6-1-2016

Testing multiple hypotheses with skewed alternatives

Naveen K. Bansal
Marquette University, naveen.bansal@marquette.edu

Gholamhossein G. Hamedani
Marquette University, gholamhoss.hamedani@marquette.edu

Mehdi Maadooliat
Marquette University

Accepted version. *Biometrics*, Vol. 72, No. 2 (June 2016): 494–502. DOI.

Testing Multiple Hypotheses with Skewed Alternatives

Naveen K. Bansal
Department of Mathematics, Statistics, and Computer Sciences, Marquette University
Milwaukee, WI
G.G. Hamedani
Department of Mathematics, Statistics, and Computer Sciences, Marquette University
Milwaukee, WI
Mehdi Maadooliat
Department of Mathematics, Statistics, and Computer Sciences, Marquette University
Milwaukee, WI

Summary: In many practical cases of multiple hypothesis problems, it can be expected that the alternatives are not symmetrically distributed. If it is known a priori that the distributions of the alternatives are skewed, we show that this information yields high power procedures as compared to the procedures based on symmetric alternatives when testing multiple hypotheses. We propose a Bayesian decision theoretic rule for multiple directional hypothesis testing, when the alternatives are distributed as skewed, under a constraint on a mixed directional false discovery rate. We compare the proposed rule

Biometrics, Vol 72, No. 2 (June 2016): pg. 494-502. DOI.

with a frequentist's rule of Benjamini and Yekutieli (2005) using simulations. We apply our method to a well-studied HIV dataset.

Key words: Bayes rule; Directional hypotheses; False discovery rate; Gene expressions; Skew normal distribution.

1. Introduction

In many biomedical applications such as in microarray data analysis, problems are stated in terms of multiple hypotheses. Since the number of hypotheses is very large in these applications, usually false discovery rate (FDR) is used to control the error rate. For directional hypotheses problems, mixed direction false discovery rate (*mdFDR*) or directional false discovery rate (DFDR) are used; see Schaffer (2002) and Benjamini and Yekutieli (2005). It turns out that the optimal procedures controlling DFDR (or *mdFDR*) use two-tailed procedures with decision about the direction made based on the sign of the test statistics. Implicitly this assumes that directional alternatives are symmetrically distributed in a random setting of null and alternative hypotheses. In many experiments where the effect of an intervention on multiple components (e.g., multiple genes) is sought, it would be unreasonable to assume that the left and right directional hypotheses are symmetrically distributed. This paper formulates and develops statistical procedure for experiments where the distribution of the alternative hypotheses is possibly skewed.

Let $\boldsymbol{X} = (X_1, X_2, \cdots, X_m)$ be a collection of test statistics such that $X_i \sim f(x_i|\theta_i)$. We consider the problem of directional hypotheses

$$H_i^{(0)} : \theta_i = 0 \; vs. H_i^{(-1)} : \theta_i < 0 \; or \; H_i^{(+1)} : \theta_i > 0, \; i = 1, 2, \cdots m. \tag{1}$$

If $f(\cdot|\theta)$ is symmetric around θ, and possesses a monotone likelihood ratio (MLR) property, then it can be expected that an optimal decision rule is of the form

Biometrics, Vol 72, No. 2 (June 2016): pg. 494-502. DOI.

Select $H_i^{(-1)}$ if $X_i < -c$; select $H_i^{(+1)}$ if $X_i > c$; otherwise select $H_i^{(0)}$, (2)

where constant c is chosen on the basis of DFDR (or *mdFDR*). Schaffer (2002) and Benjamini and Yekutieli (2005) showed that hypotheses (1) can be viewed as first testing two-sided hypotheses $H_i^{(0)} : \theta_i = 0 \, vs. H_i^{(a)} : \theta_i \neq 0, \; i = 1, 2, \cdots, m$; and upon rejection of $H_i^{(0)}$, the $H_i^{(-1)}$or $H_i^{(+1)}$ can be selected on the basis of the sign of X_i.

In a Bayesian setting, $\theta_i, i = 1, 2, \cdots, m$ would be randomly generated from some prior. If the prior is symmetric around 0, then one would expect a Bayes rule to be of the form (2). In many cases, however, it is unlikely that the prior distribution is symmetric. For example, in gene expression analysis involving microRNA, where the objective is to detect under and over expressed genes, the prior distribution is usually skewed (Bansal and Miescke, 2013). Bansal and Miescke (2013) considered the prior of the form

$$\pi(\theta_i) = p - \pi - (\theta_i) + p_0 I(\theta_i = 0) + p + \pi + (\theta_i), \quad (3)$$

where $p_- \neq p_+, \text{and } \pi_+(\theta) (= \pi_-(-\theta))$ is a density with support $(0, \infty)$. They showed that the selection rule (2) is not optimal, and the optimal rule is of the form

Select $H_i^{(-1)}$ if $X_i < -c_1$; select $H_i^{(+1)}$ if $X_i > c_2$; otherwise select $H_i^{(0)}$, (4)

where c_1 and c_2 are some positive constants, and need not be the same. This makes sense since if $p_- = 0$, then we essentially have one sided tests, and in such cases c_1 should be ∞.

In this paper, we consider a skew normal prior instead of (3) and develop Bayesian methodology for testing hypotheses (1). Skew normal priors have been shown to have many applications (Azzalini and Capitanio, 1999; Gottardo et al., 2006; Huang and Dagne, 2010; Chen et al., 2014). The following Theorem, which is essentially due to (Azzalini, 1985), gives a motivation behind the choice of skew normal priors for many practical examples.

Biometrics, Vol 72, No. 2 (June 2016): pg. 494-502. DOI.

Theorem 1: *Let* (θ, ξ) *be jointly distributed as*

$$g(\theta, \xi) = g_1(\theta) g_2(\xi - \beta\theta), \quad (5)$$

where g_1 *and* g_2 *are continuous and symmetric densities with support* $(-\infty, \infty)$*, and* β *is a parameter. Then the conditional distribution of* θ *given* $\xi > 0$ *is*

$$g+(\theta) = 2 g_1(\theta) G_2(\beta\theta), \quad (6)$$

where G_2 *is the distribution function corresponding to* g_2*.*

Proof of Theorem 1 can be seen easily by direct computation, and by noticing that the symmetry of g_1 and g_2 implies that the distribution of ξ is symmetric and thus $P(\xi > 0) = 1/2$.

Note that for bivariate normal variates (θ, ξ) with means 0 and variances and correlation parameters σ_1^2, σ_2^2, ρ, the joint density can be written in the form of (5) with

$$g_1(\theta) = \frac{1}{\sigma_1} \phi\left(\frac{\theta}{\sigma_1}\right), \text{and } g_2(\xi) = \frac{1}{\sigma_2\sqrt{1-\rho^2}} \phi\left(\frac{\xi}{\sigma_2\sqrt{1-\rho^2}}\right).$$

Here and thereafter $\phi(\cdot)$ denotes the density of $N(0,1)$.

Theorem 1 implies that the conditional distribution of θ given $\xi > 0$ is given by

$$g+(\theta) = 2\frac{1}{\sigma_1} \phi\left(\frac{\theta}{\sigma_1}\right) \Phi\left(\frac{\rho}{\sigma_1\sqrt{1-\rho^2}}\theta\right), \quad (7)$$

where $\Phi(\cdot)$ denotes the distribution function of $N(0,1)$. The distribution with density (7) is called the skew normal distribution (Azzalini, 1985).

Biometrics, Vol 72, No. 2 (June 2016): pg. 494-502. DOI.

The implication of Theorem 1 is that if a mediation is infused in a normal system and if the effect of the mediation is positive (or negative), then the effective result is a skewed distribution. Many genetic experiments involve this phenomenon. For example, consider a HIV study by van't Wout et al. (2003) in which a $CD4^+$ - T-cell line was inoculated with HIV-1 virus, and the gene expression levels of 7680 cellular RNA transcripts of the infected cells were compared with the gene expression of uninfected cells. Theorem 1 and equation (7) suggest that instead of assuming that the effects due to HIV infection is distributed as normal, we should assume them to be distributed as skew normal. The data of this study has been previously investigated; see, for example Efron (2007a,b) and Gottardo et al. (2006). We will analyze this data by using a Bayesian methodology assuming a skewed normal prior for the affected genes. We will consider the problem in the form of hypotheses testing (1), where the true gene expression levels $\theta_i, i = 1, 2, \cdots, m$ are either null $(\theta_i = 0)$ or non-null $(\theta_i \neq 0)$ with non-null $\theta_i s$ generated from a skew normal distribution with density

$$g_+(\theta) = 2\frac{1}{\sigma_1}\phi\left(\frac{\theta}{\sigma_1}\right)\Phi\left(\lambda\frac{\theta}{\sigma_1}\right), \quad (8)$$

where λ is a skewed parameter ranging from $-\infty$ to ∞. Note that $\lambda = 0$ yields the $N(0, \sigma_1^2)$ prior.

The skewed normal prior can also be justified for microRNA microarray data. MicroRNAs (miRNAs) are short non-coding RNA molecules that are believed to play an important role in regulating protein-coding genes in plant and animals including cancer genes. The miRNAs are attached to the targeted messenger RNAs (mRNAs) near their 5'-dominant region at seed position 2-8, and thus prevent the translation of the mRNAs (Lim et al., 2005). Research studies on miRNAs generally involve experiments in which specific microRNAs are either silenced or transfected. The problem of interest is to identify the targeted genes that are regulated by the microRNA. Lim et al. (2005) studied the effect of a miRNA, miR-124, by transfecting it in brain cells where it is expressed at very low level. They observed that a larger number of mRNAs were downregulated than upregulated as it can be expected. This makes sense since the transfection of miR-124 is likely

Biometrics, Vol 72, No. 2 (June 2016): pg. 494-502. DOI.

to result in gene suppression. Thus, as Theorem 1 suggests, the use of skewed prior (8) would be appropriate. We analyzed this data using methodologies developed in this paper. The results are similar to the ones for HIV data, and are presented in the supplementary materials.

In Section 2, we develop a Bayesian decision theoretic methodology for testing hypotheses (1) under a constraint on a mixed directional false discovery rate *mdFDR* (Benjamini and Yekutieli, 2005). We show that the optimal Bayes rule under skewed prior is of the form (4). We further prove theoretically that a skewed prior permits a higher power in number of correct discoveries than if the prior is symmetric. This result can be viewed analogously to one-tailed versus two-tailed hypothesis testing in which one-tailed test is more powerful than two-tailed test if it is known a priori that one-tailed alternative is true.

The rest of the paper is organized as follows. In Section 3, we discuss how to compute the Bayes rule and derive Bayes rule for the normal density $f(x|\theta)$. In Section 4, we discuss EM algorithm to estimate the parameters. A simulation study comparing the proposed rule with a frequentist's rule of Benjamini and Yekutieli (2005) is presented in Section 5. Analysis of the HIV data is presented in Section 6. We end with some concluding remarks in Section 7.

2. Decision Theoretic Formulation and Bayes rules

Let $d = (d_1, d_2, \cdots, d_m)$ with $d_i \in \{-1, 0, 1\}$ denote a selection rule, where $d_i = -1$ means that $H_i^{(-1)}$ is selected, $d_i = 0$ means $H_i^{(0)}$ is selected, and $d_i = 1$ means $H_i^{(+1)}$ is selected. Denoting $\theta = (\theta_1, \theta_2, \cdots, \theta_m)$, we consider the loss function of the form

$$L(\theta, d) = \sum_{i=1}^{m} L(\theta_i, d_i), \quad (9)$$

where $L(\theta_i, d_i)$ is the loss for testing each individual hypothesis $H_i^{(0)}$ vs. $H_i^{(-1)}$ or $H_i^{(+1)}$. If $L(\theta_i, d_i)$ is the "0-1" loss, then $L(\theta, d)$ is the total number of false discoveries. We will mainly consider "0-1" loss; however, non-"0-1" loss functions can be considered without many

Biometrics, Vol 72, No. 2 (June 2016): pg. 494-502. DOI.

alterations in our results; see Bansal and Miescke (2013) for the general approach.

We now discuss that if distribution of non-null $\theta_i s$ is skewed then this information would yield a high power in terms of correct discoveries. Suppose $\theta_i, i = 1, 2, \cdots, m$ are generated from the prior distribution

$$\pi_\lambda(\theta_i) = p\, I(\theta_i = 0) + (1 - p)\, 2g(\theta_i)\, G(\lambda\theta_i),$$

where $g(\cdot)$ is a symmetric density, $G(\cdot)$ is its distribution function and _ is the skewed parameter. Note that $\lambda = 0$ yields a symmetric prior.

The Bayes rule can be obtained by minimizing the average risk

$$r(d) = \sum_{i=1}^{m} r_i(d_i),$$

where $r_i(d_i)$ is the individual average risk of individual d_i. $r(d)$ can be written as

$$r(d) = r_-(d) + r_0(d) + r_+(d), \qquad (10)$$

where $r_-(d) = \sum_{i=1}^{n} \int_{-\infty}^{0} R(\theta_i, d_i)\pi(\theta_i)d\theta_i, r_+(d) = \sum_{i=1}^{n} \int_{0}^{\infty} (\theta_i, d_i)\pi(\theta_i)d\theta_i$ and $r_0(d) = \sum_{i=1}^{n} R(0, d_i)$. Here $R(\theta_i, d_i)$ denotes the risk function of d_i, i.e. the expected loss with respect to $\boldsymbol{X}$ given θ. For the "0-1" loss, $r_-(d)$ is the average number of falsely selected $H_i^{(0)}$ and $H_i^{(+1)}$ when $H_i^{(-1)}$ is true, $i = 1, 2, \cdots, m$; $r_0(d)$ is the average number of falsely rejected $H_i^{(0)}, i = 1, 2, \cdots, m$; and $r_+(d)$ is the average number of falsely selected $H_i^{(-1)}$ and $H_I^{(0)}$ when $H_i^{(+1)}$ is true, $i = 1, 2, \cdots, m$.

Now, suppose we want to find an optimal Bayes rule subject to the constraint that the average number of falsely rejected $H_i^{(0)}, i = 1, 2, \cdots, m$ is some predetermined value, say q, i.e, $r_0(d) = q$. Let $d_\lambda^B = (d_{\lambda 1}^B, d_{\lambda 2}^B, \cdots, d_{\lambda m}^B)$ denote the optimal Bayes rules subject to $r_0(d) = q$ under the prior $\pi\lambda$. We will use the notation $r^\lambda(d)$ (and

Biometrics, Vol 72, No. 2 (June 2016): pg. 494-502. DOI.

similarly r_{-}^{λ} , r_{0}^{λ} , and r_{+}^{λ}) to denote the Bayes risk (i.e. the average risk) of a decision rule d with respect to the prior $\pi\lambda$. Thus $r^{\lambda}(d_{\lambda}^{B})$ is the optimal Bayes risk with respect to the prior $\pi\lambda$.

The following Theorem implies that if $\theta_i s$ are generated from a skewed prior $\pi\lambda$, then a higher number of discoveries is possible than if $\theta_i s$ are generated from the symmetric prior π_0.

Theorem 2: Suppose $L(\theta_i, -1) = L(-\theta_i, 1)$*. Let* d_0^B *and* d_λ^B *be the optimal Bayes rules under the priors* π_0 *and* π_λ *respectively subject to the constraint* $r_0(d) = q$*, and let* $r^0(d_0^B)$ *and* $r^\lambda(d_\lambda^B)$ *be the corresponding Bayes risks. If* $\theta_i s$ *are generated from* π_λ*, then*

$$(i) \qquad r^{\lambda}(d_0^B) = r^0(d_0^B) \; for \; all \; \lambda.$$
$$(ii) \qquad r^{\lambda}(d_\lambda^B) \preccurlyeq r^0(d_0^B) \; for \; all \; \lambda.$$

Proof of Theorem 2 is given in the Web Appendix A.

Remarks 1:

1. Although, in Theorem 2, we are using the constraint on $r_0(d)$ (the number of falsely rejected nulls for the "0-1" loss), we believe that similar result will hold for different error rates such as *mdFDR* or other DFDRs. We will demonstrate this in our simulation study.
2. The result (i) of Theorem 2 says that the Bayes risk of optimal rule d_0^B is the same under prior π_0 or any other prior π_λ. This would imply that for the "0-1" loss, the expected number of correct discoveries would be the same for d_0^B whether $\theta_i s$ are generated from the symmetric prior π_0 or from skewed prior π_λ.

We will now discuss the optimal Bayes rule subject to the constraint on a directional false discovery rate. Note that for directional hypotheses, we need to also consider false discoveries in the wrong directions. Various types of false discovery rates have been defined to tackle the issue of false directional discoveries; see, for example Schaffer (2002), Benjamini and Yekutieli (2005) and Bansal and

Biometrics, Vol 72, No. 2 (June 2016): pg. 494-502. DOI.

Miescke (2013). We will concentrate on mixed directional false discovery rate (*mdFDR*) as defined by Benjamini and Yekutieli (2005), which is the expected proportion of falsely selecting $H_i^{(-1)}s$ or $H_i^{(+1)}s$. Formally, it can be defined as

$$mdFDR = E\left[\frac{\sum_{i=1}^{m}\{I(d_i=-1)I(\theta_i \geqslant 0) + I(d_i=+1)I(\theta i \leqslant 0)\}}{(|D_-| + |D_+|)\vee 1}\right], \quad (11)$$

where D_- is the set of indices of selected $H_i^{(-1)}s$, and D_+ is the set of indices of selected $H_i^{(+1)}s$, and $|\cdot|$ denotes the cardinality of the set. In a Bayesian setting, the expectation in (11) is with respect to the distribution of $\mathbf{X}$ given θ and the marginal distribution of θ.

If we consider the posterior version of (11), then we have.

$$mdPFDR = \frac{\sum_{i=1}^{m}\left\{I(d_i=-1)\left(v_i^{(0)}+v_i^{(+)}\right)+I(d_i=+1)\left(v_i^{(0)}+v_i^{(-)}\right)\right\}}{(|D-| + |D+|)\vee 1}, \quad (12)$$

where

$$v_i^{(0)} = P(\theta_i = 0|x),\ v_i^{(-)} = P(\theta_i < 0|x),\ v_i^{(+)} = P(\theta_i > 0|x). \quad (13)$$

Since $v_i^{(-)} + v_i^{(0)} + v_i^{(+)} = 1$, (12) can be written as

$$mdPFDR = 1 - \frac{\sum_{i=1}^{m}\left\{I(d_i = -1)v_i^{(-)} + I(d_i = +1)v_i^{(+)}\right\}}{(|D_-| + |D_+|)\vee 1}. \quad (14)$$

Without any constraint on the false discovery rate, the Bayes rule based on the "0-1" loss selects indices $D_B^- = \left\{i : v_i^{(-)} > v_i^{(0)}, v_i^{(+)}\right\}$ corresponding to $H_i^{(-1)}, D_B^+ = \left\{i : v_i^{(+)} > v_i^{(0)}, v_i^{(-)}\right\}$ corresponding to $H_i^{(+1)}$ and $D_B^0 = \left\{i : v_i^{(0)} > v_i^{(-)}, v_i^{(+)}\right\}$ corresponding to $H_i^{(0)}$.

Biometrics, Vol 72, No. 2 (June 2016): pg. 494-502. DOI.

The Bayes rule subject to the constraint that $mdPFDR \preccurlyeq \alpha$ can be described by the following procedure (see, Bansal and Miescke, 2013):

Procedure A

Define $\xi_i = v_i^{(-)}$, *for* $i \in D_B^-$ *and* $\xi_i = v_i^{(+)}$ *for* $i \in D_B^+$. *Now rank all* $\xi_i \in D_B^- \cup D_B^+$ *from the lowest to the highest. Let the ranked values be denoted by* $\xi_{[1]} \preccurlyeq \xi_{[2]} \preccurlyeq \cdots \preccurlyeq \xi_{[\hat{k}]}$, *where* $\hat{k} = |D_B^- \cup D_B^+|$. *Denote*

$$\hat{\imath}_0 = \max\left\{ j \preccurlyeq \hat{k} : \frac{1}{j}\sum_{i=1}^{j} \xi_{[\hat{k}-i+1]} \succcurlyeq 1-\alpha \right\}. \quad (15)$$

Let D_ξ *denote the set of indices corresponding to* $\xi_{[\hat{k}]} \succcurlyeq \xi_{[\hat{k}-1]} \succcurlyeq \cdots \succcurlyeq \xi_{[\hat{k}-\hat{\imath}_0+1]}$. *Now select* $H_i^{(-1)}$ *for* $i \in D_B^- \cap D_\xi$ *and* $H_i^{(+1)}$ *for* $i \in D_B^+ \cap D_\xi$.

3. Computation of the Bayes Rule

We will now assume that $X_1, X_2, \cdots, X_m$ are independent with probability density function $f(x|\theta)$, which is symmetric around θ and has MLR property.

Note that D_B^- and D_B^+ can be written as

$$D_B^- = \left\{ i : T_-(x_i) > \frac{p}{1-p}, T_-(x_i) > T_+(x_i) \right\},$$
$$D_B^+ = \left\{ i : T_+(x_i) > \frac{p}{1-p}, T_+(x_i) > T_-(x_i) \right\}, \quad (16)$$

Where

$$T_-(x_i) = \int_{-\infty}^{0} \frac{f(x_i|\theta)}{f(x_i|0)} 2g(\theta)G(\lambda\theta)d\theta\,,$$

Biometrics, Vol 72, No. 2 (June 2016): pg. 494-502. DOI.

$$T_+(x_i) = \int_0^\infty \frac{f(x_i|\theta)}{f(x_i|0)} 2g(\theta)G(\lambda\theta)d\theta \,. \quad (17)$$

Since $f(x|\theta)$ has MLR property, $\frac{f(x|\theta)}{f(x|0)}$ is increasing in x for $\theta > 0$ and decreasing in x for $\theta < 0$. This implies that $T_-(x)$ is a decreasing function and $T_+(x)$ is an increasing function. Thus $T_-(x_i) > \frac{p}{(1-p)}$ implies $x_i < c_-$ for some constant c_- and $T_+(x_i) > \frac{p}{(1-p)}$ implies $x_i > c_+$ for some constant c_+.

3.1 Normal Density $f(x|\theta)$

Suppose the density $f(x|\theta)$ is normal, i.e. $X_i \sim N(\theta_i, \sigma_2)$.We consider a skew normal prior for the non-null $\theta_i s$, $\pi_1(\theta) = 2g_1(\theta)G_1(\lambda\theta)$, where $g_1(\theta) = \sigma_1^{-1}\,\phi\left(\frac{\theta}{\sigma_1}\right)$ and G_1 is its distribution function. It is easy to see from (17) that

$$T_+(x_i) = \frac{2}{\sigma_1}\int_0^\infty exp\left(\frac{x_i\theta}{\sigma^2}\right)\phi\left(\sqrt{\frac{1}{\sigma_1^2}+\frac{1}{\sigma^2}}\,\theta\right)\Phi\left(\frac{\lambda\theta}{\sigma_1}\right)d\theta\,,$$

$$T_-(x_i) = \frac{2}{\sigma_1}\int_{-\infty}^0 exp\left(\frac{x_i\theta}{\sigma^2}\right)\phi\left(\sqrt{\frac{1}{\sigma_1^2}+\frac{1}{\sigma^2}}\,\theta\right)\Phi\left(\frac{\lambda\theta}{\sigma_1}\right)d\theta. \quad (18)$$

Figure 1 shows the plots of $T_+(x)$ and $T_-(x)$. Because of monotonicity of $T_+(x)$ and $T_-(x)$, it is easy to see from Figure 1 that D_B^- and D_B^+ can be written as

$$D_B^- = \{i : x_i < -c_1\} \text{ and } D_B^+ = \{i : x_i > c_2\}, \quad (19)$$

where $c_1 > 0$ and $c_2 > 0$ are determined as shown in Figure 1 by considering the point of intersections of $y = \frac{p}{(1-p)}$ and $y = T_-(x)$, and $y = \frac{p}{(1-p)}$ and $y = T_+(x)$ respectively. Note that when $\lambda > 0$, the

Biometrics, Vol 72, No. 2 (June 2016): pg. 494-502. DOI.

intersection point Q (as shown in the figure) will be to the left of $x = 0$, and when $\lambda < 0$, Q will be to the right of $x = 0$. Thus when $\lambda > 0$, $c_1 > c_2$ and the opposite is true when $\lambda < 0$. When $\lambda = 0, T_-(x) = T_+(-x)$ and thus $c_1 = c_2$. If $\lambda \to \infty, T_-(x) \to 0$ and thus D_B^- is an empty set which is equivalent to a one-tailed test.

[Figure 1 about here.]

To implement Procedure A, note that $v_i^{(-)}$ and $v_i^{(+)}$ (written as functions) are given by

$$v^{(-)}(x_i) = \frac{(1 - p)T_-(x_i)}{p + (1 - p)\{T_-(x_i) + T_+(x_i)\}},$$

$$v^{(+)}(x_i) = \frac{(1 - p)T_+(x_i)}{p + (1 - p)\{T_-(x_i) + T_+(x_i)\}}. \quad (20)$$

Since $T_-(x)$ is decreasing and $T_+(x)$ is increasing, it is easy to see from (20) that $v^{(-)}(x)$ is a decreasing function and $v^{(+)}(x)$ is an increasing function. Also, note that $T_-(x) + T_+(x) = \frac{m_\lambda(x)}{f(x|0)}$ where $m_\lambda(x)$ is the marginal density under the skew normal prior $SN(0, \sigma_1^2, \lambda)$. Azzalini (1985) showed that the marginal $m\lambda$ is also a skew normal density with

$$m_\lambda(x) = \frac{1}{\sqrt{\sigma^2 + \sigma_1^2}} \phi\left(\frac{x}{\sqrt{\sigma^2 + \sigma_1^2}}\right) \Phi\left(\frac{\lambda\sigma_1}{\sigma^2(1 + \lambda^2) + \sigma_1^2} \frac{x}{\sqrt{\sigma^2 + \sigma_1^2}}\right). \quad (21)$$

The following steps can be performed to implement Procedure A:

> Step 1: Determine first the values of $-c_1$ and c_2 by determining the points of intersections of $y = T_-(x)$ and $y = \frac{p}{(1-p)}$, and $y = T_+(x)$ and $y = \frac{p}{(1-p)}$ respectively.

Biometrics, Vol 72, No. 2 (June 2016): pg. 494-502. DOI.

Step 2: Rank all values $x_1, x_2, \cdots, x_m$ from lowest to the highest, say $x_{[1]} \preccurlyeq x_{[2]} \preccurlyeq \cdots \preccurlyeq x_{[m]}$. Record their indices.

Step 3: Determine all indices corresponding to $D_B^- = \{i : x_{[i]} < -c_1\}$ and $D_B^+ = \{i : x_{[i]} > c_2\}$. Suppose the indices are denoted in order by $D_B^- = \{i_1^-, i_2^-, \cdots, i_{|D_B^-|}^-\}$ and $D_B^+ = \{i_1^+, i_2^+, \cdots, i_{|D_B^+|}^+\}$.

Step 4: Compute $\xi_{i_j^-} = v^{(-)}(x_{i_j^-})$ for $i_j^- \in D_B^-$ and $\xi_{i_j^+} = v^+(x_{i_j^+})$ for $i_j^+ \in D_B^+$ using (20) and (21).

Step 5: Rank all $\xi_{i_j^-}$ and $\xi_{i_j^+}$ values from the lowest to the highest. Denote them by $\xi_{[1]} \preccurlyeq \xi_{[2]} \preccurlyeq \cdots \preccurlyeq \xi_{[\hat{k}]}$, where $\hat{k} = |D_B^-| + |D_B^+|$.

Step 6: Determine $\hat{\imath}_0$ according to (15) and find the set of indices D_ξ corresponding to $\xi_{[\hat{k}]} \succcurlyeq \xi_{[\hat{k}-1]} \succcurlyeq \cdots \succcurlyeq \xi_{[\hat{k}-\hat{\imath}_0+1]}$. Select $H_i^{(-1)}$ for $i \in D_\xi \cap D_B^-$ and select $H_i^{(+1)}$ for $i \in D_\xi \cap D_B^+$.

Biometrics, Vol 72, No. 2 (June 2016): pg. 494-502. DOI.

Remarks 2:

Note that if the non-null $\theta_i s$ are generated from the skew normal distribution (8), then *mdPFDR* $\leqslant \alpha$ also implies the Bayesian *mdFDR* $\leqslant \alpha$, where expectations are taken with respect to both X and θ. If the non-null $\theta_i s$ are generated from other distributions, then the posterior *mdPFDR* $\leqslant \alpha$ may not imply the Bayesian *mdFDR* $\leqslant \alpha$. The procedure proposed by Benjamini and Yekutieli (2005) (BY) achieves Bayesian *mdFDR* $\leqslant \alpha$ irrespective of the distribution of the non-null $\theta_i s$, and thus can be viewed as a nonparametric method. However, for the non-null skewed normal alternatives, BY procedure will be inferior to the Procedure A in terms of the power of discoveries as demonstrated by the Theorem 2 and the simulation results discussed in Section 5. Thus, we believe that in many practical situations, where the skewed normally distributed $\theta_i s$ are appropriate non-null alternatives, Procedure A would perform better than BY.

4. Parameter Estimation using EM Algorithm

Parameter estimates can be found empirically by maximizing the marginal likelihood function using EM algorithm. A hierarchical Bayesian approach can also be used along the lines of Bansal et al. (2008), but here we will only consider empirical approach. Note that, marginally, under the null, $X_i \sim N(0, \sigma^2)$, and under the non-null, $X_i \sim SN\left(0, \sigma^2 + \sigma_1^2, \frac{\lambda\sigma_1}{\sqrt{(1+\lambda^2)\sigma^2 + \sigma_1^2}}\right)$ (Azzalini, 1985). This implies that, marginally, $X_i, i = 1, 2, \cdots, m$ are *i.i.d.* with a mixture of $N(0, \sigma^2)$ and $SN\left(0, \sigma^2 + \sigma_1^2, \frac{\lambda\sigma_1}{\sqrt{(1+\lambda^2)\sigma^2 + \sigma_1^2}}\right)$ with weights p and $1-p$ respectively. Lin et al. (2007) gave an EM algorithm to estimate the parameters of a mixture of skew normal distributions. Prates et al. (2013) provided an R subroutine, *mixsmsn*, for the EM algorithm to estimate the parameters of finite scale mixture of skew normal distributions. Since normal is a special case of skew normal, *mixsmsn* can be used to estimate the parameters in our case. Note that, in both Lin et al. (2007) and *mixsmsn* subroutine of Prates et al. (2013), components of

Biometrics, Vol 72, No. 2 (June 2016): pg. 494-502. DOI.

the mixture distributions are not identified. However, in the case considered here, we know that one component is normal with mean 0, and the other component is skew normal with the location parameter 0. Thus, some adjustments in *mixsmsn* are needed to force some parameters to be 0. With these adjustments, we estimate $p, \sigma^2, \sigma^2 + \sigma_1^2$, and $\frac{\lambda\sigma_1}{\sqrt{(1+\lambda^2)\sigma^2 + \sigma_1^2}}$, and then find the estimates of σ^2, σ_1^2, and λ. For the initial estimates, to use in the subroutine *mixsmsn*, we use method of moments estimates. Since the moments of skew normals are well known (see, Lin et al., 2007), it is easy to see that the first four moments of the marginal distribution of X_i are given by

$$\mu_1' = (1 - p)\sqrt{\frac{2}{\pi}}\delta\sigma_1,$$

$$\mu_2' = p\sigma^2 + (1 - p)(\sigma^2 + \sigma_1^2),$$

$$\mu_3' = (1 - p)\sqrt{\frac{2}{\pi}}\delta\sigma_1[3\sigma^2 + (3 - \sigma^2)\sigma_1^2],$$

$$\mu_4' = 2p\sigma^4 + 2(1 - p)(\sigma^2 + \sigma_1^2)^2,$$

where $\delta = \frac{\lambda}{1+\lambda^2}$. In practice, our proposed Procedure A will thus be implemented by first estimating the parameters as described above, and then by following the steps 1-6 as described in Section 3. It should be noted that the same data is used twice: First by estimating the parameters, and second by implementing Procedure A. This is very common in empirical Bayesian methodology (Efron, 2007b). The effect of estimation may not matter much since for large m, as it will be the case in most applications, the maximum likelihood estimates are likely to be close to the true parameters.

Biometrics, Vol 72, No. 2 (June 2016): pg. 494-502. DOI.

5. Monte Carlo Simulations

Through simulation studies, we compare the performance of the proposed Bayes method with the directional BY procedure. We simulate the test statistics $z_i \sim N(\theta_i, 1), i = 1, 2, \cdots, m$ with $[mp]$ of $\theta_i = 0$ and the remaining $m - [mp]$ of $\theta_i s$ generated from skewed normal distribution $SN(0, \sigma_1^2, \lambda)$.

We consider different combinations of p, σ_1 and λ from the following ranges: (1) $p = 0.7, 0.8, 0.9$, (2) $\sigma_1 = 2, 3, 5$, and (3) $\lambda = -2, -1.75, -1.5, \cdots, 1.5, 1.75, 2$. From each of the above combinations, we simulate 1000 data sets with m = 5000. For the proposed Procedure A as described in subsection 3.1, we first use the EM algorithm, as discussed in Section 4, to estimate p, σ_1 and λ. These estimates are then plugged in for the parameters before using Procedure A. The BY procedure is performed using p-values based on z_i and the the sign of z_i as described in Benjamini and Yekutieli (2005).

For comparison, we use the following two measurements:

(a) Correct Discovery: The expected number of correct discoveries.
(b) False Discovery Rate: The expected ratio of false discoveries to total number of discoveries.

To compare the procedures in left (right) directions, we further consider the expected number of correct left (right) discoveries and the expected rate of false left (right) discoveries. Figure 2 represents the aforementioned measurements for the proposed empirical Bayes and the BY procedure at $p = 0.8$ and different combinations of $\sigma 1$ and λ. Expected number of right discoveries and the expected rate of false right discoveries are not presented since they can be interpreted from the results of the left discoveries.

[Figure 2 about here.]

Figure 2 shows that, overall, the results of the proposed Bayes rule is superior to BY procedure as the skewed parameter λ moves away from 0. There are higher number of correct discoveries by the

Biometrics, Vol 72, No. 2 (June 2016): pg. 494-502. DOI.

Bayes rule with false discovery rate close to pre-assigned rate of 0.05. Note that the expected number of false discoveries for BY are flat with varying λ as we conjectured in Remarks 1. The figures for left discoveries show a gain made in expected number of correct discoveries when λ is negative while keeping the false left discovery rates comparable with the BY procedure. Note also that the false left discovery rates are much higher under BY rule for positive λ.

Compared to BY procedure, the advantage of the proposed Bayes method is more prominent as λ gets further away from zero when the skewness becomes more evident. For further illustration, we compare the performance of the Bayes and the BY methods in Table 1 based on 1000 simulations for $\lambda = -1, 0, 2, p = 0.7, 0.8, 0.9$ and $\sigma_1 = 2, 3, 5$. Some notable results of Table 1 are singled out below:

1. While *mdFDR* is controlled by $q = 0.05$ level in both proposed Bayes and BY methods, the expected number of correct discoveries by the Bayes procedure is always higher than the BY procedure.
2. For larger values of p and/or smaller values of σ_1, expected number of discoveries are small by both methods, but the proposed Bayes method is still better.
3. Although for λ near 0, the Bayes method performs only slightly better than BY, however, as λ gets further away from zero, the directional correct discovery rates by the proposed Bayes are more balanced and more reliable than by the BY procedure. More precisely, for negative (positive) λ, the correct left (right) discovery rates are comparable by both BY and the Bayes, but the right (left) discovery rates are significantly lower by the BY procedure than by the proposed Bayes method.

[Table 1 about here.]

6. HIV Data

The HIV-data (van't Wout et al., 2003), that has been previously studied in Efron (2007a,b), is revisited here. The data consists of eight microarrays, four from cells of HIV infected subjects and four from uninfected subjects, each with expression levels of 7680 genes. For each gene we obtain a two-sample t-statistic, comparing the infected versus the uninfected subjects, which is then transformed

to a z-value, where $z_i = \Phi^{-1}\{F_6(t_i)\}$. Here $F_6(\cdot)$ denotes the cumulative distribution function (cdf) of t-distribution with 6 degrees of freedom. The histogram of the z-values is given in the left panel of Figure 3 with a fit of skew normal distribution. Note that theoretically null distribution of z_i should have a $N(0,1)$ distribution if i^{th} gene is not affected by HIV infection. However, Efron (2007b) estimated the null distribution to be $N(-0.11, 0.75^2)$. Thus, we formulate our problem as testing hypotheses (1) with test statistics $Z_i \sim N(-0.11 + \theta_i, 0.75^2)$.

[Figure 3 about here.]

BY procedure resulted in cutoffs $(-3.94, 3.94)$, which resulted in 18 total discoveries with 2 genes declared as under-expressed and 16 as over-expressed. For the proposed Bayes rule, we first used the EM algorithm to obtain the parameter estimates as: $\hat{p} = 0.9, \hat{\sigma} = 0.79$, $\hat{\sigma}_1 = 1.54$ and $\hat{\lambda} = 0.22$. Next, we followed the Procedure A, and we ended up with cut-off points $(-2.82, 2.70)$ with total 86 discoveries (under-expressed genes: 23 and over-expressed genes: 63).

In parallel to Efron's local false discovery rate (Efron, 2007a), we define it as

$$lfdr(z) = \frac{pf_0(z)}{pf_0(z) + (1-p)f_1(z)}, \quad (22)$$

where $f_0(z)$ is the null density, and $f_1(z)$ is the non-null density, which in our case is the density of the $SN\left(0.11, \sigma^2 + \sigma_1^2, \frac{\lambda\sigma_1}{\sqrt{(1+\lambda^2)\sigma^2 + \sigma_1^2}}\right)$. The estimated functions of the $lfdr$ (22) and the Efron's $lfdr$ are plotted in the right panel of Figure 3. The cutoff threshold $lfdr \leqslant 0.2$ is used to discover infected genes (Efron, 2007a). Efron's $lfdr$ approach resulted in 160 discoveries (left: 54 and right:106), while the $lfdr$ approach based on (22) resulted in 94 discoveries (left: 17 and right: 77). It is interesting to note that there is not a big difference, in terms of the total number of discoveries, between the proposed Bayes approach and the $lfdr$ approach based on (22). Perhaps the Efron's $lfdr$ approach overpredicts the number of discoveries as it was also pointed

Biometrics, Vol 72, No. 2 (June 2016): pg. 494-502. DOI.

out by Gottardo et al. (2006). Also note that the cut-off z-values of the BY procedure produces very low values of $lfdr$.

7. Concluding Remarks

Generally, the hypotheses problems are stated in terms of either one-tailed or two-tailed tests. For directional hypotheses, two-tailed tests can still be used with decision about the direction made based on the sign of the test statistics, as long as the left and the right directional hypotheses are symmetrically distributed. Many practical situations would however dictate that one direction is more likely than the other direction under the alternative hypotheses. In such cases, one-tailed or two-tailed tests will not be appropriate. Bayesian approach seems to be an ideal approach to perform hypotheses testing in such situations, and the skew normal distribution can be an appropriate choice of the prior distribution under the alternative hypotheses. We showed in this paper that the use of such skewed distribution yields a better power theoretically and empirically. We developed a Bayesian decision theoretic methodology to obtain a Bayes rule obeying a control mixed directional false discovery rate (*mdFDR*). A local false discovery rate (Efron, 2007a) approach can also be implemented as a parametric alternative to Efron's non-parametric $lfdr$ approach.

We provide a complete procedure for directional hypotheses testing including (a) estimation of the parameters of the prior, and (b) derivation of the cut-off points of the test statistics. Any type of test statistics can be used as long as it can be converted back to the normal variates through the p-values. Our approach is parametric in the sense that we assume the distribution of the alternatives to be skew normal. However, this model can be further extended to incorporate a more general family of distributions (e.g. skew-t), or even relax the distribution assumption through a non-parametric framework.

Biometrics, Vol 72, No. 2 (June 2016): pg. 494-502. DOI.

8. Supplementary Materials

The analysis of the microRNA dataset provided in Lim et al. (2005) referenced in Section 1, the Web Appendix A referenced in Section 2, and the implementation of the R codes are available with this paper at the Biometrics website on Wiley Online Library.

Acknowledgements

The authors are deeply indebted to the anonymous referees and an Associate Editor for their important suggestions, which led to a significant improvement in this paper.

References

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 12, 171–178.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 61, 579–602.

Bansal, N. K., Maadooliat, M., and Wang, X. (2008). Empirical Bayes and hierarchical Bayes estimation of skew normal populations. *Communications in Statistics – Theory and Methods* 37, 1024–1037.

Bansal, N. K. and Miescke, K. J. (2013). A Bayesian decision theoretic approach to directional multiple hypotheses problems. *Journal of Multivariate Analysis* 120, 205– 215.

Benjamini, Y. and Yekutieli, D. (2005). False discovery rate: Adjusted multiple confidence intervals for selected parameters. *Journal of the American Statistical Association* 100, 71–80.

Chen, L., Pourahmadi, M., and Maadooliat, M. (2014). Regularized multivariate regression models with skew-t error distributions. *Journal of Statistical Planning and Inference* 149, 125 – 139.

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

Efron, B. (2007b). Correlation and large-scale simultaneous significance testing. *Journal of the American Statistical Association* 102, 93–103.

Ghosh, P., Branco, M. D., and Chakraborty, H. (2007). Bivariate random effect model using skew-normal distribution with application to HIV-RNA. *Statistics in Medicine* 26, 1255–1267.

Gottardo, R., Raftery, A. E., Yee Yeung, K., and Bumgarner, R. E. (2006). Bayesian robust inference for differential gene expression in microarrays with multiple samples. *Biometrics* 62, 10–18.

Huang, Y. and Dagne, G. (2010). Skew-normal Bayesian nonlinear mixed-effects models with application to aids studies. *Statistics in Medicine* 29, 2384–2398.

Lim, L. P., Lau, N. C., Garrett-Engele, P., Grimson, A., Schelter, J. M., Castle, J., Bartel,

D. P., Linsley, P. S., and Johnson, J. M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433, 769–773.

Lin, T. I., Lee, J. C., and Yen, S. Y. (2007). Finite mixture modeling using the skew normal distribution. *Statistica Sinica* 17, 909–927.

Muller, P., Parmigiani, G., and Rice, K. (2006). FDR and Bayesian multiple comparison rules. *Proc. Vanecia/ ISBA 8th World Meeting on Bayesian Statistics*. Alicante, Spain.

Prates, M. O., Lachos, V. H., and Cabral, C. R. B. (2013). *mixsmsn*: Fitting finite mixture of scale mixture of skew-normal distributions. *Journal of Statistical Software* 54,.

Schaffer, J. P. (2002). Multiplicity, directional (type III) errors, and the null hypothesis. *Psychological Methods* pages 356–369.

van't Wout, A. B., Lehrman, G. K., Mikheeva, S. A., O'Keeffe, G. C., Katze, M. G., Bumgarner, R. E., Geiss, G. K., and Mullins, J. I. (2003). Cellular gene expression upon human immunodeficiency virus type 1 infection of CD4+-T-cell lines. *Journal of Virology* 77, 1392–1402.