A comparison study on the robustness of methods for brain connectivity analysis

Chaeree Kang, Narim Baek, Yeonkyung Lee and Donghwan Lee*

Department of Statistics, Ewha Womans University, Korea

Abstract

In the brain connectivity analysis, various methods have been developed for testing the correlations of the observations at multiple regions of interest. However, they often require specific distributional assumptions of observations. In this study, we compare the robustness of the false discovery rate (FDR) control of existing methods when the underlying distribution assumption is misspecified. We use numerical simulation to investigate how the misspecification of distribution affects the reliability of the estimates of FDR. A real data example of a resting-state functional MRI study is illustrated.

Introduction

What is brain connectivity?
The relationship between nodes(Region of Interests; ROIs)

Multiple testing problem
Testing the null hypotheses simultaneously,

$$H_{0ij}: \rho_{ij1} = \rho_{ij2}, \qquad 1 \le i \le j \le p$$

p: number of nodes (ROIs)
$N = p(p-1)/2$: number of tests
ρ_{ijm}: pairwise correlation between ROI(i) and ROI(j) of group m

	Declared as null	Declared as alternative	Total
Null	True Negative : U	False Positive : V	N_0
Alternative	False Negative : W	True Positive : S	N_1
Total	$N-R$	R	N

Table 1. Possible outcomes of multiple testing

- FDP = V/R : False discovery proportion
  c.f. FDR=$E(\text{FDP}|R>0)P(R>0)$: False discovery rate (Benjamini and Hochberg, 1995)
- FNDP = $W/(N-R)$: False nondiscovery proportion
- 1-FNDP = $U/(N-R)$: True negative among declared as null hypothesis

Methods

Test statistics for correlation testing

Under $H_{0ij}: \rho_{ij1} = \rho_{ij2}$,

$$T_{ij} = \sqrt{\frac{(n_1-3)(n_2-3)}{n_1+n_2-6}}\,(Z_{ij1} - Z_{ij2})$$

Step 1: Compute p-values by using T_{ij} and its null distribution→ $N(0,1)$.
where Fisher's z-transformation

$$Z_{ijm} = \frac{1}{2}\log\left(\frac{1+\hat{\rho}_{ijm}}{1-\hat{\rho}_{ijm}}\right)$$

with $\hat{\rho}_{ijm}$ is the sample correlation between ROI(i) and ROI(j) of group m.

Conventional testing procedure: BH & BY

Step 1: Compute p-values by using T_{ij} and its null distribution.

Step 2: Apply conventional FDR procedure such as Benjamini and Hochberg (1995) procedure (BH) or Benjamini and Yekutieli (2001) procedure (BY).

Cai and Liu (2016)'s procedure (CL)

$$T_{ij} = \frac{\hat{\rho}_{ij1} - \hat{\rho}_{ij2}}{\sqrt{\frac{\widehat{k_1}}{n_1}\left(1-\rho_{ij}\right)^2 + \frac{\widehat{k_2}}{n_2}\left(1-\rho_{ij}\right)^2}}$$

$$\widetilde{t_1} = \inf\left\{0 \le t \le b_p : \times \frac{G(t)(p^2-p)/2}{\max\left\{\sum_{1\le i<j\le p} I\{|T_{ij}| \ge t\}, 1\right\}} \le \alpha\right\} \ or\ \sqrt{4logp}$$

$$a_p = 2\log(logp) \qquad b_p = \sqrt{4logp - a_p} \qquad \mathrm{G(t)} = 2 - 2\Phi(t)$$

Step 1: compute test-statistics T_{ij} and threshold level $\widetilde{t_1}$.

Step 2: Reject H_{0ij} $whenever$ $|T_{ij}| \ge \widetilde{t_1}$.

Cai and Liu (2016)'s procedure with bootstrap (CLB)

Step 1: Draw resamples randomly with replacement form X, Y respectively.
where X_i $(i = 1, \dots, n_1)$ and Y_i $(i = 1, \dots, n_2)$ are p-dimensional random vectors with mean μ_1, μ_2 and correlation matrix $\Sigma_1 = (\rho_{ij1})$, $\Sigma_2 = (\rho_{ij2})$, respectively.

Step 2: For each resamples, compute test-statistics T_{ij} and threshold level $\widetilde{t_1}$.

Step 3: Reject H_{0ij} $whenever$ $|T_{ij}| \ge \widetilde{t_1}$.

Modification of Lee and Lee's method (LL)

Idea: The test statistics T_{ij} are also correlated. Consider the multivariate distribution of T (N-dimensional random vector).

Step 1. Under the null, T follows multivariate normal distribution with mean 0 and covariance matrix.

$$\Sigma = \frac{(n_1-3)(n_2-3)}{n_1+n_2-6}(\Sigma_1 + \Sigma_2)$$

where the elements in $\sum_m$ are computed from $Cov(Z_{ijm}, Z_{khm})$.

Step 2. To get the asymptotically consistent estimator of FDP, using Eigen-decomposition, express $\Sigma = LL^T + A$ with $L = \left(\sqrt{\lambda_1}v_1, \cdots, \sqrt{\lambda_k}v_k\right)$. Then construct the factor model

$$T = \mu + LW + e$$

with $W \sim N(0, I_k), e \sim N(0, A)$) (Fan et al., 2012).

L is computed based on the idea of PCA(Principal Component Analysis). Choose the first k $PC_1, \dots PC_K$ which explains 80% of the total variance.

$$\frac{\sum_{j=1}^{k}\lambda_j}{\sum_{j=1}^{p}\lambda_j} \ge 0.8$$

Step 3. Reject H_{0ij} if p-value $p_{ij} = 2 * (1 - \Phi(|T_{ij}|))$ is less than a given threshold α .Then, compute the consistent estimator of FDP, $\widehat{\text{FDP}}(\alpha)$.

Simulation Results

Number of simulations = 150, Type I error level: α=0.05, $p = 50$, $n = 20$, $(N = 1025, N_1 = 200)$

Distribution

- **Normal Mixture distribution**
  Generate U1 & U2 independently from U(0,1) Z1 & Z2 independently from N(0,sigma)
  X = U1*Z1, Y=U2*Z2
- **Normal distribution**
  Generate X & Y independently from N(0,sigma)
- **T distribution**
  Generate X & Y independently from t(0, sigma, df=6)
- **Exponential distribution**
  Generate X & Y independently from Exp(0, sigma)

Setting 1: Equal two groups model ($\rho_{ij} = 0.5$ for all $i \neq j$, $\rho_{ij} = 1$ for all $i = j$)

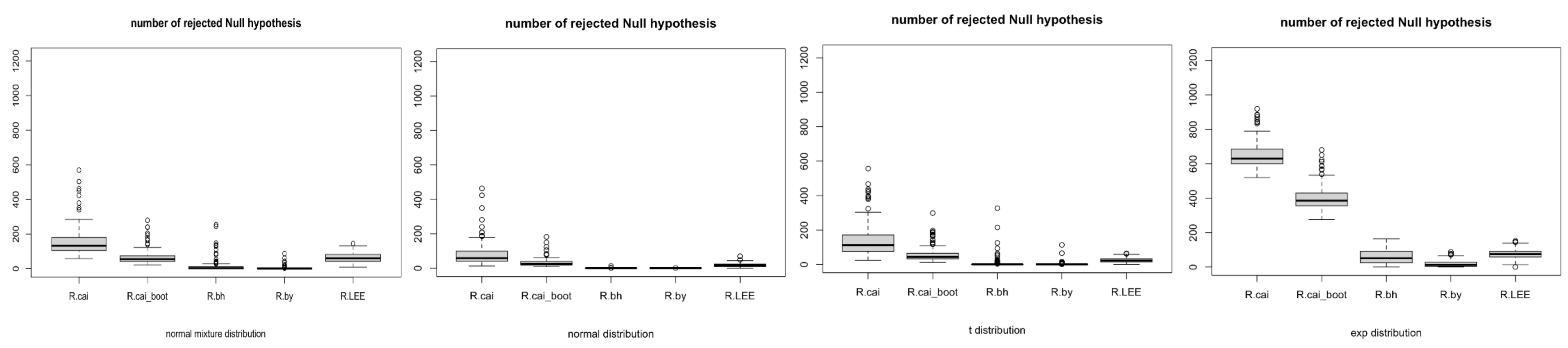


Figure 1. number of rejected in Equal two groups

Setting 2: Five nodes different group model (five nodes of X and Y are generated from U(0,0.1), U(0.9,1) respectively.)

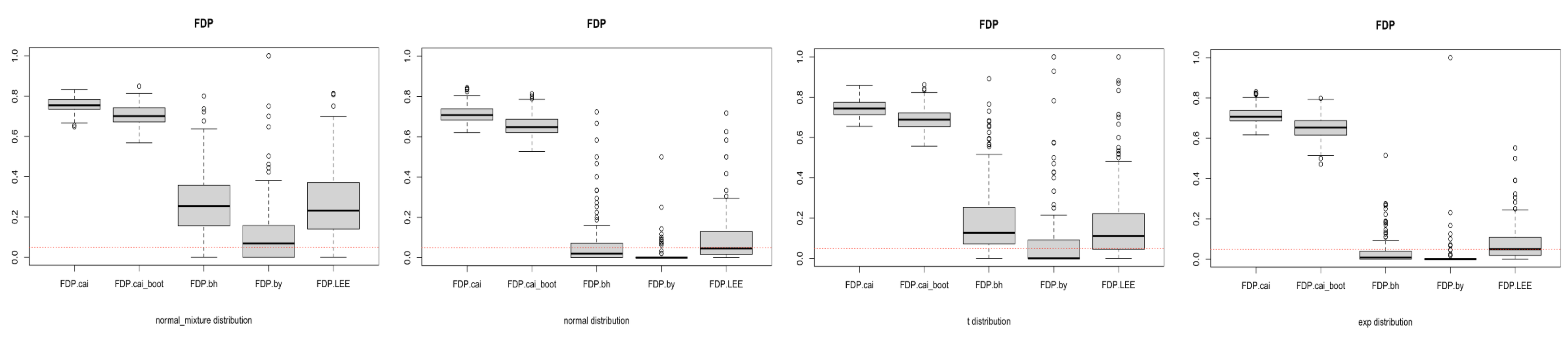
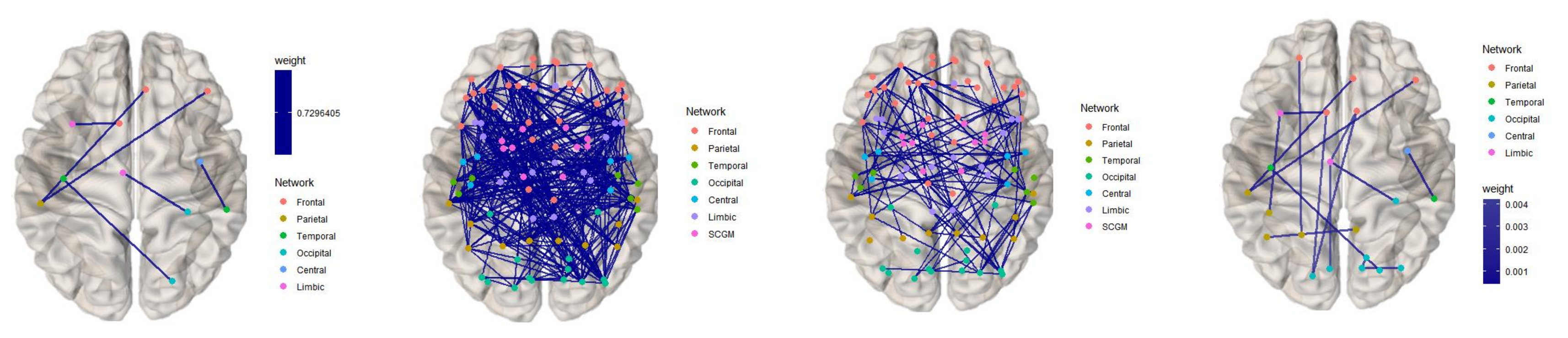


Figure 2. Actual FDP values when controlling FDP (or FDR) at 0.05

Real data analysis

Compare the brain connectivity between control group (n_1=21) and deaf group (n_2=12)
ROIs: Automated Anatomical Labeling 2 (AAL2) atlas ($p = 94$)



BH CL CLB LL

Figure 3. Significant connectivities between control group and deaf group

Among 4371 (=94*93/2) possible pairwise correlations (edges), the proposed method finds 15 significant at FDP level =0.05 (Note: BH can find significant correlations at level 0.73, and BY cannot find any significant correlations at any level)

significant pairs of edge : (Temporal pole: middle temporal gyrus_left, Olfactory cortex_left), (Gyrus rectus_right, SupraMarginal gyrus_left), (Precentral gyrus_right, Middle temporal gyrus_right), (Fusiform gyrus_right, Middle cingulate & paracingulate gyri_left), (IFG pars orbitalis_right, SupraMarginal gyrus_left), (Superior occipital gyrus_right, Heschl's gyrus_left)

Conclusion

We found that conventional methods show either conservative or liberal under the same FDR level. Taking account of the correlated test statistics, Lee and Lee's method outperforms in terms of the validity of FDR control.

References

- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society B, 57, 289-300.
- Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency, Annals of Statistics, 29, 1165-1188.
- Fan, J., Han, X. and Gu, W. (2012). Estimating false discovery proportion under arbitrary covariance dependence, Journal of the American Statistical Association, 107, 1019-1035
- Lee, D., and Lee, W. (2022+) Estimation of false discovery proportion for correlation testing in neuroimaging data, manuscript for submission.
- T. Tony Cai & Weidong Liu (2016) Large-Scale Multiple Testing of Correlations, Journal of the American Statistical Association, 111:513, 229-240.