Multi-level Thresholding Tests for High Dimensional Means and Covariance Matrices

Song Xi Chen

Guanghua School of Management

Center for Statistical Science

Peking University

Joint work with Bin Guo and Yumou Qiu

Two Sample Testing and Signal Detection

- $\mathbf{X}_1, \ldots, \mathbf{X}_{n_1} \overset{i.i.d.}{\sim} \mathsf{F}_1(\mu_1, \Sigma_1)$ and $\mathbf{Y}_1, \ldots, \mathbf{Y}_{n_2} \overset{i.i.d.}{\sim} \mathsf{F}_2(\mu_2, \Sigma_2)$
- $\mathbf{X}_k = (X_{k1}, \ldots, X_{kp})^{\mathrm{T}}$ and $\mathbf{Y}_k = (Y_{k1}, \ldots, Y_{kp})^{\mathrm{T}}$ are p-dimensional
- Means: $\mu_1 = (\mu_{11}, \ldots, \mu_{1p})^{\mathrm{T}}$ and $\mu_2 = (\mu_{21}, \ldots, \mu_{2p})^{\mathrm{T}}$
- Covariances: $\mathbf{\Sigma}_1 = (\sigma_{ij1})_{p \times p}$ and $\mathbf{\Sigma}_2 = (\sigma_{ij2})_{p \times p}$

Signals in the Mean

$$H_0 : \mu_1 = \mu_2 \quad \text{vs.} \quad H_a : \mu_1 \neq \mu_2$$

Signals in the Covariance

$$H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 \quad \text{vs.} \quad H_a : \mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$$

Tests for Means: Hotelling's T^2

$$T^2 = (\bar{X}_1 - \bar{X}_2)' \left\{ S_n (\frac{1}{n_1} + \frac{1}{n_2}) \right\}^{-1} (\bar{X}_1 - \bar{X}_2) \quad \text{where}$$

$$S_n = (n_1 + n_2 - 2)^{-1} \sum_{i=1}^{2} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)'.$$

Under $H_0 : \mu_1 = \mu_2$ and Gaussianity

$$\frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} T^2 \sim F_{p, n_1 + n_2 - p - 1}.$$

Reject H_0 at level α if

$$\frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} T^2 > F_{p, n_1 + n_2 - p - 1}(\alpha).$$

HD Tests for Means without Threshholding

- Bai and Saranadasa (BS) (1996) removed S_n^{-1} from T^2

$$BS = (\bar{X}_1 - \bar{X}_2)'(\bar{X}_1 - \bar{X}_2) - \frac{n_1 + n_2}{n_1 n_2} tr S_n$$

  Requires: (i) $\frac{p}{n} \to c \in [0, \infty)$ and (ii) $\lambda_{max} = o\left(\sqrt{tr(\Sigma^2)}\right)$.

- Srivastava (2009): replaced S_n with the diagonal matrix of S_n in T^2

  Requires: Gaussian data and $p \sim n$.

- Chen and Qin (2010): proposed U-statistic formulation

  allowing $p \gg n$, $\Sigma_1 \neq \Sigma_2$.

Chen-Qin (2010, AoS) Test

$$Q_n = \frac{\sum_{i \neq j} X_{1i}^T X_{1j}}{n_1(n_1-1)} + \frac{\sum_{i \neq j} X_{2i}^T X_{2j}}{n_2(n_2-1)} - 2\frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} X_{1i}^T X_{2j}}{n_1 n_2}$$

- A linear combination of one- and two-sample U-statistics.

$$E(Q_n) = \mu_1^T \mu_1 + \mu_2^T \mu_2 - 2\mu_1^T \mu_2 = \|\mu_1 - \mu_2\|^2.$$

- Main Assumption for Asymptotic Normality of Q_n

$$\frac{tr(\Sigma^4)}{tr^2(\Sigma^2)} \to 0 \quad \text{as } p \to \infty.$$

- Applicable for ANY p if the eigenvalues are bounded.
- Thus, allows $p \gg n$.

Asymptotic Power of Chen-Qin Test

$$\Phi\left(-z_\alpha + \frac{n\kappa(1-\kappa)\|\mu_1-\mu_2\|^2}{\sqrt{2tr(\tilde{\Sigma}^2)}}\right),$$

where $\tilde{\Sigma} = \kappa\Sigma_1 + (1-\kappa)\Sigma_2$ and $\kappa = \lim_{n_1,n_2\to\infty} n_1/(n_1+n_2)$.

- A VALID test under weak assumptions for wide range of dimensions.
- "VALID" means control of type I error.
- The power may be weak under high dimension due to inflated $tr(\tilde{\Sigma}^2)$.

Thresholding Tests for Means

- One Sample Higher Criticism (HC) Test (Tukey, 1976).
- Donoho and Jin (2004) pioneered the theory under $N_p(\mu, I_p)$.
- $\mu = (\mu_1, \cdots, \mu_p)$ and those non-zero $\mu_i = \sqrt{2r \log p}$
- Faint Signals if $r \in (0, 1)$
- $S_\beta = \{k : \mu_k \neq 0\}$, the signal set.
- $|S_\beta| = p^{1-\beta}$ – the number of signals.
- Sparse Signals if $\beta \in (0.5, 1)$.

Higher Criticism (HC)

- $X \sim N(\mu, I_p)$
- Z_i is the Z-statistic at the i-th dimension.
- $p_i = P(N(0,1) > Z_i)$ is the p-value for the i-th null.
- Sorted p-values: $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(p)}$.
- The HC statistic:

$$HC_n^* = \max_{0 \leq i \leq \alpha * p} \frac{\sqrt{p}(i/p - p_{(i)})}{\sqrt{p_{(i)}(1 - p_{(i)})}}.$$

Optimal Detection Boundary Under $N_p(\mu, I_p)$

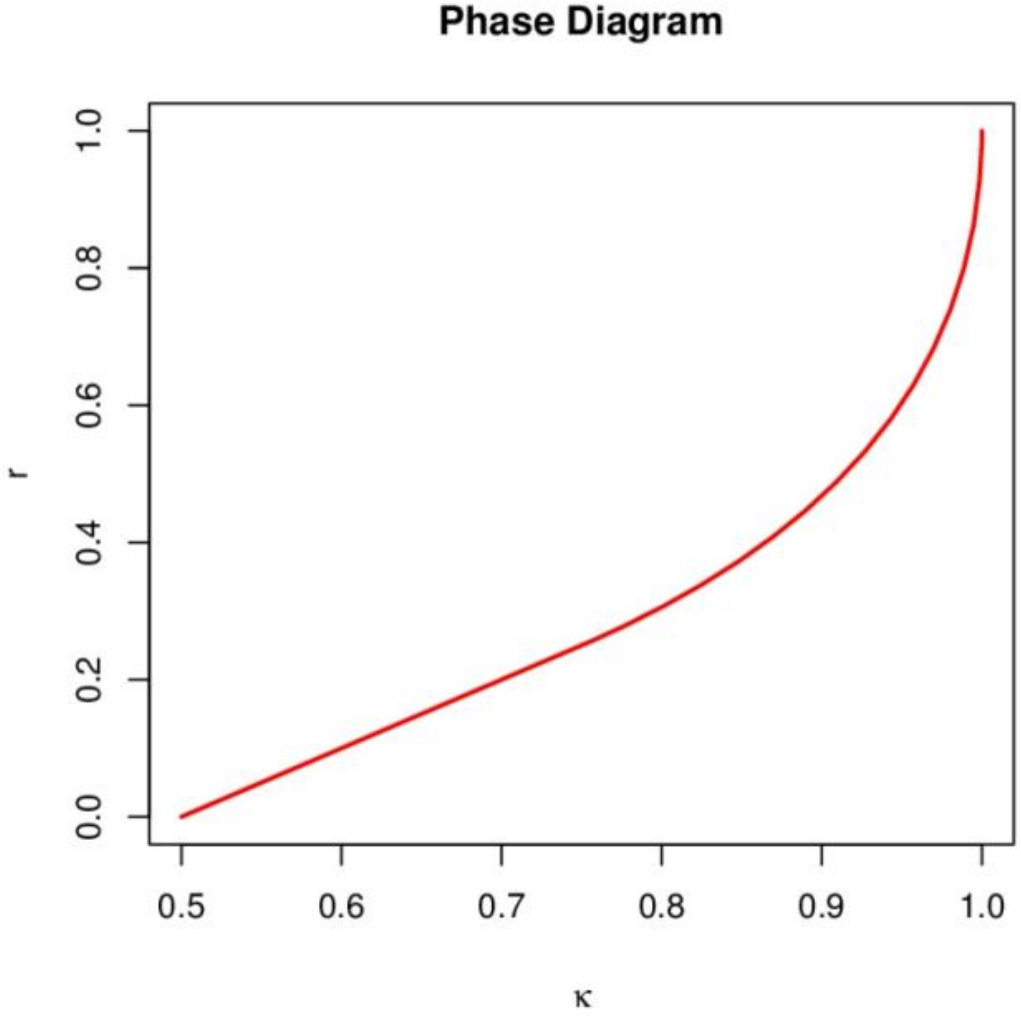
- A phase diagram in (r, β)-plane $r = \varrho(\beta)$.

- If $r > \varrho(\beta)$, H_0 and H_1 are asymptotically separable;
  If $r < \varrho(\beta)$, H_0 and H_1 are not separable.

- Donoho and Jin (2004) established the detection boundary for HC test for Gaussian data

$$\varrho^*(\beta) = \begin{cases} \beta - 1/2, & 1/2 < \beta \leq 3/4; \\ (1 - \sqrt{1-\beta})^2, & 3/4 < \beta < 1 \end{cases}$$

Same as the optimal detection boundary by Ingster (1999) without knowing the underlying signal strength r and sparsity β.



(i) For any test of the hypothesis,

$$\text{P(Type I Error)} + \text{P(Type II Error)} \to 1 \quad \text{if} \quad r < \rho(\beta) \quad \text{as } n, p \to \infty;$$

(ii) There exists a test (HC) such that

$$\text{P(Type I Error)} + \text{P(Type II Error)} \to 0 \quad \text{if} \quad r > \rho(\beta) \quad \text{as } n, p \to \infty.$$

L_γ-Thresholding for $H_0 : \mu = 0$

- Motivated by Donoho and Johnstone (1994) and Fan (1996).
- $\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$
- The threshold statistics

$$T_{\gamma n}(s) = \sum_{i=1}^{p} |n\bar{X}_i|^\gamma I\{|\sqrt{n}\bar{X}_i| > \sqrt{2s\log(p)}\} \quad \text{for} \quad s \in (0,1)$$

  - $\gamma = 0$: the HC;
  - $\gamma = 1$: the L_1-thresholding (hard thresholding) by Donoho and Johnstone (1994);
  - $\gamma = 2$: the L_2-thresholding used in Zhong, Chen and Xu (2013).

L_2-Thresholding Tests

- One sample: Zhong, Chen and Xu (2013, AoS)
- Two sample: Chen, Li and Zhong (2019, AoS)
- Can also attain Ingster "optimal" detection boundary when the underlying distributions is unknown and data are dependent ($\Sigma \neq I_p$).
- More powerful than the HC when (r, β) are above the boundary (ZCX).
- The detection boundary can be lowered by utilizing the dependence (CLZ) by first transforming data X_{ij} to $\hat{\tilde{\Sigma}}^{-1} X_{ij}$ then applying the L_2-thresholding.

Two-Sample for Means: Signals and Sparsity

- $\delta_k = \mu_{1k} - \mu_{2k}$ – signal in the k-th dimension.
- $S_\beta = \{k : \delta_k \neq 0\}$, the signal set.
- $|S_\beta| = p^{1-\beta}$ — the number of signals.
- Sparse if $\beta \in (0.5, 1)$.

Two-Sample for Means: L_2 Test Statistic

- An unbiased estimator to the signal δ_k^2: U-statistics

$$\begin{aligned} T_{nk} &= \frac{1}{n_1(n_1-1)} \sum_{i \neq j}^{n_1} X_{1i}^{(k)} X_{1j}^{(k)} + \frac{1}{n_2(n_2-1)} \sum_{i \neq j}^{n_2} X_{2i}^{(k)} X_{2j}^{(k)} \\ &- \frac{2}{n_1 n_2} \sum_{i}^{n_1} \sum_{j}^{n_2} X_{1i}^{(k)} X_{2j}^{(k)}. \end{aligned}$$

- Test statistic

$$\tilde{T}_n = n \sum T_{nk}.$$

- Chen and Qin (2010, AoS)

Two-Sample for Means: L_2 vs L_2-Thresholding Statistics

- CQ: $\tilde{T}_n = n \sum_{i \in S_\beta^c} T_{ni} + n \sum_{k \in S_\beta} T_{nk}$.
- Oracle: $n \sum_{i \in S_\beta} T_{ni}$.
- Thresholding Statistic

$$L_n(s) = \sum_{k=1}^{p} nT_{nk} I\left\{ nT_{nk} + 1 > \lambda_n(s) \right\}$$

  where $\lambda_n(s) = 2s \log(p)$.

- Try to exclude those $\delta_k = 0$ dimensions.

Two-Sample Tests for Means: Variance Comparison – Strong Signal Case of $n\delta_k^2 > 2\log(p)$

Tests	Variances
L_2	$2p + 2\sum_{i\neq j}\rho_{ij}^2 + 4n\sum_{k,l\in S_\beta}\delta_k\delta_l\rho_{kl}$
Oracle	$2p^{1-\beta} + 2\sum_{i\neq j\in S_\beta}\rho_{ij}^2 + 4n\sum_{k,l\in S_\beta}\delta_k\delta_l\rho_{kl}$
Thresholding	$2L_p + 2p^{1-\beta} + 2\sum_{i\neq j\in S_\beta}\rho_{ij}^2 + 4n\sum_{k,l\in S_\beta}\delta_k\delta_l\rho_{kl}$

- L_p denotes slowly varying functions in the form of $(a\log p)^b$.

Multi-level Thresholding: Weak Signal Case

Weak Signals: $\delta_k^2 = 2r\log p/n$ for $r < 1$.

$$M_{L_n} = \max_{s \in \mathcal{S}_n} \frac{L_n(s) - \hat{\mu}_{L_n(s),0}}{\hat{\sigma}_{L_n(s),0}},$$

$$\mathcal{S}_n = \{s_k : s_k = n(\bar{X}_1^{(k)} - \bar{X}_1^{(k)})^2/(2\log p) \quad \text{for} \quad k = 1, \cdots, p\}.$$

Theorem. Under Conditions **C1**-**C3** and H_0,

$$\mathrm{P}\left\{a(\log p)M_{L_n} - b(\log p, \eta) \le x\right\} \to \exp(-e^{-x}),$$

where $a(y) = (2\log y)^{\frac{1}{2}}$ and $b(y, \eta) = 2\log y + 2^{-1}\log\log y - 2^{-1}\log\{\frac{4\pi}{(1-\eta)^2}\}$.

Detection Boundary of Multi-level Thresholding for Means

Multi-level Thresholding test rejects H_0 if

$$M_{L_n} \geq G_\alpha = \{q_\alpha + b(\log p, \eta)\}/a(\log p),$$

q_α is the upper α quantile of the Gumbel distribution.

Define

$$\varrho(\beta) = \begin{cases} \beta - \frac{1}{2}, & \frac{1}{2} \leq \beta \leq \frac{3}{4}; \\ (1 - \sqrt{1-\beta})^2, & \frac{3}{4} < \beta < 1, \end{cases}$$

Theorem Assume Conditions **C1**-**C3**. If $r > \varrho(\beta)$, the sum of type I and II errors of the multi-level thresholding test converges to zero as $\alpha \to 0$ and $p \to \infty$.

- The same "detection boundary" as the optimal one for $N_p(\mu, I_p)$ case.
- Signal enhancement by transforming data with the precision matrix $\Omega = \Sigma^{-1}$.
- Improved detection boundary: lower than $\rho(\beta)$.
- See Chen, Li and Zhong (2019) for details.

Two Sample Tests for Covariance Matrices

- $H_0 : \Sigma_1 = \Sigma_2$ vs $\Sigma_1 \neq \Sigma_2$.
- $\mathbf{S}_{n1} = (s_{ij1})$, $\mathbf{S}_{n2} = (s_{ij2})$: two sample covariances
- $\theta_{ij1} = \mathrm{Var}\{(X_{ki} - \mu_{1i})(X_{kj} - \mu_{1j})\}$ and $\theta_{ij2} = \mathrm{Var}\{(Y_{ki} - \mu_{2i})(Y_{kj} - \mu_{2j})\}$
- $\hat{\theta}_{ij1} = \frac{1}{n_1} \sum_{k=1}^{n_1} \{(X_{ki} - \bar{X}_i)(X_{kj} - \bar{X}_j) - s_{ij1}\}^2 \xrightarrow{p} \theta_{ij1}$
- $\hat{\theta}_{ij2} = \frac{1}{n_2} \sum_{k=1}^{n_2} \{(Y_{ki} - \bar{Y}_i)(Y_{kj} - \bar{Y}_j) - s_{ij2}\}^2 \xrightarrow{p} \theta_{ij2}$

$$M_{ij} = \frac{(s_{ij1} - s_{ij2})^2}{\hat{\theta}_{ij1}/n_1 + \hat{\theta}_{ij2}/n_2}, \quad 1 \leq i \leq j \leq p.$$

Existing Work

- Bai et al. (2009, AoS): Corrected Likelihood Ratio test using RMT.
- Cai, Liu and Xia (2013): L_{max} statistic $M_n = \max_{1 \leq i \leq j \leq p} M_{ij}$
  - Only use the maximal signal
- Li and Chen (2012): L_2 statistic, sum over all M_{ij}
  - Include too many uninformative entries
- Srivastava and Yanagihara (2010): Also L_2 statistic to measure

$$tr(\Sigma_1^2)/(tr^2(\Sigma_1)) - tr(\Sigma_2^2)/(tr^2(\Sigma_2))$$

L_2-Test Statistic: Li and Chen (2012)

- Target on Square of Frobenius norm:

$$tr\{(\Sigma_1 - \Sigma_2)^2\} = tr(\Sigma_1^2) + tr(\Sigma_2^2) - 2tr(\Sigma_1\Sigma_2).$$

- Note that $\Sigma_1 = \Sigma_2$ if and only if $tr\{(\Sigma_1 - \Sigma_2)^2\} = 0$.

- Although the Frobenius norm is large, it brings two advantages.
  - (i) Relatively easier to analyze for test procedures and power formula.
  - (ii) Can target on certain sections of the covariance matrix.

Unbiased estimator of $tr(\Sigma_h^2)$ and $tr(\Sigma_1\Sigma_2)$

For $h = 1$ or 2,

$$\begin{aligned} A_{n_h} &= \frac{1}{n_h(n_h-1)}\sum_{i\neq j}(X'_{hi}X_{hj})^2 - \frac{2}{n_h(n_h-1)(n_h-2)}\sum_{i,j,k}^{\star}(X'_{hi}X_{hj}X'_{hj}X_{hk}) \\ &+ \frac{1}{n_h(n_h-1)(n_h-2)(n_h-3)}\sum_{i,j,k,l}^{\star}(X'_{hi}X_{hj}X'_{hk}X_{hl}), \end{aligned}$$

For $tr(\Sigma_1\Sigma_2)$:

$$\begin{aligned} C_{n_1n_2} &= \frac{1}{n_1n_2}\sum_{i,j}(X'_{1i}X_{2,j})^2 - \frac{1}{n_1n_2(n_1-1)}\sum_{i\neq k,j}(X'_{1i}X_{2j}X'_{2j}X_{1k}) \\ &- \frac{1}{n_1n_2(n_2-1)}\sum_{i\neq k,j}(X'_{2i}X_{1j}X'_{1j}X_{2k}) \\ &+ \frac{1}{n_1n_2(n_1-1)(n_2-1)}\sum_{i\neq k,j\neq l}(X'_{1i}X_{2j}X'_{1k}X_{2l}), \end{aligned}$$

Test statistic

$$T_{n_1,n_2} = A_{n_1} + A_{n_2} - 2C_{n_1 n_2}.$$

- $E(T_{n_1,n_2}) = tr\{(\Sigma_1 - \Sigma_2)^2\}$.
- Leading order variance:

$$\sigma^2_{n_1,n_2} = \sum_{i=1}^{2} \Big[\frac{4}{n_i^2} tr^2(\Sigma_i^2) + \frac{8}{n_i} tr\{(\Sigma_i^2 - \Sigma_1\Sigma_2)^2\} + \frac{4\Delta_i}{n_i} tr\{\Gamma_i'(\Sigma_1 - \Sigma_2)\Gamma_i \circ \Gamma_i'(\Sigma_1 - \Sigma_2)\Gamma_i\} \Big] + \frac{8}{n_1 n_2} tr^2(\Sigma_1\Sigma_2).$$

- Under H_0: $\Sigma_1 = \Sigma_2 = \Sigma$,

$$\sigma^2_{0,n_1,n_2} = 4(\frac{1}{n_1} + \frac{1}{n_2})^2 tr^2(\Sigma^2).$$

Assumptions

- A1: As $\min\{n_1, n_2\} \to \infty$, $n_1/(n_1 + n_2) \to \kappa \in (0, 1)$.

- A2: $\min\{n_1, n_2\} \to \infty$, $p(n_1, n_2) \to \infty$ and for i, j, k and $l \in \{1, 2\}$,
$$tr\{(\Sigma_i\Sigma_j)(\Sigma_k\Sigma_l)\} = o\{(tr(\Sigma_i\Sigma_j)tr(\Sigma_k\Sigma_l)\}.$$

- A3: $$X_{ij} = \Gamma_i Z_{ij} + \mu_i,$$
where $\Gamma_i\Gamma_i' = \Sigma_i$; $\mathsf{E}(Z_{ij}) = 0$, $\mathsf{Var}(Z_{ij}) = I_{m_i}$.

Asymptotic Normality of T_{n_1,n_2}

- Under Assumptions 1-3, as $\min\{n_1, n_2\} \to \infty$

$$\sigma_{n_1,n_2}^{-1}\left[T_{n_1,n_2} - tr\{(\Sigma_1 - \Sigma_2)^2\}\right] \xrightarrow{d} \mathsf{N}(0,1).$$

- Variance Estimation:

$\hat{\sigma}^2_{0,n_1,n_2} =: \frac{2}{n_2} A_{n_1} + \frac{2}{n_1} A_{n_2}$

$$\frac{A_{n_i}}{tr(\Sigma_i^2)} \xrightarrow{p} 1 \quad \text{and} \quad \frac{\hat{\sigma}_{0,n_1,n_2}}{\sigma_{0,n_1,n_2}} \xrightarrow{p} 1.$$

Test Procedure and Power

- A nominal α level test rejects H_{0a} if

$$T_{n_1,n_2} \geq \hat{\sigma}_{0,n_1,n_2} z_\alpha$$

where z_α is the upper-α quantile of $N(0,1)$.

- Power of the test

$$\Phi\left(-Z_{n_1,n_2}(\Sigma_1,\Sigma_2)z_\alpha + \frac{tr\{(\Sigma_1-\Sigma_2)^2\}}{\sigma_{n_1,n_2}}\right),$$

$$Z_{n_1,n_2}(\Sigma_1,\Sigma_2) = (\sigma_{n_1,n_2})^{-1}\{\frac{2}{n_2}tr(\Sigma_1^2) + \frac{2}{n_1}tr(\Sigma_2^2)\}.$$

- Li-Chen Test operates under weak assumptions
- but the power would not be high for sparse and faint signals.

Standardization of Sample Covariances

- $\mathbf{S}_{n1} = (s_{ij1})$, $\mathbf{S}_{n2} = (s_{ij2})$: two sample covariances
- $\theta_{ij1} = \mathrm{Var}\{(X_{ki} - \mu_{1i})(X_{kj} - \mu_{1j})\}$ and $\theta_{ij2} = \mathrm{Var}\{(Y_{ki} - \mu_{2i})(Y_{kj} - \mu_{2j})\}$
- $\hat{\theta}_{ij1} = \frac{1}{n_1} \sum_{k=1}^{n_1} \{(X_{ki} - \bar{X}_i)(X_{kj} - \bar{X}_j) - s_{ij1}\}^2 \xrightarrow{p} \theta_{ij1}$
- $\hat{\theta}_{ij2} = \frac{1}{n_2} \sum_{k=1}^{n_2} \{(Y_{ki} - \bar{Y}_i)(Y_{kj} - \bar{Y}_j) - s_{ij2}\}^2 \xrightarrow{p} \theta_{ij2}$

$$M_{ij} = \frac{(s_{ij1} - s_{ij2})^2}{\hat{\theta}_{ij1}/n_1 + \hat{\theta}_{ij2}/n_2}, \quad 1 \le i \le j \le p.$$

Thresholding for Covariance Testing

- Under the null hypothesis some assumptions, as $n, p \to \infty$,

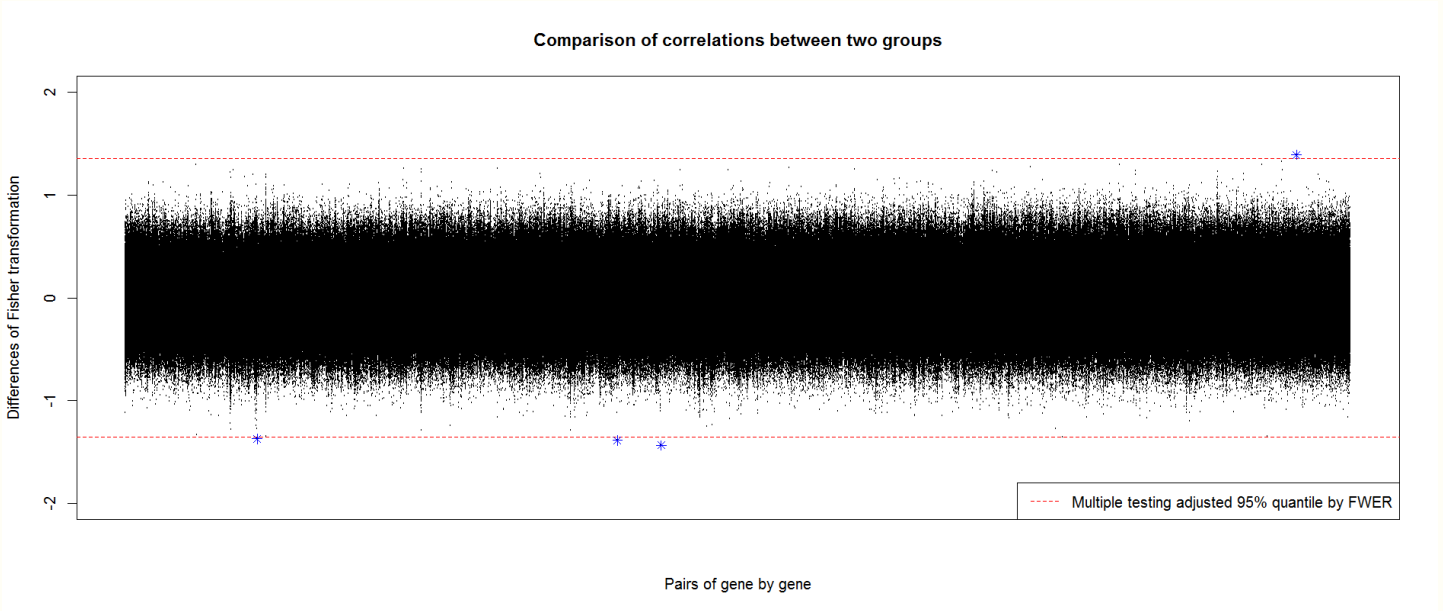
$$P\Big\{ \max_{1 \le i \le j \le p} M_{ij} > 4 \log(p) \Big\} \to 0.$$

- Thresholding on M_{ij}

$$T_n(s) = \sum_{1 \le i \le j \le p} M_{ij} \mathbb{I}\{M_{ij} > \lambda_p(s)\}$$

- $\lambda_p(s) = 4s \log(p)$ for a thresholding parameter $s \in (0, 1)$

Sparse and weak signal: A real example on Neg and Bcr



Sparse and Weak Alternative Hypothesis

- $n = n_1 n_2/(n_1 + n_2)$ and $q = p(p+1)/2$
- Number of nonzero δ_{ij}: $m_a = \lfloor q^{(1-\beta)} \rfloor$ for a $\beta \in (1/2, 1)$
- Nonzero value: $\delta_{ij} = \delta_a = \sqrt{4r \log(p)/n}$ if $\delta_{ij} \neq 0$
- $\beta \in (1/2, 1)$: sparsity signals; $r \in (0, 1)$: faint signals

$$H_0 : \delta_{ij} = 0 \quad \text{for all } 1 \leq i \leq j \leq p \quad \text{vs.}$$

$$H_a : \text{there are } m_a \text{ nonzero } \delta_{ij} \text{ with strength } \delta_a.$$

Main assumptions

Assumption 1A. Exponential rate: $\log p \sim n^{\varpi}$, $\varpi \in (0, 1/3)$.

Assumption 1B. Polynomial rate: $n \sim p^{\xi}$, $\xi \in (0, 2)$.

Assumption 2. Subgaussian distribution: $\mathrm{E}[\exp\{t(X_{kj} - \mu_{1j})^2\}] \leq C$ and $\mathrm{E}[\exp\{t(Y_{kj} - \mu_{2j})^2\}] \leq C$.

Assumption 3. β-mixing: $\{X_{kj}\}_{j=1}^{p}$ and $\{Y_{kj}\}_{j=1}^{p}$ are β-mixing after certain permutation and the mixing coefficients satisfying polynomial rate.

Remark:

- $T_n(s)$ is invariant to permutations of $\{X_{kj}\}_{j=1}^{p}$ and $\{Y_{kj}\}_{j=1}^{p}$.
- No need to know the permutation.
- The β-mixing is satisfied under weak assumptions on $\{X_{kj}\}_{j=1}^{p}$ and $\{Y_{kj}\}_{j=1}^{p}$ (Mokkadem, 1988). For example, normally distributed data with banded or block diagonal covariance matrix are special case.

Main assumptions

Assumption 1A. Exponential rate: $\log p \sim n^{\varpi}$, $\varpi \in (0, 1/3)$.

Assumption 1B. Polynomial rate: $n \sim p^{\xi}$, $\xi \in (0, 2)$.

Assumption 2. Subgaussian distribution: $\mathrm{E}[\exp\{t(X_{kj} - \mu_{1j})^2\}] \leq C$ and $\mathrm{E}[\exp\{t(Y_{kj} - \mu_{2j})^2\}] \leq C$.

Assumption 3. β-mixing: $\{X_{kj}\}_{j=1}^{p}$ and $\{Y_{kj}\}_{j=1}^{p}$ are β-mixing after certain permutation and the mixing coefficients satisfying polynomial rate.

Remark:

- $T_n(s)$ is invariant to permutations of $\{X_{kj}\}_{j=1}^{p}$ and $\{Y_{kj}\}_{j=1}^{p}$.
- No need to know the permutation.
- The β-mixing is satisfied under weak assumptions on $\{X_{kj}\}_{j=1}^{p}$ and $\{Y_{kj}\}_{j=1}^{p}$ (Mokkadem, 1988). For example, normally distributed data with banded or block diagonal covariance matrix are special case.

Mean and variance of $T_n(s)$

- $\phi(\cdot)$ and $\bar{\Phi}(\cdot)$ be the density and survival functions of the standard normal distribution, $L_p = a(\log p)^b$ with $b > 0$.
- $\mu_0(s) = \mathrm{E}\{T_n(s)|H_0\}$ and $\sigma_0^2(s) = \mathrm{Var}\{T_n(s)|H_0\}$.

Proposition

Under Assumptions **1A** *or* **1B** *and some other assumptions,*

$$\mu_0(s) = \tilde{\mu}_0(s)\{1 + O(L_p n^{-1/2})\}$$

where

$$\tilde{\mu}_0(s) = q\{2\lambda_p^{1/2}(s)\phi(\lambda_p^{1/2}(s)) + 2\bar{\Phi}(\lambda_p^{1/2}(s))\}.$$

In addition, under either (i) Assumption **1A** *with* $s > 1/2$ *or (ii) Assumption* **1B** *with* $s > 1/2 - \xi/4$, $\sigma_0^2(s) = \tilde{\sigma}_0^2(s)\{1 + o(1)\}$, *where*

$$\tilde{\sigma}_0^2(s) = q[2\{\lambda_p^{3/2}(s) + 3\lambda_p^{1/2}(s)\}\phi(\lambda_p^{1/2}(s)) + 6\bar{\Phi}(\lambda_p^{1/2}(s))].$$

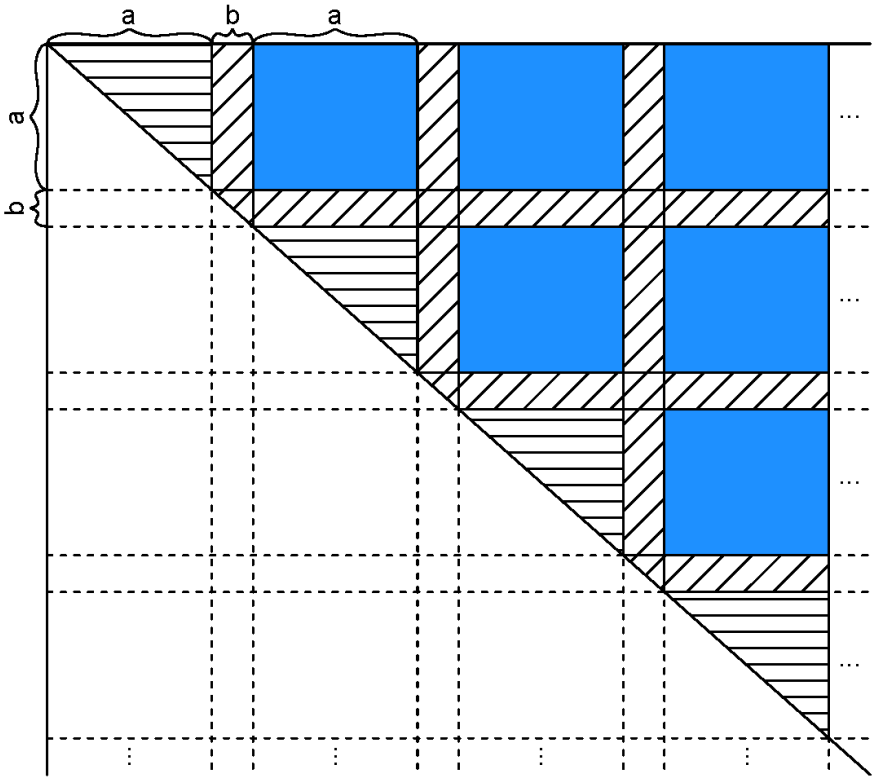
Challenge: asymptotic distribution of $T_n(s)$?

- Can't just mixing results since $\{s_{ij}\}$ are not mixing
- Can't apply Martingale CLT due to the thresholding

Coupling + Martingale CLT

Matrix Blocking

$$\{1,\ldots,a\},\{a+1,\ldots,a+b\},\{a+b+1,\ldots,2a+b\},\{2a+b+1,\ldots,2a+2b\},\ldots$$
$$S_1 \qquad R_1 \qquad S_2 \qquad R_2 \qquad \ldots$$



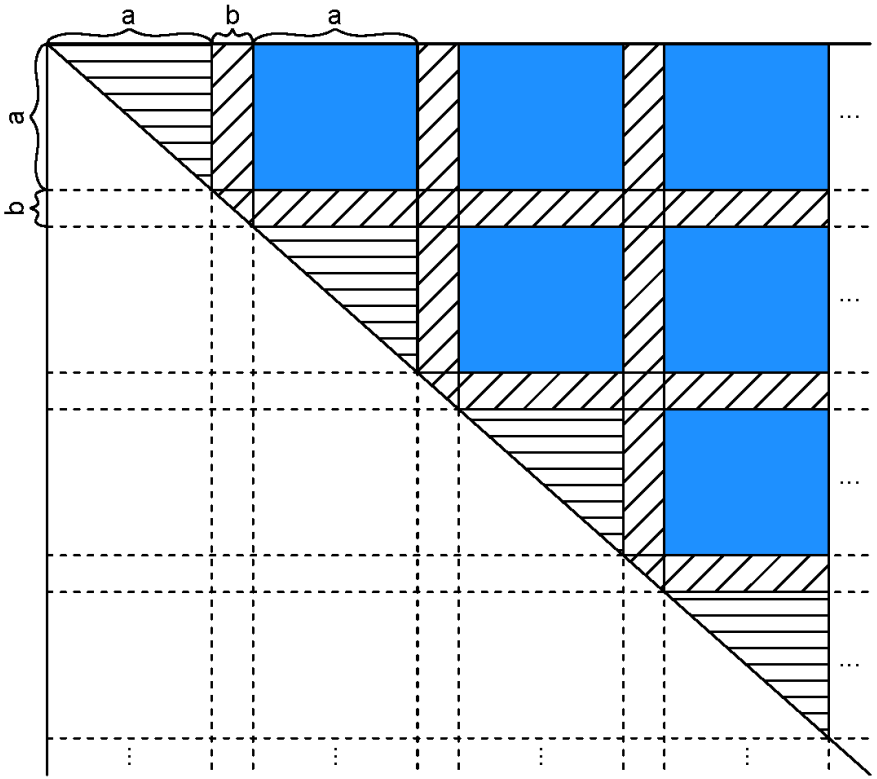
- Small blocks and triangles are negligible
- $B_{1,n}$: Summation over all big blocks
- $\mathbf{X}_{S_m}, \mathbf{Y}_{S_m}$: the segments of data matrices with the columns in S_m
- $\mathbf{Z}_{S_m} = \{\mathbf{X}_{S_m}, \mathbf{Y}_{S_m}\}$
- Coupling (Berbee 1979): $\mathbf{Z}^*_{S_m} \approx \mathbf{Z}_{S_m}$, $\{\mathbf{Z}^*_{S_m}\}$ independent
- Big blocks in different row and columns are independent

Coupling + Martingale CLT

Matrix Blocking

$\{1,\ldots,a\},\{a+1,\ldots,a+b\},\{a+b+1,\ldots,2a+b\},\{2a+b+1,\ldots,2a+2b\},\ldots$

S_1 R_1 S_2 R_2 $\ldots$



- Small blocks and triangles are negligible
- $B_{1,n}$: Summation over all big blocks
- $\mathbf{X}_{S_m}, \mathbf{Y}_{S_m}$: the segments of data matrices with the columns in S_m
- $\mathbf{Z}_{S_m} = \{\mathbf{X}_{S_m}, \mathbf{Y}_{S_m}\}$
- Coupling (Berbee 1979): $\mathbf{Z}^*_{S_m} \approx \mathbf{Z}_{S_m}$, $\{\mathbf{Z}^*_{S_m}\}$ independent
- Big blocks in different row and columns are independent

Coupling + Martingale CLT

U-statistic Equivalence

- U-statistic formulation: $B_{1,n} \sim \sum_{m_1<m_2} f(\mathbf{Z}^*_{S_{m_1}}, \mathbf{Z}^*_{S_{m_2}})$
- Apply Martingale CLT, build filtration on $\{\mathbf{Z}^*_{S_m}\}$

Theorem

Suppose Assumptions **2** *and* **3** *are satisfied. Then, under the* H_0*, and either (i) Assumption* **1A** *with* $s > 1/2$ *or (ii) Assumption* **1B** *with* $s > 1/2 - \xi/4$*, we have*

$$\sigma_0^{-1}(s)\{T_n(s) - \mu_0(s)\} \xrightarrow{d} N(0,1) \quad \text{as } n, p \to \infty.$$

Coupling + Martingale CLT

U-statistic Equivalence

- U-statistic formulation: $B_{1,n} \sim \sum_{m_1<m_2} f(\mathbf{Z}^*_{S_{m_1}}, \mathbf{Z}^*_{S_{m_2}})$
- Apply Martingale CLT, build filtration on $\{\mathbf{Z}^*_{S_m}\}$

Theorem

Suppose Assumptions **2** *and* **3** *are satisfied. Then, under the* H_0*, and either (i) Assumption* **1A** *with* $s > 1/2$ *or (ii) Assumption* **1B** *with* $s > 1/2 - \xi/4$*, we have*

$$\sigma_0^{-1}(s)\{T_n(s) - \mu_0(s)\} \xrightarrow{d} N(0,1) \quad \text{as } n, p \to \infty.$$

Multi-level Thresholding Test (MTT)

- How to choose s in reality?
- Standardization of $T_n(s)$: $U_n(s) = \tilde{\sigma}_0^{-1}(s)\{T_n(s) - \tilde{\mu}_0(s)\}$
- Maximize $U_n(s)$ over multiple thresholds

$$\mathcal{V}_n(s_0) = \sup_{s \in (s_0, 1-\eta]} U_n(s)$$

- $s_0 = 1/2$ for $\log p \sim n^{\varpi}$, or $s_0 = 1/2 - \xi/4$ for $n \sim p^{\xi}$
- η is a small positive constant, say 0.05.
- To simplify the calculation, it can be shown that

$$\mathcal{V}_n(s_0) = \sup_{s \in \mathcal{S}_n(s_0)} U_n(s).$$

where

$$\mathcal{S}_n(s_0) = \{t_{ij} : t_{ij} = M_{ij}/(4\log(p)) \text{ and } s_0 < t_{ij} < (1-\eta)\} \cup \{1-\eta\}.$$

Multi-level Thresholding Test (MTT)

Theorem

Suppose Assumptions **2** *and* **3** *are satisfied. Then, under the* H_0*, and either (i) Assumption* **1A** *with* $s > 1/2$ *or (ii) Assumption* **1B** *with* $s > 1/2 - \xi/4$*,*

$$P\{a(\log(p))\mathcal{V}_n(s_0) - b(\log(p), s_0, \eta) \leq x\} \to \exp(-e^{-x}),$$

where $a(y) = (2\log(y))^{1/2}$ *and* $b(y, s_0, \eta) = 2\log(y) + 2^{-1}\log\log(y)$ $-2^{-1}\log(\pi) + \log(1 - s_0 - \eta)$*.*

Multi-level Thresholding Test (MTT)

- Reject H_0 if

$$\mathcal{V}_n(s_0) > \{q_\alpha + b(\log(p), s_0, \eta)\}/a(\log(p)),$$

where q_α is the upper α quantile of the Gumbel distribution.

- Size distortion as the convergence to the Gumbel distribution is slow.
- A bootstrap method proposed in the simulation.

Detection boundary for covariances

Sparse and weak alternative hypothesis:

$$H_0 : \delta_{ij} = 0 \quad \text{for all } 1 \le i \le j \le p \quad \text{vs.}$$

$$H_a : \text{there are } m_a \text{ nonzero } \delta_{ij} \text{ with strength } \delta_a.$$

- $n = n_1 n_2/(n_1 + n_2)$ and $q = p(p+1)/2$
- $m_a = \lfloor q^{(1-\beta)} \rfloor$ for a $\beta \in (1/2, 1)$
- $\delta_{ij} = \delta_a = \sqrt{4r \log(p)/n}$ if $\delta_{ij} \neq 0$

Detection boundary for MTT

- The standardized signal strength

$$r_{ij} = r/\{(1-\kappa)\theta_{ij1} + \kappa\theta_{ij2}\}$$

where $\kappa = \lim_{n_1,n_2\to\infty} n_1/(n_1+n_2)$, $\theta_{ij1} = \mathsf{Var}\{(X_{ki}-\mu_{1i})(X_{kj}-\mu_{1j})\}$ and $\theta_{ij2} = \mathsf{Var}\{(Y_{ki}-\mu_{2i})(Y_{kj}-\mu_{2j})\}$

- The maximal and minimal standardized signal strength

$$\bar{r} = \max_{(i,j):\sigma_{ij1}\neq\sigma_{ij2}} r_{ij} \quad \text{and} \quad \underline{r} = \min_{(i,j):\sigma_{ij1}\neq\sigma_{ij2}} r_{ij}.$$

- The class of covariances with sparse and weak differences:

$$\mathcal{C}(\beta,\bar{r},\underline{r}) = \big\{(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2) : \text{under sparse and weak alternatives } H_a, \bar{r},\underline{r} \text{ and assumptions defined previously}\big\}$$

- For any $(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2) \in \mathcal{C}(\beta,\bar{r},\underline{r})$, the power of the MTT is

$$\mathsf{Power}_n(\mathbf{\Sigma}_1,\mathbf{\Sigma}_2) = P\big[\mathcal{V}_n(s_0) > \{q_\alpha + b(\log(p),s_0,\eta)\}/a(\log(p))|\mathbf{\Sigma}_1,\mathbf{\Sigma}_2\big]$$

Detection boundary for MTT

- Consider $\xi \in (0, 2]$ for $n \sim p^{\xi}$ and $\xi = 0$ for $\log p \sim n^{\varpi}$,

$$\rho^*(\beta, \xi) = \begin{cases} \frac{(\sqrt{4-2\xi}-\sqrt{6-8\beta-\xi})^2}{8}, & 1/2 < \beta \leq 5/8 - \xi/16, \\ \beta - 1/2, & 5/8 - \xi/16 < \beta \leq 3/4, \\ (1-\sqrt{1-\beta})^2, & 3/4 < \beta < 1. \end{cases}$$

- Under the previous assumptions, as $n, p \to \infty$,
  - If $\underline{r} > \rho^*(\beta, \xi)$, $\inf_{(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \in \mathcal{C}(\beta, \bar{r}, \underline{r})} \text{Power}_n(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \to 1$;
  - If $\bar{r} < \rho^*(\beta, \xi)$, $\sup_{(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \in \mathcal{C}(\beta, \bar{r}, \underline{r})} \text{Power}_n(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \to 0$.
- As $\xi \to 2$, $\rho^*(\beta, \xi)$ approaches to $\rho(\beta)$, which is the optimal detection boundary for testing the means with uncorrelated Gaussian data.
- Restricting $s \geq s_0 = 1/2 - \xi/4$ elevates the detection boundary $\rho^*(\beta, \xi)$ of the proposed MTT for $1/2 < \beta \leq 5/8 - \xi/16$ as a price for controlling the size of the test.

Detection boundary for MTT

- Consider $\xi \in (0, 2]$ for $n \sim p^{\xi}$ and $\xi = 0$ for $\log p \sim n^{\varpi}$,

$$\rho^*(\beta, \xi) = \begin{cases} \frac{(\sqrt{4-2\xi} - \sqrt{6-8\beta-\xi})^2}{8}, & 1/2 < \beta \leq 5/8 - \xi/16, \\ \beta - 1/2, & 5/8 - \xi/16 < \beta \leq 3/4, \\ (1 - \sqrt{1-\beta})^2, & 3/4 < \beta < 1. \end{cases}$$

- Under the previous assumptions, as $n, p \to \infty$,
  - If $\underline{r} > \rho^*(\beta, \xi)$, $\inf_{(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \in \mathcal{C}(\beta, \bar{r}, \underline{r})} \text{Power}_n(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \to 1$;
  - If $\bar{r} < \rho^*(\beta, \xi)$, $\sup_{(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \in \mathcal{C}(\beta, \bar{r}, \underline{r})} \text{Power}_n(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \to 0$.
- As $\xi \to 2$, $\rho^*(\beta, \xi)$ approaches to $\rho(\beta)$, which is the optimal detection boundary for testing the means with uncorrelated Gaussian data.
- Restricting $s \geq s_0 = 1/2 - \xi/4$ elevates the detection boundary $\rho^*(\beta, \xi)$ of the proposed MTT for $1/2 < \beta \leq 5/8 - \xi/16$ as a price for controlling the size of the test.

Detection boundary for MTT

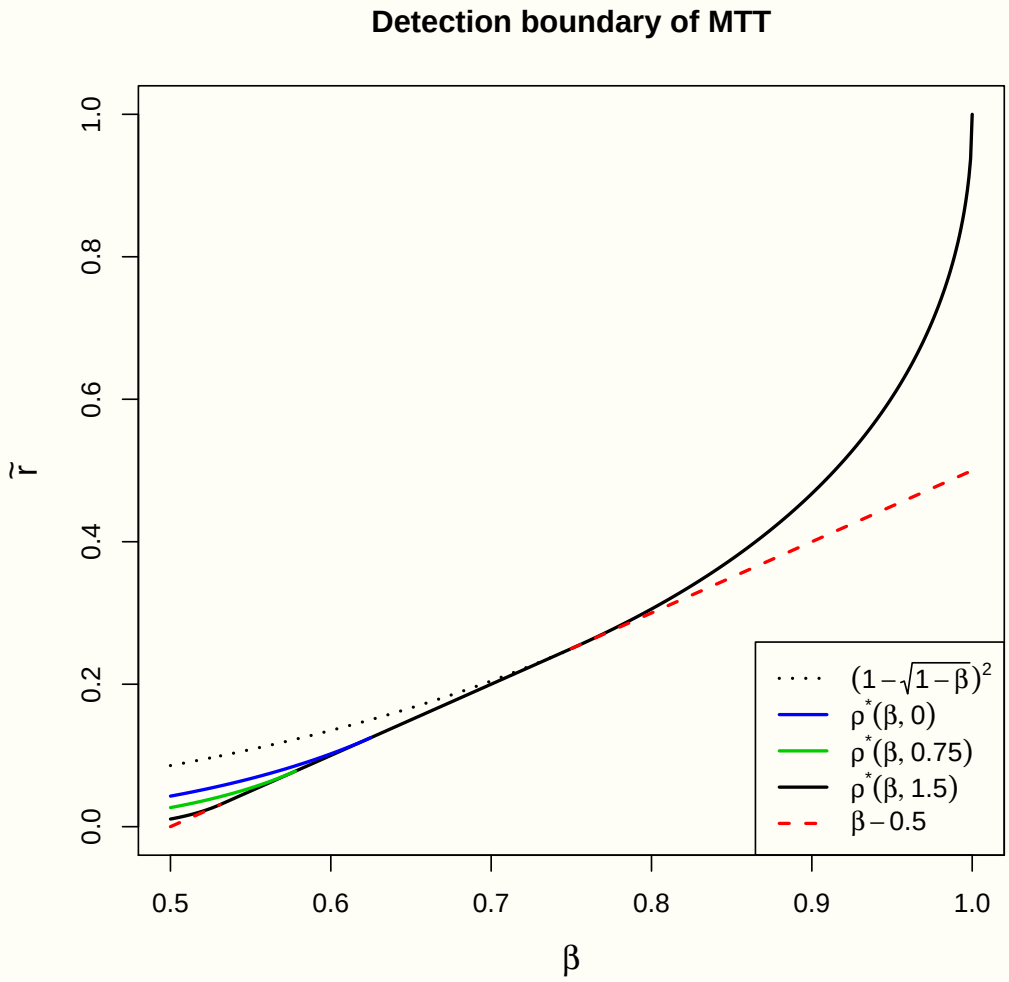


Figure: The detection boundary $\rho^*(\beta, \xi)$ of the proposed MTT for $\xi = 0, 0.75, 1.5$.

Some remarks on detection boundary

- The $L_{\max}$ test of Cai et al. (2013) is consistent if the maximal standardized signal strength $\bar{r} > 4$.
- The L_2 test of Li and Chen (2012) does not have non-trivial power when $\beta > 1/2$.
- The proposed multi-level thresholding test is more powerful in detecting sparse and weak signals as it only requires $r_{ij} \in (0, 1)$.
- Detection boundaries indicate that the differences between $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are at the order of $\sqrt{\log(p)/n}$.
- Is the order $\sqrt{\log(p)/n}$ minimax optimal?

Minimax optimality

- Let $\mathcal{W}_\alpha$ be the collection of all α-level tests for the hypotheses, i.e., $P(W_\alpha = 1|H_0) \leq \alpha$ for any $W_\alpha \in \mathcal{W}_\alpha$.
- Define

$$\underline{\mathcal{C}}(\beta, c) = \Big\{(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) : \text{under sparse and weak alternatives of } H_a \text{ with } r_{ij} \geq c \text{ for all } \sigma_{ij1} \neq \sigma_{ij2}\Big\}.$$

- Detection boundary results indicate that for sufficiently large constant c, as $n, p \to \infty$,

$$\inf_{(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \in \underline{\mathcal{C}}(\beta, c)} \{\text{Power of the MTT Test}\} \to 1.$$

- The lower bound $(\log(p)/n)^{1/2}$ for signals in $\underline{\mathcal{C}}(\beta, c)$ is optimal, i.e., there is no α-level test that can distinguish H_a from H_0 with probability approaching 1 uniformly over the class $\underline{\mathcal{C}}(\beta, c_0)$ for some $c_0 > 0$ as shown in the following theorem.

Minimax optimality

- Let $\mathcal{W}_\alpha$ be the collection of all α-level tests for the hypotheses, i.e., $P(W_\alpha = 1|H_0) \leq \alpha$ for any $W_\alpha \in \mathcal{W}_\alpha$.
- Define

$$\underline{\mathcal{C}}(\beta, c) = \Big\{(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) : \text{under sparse and weak alternatives of } H_a \text{ with } r_{ij} \geq c \text{ for all } \sigma_{ij1} \neq \sigma_{ij2}\Big\}.$$

- Detection boundary results indicate that for sufficiently large constant c, as $n, p \to \infty$,

$$\inf_{(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \in \underline{\mathcal{C}}(\beta, c)} \{\text{Power of the MTT Test}\} \to 1.$$

- The lower bound $(\log(p)/n)^{1/2}$ for signals in $\underline{\mathcal{C}}(\beta, c)$ is optimal, i.e., there is no α-level test that can distinguish H_a from H_0 with probability approaching 1 uniformly over the class $\underline{\mathcal{C}}(\beta, c_0)$ for some $c_0 > 0$ as shown in the following theorem.

Minimax optimality

- Let $\mathcal{W}_\alpha$ be the collection of all α-level tests for the hypotheses, i.e., $P(W_\alpha = 1|H_0) \leq \alpha$ for any $W_\alpha \in \mathcal{W}_\alpha$.
- Define

$$\underline{\mathcal{C}}(\beta, c) = \big\{(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) : \text{under sparse and weak alternatives of } H_a \text{ with } r_{ij} \geq c \text{ for all } \sigma_{ij1} \neq \sigma_{ij2}\big\}.$$

- Detection boundary results indicate that for sufficiently large constant c, as $n, p \to \infty$,

$$\inf_{(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \in \underline{\mathcal{C}}(\beta, c)} \{\text{Power of the MTT Test}\} \to 1.$$

- The lower bound $(\log(p)/n)^{1/2}$ for signals in $\underline{\mathcal{C}}(\beta, c)$ is optimal, i.e., there is no α-level test that can distinguish H_a from H_0 with probability approaching 1 uniformly over the class $\underline{\mathcal{C}}(\beta, c_0)$ for some $c_0 > 0$ as shown in the following theorem.

Minimax optimality

Theorem

For the Gaussian distributed data and under Assumption **1B**, *for any* $0 < \omega < 1 - \alpha$ *and* $\max\{2/3, (3-\xi)/4\} < \beta < 1$, *there exists a constant* $c_0 > 0$ *such that, as* $n, p \to \infty$,

$$\sup_{W_\alpha \in \mathcal{W}_\alpha} \inf_{(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \in \underline{\mathcal{C}}(\beta, c_0)} P(W_\alpha = 1) \leq 1 - \omega.$$

- Extend the minimax result from the highly sparse signal regime $3/4 < \beta < 1$ in Cai et al. (2013) to $\max\{2/3, (3-\xi)/4\} < \beta < 1$.
- The MTT test is at least minimax rate optimal for $\beta > \max\{2/3, (3-\xi)/4\}$ as the proposed MTT can detect signals at the rate of $\{\log(p)/n\}^{1/2}$ for $\beta > 1/2$.

Minimax optimality

Theorem

For the Gaussian distributed data and under Assumption **1B**, *for any* $0 < \omega < 1 - \alpha$ *and* $\max\{2/3, (3-\xi)/4\} < \beta < 1$, *there exists a constant* $c_0 > 0$ *such that, as* $n, p \to \infty$,

$$\sup_{W_\alpha \in \mathcal{W}_\alpha} \inf_{(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \in \underline{\mathcal{C}}(\beta, c_0)} P(W_\alpha = 1) \leq 1 - \omega.$$

- Extend the minimax result from the highly sparse signal regime $3/4 < \beta < 1$ in Cai et al. (2013) to $\max\{2/3, (3-\xi)/4\} < \beta < 1$.
- The MTT test is at least minimax rate optimal for $\beta > \max\{2/3, (3-\xi)/4\}$ as the proposed MTT can detect signals at the rate of $\{\log(p)/n\}^{1/2}$ for $\beta > 1/2$.

Simulation

- Compare the proposed test with Srivastava and Yanagihara (2010) (SY), Li and Chen (2012) (LC) and Cai, Liu and Xia (2013) (CLX).
- The data are generated from
  - $\mathbf{X}_k = \mathbf{\Sigma}_1^{\frac{1}{2}} \mathbf{Z}_{1k}$ and $\mathbf{Y}_k = \mathbf{\Sigma}_2^{\frac{1}{2}} \mathbf{Z}_{2k}$.
  - $\mathbf{Z}_{1k}$ and $\mathbf{Z}_{2k}$ are i.i.d. random vectors from a common population:
    (i) $N(0, \mathbf{I}_p)$;
    (ii) Gamma distribution where components of $\mathbf{Z}_{1k}$ and $\mathbf{Z}_{2k}$ were i.i.d. standardized Gamma(4,2) with mean 0 and variance 1.

Simulation

- Define $\mathbf{\Sigma}_1^{(0)} = \mathbf{D}_0^{\frac{1}{2}}\mathbf{\Sigma}^{(*)}\mathbf{D}_0^{\frac{1}{2}}$, $\mathbf{D}_0 = \text{diag}(d_1, \ldots, d_p)$, $d_i \overset{i.i.d.}{\sim} U(0.1, 1)$, $\mathbf{\Sigma}^{(*)} = (\sigma_{ij}^*)$:

  Design 1: $\sigma_{ij}^* = 0.4^{|i-j|}$;

  Design 2: $\sigma_{ij}^* = 0.5\mathbb{I}(i = j) + 0.5\mathbb{I}(i, j \in [4k_0 - 3, 4k_0])$.

- Under the null hypothesis, $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}_1^{(0)}$.
- Under the alternatives, $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_1^{(\star)}$ and $\mathbf{\Sigma}_2 = \mathbf{\Sigma}_2^{(\star)}$, where

$$\mathbf{\Sigma}_1^{(\star)} = \mathbf{\Sigma}_1^{(0)} + \epsilon_c \mathbf{I_p} \quad \text{and} \quad \mathbf{\Sigma}_2^{(\star)} = \mathbf{\Sigma}_1^{(0)} + \mathbf{U} + \epsilon_c \mathbf{I_p},$$

  $\mathbf{U} = (u_{kl})_{p\times p}$ is a banded symmetric matrix with $\lfloor q^{1-\beta} \rfloor$ nonzero elements ($u_{kl} = \sqrt{4r \log p/n}$ if $u_{kl} \neq 0$). $\epsilon_c = |\min\{\lambda_{\min}(\mathbf{\Sigma}_1^{(0)} + \mathbf{U}), 0\}| + 0.05$ is used to guarantee the positive definiteness of $\mathbf{\Sigma}_1^{(\star)}$ and $\mathbf{\Sigma}_2^{(\star)}$.

Simulation

- The convergence of MTT to the Gumbel distribution is slow.
- A parametric bootstrap procedure:
  - Calculate the multi-thresholding statistic $\mathcal{V}_n(s_0)$ from the original sample.
  - Under the null hypothesis, $\{\mathbf{X}_k\}_{k=1}^{n_1}$ and $\{\mathbf{Y}_k\}_{k=1}^{n_2}$ were pooled together to estimate the common covariance (Rothman(2012)), denote as $\widehat{\mathbf{\Sigma}}$.
  - For the b-th bootstrap, drew bootstrap samples of $\{\mathbf{X}_k^{*(b)}\}_{k=1}^{n_1}$ and $\{\mathbf{Y}_k^{*(b)}\}_{k=1}^{n_2}$ independently from $N(0, \widehat{\mathbf{\Sigma}})$, then calculate the bootstrapped MTT statistic $\mathcal{V}_n^{*(b)}(s_0)$.
  - Calculate p-value by using $\mathcal{V}_n(s_0)$ and the bootstrapped samples $\{\mathcal{V}_n^{*(1)}(s_0), \ldots, \mathcal{V}_n^{*(B)}(s_0)\}$.

Simulation

- The convergence of MTT to the Gumbel distribution is slow.
- A parametric bootstrap procedure:
  - Calculate the multi-thresholding statistic $\mathcal{V}_n(s_0)$ from the original sample.
  - Under the null hypothesis, $\{\mathbf{X}_k\}_{k=1}^{n_1}$ and $\{\mathbf{Y}_k\}_{k=1}^{n_2}$ were pooled together to estimate the common covariance (Rothman(2012)), denote as $\widehat{\mathbf{\Sigma}}$.
  - For the b-th bootstrap, drew bootstrap samples of $\{\mathbf{X}_k^{*(b)}\}_{k=1}^{n_1}$ and $\{\mathbf{Y}_k^{*(b)}\}_{k=1}^{n_2}$ independently from $N(0, \widehat{\mathbf{\Sigma}})$, then calculate the bootstrapped MTT statistic $\mathcal{V}_n^{*(b)}(s_0)$.
  - Calculate p-value by using $\mathcal{V}_n(s_0)$ and the bootstrapped samples $\{\mathcal{V}_n^{*(1)}(s_0), \ldots, \mathcal{V}_n^{*(B)}(s_0)\}$.

Simulation

Table: Empirical sizes for the tests of Srivastava and Yanagihara (2010) (SY), Li and Chen (2012) (LC), Cai, Liu and Xia (2013) (CLX) and the proposed multi-level thresholding based on the limiting distribution (MTT) and the bootstrap calibration (MTT-BT) for Designs 1 and 2 under the Gaussian distribution with the nominal level of 5%.

p	(n_1, n_2)	SY	LC	CLX	MTT	MTT-BT
			Gaussian	Design 1		
175	(60, 60)	0.048	0.058	0.054	0.088	0.058
277	(80, 80)	0.052	0.052	0.058	0.064	0.056
396	(100, 100)	0.042	0.046	0.058	0.064	0.054
530	(120, 120)	0.056	0.048	0.050	0.056	0.046
			Gaussian	Design 2		
175	(60, 60)	0.060	0.048	0.052	0.094	0.048
277	(80, 80)	0.040	0.060	0.040	0.064	0.052
396	(100, 100)	0.052	0.042	0.044	0.090	0.048
530	(120, 120)	0.050	0.046	0.044	0.060	0.054

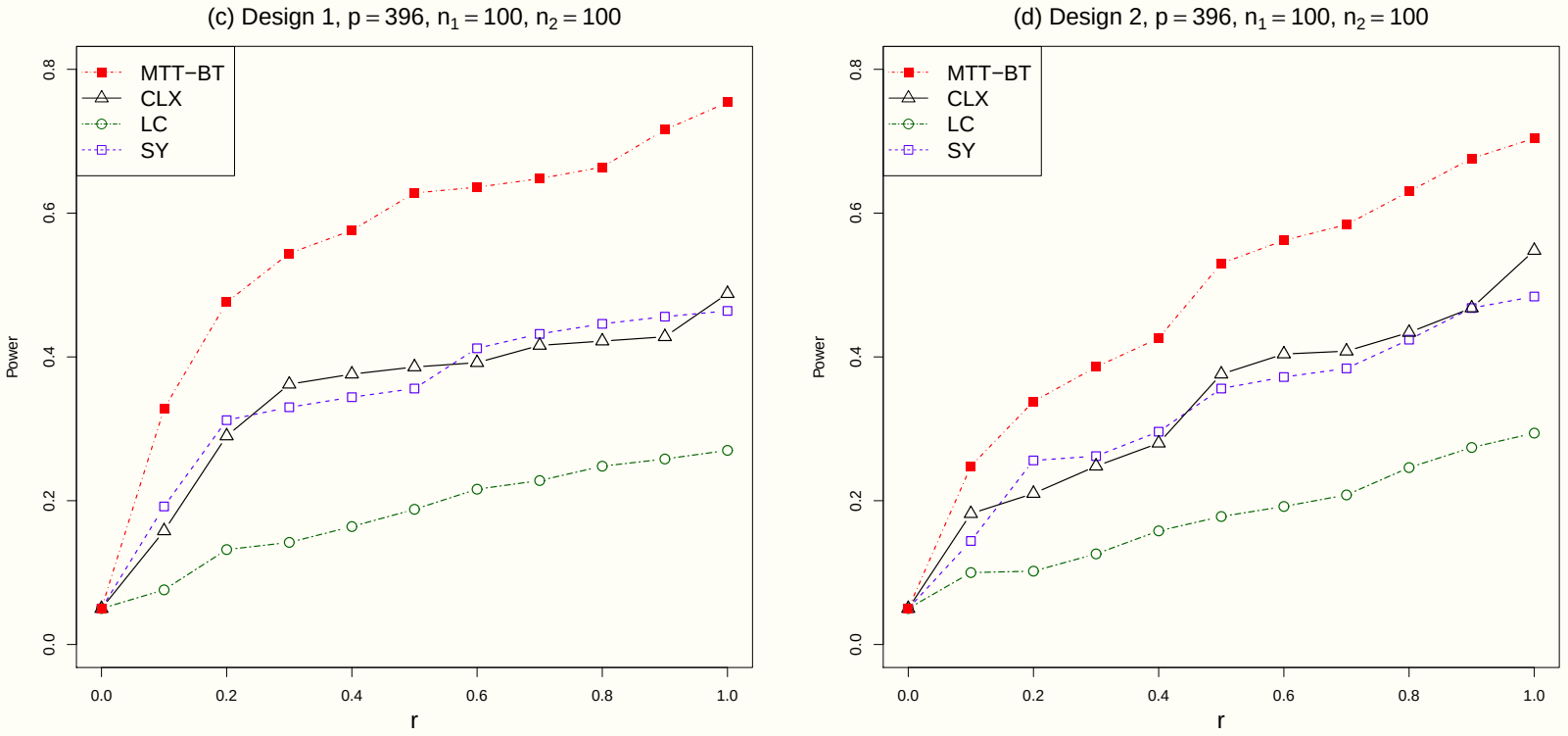
Simulation

Table: Empirical sizes for the tests of Srivastava and Yanagihara (2010) (SY), Li and Chen (2012) (LC), Cai, Liu and Xia (2013) (CLX) and the proposed multi-level thresholding based on the limiting distribution (MTT) and the bootstrap calibration (MTT-BT) for Designs 1 and 2 under the Gamma distribution with the nominal level of 5%.

p	(n_1, n_2)	SY	LC	CLX	MTT	MTT-BT
			Gamma	Design 1		
175	(60, 60)	0.046	0.060	0.066	0.110	0.056
277	(80, 80)	0.060	0.050	0.044	0.076	0.044
396	(100, 100)	0.046	0.052	0.046	0.066	0.054
530	(120, 120)	0.060	0.056	0.048	0.060	0.048
			Gamma	Design 2		
175	(60, 60)	0.070	0.056	0.066	0.108	0.056
277	(80, 80)	0.060	0.058	0.068	0.112	0.044
396	(100, 100)	0.060	0.050	0.044	0.068	0.046
530	(120, 120)	0.054	0.056	0.048	0.056	0.048

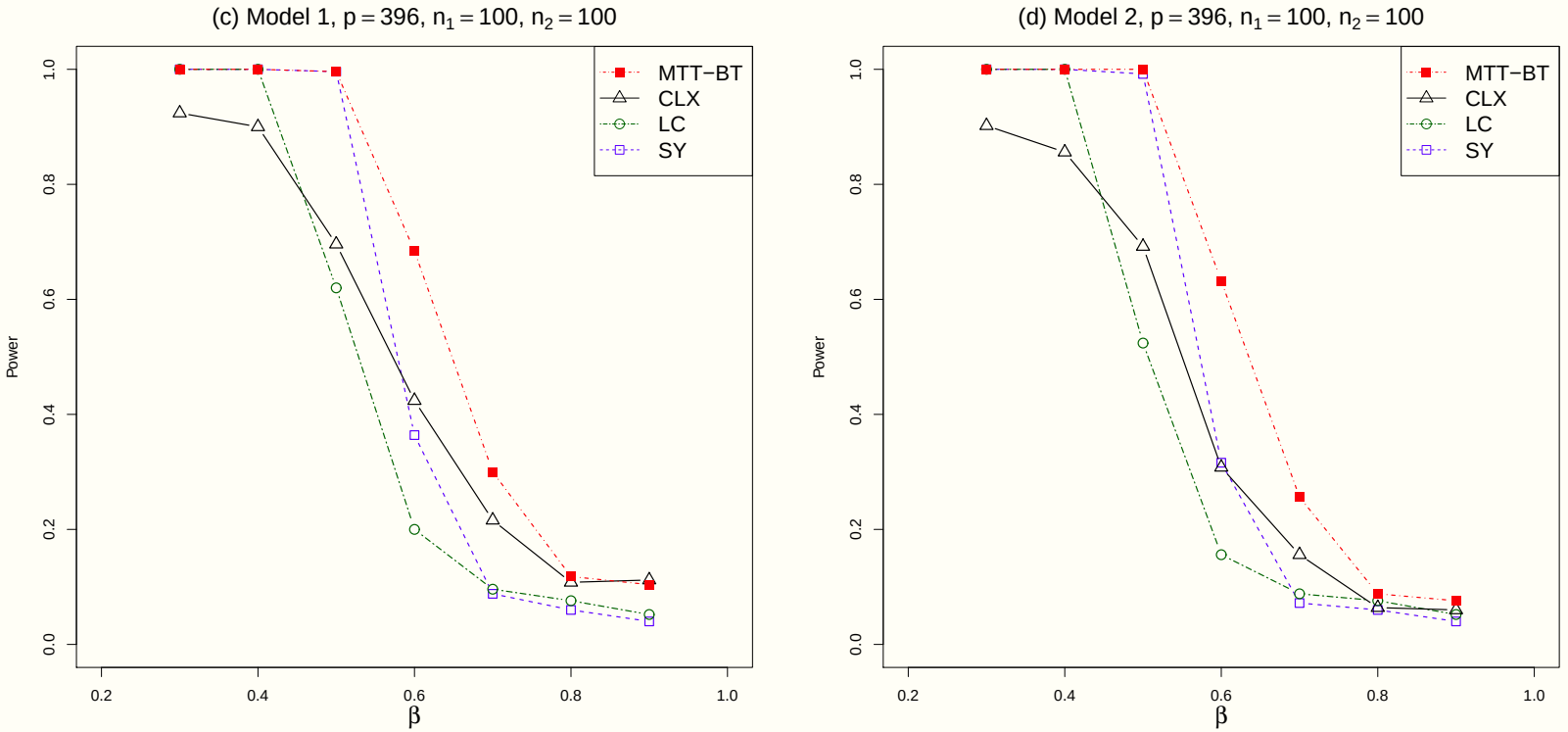
Simulation

Figure: Empirical powers with respect to the signal strength r for the tests of Srivastava and Yanagihara (2010) (SY), Li and Chen (2012) (LC), Cai, Liu and Xia (2013) (CLX) and the proposed thresholding test (MTT-BT) for Designs 1 and 2 with Gaussian innovations under $\beta = 0.6$ when $p = 396$, $n_1 = n_2 = 100$.



Simulation

Figure: Empirical powers with respect to the sparsity level β for the tests of Srivastava and Yanagihara (2010) (SY), Li and Chen (2012) (LC), Cai, Liu and Xia (2013) (CLX) and the proposed thresholding test (MTT-BT) for Designs 1 and 2 with Gaussian innovations under $r = 0.6$ when $p = 396$, $n_1 = n_2 = 100$ respectively.



Real data analysis

- A microarray dataset of large airway epithelial cells with 22283 genes.
- 187 smokers: 97 persons with lung cancer and 90 persons were healthy.
- Gene Ontology (GO) terms for sets of genes under three broad functional categories: Biological Processes (BP), Cellular Components (CC) and Molecular Functions (MF).
- 3063 unique GO terms in the BP category, 317 sets in the CC category and 442 sets in the MF category.
- We are interested in identifying gene-sets with different covariance structures between the smokers with the lung cancer and the controls.
- This will provide useful results toward identifying the differential co-expression networks and the functionally related genes.

Real data analysis

- A microarray dataset of large airway epithelial cells with 22283 genes.
- 187 smokers: 97 persons with lung cancer and 90 persons were healthy.
- Gene Ontology (GO) terms for sets of genes under three broad functional categories: Biological Processes (BP), Cellular Components (CC) and Molecular Functions (MF).
- 3063 unique GO terms in the BP category, 317 sets in the CC category and 442 sets in the MF category.
- We are interested in identifying gene-sets with different covariance structures between the smokers with the lung cancer and the controls.
- This will provide useful results toward identifying the differential co-expression networks and the functionally related genes.

Real data analysis

- We tested

$$H_{g,0} : \mathbf{\Sigma}_{1g} = \mathbf{\Sigma}_{2g} \quad \text{vs.} \quad H_{g,a} : \mathbf{\Sigma}_{1g} \neq \mathbf{\Sigma}_{2g},$$

  where $\mathbf{\Sigma}_{1g}$ and $\mathbf{\Sigma}_{2g}$ denote the population covariance matrices of the cancer and control groups for the gth gene-set.

- Controlling the FDR (false discovery rate) at 5%.
- The proposed test found more significant gene-sets than the other tests among BP and CC categories.
- The gene set GO:0001824 in the BP category was only discovered by the proposed MTT-BT test, which had been shown to be highly correlated with the lung cancer in other studies.

Real data analysis

- We tested

$$H_{g,0} : \mathbf{\Sigma}_{1g} = \mathbf{\Sigma}_{2g} \quad \text{vs.} \quad H_{g,a} : \mathbf{\Sigma}_{1g} \neq \mathbf{\Sigma}_{2g},$$

  where $\mathbf{\Sigma}_{1g}$ and $\mathbf{\Sigma}_{2g}$ denote the population covariance matrices of the cancer and control groups for the gth gene-set.

- Controlling the FDR (false discovery rate) at 5%.
- The proposed test found more significant gene-sets than the other tests among BP and CC categories.
- The gene set GO:0001824 in the BP category was only discovered by the proposed MTT-BT test, which had been shown to be highly correlated with the lung cancer in other studies.

Real data analysis

Table: Cross tabulations of the numbers of significant gene-sets with different covariance matrices by the four tests for the three Gene Ontology categories. The numbers on the diagonals show the significant gene-sets by the four tests, respectively, while the off-diagonal entries are the number of common gene-sets by any two tests.

	Biological Processes				Cellular Component			
Methods	MTT-BT	CLX	LC	SY	MTT-BT	CLX	LC	SY
MTT-BT	64	5	9	2	13	0	3	0
CLX		23	0	0		1	0	0
LC			54	4			9	1
SY				9				1
	Molecular Functions							
MTT-BT	10	2	3	0				
CLX		5	1	0				
LC			14	1				
SY				1				

Conclusion

- We proposed a powerful multi-thresholding test for high-dimensional covariances;
- The detection boundary of the MTT test;
- The minimax optimality;
- The MTT test can be extended to correlation matrices.

Thank you!

References

- Bai, Z., Jiang, D., Yao, J.-F. and Zheng, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *The Annals of Statistics*, 37, 3822-3840.
- Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica*, 6, 311 C329.
- Berbee, H. (1979). *Random Walks with Stationary Increments and Renewal Theory*, Amsterdam: Mathematical Centre.
- Cai, T., Liu, W. D. and Xia, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the Americain Statistical Association*, 108, 265-277.
- Chen, S. X., Li, J. and Zhong P. S. (2019). Two-Sample and ANOVA Tests for High Dimensional Means, *The Annals of Statistics*, 47, 1443-1474.
- Chen, S. X., Guo, B. and Qiu, Y. (2019). Multi-level thresholding test for high dimensional covariance matrices. *Manuscript*.

References

- Chen, S. X. and Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, 38 808-835.
- Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics*, 32, 962-994.
- Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81, 425-455.
- Fan, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *Journal of the Americain Statistical Association*, 91, 674-688.
- Li, J. and Chen, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *The Annals of Statistics*, 40, 908-940.
- Mokkadem, A. (1988). Mixing properties of ARMA processes. *Stochastic processes and their applications*, 29, 309-315.
- Srivastava, M. S. (2009). A test for the mean vector with fewer observations than the dimension under non-normality. *Journal of Multivariate Analysis*, 100, 518-532.

References

- Srivastava, M. S., and Yanagihara, H. (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *Journal of Multivariate Analysis*, 101, 1319-1329.
- Tukey, J. W. (1976). The higher criticism. Course Notes, Statistics 411, Princeton Univ.
- Ingster, Y. I. (1999). Minimax detection of a signal for ℓ_n^p-balls, *Math. Methods Statist*, 7, 401-428.
- Zhong, P., Chen, S. X. and Xu M. (2013). Tests alternative to higher criticism for high dimensional means under sparsity and column-wise dependence, *The Annals of Statistics*, 41, 2820-2851