Statistics and Its Interface Volume 9 (2016) 33–35

Correction to the paper "Optimal False Discovery Rate Control for Dependent Data"

Jichun Xie, T. Tony Cai, and Hongzhe Li*

We have found a mistake in the proof Theorem 6 in our published paper "Optimal False Discovery Rate Control for Dependent Data" [4]. We apologize to the readers and thank Professor Jens Ledet Jensen at Aarhus University for his question which led to identification of this mistake. We provide here a corrected proof of Theorem 6 with further clarifications of the assumptions.

In the GWAS setting that we consider the paper, X_i's are often the Z-scores with $\text{Var}(X_i) = 1$ for very large sample sizes. We assume that $\sigma_{ii} = 1$. We define the true latent parameter $\theta_{0,i}$: if $\theta_{0,i} = 0$, $X_i \sim N(0, 1)$; and if $\theta_{0,i} = 1$, $X_i \sim N(\mu_i, 1)$. Also, we denote the working latent parameter as θ_i, which is used to define the likelihood ratio $f(X_i \mid \theta_i = 1)/f(X_i \mid \theta_i = 0)$ and $f(\mathbf{X} \mid \theta_i = 1)/f(\mathbf{X} \mid \theta_i = 0)$.

Assumption (A) can be weakened as following:

Assumption (A'). The non-null proportion p satisfies $m^{-\tau_1} \le p \le 1 - m^{-\tau_1}$ for some constant $0 < \tau_1 < 1$.

Let the symbol "$\circ$" be the operator of Hadamard product. Assumption (B) can be clarified as following:

Assumption (B') The data $\mathbf{x}^{(m)} = (x_1, \ldots, x_m)$ is an observation of the random variable $\mathbf{X}^{(m)} = (X_1, \ldots, X_m)$, which follows a multivariate normal distribution given the mean $\boldsymbol{\mu}^{(m)} \circ \boldsymbol{\theta}^{(m)} = (\mu_1\theta_1, \ldots, \mu_m\theta_m)$, *i.e.*

$$\mathbf{X}^{(m)} \mid \boldsymbol{\mu}^{(m)}, \boldsymbol{\theta}^{(m)} \sim N(\boldsymbol{\mu}^{(m)} \circ \boldsymbol{\theta}^{(m)}, \boldsymbol{\Sigma}^{(m)}).$$

Here $\boldsymbol{\Sigma}^{(m)}$ is the covariance matrix with diagonal elements equal to 1, and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)^{\mathrm{T}}$ is a random vector, with each μ_i independently following a distribution with CDF $G(\mu)$. Assume for some constant $\tau_2 > \tau_1$,

$$G\{(2\tau_2 \log m)^{1/2}\} - G\{-(2\tau_2 \log m)^{-1/2}\} = 0.$$

It guarantees that

$$\text{pr}\left\{|\mu_i| \ge (2\tau_2 \log m)^{1/2}\right\} = 1, \quad i = 1, \ldots, m.$$

Remark: Compared to the original Assumption (A), Assumption (A') allows a larger range of p. The condition imposed on the CDF function $G(\mu)$ in Assumption (B') is very weak. For example, consider the case where the non-null proportion is small as $p = m^{-\tau_1}$, for some $1/2 < \tau_1 < 1$ (also known as the sparse case). By [2] and [1], if $|\mu_i| < (2\tau_1 \log m)^{1/2}$, it is not possible to test a single signal with diminishing type I and type II errors. Further, by [3], in order to almost recover all the signals, τ_2 has to be no smaller than $1 + (1 - \tau_1)^2$. Note that $1/2 < \tau_1 < 1$. Therefore, Assumption (B') imposes a weaker condition on the signal strength than what is needed for signal recovery.

Assumption (C) can be weakened as follows:

Assumption (C') The covariance matrix $\boldsymbol{\Sigma}^{(m)}$ is positive definite.

Theorem 6. Under the assumptions (A'), (B') and (C'), define $T_{OR,i}$ and $T_{MG,i}$ as in equation (6) and equation (12). Then for all $\epsilon > 0$ and for all $i = 1, \ldots, m$,

$$\lim_{m\to\infty} \text{pr}\left(|T_{MG,i} - T_{OR,i}| > \epsilon\right) = 0.$$

Proof of Theorem 6. We prove the results for $i = 1$. Let $\mathbf{X}_2 = (X_2, \ldots, X_m)^{\mathrm{T}}$ be the subvector of the random vector $\mathbf{X}$ without the first variable X_1. Correspondingly, let $\boldsymbol{\theta}_2 = (\theta_2, \ldots, \theta_m)^{\mathrm{T}}$ and $\boldsymbol{\mu}_2 = (\mu_2, \ldots, \mu_m)^{\mathrm{T}}$. Define $2\varepsilon = \tau_2 - \tau_1 > 0$. Then $\tau_2 - \tau_1 - \varepsilon = \varepsilon > 0$.

The proof has several steps.

1). We temporarily fix $\boldsymbol{\mu}$ and $\boldsymbol{\theta}_2$. WLOG, assume $\mu_1 \ge (2\tau_2 \log m)^{1/2} > 0$.

1.1) We first consider the case that the true latent variable $\theta_{0,1} = 0$. We show that with probability greater than $1 - O\{(\log m)^{-1/2}\}$,

$$f(X_1 \mid \theta_1 = 1, \mu_1) < m^{-\tau_2+\varepsilon} \cdot f(X_1 \mid \theta_1 = 0, \mu_1), \tag{1}$$

$$f(\mathbf{X} \mid \theta_1 = 1, \boldsymbol{\theta}_2, \boldsymbol{\mu}) < m^{-\tau_2+\varepsilon} \cdot f(\mathbf{X} \mid \theta_1 = 0, \boldsymbol{\theta}_2, \boldsymbol{\mu}). \tag{2}$$

Note that

$$\begin{aligned}\frac{f(X_1 \mid \theta_1 = 1, \mu_1)}{f(X_1 \mid \theta_1 = 0, \mu_1)} &= \exp\left\{-\frac{1}{2}(X_1 - \mu_1)^2 + \frac{1}{2}X_1^2\right\} \\ &= \exp\left(\mu_1 X_1 - \frac{1}{2}\mu_1^2\right). \end{aligned} \tag{3}$$

We assume $\theta_{0,1} = 0$, so $X_1 \sim N(0, 1)$ and for sufficiently large m

$$\begin{aligned}\text{pr}\left(X_1 > (\log\log m)^{1/2}\right) &= \Phi(-(\log\log m)^{1/2}) \\ &\le \frac{\varphi((\log\log m)^{1/2})}{(\log\log m)^{1/2}}\end{aligned}$$

*Corresponding author.

$$(4) \qquad < (\log m)^{-1/2}.$$

Here $\Phi(\cdot)$ is the cdf of $N(0,1)$ and $\varphi(\cdot)$ is the corresponding pdf.

Note that $\mu_1 = (2\tau_2 \log m)^{1/2}$. Then for all sufficiently large m, with probability greater than $1-(\log m)^{-1/2}$,

$$\frac{f(X_1 \mid \theta_1 = 1, \mu_1)}{f(X_1 \mid \theta_1 = 0, \mu_1)} < m^{-\tau_2+\varepsilon}.$$

Let $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$ be the precision matrix of $\mathbf{X}$. Corresponding to the partition $\mathbf{X} = (X_1, \mathbf{X}_2)$, we can write

$$\boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Omega} = \begin{pmatrix} \Omega_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix}.$$

where Σ_{11} and Ω_{11} are scalars.

Based on the model, given $\boldsymbol{\theta}_2$ and $\boldsymbol{\mu}$,

$$\mathbf{X}_2 \sim N(\boldsymbol{\mu}_2 \circ \boldsymbol{\theta}_2, \boldsymbol{\Sigma}_{22}).$$

Conditioning on $\boldsymbol{\theta}_2$ and $\boldsymbol{\mu}$, we have

$$\begin{aligned} &\frac{f(\mathbf{X} \mid \theta_1 = 1, \boldsymbol{\theta}_2, \boldsymbol{\mu})}{f(\mathbf{X} \mid \theta_1 = 0, \boldsymbol{\theta}_2, \boldsymbol{\mu})} \\ &= \exp\Big\{ -\frac{1}{2}(X_1-\mu_1)^2\Omega_{11} + (X_1-\mu_1)\boldsymbol{\Omega}_{12}(\mathbf{X}_2 - \boldsymbol{\mu}_2 \circ \boldsymbol{\theta}_2) \\ &\qquad - \frac{1}{2}(\mathbf{X}_2 - \boldsymbol{\mu}_2 \circ \boldsymbol{\theta}_2)^{\mathrm{T}}\boldsymbol{\Omega}_{22}(\mathbf{X}_2 - \boldsymbol{\mu}_2 \circ \boldsymbol{\theta}_2) \\ (5) &\qquad + \frac{1}{2}X_1^2\Omega_{11} - X_1\boldsymbol{\Omega}_{12}(\mathbf{X}_2 - \boldsymbol{\mu}_2 \circ \boldsymbol{\theta}_2) \\ &\qquad + \frac{1}{2}(\mathbf{X}_2 - \boldsymbol{\mu}_2 \circ \boldsymbol{\theta}_2)^{\mathrm{T}}\boldsymbol{\Omega}_{22}(\mathbf{X}_2 - \boldsymbol{\mu}_2 \circ \boldsymbol{\theta}_2) \Big\} \\ (6) \\ &= \exp\left\{ X_1\mu_1\Omega_{11} - \frac{1}{2}\mu_1^2\Omega_{11} - \mu_1\Omega_{11}\frac{\boldsymbol{\Omega}_{12}(\mathbf{X}_2 - \boldsymbol{\mu}_2 \circ \boldsymbol{\theta}_2)}{\Omega_{11}} \right\} \end{aligned}$$

Let $Z_2 = \boldsymbol{\Omega}_{12}(\mathbf{X}_2 - \boldsymbol{\mu}_2 \circ \boldsymbol{\theta}_2)/\Omega_{11}$. Then

$$Z_2 | (\boldsymbol{\theta}_2, \boldsymbol{\mu}_2) \sim N(0, \boldsymbol{\Omega}_{12}\boldsymbol{\Sigma}_{22}\boldsymbol{\Omega}_{21}/\Omega_{11}^2).$$

We now show that the variance of Z_2 is upper bounded by 1.

By the equality

$$\boldsymbol{\Sigma}\boldsymbol{\Omega} = \begin{pmatrix} \Sigma_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \begin{pmatrix} \Omega_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0} & \mathbf{I} \end{pmatrix},$$

we have

$$(7) \qquad \boldsymbol{\Sigma}_{12}\boldsymbol{\Omega}_{21} = 1 - \Sigma_{11}\Omega_{11}.$$

By another equality $\boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega} = \boldsymbol{\Omega}$ with the same partition, we have

$$(8) \qquad \Omega_{11}^2\Sigma_{11} + 2\Omega_{11}\boldsymbol{\Sigma}_{12}\boldsymbol{\Omega}_{21} + \boldsymbol{\Omega}_{12}\boldsymbol{\Sigma}_{22}\boldsymbol{\Omega}_{21} = \Omega_{11}.$$

By (7) and (8),

$$\begin{aligned} (9) \quad 0 &\le \boldsymbol{\Omega}_{12}\boldsymbol{\Sigma}_{22}\boldsymbol{\Omega}_{21}/\Omega_{11}^2 \\ &= (\Sigma_{11}\Omega_{11} - 1)/\Omega_{11} = (\Omega_{11} - 1)/\Omega_{11} < 1. \end{aligned}$$

Also by (9), we have $\Omega_{11} \ge 1$, and

$$\begin{aligned} (10) \quad &\mathrm{pr}\,(Z_2 < -(\log\log m)^{1/2} \mid \boldsymbol{\theta}_2, \boldsymbol{\mu}) \\ &\qquad < \Phi(-(\log\log m)^{1/2}) < (\log m)^{-1/2}. \end{aligned}$$

By (4) and (10), given $\boldsymbol{\theta}_2$ and $\boldsymbol{\mu}$, with probability greater than $1 - 2(\log m)^{-1/2}$,

$$\begin{aligned} \frac{f(\mathbf{X} \mid \theta_1 = 1, \boldsymbol{\theta}_2, \boldsymbol{\mu})}{f(\mathbf{X} \mid \theta_1 = 0, \boldsymbol{\theta}_2, \boldsymbol{\mu})} &< \exp\{2(\log\log m)^{1/2}\mu_1\Omega_{11} - \mu_1^2\Omega_{11}/2\} \\ &< m^{-\tau_2+\varepsilon}, \end{aligned}$$

where the second inequality is due to $\Omega_{11} \ge 1$ and $\mu_1 = (2\tau_2 \log m)^{1/2}$. This implies (2).

1. ii) We now turn to the case where the true latent variable $\theta_{0,1} = 1$. We show that with probability greater than $1 - 2\{(\log m)^{-1/2}\}$,

$$(11) \qquad f(X_1 \mid \theta_1 = 1, \mu_1) > c_2 m^{\tau_2-\varepsilon} \cdot f(X_1 \mid \theta_1 = 0, \mu_1),$$

$$(12) \qquad f(\mathbf{X} \mid \theta_1 = 1, \boldsymbol{\theta}_2, \boldsymbol{\mu}) > c_2 m^{\tau_2-\varepsilon} \cdot f(\mathbf{X} \mid \theta_1 = 0, \boldsymbol{\theta}_2, \boldsymbol{\mu}).$$

Since now $\theta_{0,1} = 1$, $X_1 \sim N(\mu_1, 1)$,

$$\begin{aligned} (13) \quad &\mathrm{pr}\,(X_1 - \mu_1 < -(\log\log m)^{1/2}) \\ &\qquad = \Phi(-(\log\log m)^{1/2}) < (\log m)^{-1/2}. \end{aligned}$$

By (3) and (13), with probability greater than $1 - (\log m)^{-1/2}$,

$$\frac{f(X_1 \mid \theta_1 = 1, \mu_1)}{f(X_1 \mid \theta_1 = 0, \mu_1)} = \exp\{\mu_1(x_1 - \mu_1) + \mu_1^2/2\} > m^{\tau_2-\varepsilon}.$$

In addition, given $\boldsymbol{\theta}_2$ and $\boldsymbol{\mu}_2$,

$$(14) \qquad \mathrm{pr}\,(Z_2 > (\log\log m)^{1/2} \mid \boldsymbol{\theta}_2, \boldsymbol{\mu}_2) = (\log m)^{-1/2}.$$

By (6), (13) and (14) and Assumption (C'), we can follow the proof of Step (1.i) and show that with probability greater than $1 - 2(\log m)^{-1/2}$,

$$\frac{f(\mathbf{X} \mid \theta_1 = 1, \boldsymbol{\theta}_2, \boldsymbol{\mu})}{f(\mathbf{X} \mid \theta_1 = 0, \boldsymbol{\theta}_2, \boldsymbol{\mu})} > m^{\tau_2-\varepsilon}.$$

2) Now consider $\boldsymbol{\theta}_2$ and $\boldsymbol{\mu}$ as random vectors.

When $\theta_{0,1} = 0$, for each given $\boldsymbol{\theta}_2$ and $\boldsymbol{\mu}$,

$$f(\mathbf{X} \mid \theta_1 = 1, \boldsymbol{\theta}_2, \boldsymbol{\mu}) < m^{-\tau_2+\varepsilon} \cdot f(\mathbf{X} \mid \theta_1 = 0, \boldsymbol{\theta}_2, \boldsymbol{\mu})$$

holds with probability greater than $1 - 2(\log m)^{-1/2}$. Therefore, with probability greater than $1 - 2(\log m)^{-1/2}$, we have

$$\sum_{\boldsymbol{\theta}_2} f(\mathbf{X} \mid \theta_1 = 1, \boldsymbol{\theta}_2, \boldsymbol{\mu}) \operatorname{pr}(\boldsymbol{\theta}_2) < m^{-\tau_2+\varepsilon} \cdot \sum_{\boldsymbol{\theta}_2} f(\mathbf{X} \mid \theta_1 = 0, \boldsymbol{\theta}_2, \boldsymbol{\mu}) \operatorname{pr}(\boldsymbol{\theta}_2).$$

We conclude that

$$\frac{f(\mathbf{X} \mid \theta_1 = 1, \boldsymbol{\mu})}{f(\mathbf{X} \mid \theta_1 = 0, \boldsymbol{\mu})} = \frac{\sum_{\boldsymbol{\theta}_2} f(\mathbf{X} \mid \theta_1 = 1, \boldsymbol{\theta}_2, \boldsymbol{\mu}) \operatorname{pr}(\boldsymbol{\theta}_2)}{\sum_{\boldsymbol{\theta}_2} f(\mathbf{X} \mid \theta_1 = 0, \boldsymbol{\theta}_2, \boldsymbol{\mu}) \operatorname{pr}(\boldsymbol{\theta}_2)} < m^{-\tau_2+\varepsilon} \tag{15}$$

holds with probability greater than $1 - 2(\log m)^{-1/2}$.

By (1), (15) and the following equality

$$\frac{f(X_1 \mid \theta_1 = 1)}{f(X_1 \mid \theta_1 = 0)} = \frac{\int f(X_1 \mid \theta_1 = 1, \boldsymbol{\mu}) \, \mathrm{d}G(\boldsymbol{\mu})}{\int f(X_1 \mid \theta_1 = 0, \boldsymbol{\mu}) \, \mathrm{d}G(\boldsymbol{\mu})},$$
$$\frac{f(\mathbf{X} \mid \theta_1 = 1)}{f(\mathbf{X} \mid \theta_1 = 0)} = \frac{\int f(\mathbf{X} \mid \theta_1 = 1, \boldsymbol{\mu}) \, \mathrm{d}G(\boldsymbol{\mu})}{\int f(\mathbf{X} \mid \theta_1 = 0, \boldsymbol{\mu}) \, \mathrm{d}G(\boldsymbol{\mu})},$$

we have when $\theta_{0,1} = 0$, with probability greater than $1 - 2(\log m)^{-1/2}$,

$$\frac{f(X_1 \mid \theta_1 = 1)}{f(X_1 \mid \theta_1 = 0)} < m^{-\tau_2+\epsilon} \quad \text{and} \quad \frac{f(\mathbf{X} \mid \theta_1 = 1)}{f(\mathbf{X} \mid \theta_1 = 0)} < m^{-\tau_2+\epsilon} \tag{16}$$

Similarly, when $\theta_{0,i} = 1$, we can show that with probability greater than $1 - 2(\log m)^{-1/2}$,

$$\frac{f(X_1 \mid \theta_1 = 1)}{f(X_1 \mid \theta_1 = 0)} > m^{\tau_2-\epsilon} \quad \text{and} \quad \frac{f(\mathbf{X} \mid \theta_1 = 1)}{f(\mathbf{X} \mid \theta_1 = 0)} > m^{\tau_2-\epsilon} \tag{17}$$

3) We are now ready to prove Theorem 6.

It is easy to show

$$T_{MG,1} = \frac{1-p}{(1-p) + pf(X_1 \mid \theta_1 = 0)/f(X_1 \mid \theta_1 = 0)}$$
$$T_{OR,1} = \frac{1-p}{(1-p) + pf(\mathbf{X} \mid \theta_1 = 0)/f(\mathbf{X} \mid \theta_1 = 0)}$$

By Assumption (A'), $cm^{-\tau_1} \leq p/(1-p) \leq cm^{\tau_1}$.

When $\theta_{0,1} = 0$, with probability greater than $1 - O((\log m)^{-1/2})$,

$$T_{MG,1} \geq \frac{(1-p)}{1 - p + pm^{-\tau_2+\varepsilon}} \geq \frac{1}{1 + pm^{-\tau_2+\varepsilon}/(1-p)} \geq \frac{1}{1 + cm^{\tau_1-\tau_2+\varepsilon}},$$

which yields

$$1 - T_{MG,1} \leq \frac{cm^{-(\tau_2-\tau_1-\varepsilon)}}{1 + cm^{-(\tau_2-\tau_1-\varepsilon)}}.$$

Similarly, it can be shown that this result holds for $T_{OR,1}$. By Assumption (A'), with probability greater than $1 - O((\log m)^{-1/2})$,

$$|T_{OR,1} - T_{MG,1}| \leq |1 - T_{OR,1}| + |1 - T_{MG,1}| = \frac{2cm^{-(\tau_2-\tau_1-\varepsilon)}}{1 + cm^{-(\tau_2-\tau_1-\varepsilon)}} = O(m^{-(\tau_2-\tau_1-\varepsilon)}) \to 0.$$

When $\theta_{0,1} = 1$, with probability greater than $1 - O((\log m)^{-1/2})$,

$$T_{MG,1} \leq \frac{(1-p)}{1 - p + pc_2 m^{\tau_2-\varepsilon}} \leq \frac{1}{1 + pm^{\tau_2-\varepsilon}/(1-p)} \leq \frac{1}{1 + cm^{\tau_2-\tau_1-\varepsilon}}.$$

Same result holds for $T_{OR,1}$. By Assumption (A'), with probability greater than $1 - O((\log m)^{-1/2})$,

$$|T_{OR,1} - T_{MG,1}| \leq \frac{2}{1 + cm^{\tau_2-\tau_1-\varepsilon}} = O(m^{-(\tau_2-\tau_1-\varepsilon)}) \to 0. \qquad \square$$

Received 21 August 2014

REFERENCES

[1] Ji, P. and Jin, J. (2012). UPS delivers optimal phase diagram in high-dimensional variable selection. *The Annals of Statistics* **40-3** 73–103. MR3013180

[2] Jin, J. and Donoho, D. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics* **32-3** 962–994. MR2065195

[3] Xie, J., Cai, T. and Li, H. (2011). Sample size and power analysis for sparse signal recovery in genome-wide association studies. *Biometrika* **98-2** 273–290. MR2806428

[4] Xie, J., Cai, T. T., Maris, J. and Li, H. (2011). Optimal false discovery rate control for dependent data. *Statistics and Its Interface* **4** 417–430. MR2868825

Jichun Xie
Department of Biostatistics and Bioinformatics
School of Medicine
Duke University
USA
E-mail address: jichun.xie@duke.edu

T. Tony Cai
Department of Statistics
The Wharton School
University of Pennsylvania
USA
E-mail address: tcai@wharton.upenn.edu

Hongzhe Li
Department of Biostatistics and Epidemiology
School of Medicine
University of Pennsylvania
USA
E-mail address: hongzhe@upenn.edu