同济大学 TONGJI UNIVERSITY
地址：中国上海市四平路1239号 邮编：200092
1239 SIPING ROAD SHANGHAI CHINA 200092
电话（TEL）：+86 21- 传真（FAX）：+86 21-
网址（WEB）：www.tongji.edu.cn

Uncorrelated Screening-based FDR control procedure

1. 改进目标：① 本文是针对均值检验提出的，主要改进是利用假设已知的稀疏信息，包含入统计量中，以此获得检验更高的Power. 由稀疏假设，我们更多关注两均值的支撑 $S_1 \cup S_2$. 故思想为第一步筛出(screen) $S_1 \cup S_2$，第二步再进行FDR检验

② 在以往的screening方法，要求screening步骤和检验步骤独立，本文提出的方法是screen统计量与test统计量渐进不相关，以前保证独立通常还需要额外的数据集或对样本进行分割，(也可能会导致 information loss) 且十分依赖于分割方法。

③ 本文还证明了 powerUS ≥ powerBH 的结果，还对信号大小(H_1)进行了分析证明了US方法power收敛到1的信号范围要于BH方法。NOTE：US方法可以没有稀疏假设，此时至少也和BH方法一样powerful. 但稀疏情形可以更优。

2. 方法步骤.(以 $\sigma_{i,1}^2 \neq \sigma_{i,2}^2$ 为例)

筛选 → 检验 → 调参 → 决策

Step 1: 计算 $T_i = \frac{\bar{X}_{i,1} - \bar{X}_{i,2}}{\sqrt{\frac{\hat{\sigma}_{i,1}^2}{n_1} + \frac{\hat{\sigma}_{i,2}^2}{n_2}}}$，$\hat{\sigma}_{i,y} = \frac{1}{n_y - 1}\sum_{k=1}^{n_y}(X_{k,i,y} - \bar{X}_{i,y})$，$y=1,2$. 和 $S_i = \sqrt{\frac{n_1}{\hat{\sigma}_{i,1}^2(1 + \frac{n_1\hat{\sigma}_{i,1}^2}{n_1\hat{\sigma}_{i,2}^2})}}(\bar{X}_{i,1} + \frac{n_1\hat{\sigma}_{i,1}^2}{n_1\hat{\sigma}_{i,2}^2}\bar{X}$

NOTE：用 S_i 作为screening统计量，S_i 较小较大分成两组，一般情况下，$S_1 \cup S_2$ 在小组中，大组中为 $(S_1 \cup S_2)^c$ 的元素，但可能小组中有 $U_{i,1}, U_{i,2}$ 异号的，这没关系，分成两组后，异号的和 $S_1 \cup S_2$ 在一起更易检验出来. 记 $\hat{m}_{y,\lambda} = \sum_{i \in H_0} I(|S_i| > \lambda)$，$\hat{m}_{y,\lambda} = \sum_{i=1}^{m} I(|S_i| > \lambda)$，$y=1,2$.

Step 2: 参数 λ 的选取 $\hat{\lambda} = \hat{\alpha}\sqrt{\log m}$，$\hat{\alpha} = \underset{1 \le k \le 4N}{\operatorname{argmax}} |R_{\lambda_k}|$

NOTE：相当于用网格法寻找最合适的 λ. 计算 $|R_{\lambda_k}|$ 时要用到第三步的阈值.

Step 3: 计算 threshold $\hat{t}_{1,\lambda} = \inf\{t > 0, \frac{\hat{m}_{1,\lambda}(2 - \Phi(t))}{\max(1, \sum_{i=1}^{m} I(|S_i| \ge \lambda, |T_i| \ge t))} \le \alpha\}$

$\hat{t}_{2,\lambda} = \inf\{t > 0 \frac{\hat{m}_{2,\lambda}(2 - \Phi(t))}{\max(1, \sum_{i=1}^{m} I(|S_i| < \lambda, |T_i| \ge t))} \le \alpha\}$

NOTE：$R_\lambda = \{1 \le i \le m, I(|S_i| \ge \lambda, |T_i| \ge \hat{t}_{i,\lambda}) = 1$ 或 $I(|S_i| < \lambda, |T_i| \ge \hat{t}_{i,\lambda}) = 1\}$，相当于筛选后两组各自使用 BH方法检验.

Step 4: 对给定的 $\alpha \in (0,1)$，当 $i \in R_\lambda$ 时，拒绝 H_{i0}，$i = 1, \cdots, m$.

NOTE：调参步中 λ 在 $(0, \lambda < 4\sqrt{\log m})$ 考虑是因为 $P(|S_i| > 4\sqrt{\log m}) \to 0$.

3. 假设条件.(1) $|H_1| = \alpha m$，即 $u_1 - u_2$ 的支撑是稀疏的.

(2) $E[\exp(t_0|X_{i,y} - u_{i,y}|/\sigma_{i,y})] \le K_1$，对某 $K_1 > 0, t_0 > 0$，$\forall 1 \le i \le m$，$y = 1,2$，设 $c_1 \le \frac{n_1}{n_2} \le c_2$，其中 $c_1, c_2 > 0$，以及 $\min(n_1, n_2) \ge c(\log m)^\beta$，$\beta > 5$，$c > 0$. 该条件相当于对总体的矩条件，且此在高维情形中常见.

(3). 设 M_1 为 H_0 的子集，满足 $\{(X_{i,1}, X_{i,2}), i \in M_1\}$ 与 $(X_{i,1}, X_{i,2})$ 相互独立. 对 $\forall i \in H_0$，有 $|M_1| \ge m_0 - K$. 即对任意 $(X_{i,1}, X_{i,2})$，我们允许 K 个变量强烈相关

同济大学 TONGJI UNIVERSITY
地址：中国上海市四平路1239号 邮编：200092
1239 SIPING ROAD SHANGHAI CHINA 200092
电话（TEL）：+86 21- 传真（FAX）：+86 21-
网址（WEB）：www.tongji.edu.cn

四、理论结果

定理一：设有(1),(2)，且 $|R_3| \xrightarrow{P} \infty$，$m\to\infty$，$\forall \varepsilon>0$，$P(FDP \le \alpha+\varepsilon)\to 1$．因此有 $\limsup_{m\to\infty} FDR \le \alpha$

定理二：设 $m_1\to\infty$，(1)-(3)成立，则有 $power_{OS} \ge power_{BH} + o_p(1)$，对某些 $o_p(1)$，$m\to\infty$．

接下来探讨 Power，假设 $\frac{|u_{i,1}-u_{i,2}|}{\sqrt{\sigma_{i,1}^2/n_1+\sigma_{i,2}^2/n_2}} = \theta\sqrt{\log m}$，$i\in H_1$，$\theta>0$．$\theta$ 衡量了信号 (H_1) 元素的大小（强弱）．设 $|H_1| = m^\beta$，$0<\beta<1$ 衡量信号的多少．(*)

定理三：设 (2),(3) 成立．若 $0<\theta<\sqrt{2(1-\beta)}$，有 $power_{BH} \xrightarrow{P} 0$．若 $\theta>\sqrt{2(1-\beta)}$，则 $power_{BH} \xrightarrow{P} 1$．

设两均值向量满足：对某个 $0\le r\le 1$，$0<\lambda\le 2$，$0<\rho<1$，$k>\lambda+\sqrt{r(1-\beta)}$

(4) $Card\{i\in H_0 : \sqrt{\frac{n_1}{\sigma_{i,1}^2(1+\frac{n_2\sigma_{i,1}^2}{n_1\sigma_{i,2}^2})}}\left|u_{i,1}+\frac{n_2\sigma_{i,1}^2}{n_1\sigma_{i,2}^2}u_{i,2}\right| \ge k\sqrt{\log m}\} = O(m^r)$（$u_i$ 和 σ_i 的渐近稀疏条件）

(5) $Card\{i\in H_1 : \sqrt{\frac{n_1}{\sigma_{i,1}^2(1+\frac{n_2\sigma_{i,1}^2}{n_1\sigma_{i,2}^2})}}\left|u_{i,1}+\frac{n_2\sigma_{i,1}^2}{n_1\sigma_{i,2}^2}u_{i,2}\right| \ge k\sqrt{\log m}\} \ge \rho|H_1|$（确保 S_1 可以筛选）

定理四：设 (2),(3),(*) 成立．

① 若 $\theta>\sqrt{2(1-\beta)}$，$power_{OS} \xrightarrow{P} 1$，$m\to\infty$

② (4),(5) 成立，令 $\theta>\sqrt{\max(0,2r-2\beta)}$，$N \ge \frac{10}{m^{\min(1-\beta,\frac{\theta^2}{2})}}$，有 $P(power_{OS}\ge \rho-\varepsilon)\to 1$ $\forall\varepsilon>0$，$m\to\infty$．

③ 有 $P(FDP\le\alpha+\varepsilon)\to 1$．$\forall\varepsilon>0$．$m\to\infty$．