Chapter 4

Hypothesis Testing for High-Dimensional Problems

Naveen K. Bansal

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.70210

Abstract

For high-dimensional hypothesis problems, new approaches have emerged since the publication. The most promising of them uses Bayesian approach. In this chapter, we review some of the past approaches applicable to only law-dimensional hypotheses testing and contrast it with the modern approaches of high-dimensional hypotheses testing. We review some of the new results based on Bayesian decision theory and show how Bayesian approach can be used to accommodate directional hypotheses testing and skewness in the alternatives. A real example of gene expression data is used to demonstrate a Bayesian decision theoretic approach to directional hypotheses testing with skewed alternatives.

Keywords: multiple directional hypotheses, false discovery rate, familywise error rate, gene expression, skew-normal distribution

1. Introduction

In today's world, most of the statistical inference problems involve high-dimensional multiple hypothesis testing. Whenever we collect data, we collect data on multiple features, involving very high-dimensional variables in some cases. For example, gene expression data consist of gene expressions on thousands of genes; image data consist of image expressions on multiple voxels. The statistical analysis for these types of data involves multiple hypotheses testing (MHT). It is well known that univariate methods cannot be applied to simultaneously test hypotheses on the multiple features. The reason for this is that the error rates for the univariate analysis get multiplied under MHT, and as a result the actual error rate can be very high. To understand the main issue of multiplicity, consider the following example. Suppose there are, say, 100 misspelled words in a book, and each of these words occurs in 5% of the pages. You pick a page at random. For each misspelled word, the probability is certainly 0.05 of finding that word in the page. However, the probability is much higher that you will find at least one of the 100 misspelled words. If these words were independently

distributed, then the probability of finding at least one misspelled word is $1-(0.95)^{100}\approx 0.995$. If the placements of the misspelled words were positively dependent, then the probability will be lower than 0.995. For example, if we take an extreme case of dependence that they all occur together, then the probability will be 0.05. The same phenomenon occurs in the MHT. The statistical inference, based on the error rate of individual hypothesis testing, can lead to very high error rate for the combined hypotheses. Thus, for the MHT, adjustment in the error rate needs to be made. Note that the adjustment rate may depend on the dependent structure, but due to the complexity of the dependent structure in high dimension, dependency is usually ignored in the literature [1].

The statistical inference depends on how we define the error rate for the combined hypotheses testing. Let us suppose that there are m hypotheses testing $H_0^i vs. H_a^i$, $i=1,2,\ldots,m$. If we do not want to make even one false discovery, then we should control the familywise error rate (*FWER*), which is defined as

$$FWER=\Pr\left(\text{Falsely Reject } H_0^i \text{ for at least one } i,\ i=1,2,\ldots,m\right) \tag{1}$$

There are many methods for controlling $FWER\le\alpha_F$ (=0.05, e.g.). A simplest method is the Bonferroni's procedure. Let T_i be the test statistics for testing $H_0^i vs. H_a^i$ with the corresponding p-values p_i. Then, Bonferroni's procedure rejects H_0^i if $p_i<\alpha_F/m$. To see the proof of this, suppose I_0 be the set of all i for which H_0^i is true, and suppose $p_j<\alpha_F/m$ for at least one $\in I_0$. Then using Boole's inequality, we have, from Eq. (1),

$$FWER=Pr\left\{\bigcup_{i\in I_0}\left(p_i<\alpha_F/m\right)\right\}\le\sum_{i\in I_0}Pr\left\{p_i<\alpha_F/m\right\} \tag{2}$$

Now, since, under H_0^i, $p_i\sim U(0,1)$, $\Pr\{p_i<\alpha_F/m\}=\alpha_F/m$. Then, assuming that there are m_0 number of elements in I_0, we have, from Eq. (2),

$$FWER\le\frac{m_0\alpha_F}{m}\le\alpha_F$$

Holm [2] gave a modified version of Bonferroni's procedure which also controls the familywise error rate. Holm's Bonferroni Procedure is the following: First rank all the p-values, $p_{(1)}\le p_{(2)}\le\ldots\le p_{(m)}$, and let $H_0^{(1)}, H_0^{(2)},\ldots, H_0^{(m)}$ be their associated null hypotheses. Let l be the smallest index such that $p_{(l)}>\alpha_F/(m-l+1)$. Then, reject only those null hypotheses that are associated with $H_0^{(1)}, H_0^{(2)},\ldots, H_0^{(l-1)}$. Note that the selected hypotheses have p-values with $p_{(1)}<\alpha_F/m, p_{(2)}<\alpha_F/(m-1),\ldots,p_{(l-1)}<\alpha_F/(m-l+2)$, and thus more powerful than Bonferroni's procedure, since hypotheses that are selected under Bonferroni's procedure will also be selected under Holm's procedure.

The above Bonferroni type procedures are not very satisfactory when m is very high. Let us suppose $m=10,000$ (this is actually not very high for most of the high-dimensional problems), and suppose we want to control FWER by $\alpha_F=0.05$. Then, for Holm's procedure, the smallest

p-value has to be lower than 0.000005 in order to reject at least one hypothesis, which may be very hard to achieve. The problem is not really with Holm's procedure; the problem is with the use of FWER as an error rate. For a high-dimensional problem, it is unrealistic to seek for a procedure which will not make at least one false discovery. Benjamini and Hochberg [1] proposed a new approach called false discovery rate (FDR) and proposed a procedure that works much better for high-dimensional MHT.

In Section 2, we review the FDR procedure and Bayesian procedures for two-sided alternatives. An extension of directional hypotheses is presented in Section 3. In Section 3, we also discuss Bayesian procedures under skewed alternatives. In Section 4, the problem of directional hypotheses is considered by converting p-values to normally distributed test statistics. We also discuss, in Section 4, a Bayes procedure under skew-normal alternatives. An application using real data of gene expressions is also discussed in Section 4. Some concluding remarks are made in Section 5.

2. False discovery rate (FDR), Benjamini and Hochberg's (BH) procedure, and Bayesian procedures

For each of the hypothesis testing $H_0^i vs. H_a^i$, suppose a statistical procedure either rejects the null hypothesis H_0^i or fails to reject H_0^i. For the sake of simplicity, we equate fail to reject H_0^i as accepting the null H_0^i . However, for small sample size case, it will be unwise to make a conclusion of accepting H_0^i. From now on, rejections of the null will be called discoveries. **Table 1** shows the possible outcomes by a procedure, where, for example, V is the total number of discoveries, among them V_0 is the number of false discoveries.

Thus, the proportion of the false discoveries is $V_0/\max(V,1)$. The FDR is defined as the expected proportion of false discoveries, that is,

$$FDR = E\left[\frac{V_0}{\max(V,\,1)}\right]. \tag{3}$$

If, for example, $FDR = 0.05$, then we can expect on the average 5% of all discoveries to be false. In other words, under repeated experiments on the average, we make 5% of the false discoveries (in a frequentist's sense). Note that $FDR \le FWER = P(V_0 \ge 1)$ as the following inequality shows:

	Accept H_0	**Reject H_0**	**Total**
H_0 is true	U_0	V_0	m_0
H_a is true	U_a	V_a	$m-m_0$
	U	V	m

Table 1. Total number of decisions made.

$$FDR = E\left[\frac{V_0}{\max(V,1)}\right] = E\left[\frac{V_0}{\max(V,1)} I(V_0 \geq 1)\right] \leq E[I(V_0 \geq 1)] = P(V_0 \geq 1).$$

Thus, we are likely to make a higher number of discoveries under FDR approach than under FWER, since if a procedure controls FWER ($\leq\alpha$), then it also controls FDR (($\leq\alpha$), but not vice versa.

2.1. Benjamini and Hochberg's procedure

Benjamini and Hochberg [1] proposed the following BH procedure which controls the FDR.

Let p_i be the p-value for the ith hypothesis under a test statistic T_i. Suppose $T_1, T_2, \ldots, T_m$ are independently distributed. Let $p_{[1]} < p_{[2]} < \ldots < p_{[m]}$ be the ordered p-values with the corresponding null hypotheses be denoted by $H_0^{(1)}$, $H_0^{(2)}$, …, $H_0^{(m)}$. Let

$$i_0 = \max\left\{ i : p_{[i]} \leq \frac{i}{m}\alpha \right\}$$

Then, reject $H_0^{(i)}$ for all $i \leq i_0$.

This procedure controls $FDR \leq \frac{m_0}{m}\alpha \leq \alpha$. Since m_0 is unknown, having the upper bound of $\frac{m_0}{m}\alpha$ is not very useful. If m_0 can be estimated reliably, a better bound is possible.

The above result was proven in [1], under the independence of the test statistics. Hochberg and Yekitieli [3] extended the result to positively correlated test statistics, and they also sharpened the BH procedure with new i_0 defined as

$$i_0 = \max\left\{ i : p_{[i]} \leq \frac{1}{mc(m)}\alpha \right\},$$

where $c(m) = \sum_{i=1}^{m} \frac{1}{i}$.

2.2. Bayesian procedures

Under Bayesian setting, we assume that H_0^i and H_a^i, $i = 1, 2, \ldots, m$ are generated probabilistically with

$$P\left(H_0^i\right) = p \text{ and } P\left(H_a^i\right) = 1 - p$$

Under this setting, [4] developed a concept of local false discovery rate (fdr). If $T_i, i = 1, 2, \ldots, m$ are test statistics with pdf $T_i | H_0 \sim f_0(t)$ and $T_i | H_a \sim f_a(t)$. Then, marginally, $T_i \sim f(t) = pf_0(t) + (1-p)f_a(t)$, and

$$fdr(t) = P\left(H_0^i | T_i = t\right) = \frac{pf_0(t)}{f(t)} \tag{4}$$

The idea is that if $T_i \in [t, t + \delta t]$, where $\delta t \to 0$, then $fdr(t)$ represents that the proportion of the times H_0^i will be true. If t is very high, then $fdr(t)$ will be very small indicating the probability of H_0^i to be very small (i.e., the false discovery rate will be very small). In Eq. (3), p and $f(t)$ are unknown, which can be estimated (see [4]).

Storey [5] proposed a positive false discovery rate

$$pFDR = E\left[\frac{V_0}{V} | V > 0\right], \tag{5}$$

where expectation is with respect to the distribution of $(T_i, \theta_i), i = 1, 2, \ldots, m$. Under the assumption that $T_1, T_2, \ldots T_m$ are identically and independently distributed, [6] proved that

$$pFDR(\Gamma) = P(H_0 | T \in \Gamma),$$

for a procedure that rejects H_0^i when $T_i \in \Gamma$. Based on this, $q - value$ for the multiple hypothesis (analogous to p-value for a single hypothesis) is defined as the smallest value of $pFDR(\Gamma)$ such that the observed $T_i = t_i \in \Gamma$, see [6]. Under most cases, $q - value(t_i) = P(H_0 | \ T_i > t_i)$. This gives a procedure under multiple hypothesis that rejects H_0^i if $q - value(t_i) < \alpha$.

3. Directional hypotheses testing

As described earlier, the null hypothesis H_0^i is either accepted or rejected. In most cases, however, rejection of null hypotheses is not sufficient. After rejecting H_0^i, finding the direction of the alternatives may also be important. A detailed discussion of the directional hypotheses can be found in [7].

Direction hypotheses testing involves testing H_0^i against directional hypotheses H_-^i and H_+^i, and the objective is to obtain selection region $\{T_i \in \Gamma_-\}$ for selecting H_-^i and selection region $\{T_i \in \Gamma_+\}$ for selecting H_+^i. In other words, H_0^i will be rejected if $T_i \in \Gamma_-$ or $T_i \in \Gamma_+$, and the direction H_-^i or H_+^i is determined according to $T_i \in \Gamma_-$ or $T_i \in \Gamma_+$, respectively. Analogous to **Table 1**, we now have

Table 2 illustrates the number of cases possible when accepting H_0 or selecting H_- or selecting H_+. For example, out of V times when selecting H_-, V_0 times errors are made when in fact H_0 is

	Accept H_0	**Select H_-**	**Select H_+**	**Total**
H_0 is true	U_0	V_0	W_0	m_0
H_- is true	U_-	V_-	W_-	m_-
H_+ is true	U_+	V_+	W_+	m_+
Total	U	V	W	m

Table 2. Number of decisions under directional hypotheses.

true, and V_+ times errors are made when in fact H_+ is true. In other words, when selecting H_-, not only H_0 is falsely rejected V_0 times but the direction is also falsely selected V_+ times. This leads to a concept of directional false discovery rate *DFDR* defined as

$$DFDR = E\left[\frac{V_0 + V_+ + W_0 + W_-}{\max(V + W,\, 1)}\right]. \tag{6}$$

This is analogous to *FDR* for two-sided alternatives. For most cases, [8] showed that DFDR-controlling procedures for directional hypotheses can be treated as FDR-controlling procedure for two-sided multiple hypotheses with direction determined by the sign of the test statistics.

Bansal and Miescke [9] considered a decision theoretic formulation to multiple hypotheses problems. The approach assumes parametric modeling. Suppose the model for the observed data $\boldsymbol{x}$ be represented by $P(\boldsymbol{x};\, \boldsymbol{\theta}, \eta)$, where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)'$ is a parameter vector of interest, and η is a nuisance parameter. The problem of interest is to test

$$H_0^i : \theta_i = 0 \, vs. \, H_-^i : \theta_i < 0 \; or \; H_+^i : \theta_i > 0 \tag{7}$$

Let the loss function of a decision rule $d(\boldsymbol{x}) = (d_1(\boldsymbol{x}), d_2(\boldsymbol{x}), \ldots, d_m(\boldsymbol{x}))$ is given by

$$L(\boldsymbol{\theta},\, d(\boldsymbol{x})) = \sum_{i=1}^{m} l_i(\boldsymbol{\theta},\, d_i(\boldsymbol{x})), \tag{8}$$

where $l_i(\boldsymbol{\theta}, d_i(\boldsymbol{x}))$ is an individual loss of d_i. Here, $d_i \in \{-1, 0, 1\}$ with $d_i = 0$, $d_i = -1$, and $d_i = 1$ means accepting H_0^i, selecting H_-^i and selecting H_+^i, respectively. Note that for the "0-1" loss, that is, when $l_i = 0$ for correct decision, and $l_i = 1$ for the incorrect decision, L is the total number of incorrect decisions. Thus, minimizing the $E[L(\boldsymbol{\theta}, d(\boldsymbol{X}))]$ for the "0-1" loss amounts to minimizing the expected number of incorrect decisions.

Now, suppose under the Bayesian setting, $\theta_i, i = 1, 2, \ldots, m$ are generated from

$$\pi(\theta) = p_- \pi_-(\theta) + p_0 I(\theta = 0) + p_+ \pi_+(\theta), \tag{9}$$

where π_- is the prior density over $(-\infty, 0)$ and π_+ is the prior density over $(0, \infty)$. A special case of prior (9) is that $\pi_-(\theta) = \pi_+(-\theta)$. In this case, p_- and p_+ reflect the skewness in the alternative hypotheses. For example, if $p_- = p_+$, then we have a symmetric case. In this case, the selection of H_- or H_+, after rejecting H_0, based on the sign of the test statistics makes sense. On the other hand, if $p_- < p_+$, then it reflects that more of the θ_is are positives than negatives. For many gene expressions data analyses, this presents a useful case when over-expressed genes may occur more frequently than under-expressed genes as a result of gene mutation (naturally or as a result of external factors). For specific examples, see [9, 10].

From now on, we focus on the "0-1" loss. The results can be easily extended to other loss functions. The "0-1" loss can be written as

$$L(\boldsymbol{\theta}, \boldsymbol{d}) = \sum_{i=1}^{m} \left[1 - \sum_{j=-1}^{1} I(d_i = j) I\left(v_i^{\theta} = j\right) \right],$$

where $v_i^{\theta} \in \{-1, 0, 1\}$ is an indicator variable indicating $\theta_i < 0$ when $v_i^{\theta} = -1$, $\theta_i = 0$ when $v_i^{\theta} = 0$, and $\theta_i > 0$ when $v_i^{\theta} = 1$. It is easy to see that minimizing the posterior expected loss yields the selection rule that selects H_-^i, H_0^i, *or* H_+^i according to $\max\left\{v_i^{(-)}, v_i^{(0)}, v_i^{(+)}\right\}$, where $v_i^{(-)} = P\left(H_i^{(-)}|\boldsymbol{x}\right)$, $v_i^{(0)} = P\left(H_i^{(0)}|\boldsymbol{x}\right)$, and $v_i^{(+)} = P\left(H_i^{(+)}|\boldsymbol{x}\right)$.

3.1. The constrained Bayes rule

The Bayes procedure described earlier accommodates skewness in the prior, but no type of false discovery rates is controlled. In order to control a false discovery rate, we need to obtain a constrained Bayes rule that minimizes the posterior expected loss subject to a constraint on the false discovery rate.

The directional false discovery rate (6) is defined in a frequentist's manner, in which expectation is with respect to $\boldsymbol{X}|\theta$. Let us define Eq. (6) as *BDFDR* when expectation is taken with respect to $\boldsymbol{X}|\theta$ and then further expectation is taken with respect to $\boldsymbol{\theta}$. We define posterior version of Eq. (6) as *PDFDR* when the expectation is taken with respect to the posterior distribution of $\boldsymbol{\theta}|X = x$. It can be shown that

$$PDFDR = 1 - \frac{\sum_{i=1}^{m} \left\{ I(d_i = -1) v_i^{(-)} + I(d_i = +1) v_i^{(+)} \right\}}{(|D_-| + |D_+|) \vee 1} \tag{10}$$

Here, $|D_-| = \sum_{i=1}^{m} I(d_i = -1)$ and $|D_+| = \sum_{i=1}^{m} I(d_i = 1)$.

A constrained Bayes rule can be obtained by minimizing the posterior expected loss subject to the constraint that $PDFDR \leq \alpha$. There can be many approaches to obtain the constraint minimization. We present, here, an approach given in [9], which is as follows:

Consider the sets D_-^B and D_+^B of indices that selects $H_i^{(-)}$ and $H_i^{(+)}$, respectively, according to the unconstraint Bayes rule, that is, when $v_i^{(-)} = \max\left\{v_i^{(0)}, v_i^{(+)}\right\}$ and $v_i^{(+)} = \max\left\{v_i^{(0)}, v_i^{(-)}\right\}$, respectively. Define $\xi_i = v_i^{(-)}$ for $i \in D_+^B$, and $\xi_i = v_i^{(+)}$ for $i \in D_+^B$, and then rank all ξ_i, $i \in D_-^B \cup D_+^B$ from the lowest to the highest. Let the ranked values be denoted by $\xi_{[1]} \leq \xi_{[2]} \leq \ldots \leq \xi_{\left[\widehat{k}\right]}$, where $\widehat{k} = \left|D_-^B \bigcup D_+^B\right|$. Denote

$$\widehat{i}_0 = \max\left\{ j \leq \widehat{k} : \frac{1}{j} \sum_{i=1}^{j} \xi_{\left[\widehat{k}-i+1\right]} \geq 1 - \alpha \right\}.$$

Let $\boldsymbol{D}_{\xi}$ denotes the set of indices corresponding to $\xi_{\left[\widehat{k}\right]} \geq \xi_{\left[\widehat{k}-1\right]} \geq \ldots \geq \xi_{\left[\widehat{k}-\widehat{i}_0+1\right]}$. Now, select $\boldsymbol{H}_-^{\boldsymbol{i}}$ for $\boldsymbol{i} \in \boldsymbol{D}_-^{\boldsymbol{B}} \cap \boldsymbol{D}_{\xi}$, and $\boldsymbol{H}_+^{\boldsymbol{i}}$ for $\boldsymbol{i} \in \boldsymbol{D}_+^{\boldsymbol{B}} \bigcap \boldsymbol{D}_{\xi}$.

3.2. Estimating mixture parameters

The above procedure requires estimation of the parameters (p_-, p_0, p_+) and estimation of the nuisance parameter η. Note that marginally,

$$X_i \sim p_- f_-(x_i|\eta) + p_0 f_0(x_i|\eta) + p_+ f_+(x_i|\eta),$$

where $f_0(x_i|\ \eta) = f(x_i|\ 0, \eta)$, and

$$f_-(x_i|\eta) = \int_{-\infty}^{0} f(x_i|\theta,\ \eta)\pi_-(\theta)d\theta,\ f_+(x_i|\eta) = \int_0^{\infty} f(x_i|\theta,\ \eta)\pi_+(\theta)d\theta$$

and $X_1, X_2, \ldots, X_m$ are independently distributed. Estimates of the parameters of the mixed density can be obtained by using EM algorithm. It is easy to see that the EM estimators of (p_-, p_0, p_+) follows the following iterative scheme:

$$p_-^{(j+1)} = \frac{1}{m}\sum_{i=1}^{m} \frac{p_-^{(j)} f_-(x_i|\eta)}{p_-^{(j)} f_-(x_i|\eta) + p_0^{(j)} f_0(x_i|\eta) + p_+^{(j)} f_+(x_i|\eta)},$$

$$p_0^{(j+1)} = \frac{1}{m}\sum_{i=1}^{m} \frac{p_0^{(j)} f_-(x_i|\eta)}{p_-^{(j)} f_-(x_i|\eta) + p_0^{(j)} f_0(x_i|\eta) + p_+^{(j)} f_+(x_i|\eta)},$$

$$p_+^{(j+1)} = \frac{1}{m}\sum_{i=1}^{m} \frac{p_+^{(j)} f_-(x_i|\eta)}{p_-^{(j)} f_-(x_i|\eta) + p_0^{(j)} f_0(x_i|\eta) + p_+^{(j)} f_+(x_i|\eta)}$$

Estimation of η can also be estimated iteratively by using EM algorithm or by different means. See [9] for more details.

4. Bayes rules by converting *p*-values to normally distributed test statistics

Let $T_i, i = 1, 2, .., m$ be independently and identically distributed test statistics. Let $P_i = P(T_i \leq t_i|H_0^i)$ be the corresponding p-values. Note that under H_0^i, $P_i \sim U(0, 1)$. Let $X_i = \Phi^{-1}(P_i)$ be the corresponding z-score. Then, under H_0^i, $X_i \sim N(0,1)$. Efron [11] suggested using $X_i \sim N(0, \sigma^2)$ under H_0^i with σ^2 appropriately estimated. Efron pointed out that, in practice, σ^2 may not be equal to 1 due to possible correlation among multiple components. Under the alternative, we assume that $X_i \sim N(\theta_i, \sigma^2)$, where $\theta_i s$ are generated with distribution described in Eq. (9). It is true that this is a big leap in making this assumption. In practice, this assumption can be tested, however, and if true, it can lead to very powerful results. [9] assumed that $\pi_+(\theta)$ is a truncated normal distribution $N(0,\ \sigma^2/\omega)$, and $\pi_-(\theta) = \pi_+(-\theta)$, where ω is some positive constant depending upon how inflated we believe the alternative $\theta_i s$ are. It can be seen that

$$v_i^{(-)} \propto p_- T_-(x_i),\ v_i^{(+)} \propto p_+ T_+(x_i),\ \text{and}\ v_i^{(0)} \propto p_0 \tag{11}$$

with the proportionality constant $[p_- T_-(x_i) + p_+ T_+(x_i) + p_0\}^{-1}$. Also, $T_-(x_i) = T_+(-x_i)$, and

$$T_+(x_i) = \exp\left\{\frac{x_i^2}{2(1+\omega)\sigma^2}\right\}\Phi\left(\frac{x_i}{\sigma\sqrt{1+\omega}}\right) \tag{12}$$

In order to apply the Bayes procedure as discussed in Section 3, all we need are Eqs. (11) and (12). For computation details, see [9].

4.1. Skew-normal alternatives

In the above discussions, we assumed that $\theta_i s$ are generated from distribution with pdf (9). [12] considered the case when $\theta_i s$ are generated from a skew-normal distribution under the alternative hypotheses. The skew-normal distribution was first introduced in [13]. It has an important property that if $(\xi_1, \xi_2) \sim$ Bivariate Norma with mean 0, then the distribution of $\xi_1 | \xi_2 > 0 \sim Skew - normal$. Its pdf is given by

$$g_+(\xi_1) = 2\frac{1}{\sigma_1}\phi\left(\frac{\xi_1}{\sigma_1}\right)\Phi\left(\lambda\frac{\xi_1}{\sigma_1}\right),$$

and is denoted by $SN(0, \sigma_1, \lambda)$. Here, λ is a skew parameter. If $\lambda = 0$, then this distribution is $N(0, \sigma_1)$. The implication of this result is the following: suppose within a normal system an outcome follows a normal distribution, but if a correlated factor starts exerting a positive effect, then the outcome variable will start following a skew-normal distribution. For example, consider RNAs experiments and assume that genes are in a normal state. Suppose a gene mutation occurs at a later state and it starts exerting positive effect on the affected genes. In this case, based on the above property of skew-normal distribution, we can assume that the expressions of the affected genes will follow a skew-normal distribution.

Under this formulation, we assume that $\theta_i = 1, 2, \ldots, m$ are generated from

$$\pi_\lambda(\theta_i) = pI(\theta_i = 0) + (1-p)\frac{2}{\sigma_1}\phi\left(\frac{\theta_i}{\sigma_1}\right)\Phi\left(\lambda\frac{\theta_i}{\sigma_1}\right)$$

Now, similar to Eq. (11), it can be seen that

$$v_i^{(-)} \propto (1-p)T_-(x_i),\ v_i^{(+)} \propto (1-p)T_+(x_i),\ v_i^{(0)} \propto p$$

with proportionality constant $[(1-p)(T_+(x_i) + T_-(x_i) + p]^{-1}$, where

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

and

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

The sets D_{-}^{B} and D_{+}^{B} can be written as

$$D_{-}^{B} = \{i : x_i < -c_1\} \text{ and } D_{+}^{B} = \{i : x_i > c_2\}$$

where $c_1 > 0$ and $c_2 > 0$ are determined as shown in **Figure 1** by considering the point of intersections of $y = p/(1-p)$ and $y = T_{-}(x)$, and $y = p/(1-p)$ and $y = T_{+}(x)$, respectively. Note that when $\lambda > 0$, the 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. As discussed in Section 3, the procedure based on Eq. (13) by itself does not control *BDFDR*. However, c_1 and c_2 can be further shrunk so that the resulting procedure achieves $BDFDR \le \alpha$; see [12] for details.

To illustrate the above procedure, and to compare it with the standard FDR procedure (BY) of [8], which selects the direction based on the sign of the test statistics, we consider a HIV data described in [14]. For detailed analysis, see [12]. Here, we describe the analysis very briefly. The data consist 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 obtained 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

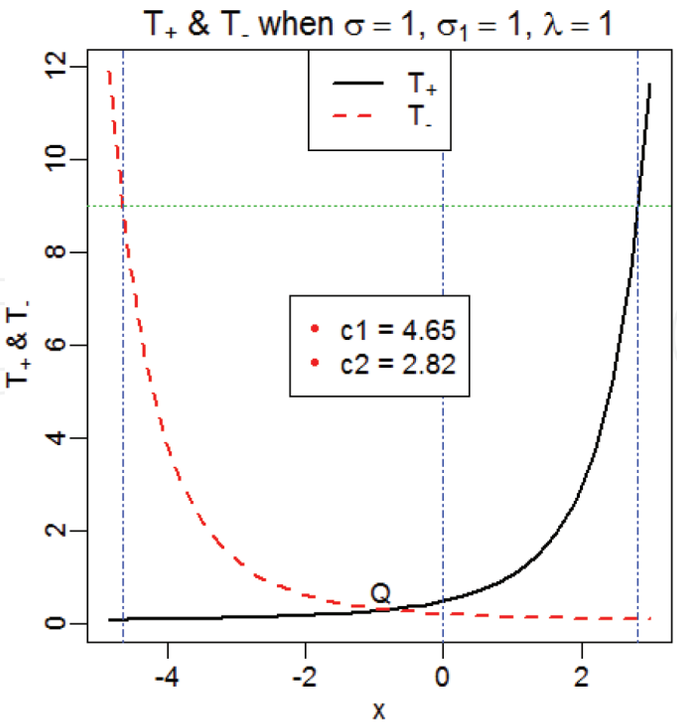

Figure 1. Graph of $T_{+}(x)$ and $T_{-}(x)$ with cutoff values $-c_1$ and c_2 such that $T_{+}(x) \ge \frac{p}{1-p}$ and $T_{-}(x) \ge \frac{p}{1-p}$.

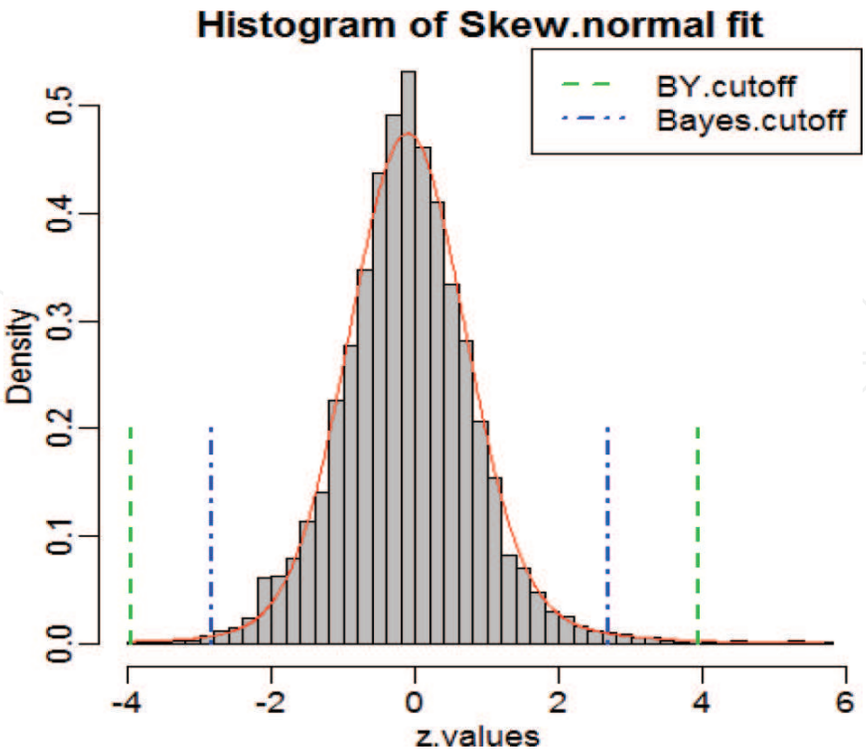

Figure 2. Histogram of the HIV data with cutoff points by BY and the Bayes method under skew-normal prior.

function (cdf) of t -distribution with six degrees of freedom. **Figure 2** shows the histogram of the z-values with a skew-normal fit. Although the null distribution of Z_i should be $N(0,1)$. However, as suggested in [11], we use the null distribution as $N(-0.11, 0.75^2)$. Thus, we formulate our problem as testing hypotheses (7) with test statistics $Z_i \sim N(-0.11 + \theta_i, 0.75^2)$.

BY procedure resulted in cutoffs $(-3.94, 3.94)$, which resulted in 18 total discoveries with two genes declared as under-expressed and 16 as over-expressed. For the constrained Bayes rule, we first used the EM algorithm to obtain the parameter estimates as $\widehat{p} = 0.9$, $\widehat{\sigma} = 0.79$, $\widehat{\sigma_1} = 1.54$, and $\widehat{\lambda} = 0.22$. The Bayes procedure ended up with cut-off points $(-2.82, 2.70)$ with a total of 86 discoveries (under-expressed genes: 23 and over-expressed genes: 63). Note that the number of discoveries by the Bayes rule is much higher than by the BY procedure.

5. Concluding remarks

There are many different methods of testing multiple hypotheses. Methodologies, however, depend on the criteria we choose. When the dimension of multiple hypotheses is not very high, the familywise error rate (FWER) is an appropriate criterion which safeguards against making even one false discovery. However, when the dimension of multiple hypotheses is very high, the FWER is not very useful; instead, a false discover rate (FDR) criterion is a good approach. Although FDR was originally defined as a frequentist's concept, it can be re-interpreted in a Bayesian framework. The Bayesian framework brings many advantages. For example, a decision-theoretic formulation is easy to implement, directional hypotheses are easy to handle,

and the skewness in the alternatives is easy to implement. Drawback is that we need to make an assumption about the prior distributions under the alternatives. Some work has been done based on nonparametric priors; however, much more work is needed.

Author details

Naveen K. Bansal

Address all correspondence to: naveen.bansal@mu.edu

Department of Mathematics, Statistics, and Computer Science, Marquette University, Milwaukee, WI, USA

References

[1] Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practice and powerful approach to multiple testing. Journal of the Royal Statistical Society B. 1995;**57**(1):289-300

[2] Holm S. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics. 1979;**6**(2):65-70

[3] Hochberg B, Yekitieli D. The control of the false discovery rate in multiple testing under dependency. Annals of Statistics. 2001;**29**(4):1165-1188

[4] Efron B, Tibshirani R, Storey JD, Tusher V. Empirical Bayes analysis of a microarray experiment. Journal of the American Statistical Association. 2001;**96**(456):1151-1160

[5] Storey JD. A direct approach to false discovery rates. Journal of the Royal Statistical Society B. 2002;**64**(3):479-498

[6] Storey JD. The positive false discovery rate: A Bayesian interpretation and the q value. The Annals of Statistics. 2003;**31**(6):2013-2035

[7] Shaffer JP. Multiplicity, directional (Type III) errors, and the null hypothesis. Psychological Methods. 2002;**7**(3):356-369

[8] Benjamini Y, Yekutieli D. False discovery rate controlling confidence intervals for selected parameters. Journal of American Statistical Association. 2005:71-80

[9] Bansal NK, Miescke KJ. A Bayesian decision theoretic approach to directional multiple hypotheses problems. Journal of Multivariate Analysis. 2013:205-215

[10] Bansal NK, Jiang H, Pradeep P. A Bayesian methodology for detecting targeted genes under two related experiments. Statistics in Medicine. 2015;**34**(25):3362-3375

[11] Efron B. Correlation and large-scale simultaneous significance testing. Journal of the American Statistical Association. 2007:93-103

[12] Bansal NK, Hamedani GG, Maadooliat M. Testing multiple hypotheses with skewed alternatives. Biometrics. 2016;**72**(2):494-502

[13] Azzalini A. A class of distributions which includes the normal ones. Scandinavian Journal of Statistics. 1985;**12**(2):171-178

[14] van't Wout AB, Lehrman GK, Mikheeva SA, O'Keeffe GC, Katze MG, Bumgarner RE, Mullins JI. Cellular gene expression upon human immunodeficiency virus type 1 infection of CD4+-T-cell lines. Journal of Virology. 2003;**77**(2):1392-1402

IntechOpen