Supplementary Material for "GAP: A General Framework for Information Pooling in Two-Sample Sparse Inference"

Yin Xia[1], T. Tony Cai[2], and Wenguang Sun[3]

This supplement contains the derivation of power formulas (Section A) and the proofs of all theoretical results in the main text (Section B).

A The power formulas for GAP and BH

In this section, we derive explicit formulas to characterize the power gain of GAP over BH in an idealized setting, where several simplifications were made including (i) the test statistics are independent; (ii) the conditional distributions of T_i are not affected by S_i; and (iii) the conditional proportions are known (in practice they can be estimated consistently under independence).

[1]Department of Statistics, School of Management, Fudan University. The research of Yin Xia was supported in part by NSFC Grants 11771094, 11690013 and "The Recruitment Program of Global Experts" Youth Project.

[2]Department of Statistics, The Wharton School, University of Pennsylvania. The research of Tony Cai was supported in part by NSF grant DMS-1712735 and NIH grants R01-GM129781 and R01-GM123056.

[3]Department of Data Sciences and Operations, University of Southern California. The research of Wenguang Sun was supported in part by NSF grant DMS-CAREER 1255406.

A.1 Theoretical setup

A key feature in the two-sample testing problem is that there exists an auxiliary statistic S_i that encodes the sparsity information. This motivates us to define the conditional proportion $\pi(s) = P(\theta_i = 1|S_i = s)$ to reflect the heterogeneity across different testing units. We further assume that $f_0(t)$ and $f_1(t)$ remain the same for all s:

$$f(t|S_i = s) = \{(1 - \pi(s)\}f_0(t) + \pi(s)f_1(t). \tag{A.1}$$

Remark 1 Condition A.1 implies that S_i only affects the distribution of T_i via the conditional proportion $\pi(s)$. We stress that (A.1) is used to simplify our analysis and that GAP only requires that T_i and S_i are conditionally independent *under the null* (Condition A3 in Section 3 of the main text), which is less stringent than (A.1). By contrast, (A.1) essentially requires that T_i and S_i are conditionally independent *under both the null and alternative*. Moreover, Condition A3 holds asymptotically by construction whereas (A.1) may be violated in some applications.

Suppose that the number of groups is K. GAP searches "optimal" cutoffs $\lambda_1, \cdots, \lambda_{K-1}$ to maximize the number of rejections subject to the FDR constraint. Let $\lambda_0 = -\infty$ and $\lambda_K = \infty$. The density functions for separate groups are thus given by:

$E(X) = \frac{X_1 + \cdots + X_n}{n} \to \int x f(x) dx$

$$f_k(t) = (1 - \pi_k)f_0(t) + \pi_k f_1(t), \quad k = 1, \cdots, K,$$

where $\pi_k = \int_{\lambda_{k-1}}^{\lambda^k} \pi(s)dF(s)$ and $F(s)$ is the CDF of S_i. When p-values are used in analysis, the corresponding CDFs for separate groups are given by

$\int_{\lambda_{k-1}}^{\lambda_k} dF(s)$

$$H_k(x) = (1 - \pi_k)x + \pi_k H_{k1}(x), \quad k = 1, \cdots, K.$$

Combining all groups (or ignoring S_i), let $H(x) = (1-\pi)x+\pi H_1(x)$ denote the marginal distribution of the p-value, where $\pi = \int \pi(s)dF(s)$ is the (overall) non-null proportion. The

定义有问题? $\pi = \pi_1 + \cdots + \pi_K$.

π变换成了 $w_k\pi_k$

$w_i = \sum_{k=1}^{K} r\pi_k/(1-\lambda_k)\ I$

asymptotic p-value threshold of the BH procedure (as $m \to \infty$) is given by

$$t_{BH} = Q^{-1}(\alpha) := \sup\{x : Q(x) \leq \alpha\}, \tag{A.2}$$

where $Q(x) = \overset{(1-\pi)}{x}/H(x)$. With this asymptotic threshold, the power of the BH procedure can be calculated as

$$\Psi_{\text{BH}}(t_{BH}) = \overset{\pi}{H_1}(t_{BH}). \tag{A.3}$$

Consider the weights $\{w_k : k = 1, \cdots, K\}$ defined in Section 2.2 in the paper. Define the ratio of total number of tests to the total non-null versus null proportion ratios

$$r = \frac{m}{\sum_{i=1}^{m} \sum_{k=1}^{K} \pi_k/(1-\pi_k) I(i \in \mathcal{G}_k)}.$$

Then we have $w_k = (r\pi_k)/(1-\pi_k)$. This ensures that the total weights sum up to m.

Under independence, consistent estimators for the conditional proportions may be constructed [e.g. using the method in Jin and Cai (2007)]. Hence, to simplify the analysis, we use the true conditional proportions in place of the estimated proportions in later calculations. Let $\pi_{i,*} = \sum_{k=1}^{K} \pi_k I(i \in \mathcal{G}_k)$ and $w_{i,*} = \sum_{k=1}^{K} w_k I(i \in \mathcal{G}_k)$.

i在哪个组取决于节点、

Remark 2 The conditional proportions and corresponding weights depend on the grouping cutoffs $(\lambda_1, \cdots, \lambda_{K-1})$. The notations $\pi_{i,*}$ and $w_{i,*}$ should be understood as $\pi_{i,*}(\lambda_1, \cdots, \lambda_{K-1})$ and $w_{i,*}(\lambda_1, \cdots, \lambda_{K-1})$. We use the simplified notations $\pi_{i,*}$ and $w_{i,*}$ in later discussions when there is no confusion.

The limiting value of the FDR of GAP with threshold t can be derived as

相当于p值变成了 $w_k t$.

$$Q_{\text{GAP}}(t; \lambda_1, \cdots, \lambda_{K-1}) = \frac{r\pi t}{r\pi t + m^{-1} \sum_{i=1}^{m} \pi_{i,*} H_1(w_{i,*}t)}. \tag{A.4}$$

$\because \pi = \pi_1 + \cdots + \pi_K,\ r\pi = r(\pi_1 + \cdots + \pi_K) = \sum_{k=1}^{K} w_k(1-\pi_k).$

$m r\pi t = m \cdot \sum_{k=1}^{K} (1-\pi_k) w_k t$

H_1部分为: $\sum_{i=1}^{m}\sum_{k=1}^{K}(1-\pi_k) w_k t I(i \in G_k) = \sum_{i=1}^{m}\sum_{k=1}^{K} r\pi_k t I(i\in G_k)$

$= rt \sum \sum_k \pi_k I(i \in G_k)$

$rt\sum \pi_{i,*}$ $= rt\sum_{k=1}^{K} m_k\pi_k = rt\cdot m\pi$

with the corresponding power given by

正确拒绝个数

$$\Psi_{\text{GAP}}(t;\lambda_1,\cdots,\lambda_{K-1}) = (m\pi)^{-1}\sum_{i=1}^{m}\pi_{i,*}H_1(w_{i,*}t)$$

$m\pi$: non-null个数

The optimal grouping for a given threshold t is therefore determined by

$$(\tilde{\lambda}_1,\cdots,\tilde{\lambda}_{K-1}) = \text{argmax}_{\lambda_1,\cdots,\lambda_{K-1}}\left\{\Psi_{\text{GAP}}(t;\lambda_1,\cdots,\lambda_{K-1}) : Q_{\text{GAP}}(t)\le\alpha\right\}.$$

The corresponding conditional proportions and weights are denoted $\tilde{\pi}_i$ and $\tilde{w}_i$, respectively. Define the FDR and power of GAP with the optimal grouping as

$$\begin{aligned}\tilde{Q}_{\text{GAP}}(t) &= \frac{r\pi t}{r\pi t + m^{-1}\sum_{i=1}^{m}\tilde{\pi}_i H_1(\tilde{w}_i t)},\\ \tilde{\Psi}_{\text{GAP}}(t) &= m^{-1}\sum_{i=1}^{m}\tilde{\pi}_i H_1(\tilde{w}_i t).\end{aligned}$$

A.2 GAP dominates BH in power asymptotically

Assume that $t\to H_1(t)$ is concave and $x\to H_1(t/x)$ is convex for $x\ge\tilde{t}$, where $\tilde{t} = \pi(1-\pi)^{-1}\min\{\pi_i^{-1}(1-\pi_i), i=1,\ldots,m\}$. Let $t_{BH}^* = \frac{1-\pi}{r\pi}t_{BH}$ denote the adjusted BH threshold for the GAP procedure. Then we have

$$\textbf{Claim A.1}\quad \tilde{Q}_{\text{GAP}}(t_{BH}^*)\le\alpha,\quad \tilde{\Psi}_{\text{GAP}}(t_{BH}^*)\ge\Psi_{\text{BH}}(t_{BH}). \tag{A.5}$$

Therefore the BH procedure is uniformly dominated by the GAP procedure in the asymptotic setup considered in Section A.1 . The second result in Claim A.1 $\tilde{\Psi}_{\text{GAP}}(t_{\text{BH}}^*) - \Psi_{\text{BH}}(t_{\text{BH}})\ge 0$ can be proved using Jensen's inequality as done in Hu et al. (2012). Meanwhile, we conclude from the second result in Claim A.1 and Equation A.4 that

$$\tilde{Q}_{\text{GAP}}(t_{\text{BH}}^*)\le Q_{\text{BH}}(t_{\text{BH}}) = (1-\pi)\alpha.$$

Note that GAP utilizes BH in the pooling step, and the operation of BH implies that

$$t_{\text{GAP}} = \sup\{x : \tilde{Q}_{\text{GAP}}(x) \leq \alpha\} \geq t^*_{\text{BH}}.$$

We conclude that $\tilde{\Psi}_{\text{GAP}}(t_{\text{GAP}}) \geq \tilde{\Psi}_{\text{GAP}}(t^*_{\text{BH}}) \geq \Psi_{\text{BH}}(t_{\text{BH}})$, with the asymptotic difference in power given by

$$(m\pi)^{-1} \sum_{i=1}^{m} \tilde{\pi}_i H_1(\tilde{w}_i t_{\text{GAP}}) - H_1(t_{\text{BH}}). \tag{A.6}$$

A.3 Quantifying the gap under a concrete model.

Consider a two-point Gaussian mixture model. Assume that $\boldsymbol{Y}_d \sim N(\boldsymbol{\beta}_d, I)$, where $\beta_{i,1} = 0$ for $1 \leq i \leq m$, $\beta_{i,2} = \mu_0$ for $1 \leq i \leq m_1$, and $\beta_{i,2} = 0$ for $m_1 + 1 \leq i \leq m$.

We first derive the asymptotic p-value threshold for the BH procedure at FDR level α. The p-value CDF for T_i is given by

$$H(x) = (1-\pi)x + \pi \left\{ \Phi\left(-z_{x/2} + \mu_0/\sqrt{2}\right) + \Phi\left(-z_{x/2} - \mu_0/\sqrt{2}\right) \right\}.$$

Let $Q(x) = x/H(x)$. Then the BH threshold is $t_{BH} = Q^{-1}(\alpha)$, which can be solved numerically using the `uniroot` function in `R`. It is easy to see that the FDR level of the BH procedure is given by $(1-\pi)\alpha$.

The primary and auxiliary statistics are

$$T_i = \frac{1}{\sqrt{2}}(Y_{i,2} - Y_{i,1}), \quad S_i = \frac{1}{\sqrt{2}}(Y_{i,2} + Y_{i,1}). \tag{A.7}$$

For this special data structure, the conditional proportion can be easily computed as

$$\pi(s) = P(\theta_i = 1 | S_i = s) = \frac{\pi f_1(s)}{f(s)},$$

where $f(s)$ is the marginal density of S_i and $f_1(s)$ is the conditional density of S_i given

$\theta_i = 1$. Suppose the cutoff is λ, then the proportion of non-null cases in $\mathcal{G}_1$ is $\pi_1 = \pi\Phi\left(\lambda - \frac{\mu_0}{\sqrt{2}}\right)/\kappa_1$, where Φ is the CDF of a standard normal variable, and κ_1 is the expected proportion of tests that is contained in $\mathcal{G}_1$:

$$\kappa_1 = \int_{-\infty}^{\lambda} f(s)ds = (1-\pi)\Phi(\lambda) + \pi\Phi\left(\lambda - \frac{\mu_0}{\sqrt{2}}\right).$$

Correspondingly, we have the proportion of non-nulls in $\mathcal{G}_2$: $\pi_2 = \pi\left\{1 - \Phi\left(\lambda - \frac{\mu_0}{\sqrt{2}}\right)\right\}/\kappa_2$, with $\kappa_2 = 1 - \kappa_1$. Now we evaluate the power of the GAP procedure with λ as the cutoff for grouping and t^*_{BH} as the threshold for the weighted p-values:

$$\Psi_{\mathrm{GAP}}(t^*_{BH}; \lambda) = \Phi\left(\lambda - \frac{\mu_0}{\sqrt{2}}\right) H_1(w_1 t^*_{BH}) + \left\{1 - \Phi\left(\lambda - \frac{\mu_0}{\sqrt{2}}\right)\right\} H_1(w_2 t^*_{BH}),$$

where

$$t^*_{BH} = \frac{(1-\pi)}{r\pi} t_{BH}, \quad r = \left\{\sum_{l=1}^{2} \frac{\kappa_l \pi_l}{1-\pi_l}\right\}^{-1},$$

H_1 is the alternative CDF of the p-value, and w_1 and w_2 are the weights according to the definitions in the paper:

$$w_l = \left\{\sum_{l=1}^{2} \frac{\kappa_l \pi_l}{1-\pi_l}\right\}^{-1} \frac{\pi_l}{1-\pi_l}, \quad l = 1, 2.$$

The optimal $\tilde{\lambda}$ can be solved using R function `optimize`.

If we use in the GAP procedure (i) the grouping based on $\tilde{\lambda}$ and (ii) t^*_{BH} as the threshold for the weighted p-values, then the corresponding FDR level is given by

$$Q_{\mathrm{GAP}}(t^*_{BH}; \tilde{\lambda}) = \frac{(1-\pi)t_{BH}}{(1-\pi)t_{BH} + \pi\Psi_{\mathrm{GAP}}(t^*_{BH}; \tilde{\lambda})}$$

It is important to note that the actual GAP procedure has a larger threshold t_{GAP}, which is intractable even in this simple model. We have used t^*_{BH} as a conservative estimate of t_{GAP}. Comparing with Equation (A.6), the actual gap in power must be bounded below by the following difference

$$\Psi_{\mathrm{GAP}}(t^*_{BH};\tilde{\lambda}) - \Psi_{\mathrm{BH}}(t_{BH}).$$

We have illustrated the power functions of GAP vs. BH numerically for a range of μ_0 and π in Figure 1 in Section 3 of the main text.

Remark 3 *To simplify the discussion, we have used two groups in this simple illustration. In practice, we recommend using* 3 *or* 4 *groups, which would lead to even large power gains.*

B Proofs

B.1 Technical Lemmas

For $d = 1, 2$, let $W_{i,d} = \hat{r}_{i,d}/\hat{\sigma}^2_{\eta_{i,d}}$ and let $U_{i,d} = n_d^{-1}\sum_{k=1}^{n_d}\{\epsilon_{k,d}\eta_{k,i,d} - \mathbb{E}(\epsilon_{k,d}\eta_{k,i,d})\}$ and $\tilde{U}_{i,d} = \beta_{i,d} + U_{i,d}/\sigma^2_{\eta_{i,d}}$. The following lemma is essentially proved in Liu and Luo (2014).

Lemma 1 *Suppose that Conditions (C1), (C3), (4.14) and (4.15) hold. Then for any constant $M > 0$, there exists some $b_{m,n}$ satisfying $b_{m,n} = o\{(n_d \log m)^{-1/2}\}$, such that,*

$$\mathbb{P}(|W_{i,d} - \{\tilde{U}_{i,d} + (\tilde{\sigma}^2_{\epsilon_d}/\sigma^2_{\epsilon_d} + \tilde{\sigma}^2_{\eta_{i,d}}/\sigma^2_{\eta_{i,d}} - 2)\beta_{i,d}\}| \geq b_{m,n}) = O(m^{-M}),$$

where $\tilde{\sigma}^2_{\epsilon_d} = n_d^{-1}\sum_{k=1}^{n_d}(\epsilon_{k,d} - \bar{\epsilon}_{k,d})^2$ and $\tilde{\sigma}^2_{\eta_{i,d}} = n_d^{-1}\sum_{k=1}^{n_d}(\eta_{k,i,d} - \bar{\eta}_{k,i,d})^2$ with $\bar{\epsilon}_{k,d} = n_d^{-1}\sum_{k=1}^{n_d}\epsilon_{k,d}$ and $\bar{\eta}_{k,i,d} = n_d^{-1}\sum_{k=1}^{n_d}\eta_{k,i,d}$.

For $d = 1, 2$, let $U_{i,j,d} = {n_d}^{-1}\sum_{k=1}^{n_d}(\epsilon_{k,i,d}\epsilon_{k,j,d} - \mathbb{E}\epsilon_{k,i,d}\epsilon_{k,j,d})$, and define $\tilde{U}_{i,j,d} = (r_{i,j,d} - U_{i,j,d})/(r_{i,i,d}r_{j,j,d})$ for $1 \leq i < j \leq p$ and $\tilde{U}_{i,i,d} = (r_{i,i,d} + U_{i,i,d})/(r_{i,i,d}r_{i,i,d})$. The following Lemma is proved in Xia et al. (2015).

Lemma 2 *Under the regularity conditions in Xia et al. (2015) such that equations (4) and (5) in Xia et al. (2015) are satisfied with probability greater or equal than $1 - O(m^{-M})$ for any constant $M > 0$. Then for any constant $M > 0$, there exists some $b_{m,n}$ satisfying*

$b_{m,n} = o\{(n_d \log m)^{-1/2}\}$, *such that,*

$$\mathbb{P}(|\hat{r}_{i,j,d} - \{U_{i,j,d} + (\omega_{i,i,d}\hat{\sigma}_{i,i,d,\epsilon} + \omega_{j,j,d}\hat{\sigma}_{j,j,d,\epsilon} - 1)r_{i,j,d}\}| \geq b_{m,n}) = O(m^{-M}),$$

and

$$\mathbb{P}(|\hat{r}_{i,i,d}^{-1} - \tilde{U}_{i,i,d}| \geq b_{m,n}) = O(m^{-M}),$$

where $(\hat{\sigma}_{i,j,d,\epsilon}) = (1/n_d)\sum_{k=1}^{n_d}(\boldsymbol{\epsilon}_{k,d} - \bar{\boldsymbol{\epsilon}}_d)(\boldsymbol{\epsilon}_{k,d} - \bar{\boldsymbol{\epsilon}}_d)^{\intercal}$, $\boldsymbol{\epsilon}_{k,d} = (\epsilon_{k,1,d}, \ldots, \epsilon_{k,p,d})$ *and* $\bar{\boldsymbol{\epsilon}}_d = {n_d}^{-1}\sum_{k=1}^{n_d} \boldsymbol{\epsilon}_{k,d}$.

We next introduce a lemma which provides an equivalent procedure to the BH procedure, based on the z-values, so that the dependence among the tests can be more easily analyzed. Define $G(t) = 2(1 - \Phi(t))$ and let the z-value $Z_i = \Phi^{-1}(1 - p_i/2)$, $i = 1, \ldots, m$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

Algorithm 1 Equivalent BH Procedure

1: For given $0 \leq \alpha \leq 1$, calculate

$$\hat{t} = \inf\{t \geq 0 : \frac{mG(t)}{\max\{\sum_{i=1}^{m} I(Z_i \geq t), 1\}} \leq \alpha\}. \tag{B.8}$$

2: For $1 \leq i \leq m$, reject the null hypotheses for which $Z_i \geq \hat{t}$.

Lemma 3 *Algorithm 1 is equivalent to the BH procedure.*

Proof of Lemma 3: Recall that, the BH procedure is based on the p-values $p_1, p_2, \ldots, p_m$. To determine which null hypotheses are true and which are false, it first orders the p-values $p_{(1)} \leq \ldots \leq p_{(m)}$ and then rejects all the null hypotheses $H_{0,i}$ for which $p_i \leq p_{(\hat{k})}$, where

$$\hat{k} = \max\{1 \leq i \leq m : p_{(i)} \leq \alpha i/m\}. \tag{B.9}$$

If $\hat{k}$ does not exist, then no hypothesis is rejected. By defining $p_{(0)} = 0$, we have

$$\hat{k} = \max\{0 \leq i \leq m : mp_{(i)}/\max\{i, 1\} \leq \alpha\}.$$

$1 - \frac{G(t)}{2} = \Phi(t)$ Φ为标准正态的cdf. 为增函数.

$G(t) = P$

Let $t_{\hat{k}} = \Phi^{-1}(1 - p_{(\hat{k})}/2)$. Then we have $p_{(\hat{k})} = G(t_{\hat{k}})$. Thus we have

$$\frac{mG(t_{\hat{k}})}{\max\{\sum_{i=1}^m I(Z_i \geq t_{\hat{k}}), 1\}} = \frac{mp_{(\hat{k})}}{\max\{\hat{k}, 1\}} \leq \alpha.$$

Similarly, let $t_{\hat{k}+1} = \Phi^{-1}(1 - p_{(\hat{k}+1)}/2) < t_{\hat{k}}$, and we have $p_{(\hat{k}+1)} = G(t_{\hat{k}+1})$. Then it can be shown that

$$\frac{mG(t_{\hat{k}+1})}{\max\{\sum_{i=1}^m I(Z_i \geq t_{\hat{k}+1}), 1\}} \geq \min_{j=1,\ldots,m-\hat{k}} \frac{mp_{(\hat{k}+j)}}{\max\{\hat{k}+j, 1\}} > \alpha.$$

Based on the definition of $\hat{t}$ in (B.8) of Algorithm 1, there exists a sequence $\{t_l\}$ with $t_l \geq \hat{t}$ and $t_l \to \hat{t}$, such that

$$\frac{mG(t_l)}{\max\{\sum_{i=1}^m I(Z_i \geq t_l), 1\}} \leq \alpha.$$

Thus we have $\sum_{i=1}^m I(Z_i \geq t_l) \leq \sum_{i=1}^m I(Z_i \geq \hat{t})$, which implies

$$\frac{mG(t_l)}{\max\{\sum_{i=1}^m I(Z_i \geq \hat{t}), 1\}} \leq \alpha.$$

G 右连续.

Let $t_l \to \hat{t}$, we have

$$\frac{mG(\hat{t})}{\max\{\sum_{i=1}^m I(Z_i \geq \hat{t}), 1\}} \leq \alpha.$$

Thus we have $t_{\hat{k}+1} < \hat{t} \leq t_{\hat{k}}$, where $\hat{t}$ is defined in (B.8). Hence, Algorithm 1 rejects the null hypotheses for which $Z_i \geq t_{\hat{k}}$ and does not reject other nulls, and is thus equivalent to the BH procedure. ∎

Lemma 4 *Let $p_1, \ldots, p_m$ be the p-values for testing m null hypotheses, $H_{0,1}, H_{0,2}, \ldots, H_{0,m}$. Define $p_i^w = \min\{p_i/q, 1\}$ for $q > 0$, and let z_i^w be the corresponding z-values, for $i = 1, \ldots, m$. Then, if $0 < q < 1$, we have the density of z_i^w equal to*

$$g(z_i^w) = \begin{cases} q\phi(z_i^w), & \text{if } z_i^w \neq 0, \\ \infty, & \text{if } z_i^w = 0, \end{cases}$$

$$A = \begin{pmatrix}1 & 1\\ & 1\end{pmatrix} - \begin{pmatrix}1 & \\ & 1\end{pmatrix}$$

with $\int_{-\infty}^{\infty} g(z_i^w)dz_i^w = 1$, *and if* $q \geq 1$,

$$\begin{pmatrix}0 & 1\\ & 0\end{pmatrix}\begin{pmatrix}0\\ 1\end{pmatrix} = \begin{pmatrix}1\\ 0\end{pmatrix}$$

$$g(z_i^w) = \begin{cases} q\phi(z_i^w), & \textit{if } |z_i^w| > \Phi^{-1}(1 - 1/(2q)), \\ 0, & \textit{otherwise}, \end{cases}$$

$$\begin{pmatrix}0 & 1\\ & 0\end{pmatrix}$$

where $\phi(\cdot)$ *is the standard normal probability density function.*

Proof of Lemma 4: Because $p_i \sim \text{Unif}(0, 1)$, if $0 < q < 1$, we have the density of p_i^w satisfies

$$f(p_i^w) = \begin{cases} q, & \text{if } 0 \leq p_i^w < 1, \\ \infty, & \text{if } p_i^w = 1, \end{cases}$$

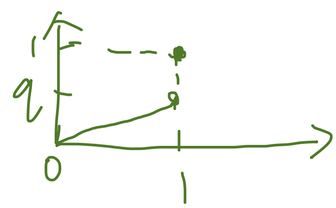
$P(p_i^w \leq x) = P(\min(\frac{p_i}{q}, 1) \leq x)$

$= 1 - P(\frac{p_i}{q} > x,\ 1 > x)$

$= 1 - P(p_i > qx,\ 1 > x)$

① $x < 1$, $1 - (1 - qx) = qx$

with $\int_{-\infty}^{\infty} f(p_i^w)dp_i^w = 1$. If $q \geq 1$, we have



$$f(p_i^w) = \begin{cases} q, & \text{if } 0 \leq p_i^w < 1/q, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, based on the definition of z_i^w, Lemma 4 is proved. ∎

B.2 Proof of Proposition 2

Define

$$V_i = \frac{U_{i,1}/\sigma_{\eta_{i,1}}^2 - U_{i,2}/\sigma_{\eta_{i,2}}^2}{(\sigma_{w,i,1}^2 + \sigma_{w,i,2}^2)^{1/2}},$$

where $\sigma_{w,i,d}^2 = \mathsf{Var}(\tilde{U}_{i,d}) = \mathsf{Var}(\epsilon_{k,d}\eta_{k,i,d}/\sigma_{\eta_{i,d}}^2)/n_d = (\sigma_{\epsilon_d}^2/\sigma_{\eta_{i,d}}^2 + \beta_{i,d}^2)/n_d$, for $d = 1, 2$. By Lemma 2 in Xia et al. (2015), under conditions (4.14) and (4.15), we have

$$\mathbb{P}\left(|\hat{\sigma}_{\epsilon_d}^2 - \sigma_{\epsilon_d}^2| \geq C\sqrt{\frac{\log m}{n_d}}\right) = O(m^{-M}),$$

and

$$\mathbb{P}\left(\max_i |\hat{\sigma}^2_{\eta_{i,d}} - \sigma^2_{\eta_{i,d}}| \geq C\sqrt{\frac{\log m}{n_d}}\right) = O(m^{-M}).$$

Thus we have

$$\mathbb{P}\left(\max_i |\hat{\sigma}^2_{w,i,d} - \sigma^2_{w,i,d}| \geq C\sqrt{\frac{\log m}{n_d}}\right) = O(m^{-M}).$$

By Lemma 1 and conditions (4.14) and (4.15), we have

$$\mathbb{P}\left(\left|T_i - \left\{V_i + \frac{f_i}{(\sigma^2_{w,i,1} + \sigma^2_{w,i,2})^{1/2}}\right\}\right| \geq Cb_m\right) = O(m^{-M}),$$

for some constant $C > 0$, where $b_m = o\{(\log m)^{-1/2}\}$. Thus under (C1) and (C2), Proposition 2 is proved by central limit theorem. ∎

B.3 Proof of Propositions 1 and 3

We prove Proposition 3 in this section, and Proposition 1 can be shown similarly. For Proposition 3, it is enough to show that

$$\mathbb{P}(|T_i - \frac{b_i}{(\sigma^2_{w,i,1} + \sigma^2_{w,i,2})^{1/2}}| \geq t, |S_i| \geq \lambda) = (1 + o(1))G(t)\mathbb{P}(|N(0,1) + s_i| \geq \lambda) + O(m^{-M}),$$

uniformly for $0 \leq t \leq 4\sqrt{\log m}$, $0 \leq \lambda \leq 4\sqrt{\log m}$ and $i = 1, \ldots, m$. The second part then directly follows due to the fact that N is fixed. Note that $G(t + o((\log m)^{-1/2}))/G(t) = 1 + o(1)$ uniformly in $0 \leq t \leq c(\log m)^{1/2}$ for any constant c. By the proof of Proposition 2, it suffices to show that,

$$\mathbb{P}(|V_i| \geq t, |\tilde{S}_i| \geq \lambda) = (1 + o(1))G(t)\mathbb{P}(|N(0,1) + s_i| \geq \lambda) + O(m^{-M}),$$

where

$$\tilde{S}_i = \frac{\hat{r}_{i,1}/\hat{\sigma}^2_{\eta_{i,1}} + (\sigma^2_{w,i,1}/\sigma^2_{w,i,2})(\hat{r}_{i,2}/\hat{\sigma}^2_{\eta_{i,2}})}{\sqrt{\sigma^2_{w,i,1}(1 + \sigma^2_{w,i,1}/\sigma^2_{w,i,2})}}.$$

By Lemma 1, it is enough to show that

$$\mathbb{P}(|V_i| \geq t, |Q_i| \geq \lambda) = (1 + o(1))G(t)\mathbb{P}(|N(0,1)| \geq \lambda) + O(m^{-M}),$$

uniformly for $0 \leq t \leq 4\sqrt{\log m}$ and $0 \leq \lambda \leq 4\sqrt{\log m}$, where

$$Q_i = \frac{U_{i,1}/\sigma^2_{\eta_{i,1}} + (\sigma^2_{w,i,1}/\sigma^2_{w,i,2})(U_{i,2}/\sigma^2_{\eta_{i,2}})}{\sqrt{\sigma^2_{w,i,1}(1 + \sigma^2_{w,i,1}/\sigma^2_{w,i,2})}},$$

Note that V_i and Q_i are uncorrelated with each other.

Let $n_2/n_1 \leq K_1$ with $K_1 \geq 1$. Define $Z_{k,i} = (n_2/n_1)\{\epsilon_{k,1}\eta_{k,i,1} - \mathbb{E}(\epsilon_{k,1}\eta_{k,i,1})\}/\sigma^2_{\eta_{i,1}}$ for $1 \leq k \leq n_1$ and $Z_{k,i} = -\{\epsilon_{k,2}\eta_{k,i,2} - \mathbb{E}(\epsilon_{k,2}\eta_{k,i,2})\}/\sigma^2_{\eta_{i,2}}$ for $n_1 + 1 \leq k \leq n_2$. Thus we have

$$V_i = \frac{\sum_{k=1}^{n_1+n_2} Z_{k,i}}{(n_2^2\sigma^2_{w,i,1} + n_2^2\sigma^2_{w,i,2})^{1/2}}.$$

Without loss of generality, we assume $\sigma^2_{\epsilon_d} = \sigma^2_{\eta_{i,d}} = 1$. Define

$$\hat{V}_i = \frac{\sum_{k=1}^{n_1+n_2} \hat{Z}_{k,i}}{(n_2^2\sigma^2_{w,i,1} + n_2^2\sigma^2_{w,i,2})^{1/2}},$$

where $\hat{Z}_{k,i} = Z_{k,i}I(|Z_{k,i}| \leq \tau_n) - \mathbb{E}\{Z_{k,i}I(|Z_{k,i}| \leq \tau_n)\}$, and $\tau_n = (4K_1/K)(\log(m+n))^{1+\epsilon}$ for any sufficiently small $\epsilon > 0$. Note that, for any $M > 0$

$$\begin{aligned}
\max_{1\leq i\leq m} n^{-1/2} \sum_{k=1}^{n_1+n_2} \quad & \mathbb{E}[|Z_{k,i}|I\{|Z_{k,i}| \geq \tau_n\}] \\
& \leq Cn^{1/2} \max_{1\leq k\leq n_1+n_2} \max_{1\leq i\leq m} \mathbb{E}[|Z_{k,i}|I\{|Z_{k,i}| \geq \tau_n\}] \\
& \leq Cn^{1/2}(m+n)^{-M} \max_{1\leq k\leq n_1+n_2} \max_{1\leq i\leq m} \mathbb{E}[|Z_{k,i}| \exp\{(K/2)|Z_{k,i}|\}] \\
& \leq Cn^{1/2}(m+n)^{-M}.
\end{aligned}$$

Hence we have,

$$\mathbb{P}\Big\{\max_{1\le i\le m}|V_i-\hat{V}_i|\ge(\log m)^{-1}\Big\}\le\mathbb{P}\Big(\max_{1\le i\le m}\max_{1\le k\le n_1+n_2}|Z_{k,i}|\ge\tau_n\Big)=O(m^{-M}).$$

Similarly, define $F_{k,i}=(n_2/n_1)\{\epsilon_{k,1}\eta_{k,i,1}-\mathbb{E}(\epsilon_{k,1}\eta_{k,i,1})\}/\sigma^2_{\eta_{i,1}}$ for $1\le k\le n_1$ and $F_{k,i}=(\sigma^2_{w,i,1}/\sigma^2_{w,i,2})\{\epsilon_{k,2}\eta_{k,i,2}-\mathbb{E}(\epsilon_{k,2}\eta_{k,i,2})\}/\sigma^2_{\eta_{i,2}}$ for $n_1+1\le k\le n_2$. Then we have

$$Q_i=\frac{\sum_{k=1}^{n_1+n_2}F_{k,i}}{(n_2^2\sigma^2_{w,i,1}(1+\sigma^2_{w,i,1}/\sigma^2_{w,i,2})^{1/2}}.$$

Without loss of generality, we assume $\sigma^2_{w,i,1}=\sigma^2_{w,i,2}$. Define

$$\hat{Q}_i=\frac{\sum_{k=1}^{n_1+n_2}\hat{F}_{k,i}}{(n_2^2\sigma^2_{w,i,1}(1+\sigma^2_{w,i,1}/\sigma^2_{w,i,2})^{1/2}}.$$

where $\hat{F}_{k,i}=F_{k,i}I(|F_{k,i}|\le\tau_n)-\mathbb{E}\{F_{k,i}I(|F_{k,i}|\le\tau_n)\}$. Then we can similarly obtain that

$$\mathbb{P}\Big\{\max_{1\le i\le m}|Q_i-\hat{Q}_i|\ge(\log m)^{-1}\Big\}=O(m^{-M}).$$

Thus, it suffices it is to show that

$$\mathbb{P}(|\hat{V}_i|\ge t,|\hat{Q}_i|\ge\lambda)=(1+o(1))G(t)G(\lambda)+O(m^{-M}),\tag{B.10}$$

uniformly for $0\le t\le 4\sqrt{\log m}$ and $0\le\lambda\le 4\sqrt{\log m}$. Let

$$\boldsymbol{W}_k=\left\{\frac{\hat{Z}_{k,i}}{(n_2\sigma^2_{w,i,1}+n_2\sigma^2_{w,i,2})^{1/2}},\frac{\hat{F}_{k,i}}{(n_2\sigma^2_{w,i,1}(1+\sigma^2_{w,i,1}/\sigma^2_{w,i,2})^{1/2}}\right\}.$$

Then we have

$$\mathbb{P}(|\hat{V}_i|\ge t,|\hat{Q}_i|\ge\lambda)=\mathbb{P}(|n_2^{-1/2}\sum_{k=1}^{n_1+n_2}W_{k,1}|\ge t,|n_2^{-1/2}\sum_{k=1}^{n_1+n_2}W_{k,2}|\ge\lambda).$$

Then it follows from Theorem 1 in Zaïtsev (1987) that

$$\mathbb{P}(|n_2^{-1/2}\sum_{k=1}^{n_1+n_2} W_{k,1}| \geq t, |n_2^{-1/2}\sum_{k=1}^{n_1+n_2} W_{k,2}| \geq \lambda)$$
$$\leq \mathbb{P}(|N_1| \geq t - \epsilon_n(\log m)^{-1/2}, |N_2| \geq \lambda - \epsilon_n(\log m)^{-1/2}) + c_1 \exp\left\{-\frac{n^{1/2}\epsilon_n}{c_2\tau_n(\log m)^{1/2}}\right\},$$

where $c_1 > 0$ and $c_2 > 0$ are constants, $\epsilon_n \to 0$ which will be specified later and $\boldsymbol{N} = (N_1, N_2)$ is a normal random vector with $\mathbb{E}(\boldsymbol{N}) = 0$ and $\mathsf{Cov}(N_1, N_2) = 0$. Because $\log m = o(n^{1/C})$ for some $C > 5$, we can let $\epsilon_n \to 0$ sufficiently slowly that, for any large $M > 0$

$$c_1 \exp\left\{-\frac{n^{1/2}\epsilon_n}{c_2\tau_n(\log m)^{1/2}}\right\} = O(m^{-M}).$$

Thus, we have

$$\mathbb{P}(|\hat{V}_i| \geq t, |\hat{Q}_i| \geq \lambda) \leq \mathbb{P}(|N_1| \geq t - \epsilon_n(\log m)^{-1/2}, |N_2| \geq \lambda - \epsilon_n(\log m)^{-1/2}) + O(m^{-M}).$$

Similarly, using Theorem 1 in Zaïtsev (1987) again, we have

$$\mathbb{P}(|\hat{V}_i| \geq t, |\hat{Q}_i| \geq \lambda) \geq \mathbb{P}(|N_1| \geq t + \epsilon_n(\log m)^{-1/2}, |N_2| \geq \lambda + \epsilon_n(\log m)^{-1/2}) - O(m^{-M}).$$

Thus (B.10) is proved, and thus Proposition 3 follows. ∎

B.4 Proof of Theorem 1

Let $z_i^w = \Phi^{-1}(1 - p_i^w/2)$, for $i = 1, \ldots, m$. Let m_0 be the total number of true nulls and $m_{01}, \ldots, m_{0K}$ be the number of true nulls for each group. Theorem 1 in Genovese et al. (2006) states that, if the sum of the weights for independent p-values is equal to the total number of hypotheses, the FDR can be controlled by applying the BH procedure. Since

lemma3 Algorithm 1 is equivalent to the BH procedure, under the asymptotic independency in (A3), by Theorem 1 in Genovese et al. (2006), we shall first show that it is enough to prove

$FDP_{GAP}(t) = \frac{\sum_{i\in\mathcal{H}_0} I(z_i^w \ge t)}{\max\{\sum_{i=1}^m I(z_i^w \ge t), 1\}}$, 分子 $= \sum_{i\in\mathcal{H}_0}\sum_{l=1}^{K} I(z_i^w \ge t, \lambda_{l-1} < S_i < \lambda_l)$

不相关，与分组相关的形式.

$FDR = E(FDP)$ $mFDR = FDR + o(m^{-\frac{1}{2}})$

用分组的tail概率逼近分子

that $\sum_{l=1,\ldots,K}\sum_{i\in\mathcal{H}_0} I(\lambda_{l-1} < S_i < \lambda_l, z_i^w \ge t)$ is close to $\sum_{l=1,\ldots,K} m_{0l}\mathbb{P}(z_i^w \ge t, i \in \mathcal{G}_l)$, for each group $\mathcal{G}_l$, $l = 1, \ldots, K$. That is, the method by using S_i as grouping statistics performs asymptotically the same as the case when the group information is known. We will start with the independent case, and then show the dependent case. To prove the dependent case, we will show that, within each group, the weighted z-values have the same dependence structure as the original z-values, up to a constant. Finally, we divide the null sets into small subsets and show that $\sum_{l=1,\ldots,K}\sum_{i\in\mathcal{H}_0} I(\lambda_{l-1} < S_i < \lambda_l, z_i^w \ge t)$ is close to $\sum_{l=1,\ldots,K}\sum_{i\in\mathcal{H}_0}\mathbb{P}(\lambda_{l-1} < S_i < \lambda_l, z_i^w \ge t)$.

此时等价方法用的是w(act) 无分组tail逼近

Step 1: Let $t_m = (2\log m - 4\log\log m)^{1/2}$. We show below that, based on the condition on $\mathcal{S}_\rho$, the $\hat{t}$ in Algorithm 1 is attained in the range $[0, t_m]$. This range is essential for showing the FDP control in equation (B.21). By the conditions of Theorem 1, we have

$S_\rho \subseteq \mathcal{H}_1$ c是某个常数, c取小, 全为1, $|\mathcal{H}_1|$, c取大, 为0. 故上式 $\in [0, |\mathcal{H}_1|]$

$$\sum_{i\in\mathcal{H}_1} I\{|T_i| \ge (c\log m)^{1/2+\rho/4}\} \ge \{1/(\pi^{1/2}\alpha) + \delta\}(\log m)^{3/2}$$

$\ge \sum_{i\in S_\rho}$

with probability going to one, for some constant $c > 0$. Recall that $\hat{\pi}_l = (\epsilon \vee \hat{\pi}_l^o) \wedge (1-\epsilon)$ with $\epsilon > m^{-C}$ for some constant $C > 0$ and that $w_l = \left\{\sum_{l=1}^K \frac{m_l\hat{\pi}_l}{1-\hat{\pi}_l}\right\}^{-1} \frac{m\hat{\pi}_l}{(1-\hat{\pi}_l)}$, $1 \le l \le K$. $\sum w_i = m$. !!
Then there exist constants $C > 0$ such that $w_l > m^{-C}$. Thus, for those indices $i \in \mathcal{H}_1$ such that $I\{|T_i| \ge (c\log m)^{1/2+\rho/4}\}$, we have

$m^{-\frac{C}{2}}$

渐近正态性 渐近正态性

$$p_i^w = p_i/w_l \le (1 - \Phi((c\log m)^{1/2+\rho/4}))/w_l = o(m^{-M}),$$

for any constant $M > 0$. Thus we have

左边前面的p值调整后很小. 之后

$$\sum_{1\le i\le m} I\{z_i^w \ge (2\log m)^{1/2}\} \ge \{1/(\pi^{1/2}\alpha) + \delta\}(\log m)^{3/2},$$

with probability going to one. Hence, with probability tending to one, we have

$$\frac{2m}{\sum_{1\le i\le m} I\{z_i^w \ge (2\log m)^{1/2}\}} \le 2m\{1/(\pi^{1/2}\alpha) + \delta\}^{-1}(\log m)^{-3/2}.$$

Because $1-\Phi(t_m) \sim 1/\{(2\pi)^{1/2}t_m\}\exp(-t_m^2/2)$, by Lemma 3, it suffices to show that, uniformly in $0 \le t \le t_m$ and $-4\sqrt{\log m} \le \lambda_1 < \cdots < \lambda_{K-1} \le 4\sqrt{\log m}$, there exists a constant $0 < c \le 1$, such that

要证.

$$\left| \frac{\sum_{l=1,\ldots,K}\sum_{i\in\mathcal{H}_0} I(\lambda_{l-1} < S_i < \lambda_l, z_i^w \ge t) - cm_0G(t)}{cm_0G(t)} \right| \to 0, \qquad \text{(B.11)}$$

in probability, where $G(t) = 2(1-\Phi(t))$.

Step 2: We shall show below that, by the asymptotic independency between T_i and S_i as described in (A3), it suffices to prove that ① $\sum_{l=1,\ldots,K}\sum_{i\in\mathcal{H}_0} I(\lambda_{l-1} < S_i < \lambda_l, z_i^w \ge t)$ is close to $\sum_{l=1,\ldots,K} m_{0l}\mathbb{P}(z_i^w \ge t, i \in \mathcal{G}_l)$, for each group $\mathcal{G}_l$, $l = 1,\ldots,K$. By Assumption (A3), we have that, uniformly in $0 \le t \le t_m$ and $-4\sqrt{\log m} \le \lambda_1 < \cdots < \lambda_{K-1} \le 4\sqrt{\log m}$, for each $l = 1,\ldots,K$,

在证 B(11)

$$\left| \frac{\sum_{i\in\mathcal{H}_0} \mathbb{P}(\lambda_{l-1} < S_i < \lambda_l, z_i^w \ge t) - m_{0l}\mathbb{P}(z_i^w \ge t, i \in \mathcal{G}_l)}{m_{0l}\mathbb{P}(z_i^w \ge t, i \in \mathcal{G}_l)} \right| \to 0.$$

Thus we have

$$\left| \frac{\sum_{l=1,\ldots,K}\sum_{i\in\mathcal{H}_0} \mathbb{P}(\lambda_{l-1} < S_i < \lambda_l, z_i^w \ge t) - \sum_{l=1,\ldots,K} m_{0l}\mathbb{P}(z_i^w \ge t, i \in \mathcal{G}_l)}{\sum_{l=1,\ldots,K} m_{0l}\mathbb{P}(z_i^w \ge t, i \in \mathcal{G}_l)} \right| \to 0. \qquad \text{(B.12)}$$

This shows that, by using S_i as grouping statistics, the method performs asymptotically the same as the case where the group information is known. We shall show below that ② $\sum_{l=1,\ldots,K} m_{0l}\mathbb{P}(z_i^w \ge t, i \in \mathcal{G}_l)$ is close to $cm_0G(t)$, and thus by (B.12), $\sum_{l=1,\ldots,K}\sum_{i\in\mathcal{H}_0} \mathbb{P}(\lambda_{l-1} < S_i < \lambda_l, z_i^w \ge t)$ is close to $cm_0G(t)$. Hence it remains to prove that

$$\left| \frac{\sum_{l=1,\ldots,K}\sum_{i\in\mathcal{H}_0} I(\lambda_{l-1} < S_i < \lambda_l, z_i^w \ge t)}{\sum_{l=1,\ldots,K}\sum_{i\in\mathcal{H}_0} \mathbb{P}(\lambda_{l-1} < S_i < \lambda_l, z_i^w \ge t)} - 1 \right| \to 0$$

in probability.

Step 2.1: We shall first show that $\sum_{l=1,\ldots,K} m_{0l}\mathbb{P}(z_i^w \ge t, i \in \mathcal{G}_l)$ is close to $cm_0G(t)$. This can be done by first considering the case when z_i^w are independent; the argument is then applied to the dependent case in Step 2.2.

According to Theorem 1 in Genovese et al. (2006), we have that, with the known group information and assume the original p-values of the null hypotheses are uniformly distributed, the procedure by applying BH procedure on the weighted p-values controls the FDR at level $\alpha m_0/m$. That is, if

$$k = \max\{i : p^w_{(i)} \leq i\alpha/m\},$$

and we reject all k hypotheses associated with $p^w_{(1)}, \ldots, p^w_{(k)}$, then we have

$$\mathbb{E}\left(\frac{\sum_{i\in\mathcal{H}_0} I(p_i^w \leq p^w_{(k)})}{\max\{\sum_{1\leq i\leq m} I(p_i^w \leq p^w_{(k)}), 1\}}\right) \leq \alpha m_0/m.$$

已知的

By the definition of z_i^w, it is equivalent to find

$$\hat{t} = \inf\{t \geq 0, \frac{2m(1-\Phi(t))}{\max\{\sum_{1\leq i\leq m} I(z_i^w \geq t), 1\}} \leq \alpha\}, \tag{B.13}$$

and reject all hypotheses with $z_i^w \geq \hat{t}$. This yields that

$$\mathbb{E}\left(\frac{\sum_{i\in\mathcal{H}_0} I(z_i^w \geq \hat{t})}{\max\{\sum_{1\leq i\leq m} I(z_i^w \geq \hat{t}), 1\}}\right) \leq \alpha m_0/m.$$

The ideal choice of the thresholding value t^o in order to control the above FDR is that

(改)

$$t^o = \inf\{t \geq 0, \frac{\sum_{i\in\mathcal{H}_0} I(z_i^w \geq t)}{\max\{\sum_{1\leq i\leq m} I(z_i^w \geq t), 1\}} \leq \alpha\}.$$

It is easy to show that, under independence of z_i^w,

有 $m - m_{0l}$ 个 0. $= \sum_{i\in G_l} I(z_i^w \geq t, i \in G_l) \to m_{0l}P$(独立 $i \in G_l$)

$$\left|\frac{\sum_{l=1,\ldots,K}\sum_{i\in\mathcal{H}_0} I(z_i^w \geq t, i\in\mathcal{G}_l) - \sum_{l=1,\ldots,K} m_{0l}\mathbb{P}(z_i^w \geq t, i\in\mathcal{G}_l)}{\sum_{l=1,\ldots,K} m_{0l}\mathbb{P}(z_i^w \geq t, i\in\mathcal{G}_l)}\right| \to 0$$

全概率公式，按 G_l 分组逼近

in probability, a good estimate of t^o would be

$$\hat{t}^o = \inf\{t \geq 0, \frac{\sum_{l=1,\ldots,K} m_{0l}\mathbb{P}(z_i^w \geq t, i \in \mathcal{G}_l)}{\sum_{1\leq i\leq m} I(z_i^w \geq t, i \in \mathcal{G}_l)} \leq \alpha\}. \tag{B.14}$$

和(*)是渐近相等的，$t^o \sim \hat{t}^o$

By rejecting all hypotheses with $z_i^w \geq \hat{t}^o$, we have

$$\mathbb{E}\left(\frac{\sum_{i\in\mathcal{H}_0} I(z_i^w \geq \hat{t}^o)}{\max\{\sum_{1\leq i\leq m} I(z_i^w \geq \hat{t}^o), 1\}}\right) \to \alpha.$$

This shows that the procedure (B.13) by applying the normal tail approximation on all hypotheses, is more conservative than the procedure (B.14), which uses different tail probability of z_i^w for each individual group. Thus, for any grouping method with number of true nulls $m_{01}, \ldots, m_{0K}$, there exists a constant $0 < c \leq 1$, such that, uniformly in $0 \leq t \leq t_m$ and $-4\sqrt{\log m} \leq \lambda_1 < \cdots < \lambda_{K-1} \leq 4\sqrt{\log m}$,

$\sum_{l=1}^{K} \sum_{i\in H_0} I(z_i^w \geq t, i \in G_l)$ 与 $\sum_{i\in H_0} I(z^w \geq t)$ ⟺ $c m_0 G(t)$ ∵ z^w的分布为 $cG(t)$ 引理4, $q\in(0,1)$

$$\left|\frac{\sum_{l=1,\ldots,K} m_{0l}\mathbb{P}(z_i^w \geq t, i \in \mathcal{G}_l) - cm_0G(t)}{cm_0G(t)}\right| \to 0,$$

where $G(t) = 2(1 - \Phi(t))$.

Step 2.2: Based on the above result obtained through the independent case, we then show that it is enough to show that $\sum_{l=1,\ldots,K}\sum_{i\in\mathcal{H}_0} I(\lambda_{l-1} < S_i < \lambda_l, z_i^w \geq t)$ is close to $\sum_{l=1,\ldots,K}\sum_{i\in\mathcal{H}_0} \mathbb{P}(\lambda_{l-1} < S_i < \lambda_l, z_i^w \geq t)$. Under (A1), the p-values are asymptotically uniformly distributed under the null. Thus, by (B.12), there exists a constant $0 < c \leq 1$, such that, uniformly in $0 \leq t \leq t_m$ and $-4\sqrt{\log m} \leq \lambda_1 < \cdots < \lambda_{K-1} \leq 4\sqrt{\log m}$, we have

$$\left|\frac{\sum_{l=1,\ldots,K}\sum_{i\in\mathcal{H}_0} \mathbb{P}(\lambda_{l-1} < S_i < \lambda_l, z_i^w \geq t) - cm_0G(t)}{cm_0G(t)}\right| \to 0.$$

Hence, to prove the FDR control of Theorem 1, it suffices to show that, under the conditions of Theorem 1, that is, when $\{Z_i, i = 1, \ldots, m\}$, before applying weighting, are weakly dependent with each other,

$$\left|\frac{\sum_{l=1,\ldots,K}\sum_{i\in\mathcal{H}_0} I(\lambda_{l-1}<S_i<\lambda_l, z_i^w\geq t)}{\sum_{l=1,\ldots,K}\sum_{i\in\mathcal{H}_0} \mathbb{P}(\lambda_{l-1}<S_i<\lambda_l, z_i^w\geq t)}-1\right|\to 0, \tag{B.15}$$

in probability, uniformly in $0\leq t\leq t_m$ and $-4\sqrt{\log m}\leq\lambda_1<\cdots<\lambda_{K-1}\leq 4\sqrt{\log m}$. Let $\tilde{\mathcal{H}}_0$ be any subset of $\mathcal{H}_0$ such that $\tilde{\mathcal{H}}_0=\mathcal{H}_0\backslash A_\tau$, with any set A_τ satisfying $|A_\tau\cap\mathcal{H}_0|=o(m^\nu)$ for any $\nu>0$. Let $\tilde{m}_{0l}=|\tilde{\mathcal{H}}_0\cap\mathcal{G}_l|$. By the proof of Theorem 4 in Xia et al. (2018), it suffices to show that

$$\left|\frac{\sum_{l=1,\ldots,K}\sum_{i\in\tilde{\mathcal{H}}_0}\{I(\lambda_{l-1}<S_i<\lambda_l, z_i^w\geq t)-\mathbb{P}(\lambda_{l-1}<S_i<\lambda_l, z_i^w\geq t)\}}{\sum_{l=1,\ldots,K}\sum_{i\in\tilde{\mathcal{H}}_0} \mathbb{P}(\lambda_{l-1}<S_i<\lambda_l, z_i^w\geq t)}\right|\to 0, \tag{B.16}$$

in probability, uniformly in $0\leq t\leq t_m$ and $-4\sqrt{\log m}\leq\lambda_1<\cdots<\lambda_{K-1}\leq 4\sqrt{\log m}$.

Step 3: To show equation (B.16), we further develop it by the following steps.

Step 3.1: We first work on the truncation of the statistics, so that it is close to the original statistics and at the mean time the normal approximation can be applied. Define

$$V_i=\frac{\sum_{k=1}^n Z_{k,i}}{\mathsf{Var}(\sum_{k=1}^n Z_{k,i})^{1/2}},$$

and define

$$\hat{V}_i=\frac{\sum_{k=1}^{n_1+n_2}\hat{Z}_{k,i}}{\mathsf{Var}(\sum_{k=1}^n Z_{k,i})^{1/2}},$$

零均值

where $\hat{Z}_{k,i}=Z_{k,i}I(|Z_{k,i}|\leq\tau_n)-\mathbb{E}\{Z_{k,i}I(|Z_{k,i}|\leq\tau_n)\}$, and τ_n can be chosen such that, under the conditions of Theorem 1,

$$\max_{i\in\tilde{\mathcal{H}}_0}|Z_i-|\hat{V}_i||=o_{\mathbb{P}}\{(\log m)^{-1/2}\},$$

similarly as shown in the proof of Proposition 3.

Step 3.2: We show next that, within each group, the weighted z-values share the same correlation structure as the original z-values. For the simplicity of notation, we let $V_i=\hat{V}_i$.

Define $V_i^w = \Phi^{-1}(1-(1-\Phi(|V_i|))/w_l)$, for $i \in \mathcal{G}_l$, $l = 1, \ldots, K$. Due to the fact that $\epsilon < \hat{\pi}_l < 1-\epsilon$, we have,

$$m^{-C} < w_l < m.$$

Thus we have $V_i^w = O_{\mathbb{P}}((\log m)^{1/2})$. Because $G(t + o((\log m)^{-1/2}))/G(t) = 1 + o(1)$ uniformly in $0 \le t \le c(\log m)^{1/2}$ for any constant c, we have

$$\max_{i \in \tilde{\mathcal{H}}_0} |z_i^w - |V_i^w|| = o_{\mathbb{P}}\{(\log m)^{-1/2}\}.$$

Thus, by the proofs of Propositions 2 and 3, it suffices to show that

$$\left| \frac{\sum_{l=1,\ldots,K} \sum_{i \in \tilde{\mathcal{H}}_0} \{I(\lambda_{l-1} < S_i < \lambda_l, |V_i^w| \ge t) - \mathbb{P}(\lambda_{l-1} < S_i < \lambda_l, |V_i^w| \ge t)\}}{\sum_{l=1,\ldots,K} \sum_{i \in \tilde{\mathcal{H}}_0} \mathbb{P}(\lambda_{l-1} < S_i < \lambda_l, |V_i^w| \ge t)} \right| \to 0, \tag{B.17}$$

in probability, uniformly in $0 \le t \le t_m$ and $-4\sqrt{\log m} \le \lambda_1 < \cdots < \lambda_{K-1} \le 4\sqrt{\log m}$. Define $\mathbf{\Sigma}_{V_l}$ be the covariance matrix of $\{V_i, i \in \tilde{\mathcal{H}}_0 \cap \mathcal{G}_l\}$, and $\mathbf{\Sigma}_{V_l^w}$ be the covariance matrix of $\{V_i^w, i \in \tilde{\mathcal{H}}_0 \cap \mathcal{G}_l\}$. By Lemma 4, we have that, $\mathbf{\Sigma}_{V_l^w} = D_l \mathbf{\Sigma}_{V_l} D_l$, where $D_l = diag(d_1, \ldots, d_{\tilde{m}_{0l}})$ is a diagonal matrix, with $0 < d_i < \infty$. Thus, $\{V_i^w, i \in \tilde{\mathcal{H}}_0 \cap \mathcal{G}_l\}$ has the same correlation structure as $\{V_i, i \in \tilde{\mathcal{H}}_0 \cap \mathcal{G}_l\}$, for each $l = 1, \ldots, K$.

Step 3.3: Finally, we divide pairs of the null sets into small subsets: the pairs share same indices $\tilde{\mathcal{H}}_{01}$, the set of highly correlated pairs $\tilde{\mathcal{H}}_{02}$ and the set of weakly correlated pairs $\tilde{\mathcal{H}}_{03}$. We shall show that the first two subsets are negligible, while $\tilde{\mathcal{H}}_{03}$ performs the dominant role, see (B.21).

Let $0 \le t_0 < t_1 < \cdots < t_b = t_m$ such that $t_\iota - t_{\iota-1} = v_m$ for $1 \le \iota \le b-1$ and $t_b - t_{b-1} \le v_m$, where $v_m = 1/\sqrt{\log m (\log_4 m)}$. Thus we have $b \sim t_m/v_m$. Let $\Psi_l(t) = \mathbb{P}(|V_i^w| \ge t, i \in \mathcal{G}_l)$. For any t such that $t_{\iota-1} \le t \le t_\iota$, due to the fact that $G(t + o((\log m)^{-1/2}))/G(t) = 1 + o(1)$ uniformly in $0 \le t \le c(\log m)^{1/2}$ for any constant c, by Lemma 4 and (B.12), we have

$$\frac{\sum_{i\in\tilde{\mathcal{H}}_0} I(\lambda_{l-1} < S_i < \lambda_l, |V_i^w| \geq t_\iota)}{\tilde{m}_{0l}\Psi_l(t_\iota)} \frac{\Psi_l(t_\iota)}{\Psi_l(t_{\iota-1})} \leq \frac{\sum_{i\in\tilde{\mathcal{H}}_0} I(\lambda_{l-1} < S_i < \lambda_l, |V_i^w| \geq t)}{\tilde{m}_{0l}\Psi_l(t)}$$
$$\leq \frac{\sum_{i\in\tilde{\mathcal{H}}_0} I(\lambda_{l-1} < S_i < \lambda_l, |V_i^w| \geq t_{\iota-1})}{\tilde{m}_{0l}\Psi_l(t_{\iota-1})} \frac{\Psi_l(t_{\iota-1})}{\Psi_l(t_\iota)}.$$

Thus it suffices to prove that, for each $l = 1, \ldots, K$,

$$\max_{0\leq\iota\leq b} \left| \frac{\sum_{i\in\tilde{\mathcal{H}}_0} I(\lambda_{l-1} < S_i < \lambda_l, |V_i^w| \geq t_\iota) - \tilde{m}_{0l}\Psi_l(t_\iota)}{\sum_{l=1,\ldots,K} \tilde{m}_{0l}\Psi_l(t_\iota)} \right| \to 0,$$

in probability. Thus, by assumption (A3), it suffices to show, for any $\epsilon > 0$,

$$\int_0^{t_m} \mathbb{P}\left\{ \left| \frac{\sum_{i\in\tilde{\mathcal{H}}_0} \{I(\lambda_{l-1} < S_i < \lambda_l, |V_i^w| \geq t) - \mathbb{P}(\lambda_{l-1} < S_i < \lambda_l, |V_i^w| \geq t)\}}{\sum_{l=1,\ldots,K} \tilde{m}_{0l}\Psi_l(t)} \right| \geq \epsilon \right\} dt$$
$$= o(v_m). \tag{B.18}$$

Note that

$$\mathbb{E}\left| \frac{\sum_{i\in\tilde{\mathcal{H}}_0} \{I(\lambda_{l-1} < S_i < \lambda_l, |V_i^w| \geq t) - \mathbb{P}(\lambda_{l-1} < S_i < \lambda_l, |V_i^w| \geq t)\}}{\tilde{m}_{0l}\Psi_l(t)} \right|^2$$
$$= \sum_{i,j\in\tilde{\mathcal{H}}_0} \Big\{ \mathbb{P}(\lambda_{l-1} < S_i < \lambda_l, |V_i^w| \geq t, \lambda_{l-1} < S_j < \lambda_l, |V_j^w| \geq t)$$
$$- \prod_{b=i,j} \mathbb{P}(\lambda_{l-1} < S_b < \lambda_l, |V_b^w| \geq t) \Big\} \Big/ \Big\{ \sum_{l=1,\ldots,K} \tilde{m}_{0l}\Psi_l(t) \Big\}^2.$$

Recall that $\Gamma_i(\gamma) = \{j : 1 \leq j \leq p, |\rho_{i,j}| \geq (\log m)^{-2-\gamma}\}$. Then we divides the indices $i, j \in \tilde{\mathcal{H}}_0$ into the following three subsets:

$$\begin{aligned}
\tilde{\mathcal{H}}_{01} &= \{i, j \in \tilde{\mathcal{H}}_0, i = j\}, \\
\tilde{\mathcal{H}}_{02} &= \{i, j \in \tilde{\mathcal{H}}_0, i \in \Gamma_j(\gamma), \text{ or } j \in \Gamma_i(\gamma)\}, \\
\tilde{\mathcal{H}}_{03} &= \{(i, j), i, j \in \tilde{\mathcal{H}}_0\} \setminus (\tilde{\mathcal{H}}_{01} \cup \tilde{\mathcal{H}}_{02}).
\end{aligned}$$

Then we have

$$
\begin{aligned}
&\sum_{i,j\in\tilde{\mathcal{H}}_{01}} \Big\{\mathbb{P}(\lambda_{l-1}<S_i<\lambda_l, |V_i^w|\geq t, \lambda_{l-1}<S_j<\lambda_l, |V_j^w|\geq t) \\
&\qquad -\mathbb{P}^2(\lambda_{l-1}<S_i<\lambda_l, |V_i^w|\geq t)\Big\}\Big/\Big\{\sum_{l=1,\ldots,K}\tilde{m}_{0l}\Psi_l(t)\Big\}^2 \\
&\quad \leq \frac{C}{\sum_{l=1,\ldots,K}\tilde{m}_{0l}\Psi_l(t)}.
\end{aligned}
\tag{B.19}
$$

By the condition that $\max_{1\leq i\leq m}|\Gamma_i(\gamma)|\leq C$ for some constant $C>0$, we have

$$
\begin{aligned}
&\sum_{i,j\in\tilde{\mathcal{H}}_{02}} \Big\{\mathbb{P}(\lambda_{l-1}<S_i<\lambda_l, |V_i^w|\geq t, \lambda_{l-1}<S_j<\lambda_l, |V_j^w|\geq t) \\
&\qquad -\prod_{b=i,j}\mathbb{P}(\lambda_{l-1}<S_b<\lambda_l, |V_b^w|\geq t)\Big\}\Big/\Big\{\sum_{l=1,\ldots,K}\tilde{m}_{0l}\Psi_l(t)\Big\}^2 \\
&\quad \leq \frac{C}{\sum_{l=1,\ldots,K}\tilde{m}_{0l}\Psi_l(t)}.
\end{aligned}
\tag{B.20}
$$

It remains to consider the subset $\tilde{\mathcal{H}}_{03}$, in which (S_i, V_i^w) and (S_j, V_j^w) are weakly correlated with each other. It is easy to check that,

$$
\begin{aligned}
&\max_{i,j\in\tilde{\mathcal{H}}_{03}} \mathbb{P}(\lambda_{l-1}<S_i<\lambda_l, |V_i^w|\geq t, \lambda_{l-1}<S_j<\lambda_l, |V_j^w|\geq t) \\
&\quad = (1+O\{(\log m)^{-1-\gamma}\})\prod_{b=i,j}\mathbb{P}(\lambda_{l-1}<S_b<\lambda_l, |V_b^w|\geq t).
\end{aligned}
$$

Thus we have

$$
\begin{aligned}
&\sum_{i,j\in\tilde{\mathcal{H}}_{03}} \Big\{\mathbb{P}(\lambda_{l-1}<S_i<\lambda_l, |V_i^w|\geq t, \lambda_{l-1}<S_j<\lambda_l, |V_j^w|\geq t) \\
&\qquad -\prod_{b=i,j}\mathbb{P}(\lambda_{l-1}<S_b<\lambda_l, |V_b^w|\geq t)\Big\}\Big/\Big\{\sum_{l=1,\ldots,K}\tilde{m}_{0l}\Psi_l(t)\Big\}^2 \\
&\quad = O\{(\log m)^{-1-\gamma}\}.
\end{aligned}
\tag{B.21}
$$

Combining (B.19), (B.20) and (B.21), we prove (B.18). So we have

$$\max_{0\le\iota\le b}\left|\frac{\sum_{l=1,\dots,K}\sum_{i\in\tilde{\mathcal{H}}_0} I(\lambda_{l-1}<S_i<\lambda_l, |V_i^w|\ge t_\iota)-\sum_{l=1,\dots,K}\tilde{m}_{0l}\Psi_l(t_\iota)}{\sum_{l=1,\dots,K}\tilde{m}_{0l}\Psi_l(t_\iota)}\right|\to 0,$$

Thus Theorem 1 is proved. ∎

B.5 Proof of Theorem 2

BH is the special one group scenario in the GAP procedure, in which case, we have $p_i^w = p_i$, for $i = 1,\dots,m$. Due to the fact that the proof of Theorem 1 also holds for the one group case, we have, for any $\epsilon > 0$,

$$\mathbb{P}\Big(\sup_{0\le t\le t_m}\Big|\frac{\sum_{i\in\mathcal{H}_0} I(|Z_i|\ge t)}{pG(t)}-1\Big|\ge\epsilon\Big)\to 0.$$

Let $\alpha' = \alpha m_0/m$. Based on the definition of BH procedure, we then have, for any $\epsilon > 0$,

$$\mathbb{P}\Big(\Big|\frac{FDP_{BH}}{\alpha'}-1\Big|\ge\epsilon\Big)\to 0.$$

By (B.18) in Theorem 1, we also have, for any $\epsilon > 0$,

$$\mathbb{P}\Big(\Big|\frac{FDP_{GAP}}{c\alpha'}-1\Big|\ge\epsilon\Big)\to 0,$$ Thm 1.

for some constant $0 < c \le 1$. Because GAP searches on all possible $\{\lambda_1,\dots,\lambda_{K-1}\}$ and choose the one with largest number of rejections, namely,

相差 $o_P(1)$

$FDP_{GAP} \to c\alpha'$

$$\sum_{i\in\mathcal{H}} I(p_i^w\le p^w_{(\hat{k}^w)}) \ge \sum_{i\in\mathcal{H}} I(p_i\le p_{(\hat{k})}), \qquad \text{(B.22)}$$

$FDP_{BH} \to \alpha'$

$FDP_{BH} \geqslant FDP_{GAP}$

$\frac{\sum_{H_0} I(p_i^w \le p_{(\hat{k})})}{\sum_H I(p_i^w \le p_{(\hat{k})})} \geqslant \frac{\sum_{H_0} I(p_i^w \le p_{(\hat{k})})}{\sum_H I(p_i^w \le p_{(\hat{k})})}$

thus, we have

$$\frac{\sum_{i\in\mathcal{H}_1} I(p_i^w\le p^w_{(\hat{k}^w)})}{|\mathcal{H}_1|} \ge \frac{\sum_{i\in\mathcal{H}_1} I(p_i\le p_{(\hat{k})})}{|\mathcal{H}_1|} + o_{\mathbb{P}}(1). \qquad \text{(B.23)}$$

$\Downarrow$

$\sum_{H_0} I(p_i \le p_{(\hat{k})}) \geqslant \sum_{H_0} I(p_i^w \le p_{(\hat{k})})$

有 $A \geqslant B$

要证. $(1-c\alpha')A \geqslant (1-\alpha)B$

$A - B - c\alpha' A + \alpha B \geqslant 0$ 是c的↓. c取1时.

$A - B - \alpha' A + \alpha B \geq 0.$

$(1-\alpha')A - (1-\alpha')B = (1-\alpha')(A-B) \geq 0.$

This yields that

$$\Psi_{GAP} \geq \Psi_{BH} + o(1).$$

B.6 Additional Propositions and Proofs

Proposition 1 *Under regularity conditions in Cai et al. (2013), $\{(T_i, S_i), 1 \leq i \leq m\}$ defined in* (4.17) *satisfy Assumptions (A1) and (A3), namely, T_i is asymptotically standard normal under the null, and for any constant $M > 0$,*

$$\mathbb{P}_{H_{0,i}}(|T_i| \geq t, |S_i| \geq \lambda) = (1 + o(1))G(t)\mathbb{P}(|N(0,1) + s_i| \geq \lambda) + O(m^{-M}),$$

uniformly for $0 \leq t \leq 4\sqrt{\log m}$, $0 \leq \lambda \leq 4\sqrt{\log m}$ and $i = 1, \ldots, m$, where $s_i = \mathbb{E}(S_i)$, and for all $0 \leq j \leq 4N$ with fixed N,

$$\mathbb{P}_{H_{0,i}}(|T_i| \geq t, |S_i| < \lambda_j) = (1 + o(1))G(t)\mathbb{P}(|N(0,1) + s_i| < \lambda_j) + O(m^{-M}),$$

uniformly for $0 \leq t \leq 4\sqrt{\log m}$ and $i \in \tilde{\mathcal{H}}_0$, where $\lambda_j = (j/N)\sqrt{\log m}$.

Proof: Recall that

$$T_i = \frac{\hat{\beta}_{i,1} - \hat{\beta}_{i,2}}{(\hat{\sigma}^2_{w,i,1} + \hat{\sigma}^2_{w,i,2})^{1/2}}, \text{ and } S_i = \frac{\hat{\beta}_{i,1} + (\hat{\sigma}^2_{w,i,1}/\hat{\sigma}^2_{w,i,2})\hat{\beta}_{i,2}}{\{\hat{\sigma}^2_{w,i,1}(1 + \hat{\sigma}^2_{w,i,1}/\hat{\sigma}^2_{w,i,2})\}^{1/2}} \quad 1 \leq i \leq m,$$

with

$$\hat{\beta}_{i,d} = \sum_{k=1}^{n_d} (Y_{k,a_i,d} - \bar{Y}_{a_i,d})(Y_{k,b_i,d} - \bar{Y}_{b_i,d}).$$

The result of Lemma 3 in Cai et al. (2013) yields that T_i satisfies Assumption (A1). For Assumption (A3), it is enough to show that

$$\mathbb{P}_{H_{0,i}}(|T_i| \geq t, |S_i| \geq \lambda) = (1 + o(1))G(t)\mathbb{P}(|N(0,1) + s_i| \geq \lambda) + O(m^{-M}),$$

uniformly for $0 \leq t \leq 4\sqrt{\log m}$, $0 \leq \lambda \leq 4\sqrt{\log m}$ and $i = 1, \ldots, m$. The second part then directly follows due to the fact that N is fixed. Note that $G(t + o((\log m)^{-1/2}))/G(t) = 1 + o(1)$ uniformly in $0 \le t \le c(\log m)^{1/2}$ for any constant c. By Lemma 3 in Cai et al. (2013), it suffices to show that,

$$\mathbb{P}(|V_i| \ge t, |Q_i| \ge \lambda) = (1 + o(1))G(t)\mathbb{P}(|N(0,1)| \ge \lambda) + O(m^{-M}),$$

where

$$V_i = \frac{\hat{\beta}_{i,1} - \hat{\beta}_{i,2}}{(\sigma^2_{w,i,1} + \sigma^2_{w,i,2})^{1/2}}, \text{ and } Q_i = \frac{\hat{\beta}_{i,1} - \beta_{i,1} + (\sigma^2_{w,i,1}/\sigma^2_{w,i,2})(\hat{\beta}_{i,2} - \beta_{i,2})}{\sqrt{\sigma^2_{w,i,1}(1 + \sigma^2_{w,i,1}/\sigma^2_{w,i,2})}}.$$

Note that V_i and Q_i are uncorrelated with each other.

Let $n_2/n_1 \le K_1$ with $K_1 \ge 1$. Define $Z_{k,i} = (n_2/n_1)\{Y_{k,a_i,1}Y_{k,b_i,1} - \mathbb{E}(Y_{k,a_i,1}Y_{k,b_i,1})\}$ for $1 \le k \le n_1$ and $Z_{k,i} = -\{Y_{k,a_i,2}Y_{k,b_i,2} - \mathbb{E}(Y_{k,a_i,2}Y_{k,b_i,2})\}$ for $n_1 + 1 \le k \le n_2$. Thus we have

$$V_i = \frac{\sum_{k=1}^{n_1+n_2} Z_{k,i}}{(n_2^2\sigma^2_{w,i,1} + n_2^2\sigma^2_{w,i,2})^{1/2}}.$$

Without loss of generality, we assume $\sigma^2_{\epsilon_d} = \sigma^2_{\eta_{i,d}} = 1$. Define

$$\hat{V}_i = \frac{\sum_{k=1}^{n_1+n_2} \hat{Z}_{k,i}}{(n_2^2\sigma^2_{w,i,1} + n_2^2\sigma^2_{w,i,2})^{1/2}},$$

where $\hat{Z}_{k,i} = Z_{k,i}I(|Z_{k,i}| \le \tau_n) - \mathbb{E}\{Z_{k,i}I(|Z_{k,i}| \le \tau_n)\}$, and $\tau_n = (4K_1/K)(\log(m+n))^{1+\epsilon}$ for any sufficiently small $\epsilon > 0$. Note that, for any $M > 0$

$$\begin{aligned}
\max_{1\le i\le m} n^{-1/2} \sum_{k=1}^{n_1+n_2} \quad & \mathbb{E}[|Z_{k,i}|I\{|Z_{k,i}| \ge \tau_n\}] \\
& \le Cn^{1/2} \max_{1\le k\le n_1+n_2} \max_{1\le i\le m} \mathbb{E}[|Z_{k,i}|I\{|Z_{k,i}| \ge \tau_n\}] \\
& \le Cn^{1/2}(m+n)^{-M} \max_{1\le k\le n_1+n_2} \max_{1\le i\le m} \mathbb{E}[|Z_{k,i}| \exp\{(K/2)|Z_{k,i}|\}] \\
& \le Cn^{1/2}(m+n)^{-M}.
\end{aligned}$$

Hence we have,

$$\mathbb{P}\Big\{\max_{1\le i\le m}|V_i-\hat{V}_i|\ge(\log m)^{-1}\Big\}\le\mathbb{P}\Big(\max_{1\le i\le m}\max_{1\le k\le n_1+n_2}|Z_{k,i}|\ge\tau_n\Big)=O(m^{-M}).$$

Similarly, define $F_{k,i}=(n_2/n_1)\{Y_{k,a_i,1}Y_{k,b_i,1}-\mathbb{E}(Y_{k,a_i,1}Y_{k,b_i,1})\}$ for $1\le k\le n_1$ and $F_{k,i}=(\sigma^2_{w,i,1}/\sigma^2_{w,i,2})\{Y_{k,a_i,2}Y_{k,b_i,2}-\mathbb{E}(Y_{k,a_i,2}Y_{k,b_i,2})\}$ for $n_1+1\le k\le n_2$. Then we have

$$Q_i=\frac{\sum_{k=1}^{n_1+n_2}F_{k,i}}{(n_2^2\sigma^2_{w,i,1}(1+\sigma^2_{w,i,1}/\sigma^2_{w,i,2})^{1/2}}.$$

Without loss of generality, we assume $\sigma^2_{w,i,1}=\sigma^2_{w,i,2}$. Define

$$\hat{Q}_i=\frac{\sum_{k=1}^{n_1+n_2}\hat{F}_{k,i}}{(n_2^2\sigma^2_{w,i,1}(1+\sigma^2_{w,i,1}/\sigma^2_{w,i,2})^{1/2}}.$$

where $\hat{F}_{k,i}=F_{k,i}I(|F_{k,i}|\le\tau_n)-\mathbb{E}\{F_{k,i}I(|F_{k,i}|\le\tau_n)\}$. Then we can similarly obtain that

$$\mathbb{P}\Big\{\max_{1\le i\le m}|Q_i-\hat{Q}_i|\ge(\log m)^{-1}\Big\}=O(m^{-M}).$$

Thus, it suffices it is to show that

$$\mathbb{P}(|\hat{V}_i|\ge t,|\hat{Q}_i|\ge\lambda)=(1+o(1))G(t)G(\lambda)+O(m^{-M}),\tag{B.24}$$

uniformly for $0\le t\le 4\sqrt{\log m}$ and $0\le\lambda\le 4\sqrt{\log m}$. Let

$$\boldsymbol{W}_k=\left\{\frac{\hat{Z}_{k,i}}{(n_2\sigma^2_{w,i,1}+n_2\sigma^2_{w,i,2})^{1/2}},\frac{\hat{F}_{k,i}}{(n_2\sigma^2_{w,i,1}(1+\sigma^2_{w,i,1}/\sigma^2_{w,i,2})^{1/2}}\right\}.$$

Then we have

$$\mathbb{P}(|\hat{V}_i|\ge t,|\hat{Q}_i|\ge\lambda)=\mathbb{P}(|n_2^{-1/2}\sum_{k=1}^{n_1+n_2}W_{k,1}|\ge t,|n_2^{-1/2}\sum_{k=1}^{n_1+n_2}W_{k,2}|\ge\lambda).$$

Then it follows from Theorem 1 in Zaïtsev (1987) that

$$\begin{aligned}
&\mathbb{P}(|n_2^{-1/2}\sum_{k=1}^{n_1+n_2} W_{k,1}| \geq t, |n_2^{-1/2}\sum_{k=1}^{n_1+n_2} W_{k,2}| \geq \lambda) \\
&\quad \leq \mathbb{P}(|N_1| \geq t - \epsilon_n(\log m)^{-1/2}, |N_2| \geq \lambda - \epsilon_n(\log m)^{-1/2}) + c_1 \exp\left\{-\frac{n^{1/2}\epsilon_n}{c_2\tau_n(\log m)^{1/2}}\right\},
\end{aligned}$$

where $c_1 > 0$ and $c_2 > 0$ are constants, $\epsilon_n \to 0$ which will be specified later and $\boldsymbol{N} = (N_1, N_2)$ is a normal random vector with $\mathbb{E}(\boldsymbol{N}) = 0$ and $\mathsf{Cov}(N_1, N_2) = 0$. Because $\log m = o(n^{1/C})$ for some $C > 5$, we can let $\epsilon_n \to 0$ sufficiently slowly that, for any large $M > 0$

$$c_1 \exp\left\{-\frac{n^{1/2}\epsilon_n}{c_2\tau_n(\log m)^{1/2}}\right\} = O(m^{-M}).$$

Thus, we have

$$\mathbb{P}(|\hat{V}_i| \geq t, |\hat{Q}_i| \geq \lambda) \leq \mathbb{P}(|N_1| \geq t - \epsilon_n(\log m)^{-1/2}, |N_2| \geq \lambda - \epsilon_n(\log m)^{-1/2}) + O(m^{-M}).$$

Similarly, using Theorem 1 in Zaïtsev (1987) again, we have

$$\mathbb{P}(|\hat{V}_i| \geq t, |\hat{Q}_i| \geq \lambda) \geq \mathbb{P}(|N_1| \geq t + \epsilon_n(\log m)^{-1/2}, |N_2| \geq \lambda + \epsilon_n(\log m)^{-1/2}) - O(m^{-M}).$$

Thus (B.24) is proved, and thus Proposition 1 follows. ∎

Proposition 2 *Under regularity conditions in Lemma 2, $\{(T_i, S_i), 1 \leq i \leq m\}$ defined in (4.19) satisfy that, as $n_1, n_2, m \to \infty$,*

$$T_i - \frac{f_i}{(\sigma_{w,i,1}^2 + \sigma_{w,i,2}^2)^{1/2}} \Rightarrow N(0,1),$$

uniformly in $i = 1, \ldots, m$, where $f_i = f_{a_i,b_i}$ with $f_{i,j} = 2\{\omega_{i,j}(\hat{\sigma}_{i,i,1,\epsilon} - \hat{\sigma}_{i,i,2,\epsilon}) + \omega_{i,j}(\hat{\sigma}_{j,j,1,\epsilon} - \hat{\sigma}_{j,j,2,\epsilon})\}$, $(\hat{\sigma}_{i,j,d,\epsilon}) = n_d^{-1}\sum_{k=1}^{n_d}(\boldsymbol{\epsilon}_{k,d} - \bar{\boldsymbol{\epsilon}}_d)(\boldsymbol{\epsilon}_{k,d} - \bar{\boldsymbol{\epsilon}}_d)^{\intercal}$, $\boldsymbol{\epsilon}_{k,d} = (\epsilon_{k,1,d}, \ldots, \epsilon_{k,p,d})$, and $\bar{\boldsymbol{\epsilon}}_d = (1/n_d)\sum_{k=1}^{n_d}\boldsymbol{\epsilon}_{k,d}$, *and for any constant $M > 0$,*

$$\mathbb{P}\left(\left|T_i - \frac{f_i}{(\sigma^2_{w,i,1} + \sigma^2_{w,i,2})^{1/2}}\right| \geq t, |S_i| \geq \lambda\right) = (1+o(1))G(t)\mathbb{P}(|N(0,1)+s_i| \geq \lambda)+O(m^{-M}),$$

uniformly for $0 \leq t \leq 4\sqrt{\log m}$, $0 \leq \lambda \leq 4\sqrt{\log m}$ and $i = 1, \ldots, m$. Furthermore, for all $0 \leq j \leq 4N$ with fixed N,

$$\mathbb{P}\left(\left|T_i - \frac{f_i}{(\sigma^2_{w,i,1} + \sigma^2_{w,i,2})^{1/2}}\right| \geq t, |S_i| < \lambda_j\right) = (1+o(1))G(t)\mathbb{P}(|N(0,1)+s_i| < \lambda_j)+O(m^{-M}),$$

uniformly for $0 \leq t \leq 4\sqrt{\log m}$ and $i = 1, \ldots, m$, where $\lambda_j = (j/N)\sqrt{\log m}$.

Proof: Recall that

$$T_i = \frac{\hat{\beta}_{i,1} - \hat{\beta}_{i,2}}{(\hat{\sigma}^2_{w,i,1} + \hat{\sigma}^2_{w,i,2})^{1/2}}, \text{ and } S_i = \frac{\hat{\beta}_{i,1} + (\hat{\sigma}^2_{w,i,1}/\hat{\sigma}^2_{w,i,2})\hat{\beta}_{i,2}}{\{\hat{\sigma}^2_{w,i,1}(1 + \hat{\sigma}^2_{w,i,1}/\hat{\sigma}^2_{w,i,2})\}^{1/2}} \quad 1 \leq i \leq m,$$

where $m = p(p-1)/2$, $\hat{\beta}_{i,d} = \hat{r}_{a_i,b_i,d}/(\hat{r}_{a_i,a_i}\hat{r}_{b_i,b_i})$ and $\hat{\sigma}^2_{w,i,d} = \hat{\sigma}^2_{a_i,b_i,d} = (1+\hat{\gamma}^2_{i,j,d}\hat{r}_{i,i,d}/\hat{r}_{j,j,d})/(n_d\hat{r}_{i,i,d}\hat{r}_{j,j,d})$. Let $V_{i,j} = (U_{i,j,2}-U_{i,j,1})/\{\mathsf{Var}(\epsilon_{k,i,1}\epsilon_{k,j,1})/n_1 + \mathsf{Var}(\epsilon_{k,i,2}\epsilon_{k,j,2})/n_2\}^{1/2}$, where $\mathsf{Var}(\epsilon_{k,i,d}\epsilon_{k,j,d}) = r_{i,i,d}r_{j,j,d}(1 + \rho^2_{i,j,d})$ with $\rho^2_{i,j,d} = \gamma^2_{i,j,d}r_{i,i,d}/r_{j,j,d}$. Note that the proof Lemma 2 yields that

$$\mathbb{P}\left(\max_i |\hat{\sigma}^2_{w,i,d} - \sigma^2_{w,i,d}| \geq C\sqrt{\frac{\log m}{n_d}}\right) = O(m^{-M}).$$

Hence, by Lemma 2, we have

$$\mathbb{P}\left(\left|T_i - \left\{V_i + \frac{f_i}{(\sigma^2_{w,i,1} + \sigma^2_{w,i,2})^{1/2}}\right\}\right| \geq Cb_m\right) = O(m^{-M}),$$

for some constant $C > 0$, where $b_m = o\{(\log m)^{-1/2}\}$. The first part of Proposition 2 is then proved by central limit theorem, based on which, the second part follows based on the proof of Proposition 1 by replacing $\boldsymbol{Y}_{k,d}$ by $\boldsymbol{\epsilon}_{k,d}$ for $d = 1, 2$. ∎

References

Cai, T. T., W. Liu, and Y. Xia (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Am. Statist. Assoc. 108*, 265–277.

Genovese, C. R., K. Roeder, and L. Wasserman (2006). False discovery control with p-value weighting. *Biometrika 93*(3), 509–524.

Hu, J. X., H. Zhao, and H. H. Zhou (2012). False discovery rate control with groups. *J. Am. Statist. Assoc.*.

Jin, J. and T. T. Cai (2007). Estimating the null and the proportion of nonnull effects in large-scale multiple comparisons. *J. Am. Statist. Assoc. 102*(478), 495–506.

Liu, W. and S. Luo (2014). Hypothesis testing for high-dimensional regression models. *Technical Report*.

Xia, Y., T. Cai, and T. T. Cai (2015). Testing differential networks with applications to the detection of gene-gene interactions. *Biometrika 102*, 247–266.

Xia, Y., T. Cai, and T. T. Cai (2018). Two-sample tests for high-dimensional linear regression with an application to detecting interactions. *Stat. Sinica, 28*(1), 63–92.

Zaïtsev, A. Y. (1987). On the Gaussian approximation of convolutions under multidimensional analogues of SN Bernstein's inequality conditions. *Probab. Theory Rel. 74*, 535–566.