$
GPVI-mediated activation cascade	$4.00 \cdot 10^{-1}$	$2.10 \cdot 10^{-3}$	$1.69 \cdot 10^{-3}$	$7.85 \cdot 10^{-2}$
amb2 Integrin signaling	$2.76 \cdot 10^{-1}$	$1.90 \cdot 10^{-6}$	$4.05 \cdot 10^{-2}$	$8.57 \cdot 10^{-2}$
Platelet homeostasis	$2.41 \cdot 10^{-2}$	$6.56 \cdot 10^{-1}$	$3.28 \cdot 10^{-3}$	$9.14 \cdot 10^{-2}$
Inflammatory bowel disease	$4.91 \cdot 10^{-3}$	$3.10 \cdot 10^{-2}$	$5.80 \cdot 10^{-4}$	$9.30 \cdot 10^{-2}$

was supported by the research fund of Hankuk University of Foreign Studies and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012R1A1A2006112).

SUPPLEMENTARY MATERIAL

A semiparametric mixture method for local false discovery rate estimation from multiple studies (DOI: 10.1214/20-AOAS1341SUPP; .pdf). We provide additional simulation results, including R codes, that support our claims on the robustness of the proposed method.

REFERENCES

ALCORTA, D. A., BARNES, D. A., DOOLEY, M. A., SULLIVAN, P., JONAS, B., LIU, Y., LIONAKI, S., REDDY, C. B., CHIN, H. et al. (2007). Leukocyte gene expression signatures in antineutrophil cytoplasmic autoantibody and lupus glomerulonephritis. *Kidney Int*. **72** 853–64.

BAGNOLI, M. and BERGSTROM, T. (2005). Log-concave probability and its applications. *Econom. Theory* **26** 445–469. MR2213177 https://doi.org/10.1007/s00199-004-0514-4

BALABDAOUI, F. and DOSS, C. R. (2018). Inference for a two-component mixture of symmetric distributions under log-concavity. *Bernoulli* **24** 1053–1071. MR3706787 https://doi.org/10.3150/16-BEJ864

BORDES, L., DELMAS, C. and VANDEKERKHOVE, P. (2006). Semiparametric estimation of a two-component mixture model where one component is known. *Scand. J. Stat.* **33** 733–752. MR2300913 https://doi.org/10.1111/j.1467-9469.2006.00515.x

CHANG, G. T. and WALTHER, G. (2007). Clustering with mixtures of log-concave distributions. *Comput. Statist. Data Anal.* **51** 6242–6251. MR2408591 https://doi.org/10.1016/j.csda.2007.01.008

COORDINATORS, N. R. (2017). Database resources of the national center for biotechnology information. *Nucleic Acids Res*. **45** (Database issue) D12–D17.

CULE, M., GRAMACY, R. and SAMWORTH, R. (2009). LogConcDEAD: An R package for maximum likelihood estimation of a multivariate log-concave density. *J. Stat. Softw.* **29**.

CULE, M., SAMWORTH, R. and STEWART, M. (2010). Maximum likelihood estimation of a multi-dimensional log-concave density. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **72** 545–607. MR2758237 https://doi.org/10.1111/j.1467-9868.2010.00753.x

DÜMBGEN, L. and RUFIBACH, K. (2009). Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli* **15** 40–68. MR2546798 https://doi.org/10.3150/08-BEJ141

DÜMBGEN, L. and RUFIBACH, K. (2011). logcondens: Computations related to univariate log-concave density estimation. *J. Stat. Softw.* **39** 1–28.

EDGAR, R., DOMRACHEV, M. and LASH, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. **30** 207–210.

EFRON, B. (2008). Microarrays, empirical Bayes and the two-groups model. *Statist. Sci*. **23** 1–22. MR2431866 https://doi.org/10.1214/07-STS236

EFRON, B. (2009). Empirical Bayes estimates for large-scale prediction problems. *J. Amer. Statist. Assoc*. **104** 1015–1028. MR2562003 https://doi.org/10.1198/jasa.2009.tm08523

EFRON, B. (2010). *Large-Scale Inference*: *Empirical Bayes Methods for Estimation*, *Testing*, *and Prediction*. *Institute of Mathematical Statistics* (*IMS*) *Monographs* **1**. Cambridge Univ. Press, Cambridge. MR2724758 https://doi.org/10.1017/CBO9780511761362

EFRON, B., TURNBULL, B. and NARASIMHAN, B. (2015). locfdr: Computes local false discovery rates. R package version 1.1-8. https://CRAN.R-project.org/package=locfdr.

FABREGAT, A., SIDIROPOULOS, K., GARAPATI, P., GILLESPIE, M., HAUSMANN, K., HAW, R., JASSAL, B., JUPE, S., KORNINGER, F. et al. (2016). The reactome pathway knowledgebase. *Nucleic Acids Res*. **44** (Database issue) D481–D487.

FABRIZIO, M., NONINO, M., BONO, G., FERRARO, I., FRANÇOIS, P., IANNICOLA, G., MONELLI, M., THÉVENIN, F., STETSON, P. B. et al. (2011). The Carina Project. IV. Radial velocity distribution. *Publ. Astron. Soc. Pac*. **123** 384–401.

FRIEDMAN, M. A., CHOI, D., PLANCK, S. R., ROSENBAUM, J. T. and SIBLEY, C. (2019). Gene expression pathways across multiple tissues in anti-neutrophil cytoplasmic antibody-associated vasculitis reveal core pathways of disease pathology. *J. Rheumatol*. https://doi.org/10.3899/jrheum.180455

GENOVESE, C. and WASSERMAN, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist*. **32** 1035–1061. MR2065197 https://doi.org/10.1214/009053604000000283

GRAYSON, P. C., STEILING, K., PLATT, M., BERMAN, J. S., ZHANG, X., XIAO, J., ALEKSEYEV, Y. O., LIU, G., MONACH, P. A. et al. (2015). Defining the nasal transcriptome in granulomatosis with polyangiitis (Wegener's). *Arthritis Rheumatol*. **67** 2233–2239.

HELD, L. (2019). On the Bayesian interpretation of the harmonic mean p-value. *Proc. Natl. Acad. Sci. USA* **116** 5855–5856. MR3939777 https://doi.org/10.1073/pnas.1900671116

HU, H., WU, Y. and YAO, W. (2016). Maximum likelihood estimation of the mixture of log-concave densities. *Comput. Statist. Data Anal*. **101** 137–147. MR3504841 https://doi.org/10.1016/j.csda.2016.03.002

HUNG, H. M. J., O'NEILL, R. T., BAUER, P. and KÖHNE, K. (1997). The behavior of the P-value when the alternative hypothesis is true. *Biometrics* **53** 11–22. MR1450180 https://doi.org/10.2307/2533093

HUNTER, D. R., WANG, S. and HETTMANSPERGER, T. P. (2007). Inference for mixtures of symmetric distributions. *Ann. Statist*. **35** 224–251. MR2332275 https://doi.org/10.1214/009053606000001118

JEONG, S.-O., CHOI, D. and JANG, W. (2020). Supplement to "A semiparametric mixture method for local false discovery rate estimation from multiple studies." https://doi.org/10.1214/20-AOAS1341SUPP

JOHNSTONE, I. M. and SILVERMAN, B. W. (2005). Empirical Bayes selection of wavelet thresholds. *Ann. Statist*. **33** 1700–1752. MR2166560 https://doi.org/10.1214/009053605000000345

KALLENBERG, C. G. M. (2014). Key advances in the clinical approach to ANCA-associated vasculitis. *Nat. Rev. Rheumatol*. **10** 484–493.

KUMAR PATRA, R. and SEN, B. (2016). Estimation of a two-component mixture model with applications to multiple testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **78** 869–893. MR3534354 https://doi.org/10.1111/rssb.12148

MACFARLANE, D. G., BOURNE, J. T., DIEPPE, P. A. and EASTY, D. L. (1983). Indolent Wegener's granulomatosis. *Ann. Rheum. Dis*. **42** 398–407.

PLONER, A., CALZA, S., GUSNANTO, A. and PAWITAN, Y. (2006). Multidimensional local false discovery rate for microarray studies. *Bioinformatics* **22** 556–565.

RITCHIE, M. E., PHIPSON, B., WU, D., HU, Y., LAW, C. W., SHI, W. and SMYTH, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. **43** e47. https://doi.org/10.1093/nar/gkv007

ROSENBAUM, J. T., CHOI, D., WILSON, D. J., GROSSNIKLAUS, H. E., HARRINGTON, C. A., SIBLEY, C. H. et al. (2015). Orbital pseudotumor can be a localized form of granulomatosis with polyangiitis as revealed by gene expression profiling. *Exp*. *Mol*. *Pathol*. **99** 271–278.

SELLKE, T., BAYARRI, M. J. and BERGER, J. O. (2001). Calibration of p values for testing precise null hypotheses. *Amer. Statist*. **55** 62–71. MR1818723 https://doi.org/10.1198/000313001300339950

SODERBERG, D. and SEGELMARK, M. (2016). Neutrophil extracellular traps in ANCA-associated vasculitis. *Front*. *Immunol*. **7** 256.

WALTHER, G. (2002). Detecting the presence of mixing with multiscale maximum likelihood. *J. Amer. Statist. Assoc*. **97** 508–513. MR1941467 https://doi.org/10.1198/016214502760047032

WALTHER, G. (2009). Inference and modeling with log-concave distributions. *Statist*. *Sci*. **24** 319–327. MR2757433 https://doi.org/10.1214/09-STS303

WILSON, D. J. (2019). The harmonic mean p-value for combining dependent tests. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **116** 1195–1200. MR3904688 https://doi.org/10.1073/pnas.1814092116

YAN, J. (2007). Enjoy the joy of copulas: With a package copula. *J*. *Stat*. *Softw*. **21**.