

# ASE o MLqE Story Latest

Runze

May 29, 2016

## 1 Problem Description

### 1.1 Uncontaminated Model

Let  $F$  be a distribution on  $\mathcal{X} \in \mathbb{R}^d$ , satisfying  $x^T y \geq 0$  for all  $x, y \in \mathcal{X}$ . We now generate  $m$  i.i.d. graphs under the RDPG( $F$ ) model. First sample  $X_1, \dots, X_n$  independently from distribution  $F$ , and define  $X = [X_1, \dots, X_n]^T \in \mathbb{R}^{n \times d}$ ,  $P = XX^T \in [0, R]^{n \times n}$ , where  $R$  is a constant. Then we can sample  $m$  conditionally i.i.d. symmetric and hollow graphs  $G^{(1)}, \dots, G^{(m)}$ , such that conditioned on  $X$ ,  $G_{ij}^{(t)} \stackrel{\text{ind}}{\sim} \text{Exp}(P_{ij})$  for each  $1 \leq t \leq m$ ,  $1 \leq i < j \leq n$ .

### 1.2 Contaminated Observations

Now we assume the observed edges are contaminated with probability  $\epsilon$ .

Let  $G$  be a distribution on  $\mathcal{Y} \in \mathbb{R}^{d'}$ , satisfying  $x^T y \geq 0$  for all  $x, y \in \mathcal{Y}$ . First sample  $X$  from  $F$  and  $Y$  from  $G$ . Then we sample  $m$  conditionally i.i.d. symmetric and hollow graphs  $A^{(1)}, \dots, A^{(m)}$  such that conditioning on  $X$  and  $Y$ ,  $A_{ij}^{(t)} \stackrel{\text{ind}}{\sim} (1 - \epsilon)\text{Exp}(P_{ij}) + \epsilon\text{Exp}(C_{ij})$  for each  $1 \leq t \leq m$ ,  $1 \leq i < j \leq n$ , where the contamination is a rank- $d'$  matrix  $C = YY^T \in [0, R]^{n \times n}$ ,  $Y \in \mathbb{R}^{n \times d'}$ .

### 1.3 Goal

Given the contaminated observation of adjacency matrices of  $m$  graphs, i.e.  $A^{(1)}, \dots, A^{(m)}$ , we want to estimate the mean of the collection of uncontaminated graphs  $P$ .

## 2 Candidate Estimators

After observing contaminated adjacency matrices of  $m$  graphs  $A^{(1)}, \dots, A^{(m)}$ , we want to propose a good estimator for the mean of the collection of graphs  $P$ .

### 2.1 $\hat{P}^{(1)}$ based on entry-wise MLE

Under the independent edge setting, we can simplify the problem to finding an entry-wise estimate of  $P$ . And MLE is always our first choice, which exists and happen to be  $\bar{A}$ , the entry-wise mean in this case. For consistency, we define  $\hat{P}^{(1)} = \bar{A}$ .

## 2.2 $\hat{P}^{(q)}$ based on entry-wise MLqE

Since the observations are contaminated, robust estimators are preferred. A modified MLE estimator, the maximum likelihood L- $q$  estimator, is considered in this case. Define  $\hat{P}^{(q)}$  as the entry-wise MLqE.

**Remark:** MLE is a special case of MLqE when  $q = 1$ . So we notate the entry-wise MLE to be  $\hat{P}^{(1)}$  in consistent with entry-wise MLqE  $\hat{P}^{(q)}$ .

## 2.3 $\tilde{P}^{(1)}$ based on ASE of entry-wise MLE

By taking advantages of the graph structure, we expect a better performance after applying a rank-reduction procedure to the entry-wise MLE  $\hat{P}^{(1)}$  under the SBM. So we first apply ASE to  $\hat{P}^{(1)}$  to get the latent positions  $\hat{X}^{(1)}$  in dimension  $d^{(1)}$ , and then define  $\tilde{P}^{(1)} = \hat{X}^{(1)} \hat{X}^{(1)T}$ .

## 2.4 $\tilde{P}^{(q)}$ based on ASE of entry-wise MLqE

Similarly, we also expect a better performance after applying a rank-reduction procedure to the entry-wise MLqE  $\hat{P}^{(q)}$  under the SBM. So we first apply ASE to  $\hat{P}^{(q)}$  to get the latent positions  $\hat{X}^{(q)}$  in dimension  $d^{(q)}$ , and then define  $\tilde{P}^{(q)} = \hat{X}^{(q)} \hat{X}^{(q)T}$ .

# 3 Compare Estimators

## 3.1 $\hat{P}^{(q)}$ is better than $\hat{P}^{(1)}$

**Lemma 3.1** *For any  $0 < q, \epsilon < 1$ , there exists  $C_0(P_{ij}, \epsilon, q) > 0$  such that under the contaminated model with  $C > C_0(P_{ij}, \epsilon, q)$ ,*

$$\lim_{m \rightarrow \infty} \left| E[\hat{P}_{ij}^{(q)}] - P_{ij} \right| < \lim_{m \rightarrow \infty} \left| E[\hat{P}_{ij}^{(1)}] - P_{ij} \right|,$$

for  $1 \leq i, j \leq n$  and  $i \neq j$ .

**Lemma 3.2** *For  $1 \leq i, j \leq n$ , we have*

$$\lim_{m \rightarrow \infty} \text{Var}(\hat{P}_{ij}^{(1)}) = \lim_{m \rightarrow \infty} \text{Var}(\hat{P}_{ij}^{(q)}) = 0,$$

Thus,

- By Lemma 3.1, when  $C$  is large enough, for every  $1 \leq i, j \leq n$  and  $i \neq j$ ,  $\hat{P}_{ij}^{(q)}$  has smaller asymptotic bias in absolute value than  $\hat{P}_{ij}^{(1)}$  as  $m \rightarrow \infty$ ;
- By Lemma 3.2, all entry-wise variances go to 0 for estimating  $P$  as  $m \rightarrow \infty$ ;
- In terms of MSE,  $\hat{P}^{(q)}$  is better than  $\hat{P}^{(1)}$  when  $m$  and  $C$  are large enough.

### 3.2 $\tilde{P}^{(1)}$ is better than $\hat{P}^{(1)}$

**Theorem 3.3** For fixed  $m$ ,  $1 \leq i, j \leq n$ ,

$$\frac{\text{Var}(\tilde{P}_{ij}^{(1)})}{\text{Var}(\hat{P}_{ij}^{(1)})} = O(mn^{-1}(\log n)^3).$$

Thus

$$\text{ARE}(\hat{P}_{ij}^{(1)}, \tilde{P}_{ij}^{(1)}) = 0.$$

Then

- For each  $1 \leq i, j \leq n$ , both  $\hat{P}_{ij}^{(1)}$  and  $\tilde{P}_{ij}^{(1)}$  have the same asymptotic bias as  $n \rightarrow \infty$ ;
- Fix  $m$ , for every  $1 \leq i, j \leq n$ ,  $\text{ARE}(\hat{P}_{ij}^{(1)}, \tilde{P}_{ij}^{(1)}) = \lim_{n \rightarrow \infty} \text{Var}(\tilde{P}_{ij}^{(1)})/\text{Var}(\hat{P}_{ij}^{(1)}) = 0$ , which means  $\tilde{P}^{(1)}$  is better than  $\hat{P}^{(1)}$ ;
- Actually when fixing  $m$ , for every  $1 \leq i, j \leq n$ ,  $\text{Var}(\tilde{P}_{ij}^{(1)})/\text{Var}(\hat{P}_{ij}^{(1)})$  is of order  $O(n^{-1}(\log n)^3)$  as  $n \rightarrow \infty$ .

### 3.3 $\tilde{P}^{(q)}$ is better than $\hat{P}^{(q)}$

Define  $H^{(q)} = E[\hat{P}^{(q)}]$ . Let  $d^{(q)} = \text{rank}(H^{(q)})$  be the dimension in which we are going to embed  $\hat{P}^{(q)}$ . Then

- For each  $1 \leq i, j \leq n$ , both  $\hat{P}_{ij}^{(q)}$  and  $\tilde{P}_{ij}^{(q)}$  have the same asymptotic bias as  $n \rightarrow \infty$ ;
- Fix  $m$ , for every  $1 \leq i, j \leq n$ ,  $\text{ARE}(\hat{P}_{ij}^{(q)}, \tilde{P}_{ij}^{(q)}) = \lim_{n \rightarrow \infty} \text{Var}(\tilde{P}_{ij}^{(q)})/\text{Var}(\hat{P}_{ij}^{(q)}) = 0$ , which means  $\tilde{P}^{(q)}$  is better than  $\hat{P}^{(q)}$ ;
- Actually, even if  $m$  is not fixed, as long as  $m$  is growing with order  $o(n^{1/2}(\log n)^{-3/2})$ , we still have  $\text{ARE}(\hat{P}_{ij}^{(q)}, \tilde{P}_{ij}^{(q)}) = 0$ ,

### 3.4 $\tilde{P}^{(q)}$ is better than $\tilde{P}^{(1)}$

- When  $n$  is large enough, for every  $1 \leq i, j \leq n$ ,  $E[\tilde{P}_{ij}^{(1)}]$  will be close to  $E[\hat{P}_{ij}^{(1)}]$  and  $E[\tilde{P}_{ij}^{(q)}]$  will be close to  $E[\hat{P}_{ij}^{(q)}]$ . Combined with  $\hat{P}_{ij}^{(q)}$  has smaller asymptotic bias (as  $m \rightarrow \infty$ ) than  $\hat{P}_{ij}^{(1)}$  when  $C$  is large enough, we have for sufficiently large  $m$  and  $n$ ,  $C$  large enough,  $\lim_{m \rightarrow \infty} \text{Bias}(\tilde{P}_{ij}^{(1)}) > \lim_{m \rightarrow \infty} \text{Bias}(\tilde{P}_{ij}^{(q)})$ ;
- Fix  $m$ , for any  $1 \leq i, j \leq n$ , when  $n$  is large enough,  $\text{Var}(\tilde{P}_{ij}^{(1)})$  is less than  $\text{Var}(\hat{P}_{ij}^{(1)})$  times  $O(n^{-1})$  and  $\text{Var}(\tilde{P}_{ij}^{(q)})$  is less than  $\text{Var}(\hat{P}_{ij}^{(q)})$  times  $O(n^{-1/2}(\log n)^{3/2})$ . Thus  $\lim_{n \rightarrow \infty} \text{Var}(\tilde{P}_{ij}^{(1)}) = \lim_{n \rightarrow \infty} \text{Var}(\tilde{P}_{ij}^{(q)}) = 0$ ;
- In terms of MSE,  $\tilde{P}^{(q)}$  is better than  $\tilde{P}^{(1)}$  when  $m$ ,  $n$  and  $C$  are large enough.

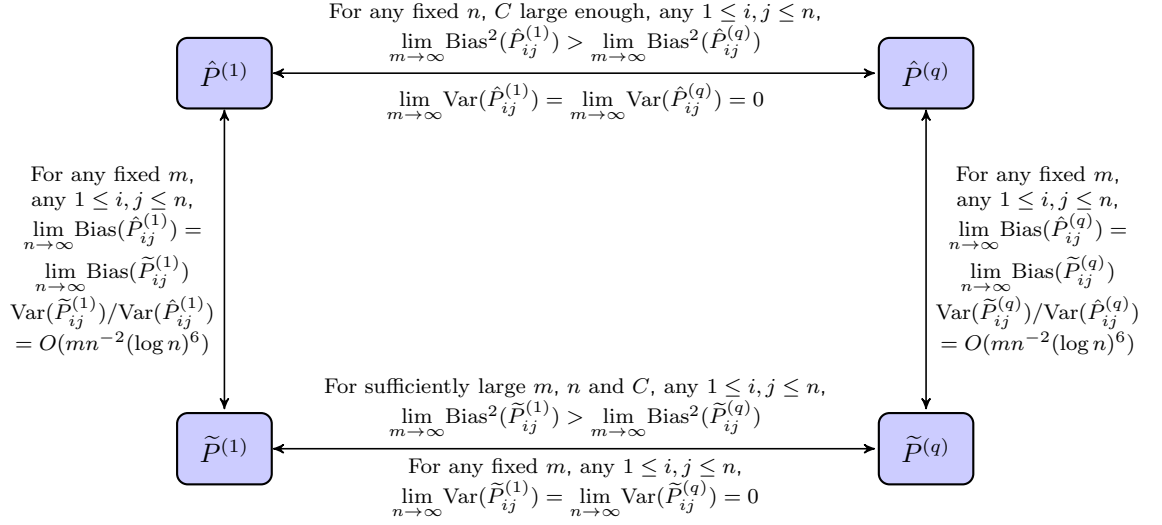


Figure 1: Relationship between four estimators.

### 3.5 Summary

Thus, we should choose the estimator  $\tilde{P}^{(q)}$ .

RT: The figure of relationship is NOT right.

## 4 Proof

### 4.1 $\hat{P}^{(q)}$ better than $\hat{P}^{(1)}$

**Lemma 4.1 (Lemma 3.1)** For any  $0 < q < 1$ , there exists  $C_0(P_{ij}, \epsilon, q) > 0$  such that under the contaminated model with  $C > C_0(P_{ij}, \epsilon, q)$ ,

$$\lim_{m \rightarrow \infty} |E[\hat{P}_{ij}^{(q)}] - P_{ij}| < \lim_{m \rightarrow \infty} |E[\hat{P}_{ij}^{(1)}] - P_{ij}|,$$

for  $1 \leq i, j \leq n$  and  $i \neq j$ .

**Proof:** For the MLE  $\hat{P}_{ij}^{(1)} = \bar{A}_{ij}$ ,

$$E[\hat{P}_{ij}^{(1)}] = E[\bar{A}_{ij}] = \frac{1}{m} \sum_{t=1}^m E[A_{ij}^{(t)}] = E[A_{ij}^{(1)}] = (1 - \epsilon)P_{ij} + \epsilon C_{ij}.$$

For the MLqE  $\hat{P}_{ij}^{(q)}$ ,

$$E[\hat{P}_{ij}^{(q)}]$$

solves a cubic equation.

■

RT: Shown in Mathematica

**Lemma 4.2 (Lemma 3.2)**

$$\lim_{m \rightarrow \infty} \text{Var}(\hat{P}_{ij}^{(1)}) = \lim_{m \rightarrow \infty} \text{Var}(\hat{P}_{ij}^{(q)}) = 0,$$

for  $1 \leq i, j \leq n$ .

**Proof:** Both MLE and MLqE follows a central limit theorem, which means their variances goes to 0 as  $m \rightarrow \infty$ .  $\blacksquare$

## 4.2 $\tilde{P}^{(1)}$ better than $\hat{P}^{(1)}$

**Theorem 4.3** (*Matrix Bernstein: Subexponential Case*). Consider a finite sequence  $\{X_k\}$  of independent, random, self-adjoint matrices with dimension  $d$ . Assume that

$$E[X_k] = 0 \quad \text{and} \quad E[X_k^p] \preceq \frac{p!}{2} R^{p-2} A_k^2 \quad \text{for } p = 2, 3, 4, \dots$$

Compute the variance parameter

$$\sigma^2 := \left\| \sum_k A_k^2 \right\|.$$

Then the following chain of inequalities holds for all  $t \geq 0$ .

$$P \left( \lambda_{\max} \left( \sum_k X_k \right) \geq t \right) \leq d \cdot \exp \left( \frac{-t^2/2}{\sigma^2 + Rt} \right).$$

**Remark:** Theorem 6.2 in [8].

**Theorem 4.4 (Theorem 3.3)** Let  $P$  and  $C$  be two  $n$ -by- $n$  symmetric matrices satisfying element-wise conditions  $0 < P_{ij} \leq C_{ij} \leq R$  for some constant  $R > 0$ . For  $0 < \epsilon < 1$ , we define  $m$  symmetric and hollow matrices as

$$A^{(t)} \stackrel{iid}{\sim} (1 - \epsilon) \text{Exp}(P) + \epsilon \text{Exp}(C),$$

for  $1 \leq t \leq m$ . Let  $\hat{P}^{(1)}$  be the element-wise MLE based on exponential distribution with  $m$  observations. Define  $H_{ij}^{(1)} = E[\hat{P}_{ij}^{(1)}] = (1 - \epsilon)P_{ij} + \epsilon C_{ij}$ , then for any constant  $c > 0$ , there exists another constant  $n_0(c)$ , independent of  $n, P, C$  and  $\epsilon$ , such that if  $n > n_0$ , then for all  $\eta$  satisfying  $n^{-c} \leq \eta \leq 1/2$ ,

$$P \left( \|\hat{P}^{(1)} - H^{(1)}\|_2 \leq 4R\sqrt{n \ln(n/\eta)} \right) \geq 1 - \eta.$$

**Remark:** This is the extended version of Theorem 3.1 in [4].

**Proof:** Let  $\{e_i\}_{i=1}^n$  be the canonical basis for  $\mathbb{R}^n$ . For each  $1 \leq i, j \leq n$ , define a corresponding matrix  $G_{ij}$ :

$$G_{ij} \equiv \begin{cases} e_i e_j^T + e_j e_i^T, & i \neq j; \\ e_i e_i^T, & i = j. \end{cases}$$

Thus  $\hat{P}^{(1)} = \sum_{1 \leq i < j \leq n} \hat{P}_{ij}^{(1)} G_{ij}$  and  $H^{(1)} = \sum_{1 \leq i < j \leq n} H_{ij}^{(1)} G_{ij}$ . Then we have  $\hat{P}^{(1)} - H^{(1)} = \sum_{1 \leq i < j \leq n} X_{ij}$ , where  $X_{ij} \equiv \left( \hat{P}_{ij}^{(1)} - H_{ij}^{(1)} \right) G_{ij}$ ,  $1 \leq i < j \leq n$ .

First consider the  $k$ -th moment of  $X_{ij}$  for  $1 \leq i < j \leq n$ . Since

$$\begin{aligned} E[(A_{ij}^{(1)} - H_{ij}^{(1)})^k] &\leq (1 - \epsilon) \cdot \exp(-H_{ij}/P_{ij}) P_{ij}^k \Gamma(1 + k, -H_{ij}/P_{ij}) \\ &\quad + \epsilon \cdot \exp(-H_{ij}/C_{ij}) C_{ij}^k \Gamma(1 + k, -H_{ij}/C_{ij}) \\ &\leq ((1 - \epsilon) \cdot \exp(-H_{ij}/P_{ij}) P_{ij}^k + \epsilon \cdot \exp(-H_{ij}/C_{ij}) C_{ij}^k) k! \\ &\leq ((1 - \epsilon) \cdot P_{ij}^k + \epsilon \cdot C_{ij}^k) k! \\ &\leq R^k k!, \end{aligned}$$

we have

$$\begin{aligned}
E[(\hat{P}_{ij}^{(1)} - H_{ij}^{(1)})^k] &= E\left[\left(\frac{1}{m} \sum_{t=1}^m A_{ij}^{(t)} - H_{ij}^{(1)}\right)^k\right] \\
&= E\left[\left(\frac{1}{m} \sum_{t=1}^m (A_{ij}^{(t)} - H_{ij}^{(1)})\right)^k\right] \\
&= \frac{1}{m^k} E\left[\left(\sum_{t=1}^m (A_{ij}^{(t)} - H_{ij}^{(1)})\right)^k\right] \\
&\leq R^k k!.
\end{aligned} \tag{1}$$

Combined with

$$G_{ij}^k \equiv \begin{cases} e_i e_i^T + e_j e_j^T, & k \text{ is even;} \\ e_i e_j^T + e_j e_i^T, & k \text{ is odd,} \end{cases}$$

thus we have

1. When  $k$  is even,

$$E[X_{ij}^k] = E[(\hat{P}_{ij}^{(1)} - H_{ij}^{(1)})^k] G_{ij}^2 \preceq k! R^k G_{ij}^2;$$

2. When  $k$  is odd,

$$E[X_{ij}^k] = E[(\hat{P}_{ij}^{(1)} - H_{ij}^{(1)})^k] G_{ij} \preceq k! R^k G_{ij}^2.$$

So

$$E[X_{ij}^k] \preceq k! R^k G_{ij}^2.$$

Let

$$\sigma^2 := \left\| \sum_{1 \leq i < j \leq n} (\sqrt{2} R G_{ij})^2 \right\|_2 = 2R^2 \|(n-1)I\|_2 = 2R^2(n-1).$$

Notice that random matrices  $X_{ij}$  are independent, self-adjoint and have mean zero, apply Theorem 4.3 we have

$$\begin{aligned}
P\left(\lambda_{\max}(\hat{P}^{(1)} - H^{(1)}) \geq t\right) &\leq n \exp\left(-\frac{t^2/2}{\sigma^2 + Rt}\right) \\
&\leq n \exp\left(-\frac{t^2/2}{2R^2n + Rt}\right).
\end{aligned}$$

Now consider  $Y_{ij} \equiv (H^{(1)} - \hat{P}^{(1)}) G_{ij}$ ,  $1 \leq i < j \leq n$ . Then we have  $H^{(1)} - \hat{P}^{(1)} = \sum_{1 \leq i < j \leq n} Y_{ij}$ . Since

$$E[(H^{(1)} - \hat{P}^{(1)})^k] = (-1)^k E[(\hat{P}^{(1)} - H^{(1)})^k],$$

1. When  $k$  is even,

$$E[Y_{ij}^k] = E[(\hat{P}^{(1)} - H^{(1)})^k] G_{ij}^2 \preceq k! R^k G_{ij}^2;$$

2. When  $k$  is odd,

$$E[Y_{ij}^k] = -E[(\hat{P}^{(1)} - H^{(1)})^k]G_{ij} \preceq k!R^k G_{ij}^2.$$

Thus by similar arguments,

$$\begin{aligned} P\left(\lambda_{\min}(\hat{P}^{(1)} - H^{(1)}) \leq -t\right) &= P\left(\lambda_{\max}(H^{(1)} - \hat{P}^{(1)}) \geq t\right) \\ &\leq n \exp\left(-\frac{t^2/2}{2R^2n + Rt}\right). \end{aligned}$$

Therefore we have

$$P\left(\|\hat{P}^{(1)} - H^{(1)}\|_2 \geq t\right) \leq n \exp\left(-\frac{t^2/2}{2R^2n + Rt}\right).$$

Now let  $c > 0$  be given and assume  $n^{-c} \leq \eta \leq 1/2$ . Then there exists a  $n_0(c)$  independent of  $n, P, C$  and  $\epsilon$  such that whenever  $n > n_0(c)$ ,

$$t = 4R\sqrt{n \ln(n/\eta)} \leq 6Rn.$$

Plugging this  $t$  into the equation above, we get

$$P(\|\hat{P}^{(1)} - H^{(1)}\|_2 \geq 4R\sqrt{n \ln(n/\eta)}) \leq n \exp\left(-\frac{t^2}{16R^2n}\right) = \eta.$$

Define  $H^{(1)} = E[\hat{P}^{(1)}] = (1 - \epsilon)P + \epsilon C$ , where  $P = XX^T$ ,  $X \in \mathbb{R}^{n \times d}$ ,  $C = YY^T$ ,  $Y \in \mathbb{R}^{n \times d'}$ . Let  $d^{(1)} = \text{rank}(H^{(1)})$  be the dimension in which we are going to embed  $\hat{P}^{(1)}$ . Then we can define  $H^{(1)} = ZZ^T$  where  $Z \in \mathbb{R}^{n \times d^{(1)}}$ . Since  $H^{(1)} = [\sqrt{1 - \epsilon}X, \sqrt{\epsilon}Y][\sqrt{1 - \epsilon}X, \sqrt{\epsilon}Y]^T$ , we have  $d^{(1)} \leq d + d'$ . ■

For simplicity, from now on, we will use  $\hat{P}$  to represent  $\hat{P}^{(1)}$ , use  $H$  to represent  $H^{(1)}$  and use  $k$  to represent the dimension  $d^{(1)}$  we are going to embed. Assume  $H = USU^T = ZZ^T$ , where  $Z = [Z_1, \dots, Z_n]^T$  is a  $n$ -by- $k$  matrix. Then our estimate for  $Z$  up to rotation is  $\hat{Z} = \hat{U}\hat{S}^{1/2}$ , where  $\hat{U}\hat{S}\hat{U}^T$  is the rank- $k$  spectral decomposition of  $|\hat{P}| = (\hat{P}^T \hat{P})^{1/2}$ .

Furthermore, we assume that the second moment matrix  $E[Z_1 Z_1^T]$  is rank  $k$  and has distinct eigenvalues  $\lambda_i(E[Z_1 Z_1^T])$ . In particular, we assume that there exists  $\delta > 0$  such that

$$\delta < \min\left(\min_{i \neq j} |\lambda_i(E[Z_1 Z_1^T]) - \lambda_j(E[Z_1 Z_1^T])|, \lambda_k(E[Z_1 Z_1^T])\right)$$

**Lemma 4.5** *Under the above assumptions,  $\lambda_i(H) = \Theta(n)$  with high probability when  $i \leq k$ , i.e. the largest  $k$  eigenvalues of  $H$  is of order  $n$ . Moreover, we have  $\|S\|_2 = \Theta(n)$  and  $\|\hat{S}\|_2 = \Theta(n)$  with high probability.*

**Remark:** This is an extended version of Proposition 4.3 in [7].

**Proof:** Note that  $\lambda_i(H) = \lambda_i(ZZ^T) = \lambda_i(Z^T Z)$  when  $i \leq k$ . Since each entry of  $Z^T Z$  is a sum of  $n$  independent random variables each in  $[0, R]$ , i.e.  $(Z^T Z)_{ij} = \sum_{l=1}^n Z_{li} Z_{lj}$ . By Hoeffding's inequality, for each entry we have

$$P(|(Z^T Z - nE[Z_1 Z_1^T])_{ij}| \geq R\sqrt{n \log n}) \leq \frac{2}{n^2}.$$

By the union bound, we have

$$P(\|(Z^T Z - nE[Z_1 Z_1^T])_{ij}\|_F \geq kR\sqrt{n \log n}) \leq \frac{2k^2}{n^2}.$$

Then by Weyl's Theorem [3], we have

$$|\lambda_i(H) - n\lambda_i(Z_1 Z_1^T)| \leq \|Z^T Z - nE[Z_1 Z_1^T]\|_2 \leq kR\sqrt{n \log n}$$

with probability at least  $1 - \frac{2k^2}{n^2}$ . Thus  $\lambda_i(H) = S_{ii} = \Theta(n)$  with probability at least  $1 - \frac{2k^2}{n^2}$  when  $i \leq k$ .

Moreover,

$$\|H\|_2 - \|H - \hat{P}\|_2 \leq \|\hat{S}\|_2 \leq \|\hat{P} - H\|_2 + \|H\|_2.$$

Combined with Theorem 4.4, with high probability we have  $\|\hat{S}\|_2 = \Theta(n)$ .  $\blacksquare$

**Lemma 4.6** *Let  $W_1 \Sigma W_2^T$  be the singular value decomposition of  $U^T \hat{U}$ . Then for sufficiently large  $n$ ,*

$$\|U^T \hat{U} - W_1 W_2^T\|_F = O(n^{-1} \log n)$$

*with high probability.*

**Proof:** Let  $\sigma_1, \dots, \sigma_d$  denote the singular values of  $U^T \hat{U}$ . Then  $\sigma_i = \cos(\theta_i)$  where the  $\theta_i$  are the principal angles between the subspaces spanned by  $\hat{U}$  and  $U$ . Furthermore, by the Davis-Kahan  $\sin(\Theta)$  theorem [1], combined with Theorem 4.4 and Lemma 4.5,

$$\|\hat{U} \hat{U}^T - U U^T\|_2 = \max_i |\sin(\theta_i)| \leq \frac{\|\hat{P} - H\|_2}{\lambda_k(H)} \leq \frac{C\sqrt{n \log n}}{n} = O(n^{-1/2} \sqrt{\log n}) \quad (2)$$

for sufficiently large  $n$ . Here  $\lambda_k(H)$  denotes the  $k$ -th largest eigenvalue of  $H$ . We thus have

$$\begin{aligned} \|U^T \hat{U} - W_1 W_2^T\|_F &= \|\Sigma - I\|_F = \sqrt{\sum_{i=1}^k (1 - \sigma_i)^2} \\ &\leq \sum_{i=1}^k (1 - \sigma_i) \leq \sum_{i=1}^k (1 - \sigma_i^2) \\ &= \sum_{i=1}^k \sin^2(\theta_i) \leq k \|\hat{U} \hat{U}^T - U U^T\|_2^2 \\ &= O(n^{-1} \log n). \end{aligned}$$

$\blacksquare$

We will denote the orthogonal matrix  $W_1 W_2^T$  by  $W^*$ .

**Lemma 4.7** *For sufficiently large  $n$ ,*

$$\|W^* \hat{S} - S W^*\|_F = O(\log n),$$



$$\|W^* \hat{S}^{1/2} - S^{1/2} W^*\|_F = O(n^{-1/2} \log n)$$

and

$$\|W^* \hat{S}^{-1/2} - S^{-1/2} W^*\|_F = O(n^{-3/2} \log n)$$

with high probability.

**Proof:** By Proposition 2.1 in [6] and Equation (2), we have for some orthogonal matrix  $W$ ,

$$\|\hat{U} - UW\|_F^2 \leq \frac{2\|\hat{U}\hat{U}^T - UU^T\|_F^2}{\delta^2} = O(n^{-1/2} \sqrt{\log n}).$$

Let  $Q = \hat{U} - UU^T \hat{U}$ . And  $Q$  is the residual after projecting  $\hat{U}$  orthogonally onto the column space of  $U$ , we have

$$\|Q\|_F = \|\hat{U} - UU^T \hat{U}\|_F \leq \|\hat{U} - UT\|_F = O(n^{-1/2} \sqrt{\log n}). \quad (3)$$

for all  $k \times k$  matrices  $T$ .

Then

$$\begin{aligned} W^* \hat{S} &= (W^* - U^T \hat{U}) \hat{S} + U^T \hat{U} \hat{S} = (W^* - U^T \hat{U}) \hat{S} + U^T \hat{P} \hat{U} \\ &= (W^* - U^T \hat{U}) \hat{S} + U^T (\hat{P} - H) \hat{U} + U^T H \hat{U} \\ &= (W^* - U^T \hat{U}) \hat{S} + U^T (\hat{P} - H) Q + U^T (\hat{P} - H) UU^T \hat{U} + U^T H \hat{U} \\ &= (W^* - U^T \hat{U}) \hat{S} + U^T (\hat{P} - H) Q + U^T (\hat{P} - H) UU^T \hat{U} + SU^T \hat{U}. \end{aligned}$$

Combined with Theorem 4.4, Lemma 4.5, Lemma 4.6, we have

$$\begin{aligned} &\|W^* \hat{S} - SW^*\|_F \\ &= \|(W^* - U^T \hat{U}) \hat{S} + U^T (\hat{P} - H) Q + U^T (\hat{P} - H) UU^T \hat{U} + S(U^T \hat{U} - W^*)\|_F \\ &\leq \|W^* - U^T \hat{U}\|_F (\|\hat{S}\|_2 + \|S\|_2) + \|U^T\|_F \|\hat{P} - H\|_2 \|Q\|_F + \|U^T (\hat{P} - H) U\|_F \\ &\leq O(\log n) + O(\log n) + \|U^T (\hat{P} - H) U\|_F \end{aligned}$$

with high probability. And we know  $U^T (\hat{P} - H) U$  is a  $k \times k$  matrix with  $ij$ -th entry to be

$$u_i^T (\hat{P} - H) u_j = \sum_{s=1}^n \sum_{t=1}^n (\hat{P}_{st} - H_{st}) u_{is} u_{jt} = 2 \sum_{s < t} (\hat{P}_{st} - H_{st}) u_{is} u_{jt}$$

where  $u_i$  and  $u_j$  are the  $i$ -th and  $j$ -th columns of  $U$ . Thus, conditioned on  $H$ ,  $U$  is fixed and  $u_i^T (\hat{P} - H) u_j$  is a sum of independent mean 0 random variables.

By Equation (1), we have

$$\begin{aligned} &E \left[ \left( (\hat{P}_{st} - H_{st}) u_{is} u_{jt} \right)^k \right] \\ &\leq k! R^k u_{is}^k u_{jt}^k \\ &\leq \frac{k!}{2} R^{k-2} (\sqrt{2} u_{is} u_{jt} R)^2. \end{aligned}$$

Also we have

$$\sigma^2 := \left| \sum_{s < t} 2R^2 u_{is}^2 u_{jt}^2 \right| \leq R^2,$$

RT: Not right here, I am using the result for spectral norm as frobenius norm.

then by Theorem 4.3, we have

$$P \left( \left| 2 \sum_{s < t} (\hat{P}_{st} - H_{st}) u_{is} u_{jt} \right| \geq t \right) \leq \exp \left( \frac{-t^2/8}{R^2 + Rt} \right).$$

Let  $t = \log n$ , we have

$$P \left( \left| 2 \sum_{s < t} (\hat{P}_{st} - H_{st}) u_{is} u_{jt} \right| \geq \log n \right) \leq n^{-c}$$

for some constant  $c$ . Thus each entry of  $U^T(\hat{P} - H)U$  is of order  $O(\log n)$  with high probability and

$$\|U^T(\hat{P} - H)U\|_F = O(\log n) \quad (4)$$

with high probability. Hence

$$\|W^* \hat{S} - SW^*\|_F = O(\log n)$$

with high probability. Also, since

$$W_{ij}^* (\lambda_j^{1/2}(\hat{P}) - \lambda_i^{1/2}(H)) = W_{ij}^* \frac{\lambda_j(\hat{P}) - \lambda_i(H)}{\lambda_j^{1/2}(\hat{P}) + \lambda_i^{1/2}(H)}$$

and the eigenvalues  $\lambda_j^{1/2}(\hat{P})$  and  $\lambda_i^{1/2}(H)$  are both of order  $\Theta(\sqrt{n})$ , we have

$$\|W^* \hat{S}^{1/2} - S^{1/2} W^*\|_F = O(n^{-1/2} \log n).$$

Similarly, since

$$W_{ij}^* (\lambda_j^{-1/2}(\hat{P}) - \lambda_i^{-1/2}(H)) = W_{ij}^* \frac{\lambda_i(H) - \lambda_j(\hat{P})}{(\lambda_j^{-1/2}(\hat{P}) + \lambda_i^{-1/2}(H)) \lambda_j(\hat{P}) \lambda_i(H)}$$

and the eigenvalues  $\lambda_j(\hat{P})$  and  $\lambda_i(H)$  are both of order  $\Theta(n)$ , we have

$$\|W^* \hat{S}^{-1/2} - S^{-1/2} W^*\|_F = O(n^{-3/2} \log n).$$

■

**Lemma 4.8** *There exists a rotation matrix  $W$  such that for sufficiently large  $n$ ,*

$$\|\hat{Z} - ZW\|_F = \|(\hat{P} - H)US^{-1/2}\|_F + O(n^{-1/2}(\log n)^{3/2})$$

*with high probability.*

**Proof:** Let  $Q_1 = UU^T \hat{U} - UW^*$ ,  $Q_2 = W^* \hat{S}^{1/2} - S^{1/2} W^*$  and  $Q_3 = \hat{U} - UW^* = \hat{U} - UU^T \hat{U} + Q_1 = Q + Q_1$ . Then since  $UU^T P = P$  and  $\hat{U} \hat{S}^{1/2} = \hat{P} \hat{U} \hat{S}^{-1/2}$ ,

$$\begin{aligned} \hat{Z} - US^{1/2} W^* &= \hat{U} \hat{S}^{1/2} - UW^* \hat{S}^{1/2} + U(W^* \hat{S}^{1/2} - S^{1/2} W^*) \\ &= (\hat{U} - UU^T \hat{U}) \hat{S}^{1/2} + Q_1 \hat{S}^{1/2} + UQ_2 \\ &= (\hat{P} - H) \hat{U} \hat{S}^{-1/2} - UU^T (\hat{P} - H) \hat{U} \hat{S}^{-1/2} + Q_1 \hat{S}^{1/2} + UQ_2 \\ &= (\hat{P} - H) UW^* \hat{S}^{-1/2} - UU^T (\hat{P} - H) UW^* \hat{S}^{-1/2} \\ &\quad + (I - UU^T) (\hat{P} - H) Q_3 \hat{S}^{-1/2} + Q_1 \hat{S}^{1/2} + UQ_2. \end{aligned}$$

By Lemma 4.6,

$$\|Q_1\|_F \leq \|U\|_F \|U^T \hat{U} - W^*\|_F = O(n^{-1} \log n).$$

By Lemma 4.7,

$$\|Q_2\|_F = O(n^{-1/2} \log n).$$

By Equation (3),

$$\|Q_3\|_F \leq \|Q\|_F + \|Q_1\|_F = O(n^{-1/2} (\log n)^{1/2}).$$

By Equation (4),

$$\|UU^T(\hat{P} - H)UW^*\hat{S}^{-1/2}\|_F \leq \|U^T(\hat{P} - H)U\|_F \|\hat{S}^{-1/2}\|_2 = O(n^{-1/2} \log n).$$

By Lemma 4.7,

$$\|W^*\hat{S}^{-1/2} - S^{-1/2}W^*\|_F = O(n^{-3/2} \log n).$$

Therefore,

$$\begin{aligned} & \|\hat{Z} - US^{1/2}W^*\|_F \\ &= \|(\hat{P} - H)UW^*\hat{S}^{-1/2}\|_F + O(n^{-1/2} \log n) + \|I - UU^T\|_2 \|\hat{P} - H\|_2 O(n^{-1} (\log n)^{1/2}) \\ & \quad + O(n^{-1/2} \log n) + O(n^{-1/2} \log n) \\ &= \|(\hat{P} - H)UW^*\hat{S}^{-1/2}\|_F + O(n^{-1/2} \log n) \\ &\leq \|(\hat{P} - H)US^{-1/2}W^*\|_F + \|(\hat{P} - H)U(W^*\hat{S}^{-1/2} - S^{-1/2}W^*)\|_F + O(n^{-1/2} \log n) \\ &= \|(\hat{P} - H)US^{-1/2}\|_F + O(n^{-1} (\log n)^{3/2}) + O(n^{-1/2} (\log n)^{3/2}) \\ &= \|(\hat{P} - H)US^{-1/2}\|_F + O(n^{-1/2} (\log n)^{3/2}). \end{aligned}$$

Note that  $Z = US^{1/2}W$  for some orthogonal matrix  $W$ . As  $W^*$  is also orthogonal, therefore  $Z\tilde{W} = US^{1/2}W^*$  for some orthogonal  $\tilde{W}$ , which completes the proof. RT:  $\|I - UU^T\|_2 = O(1)$

**Theorem 4.9** *There exists a rotation matrix  $W$  such that for sufficiently large  $n$ ,*

$$\max_i \|\hat{Z}_i - WZ_i\|_2 = O(n^{-1/2} (\log n)^{3/2})$$

*with high probability.*

**Proof:** By Lemma 4.8, we have

$$\|\hat{Z} - ZW\|_F = \|(\hat{P} - H)US^{-1/2}\|_F + O(n^{-1/2} (\log n)^{3/2})$$

and similarly we could have the bound for each column vector

$$\begin{aligned} \max_i \|\hat{Z}_i - WZ_i\|_2 &\leq \frac{1}{\lambda_k^{1/2}(H)} \max_i \|(\hat{P} - H)U\|_2 + O(n^{-1/2} (\log n)^{3/2}) \\ &\leq \frac{k^{1/2}}{\lambda_k^{1/2}(H)} \max_j \|(\hat{P} - H)u_j\|_\infty + O(n^{-1/2} (\log n)^{3/2}) \end{aligned}$$

where  $((\hat{P} - H)U)_i$  represents the  $i$ -th row of  $(\hat{P} - H)U$  and  $u_j$  denotes the  $j$ -th column of  $U$ . Now given  $i$  and  $j$ , the  $i$ -th element of the vector  $(\hat{P} - H)u_j$  is of the form

$$\sum_{s=1}^n (\hat{P}_{is} - H_{is})u_{js} = \sum_{s \neq i} (\hat{P}_{is} - H_{is})u_{js}.$$

Thus, conditioned on  $H$ , the  $i$ -th element of the vector  $(\hat{P} - H)u_j$  is a sum of independent mean 0 random variables. By Equation (1), we have

$$\begin{aligned} & E \left[ \left( (\hat{P}_{is} - H_{is})u_{js} \right)^k \right] \\ & \leq k! R^k u_{js}^k \\ & \leq \frac{k!}{2} R^{k-2} (\sqrt{2} R u_{js})^2. \end{aligned}$$

Also we have

$$\sigma^2 := \left| \sum_{s \neq i} 2R^2 u_{js}^2 \right| \leq 2R^2,$$

then by Theorem 4.3, we have

$$P \left( \left| \sum_{s \neq i} (\hat{P}_{is} - H_{is})u_{js} \right| \geq t \right) \leq \exp \left( \frac{-t^2/2}{2R^2 + Rt} \right),$$

i.e. it is of order  $O(\log n)$  with high probability. Taking the union bound over all  $i$  and  $j$ , with high probability we have

$$\begin{aligned} \max_i \|\hat{Z}_i - WZ_i\|_2 & \leq \frac{Ck^{1/2}}{\lambda_k^{1/2}(H)} (\log n)^{3/2} + O(n^{-1/2}(\log n)^{3/2}) \\ & = O(n^{-1/2}(\log n)^{3/2}). \end{aligned}$$

■

**Lemma 4.10**  $\left| \hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j \right| = O(n^{-1}(\log n)^3)$  with high probability.

**Proof:** Let  $W$  be the rotation matrix in Theorem 4.9, then

$$\begin{aligned} \left| \hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j \right| & = \left| \hat{Z}_i^T \hat{Z}_j - \hat{Z}_i^T WZ_j + \hat{Z}_i^T WZ_j - (WZ_i)^T WZ_j \right| \\ & \leq \left| \hat{Z}_i^T (\hat{Z}_j - WZ_j) + (\hat{Z}_i^T - (WZ_i)^T) WZ_j \right| \\ & \leq \|\hat{Z}_i\|_2 \|\hat{Z}_j - WZ_j\|_2 + \|Z_j\|_2 \|\hat{Z}_i^T - (WZ_i)^T\|_2. \end{aligned}$$

Since  $\|Z_i\|_2^2 = Z_i^T Z_i = H_{ii}^{(1)} = E[\hat{P}_{ii}^{(1)}] = (1 - \epsilon)P_{ij} + \epsilon C_{ij} \leq R$ , we have  $\|Z_i\|_2 = O(1)$ . Combined with Theorem 4.9,

$$\begin{aligned} \left| \hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j \right| & = (\|\hat{Z}_i\|_2 + \|Z_j\|_2) O(n^{-1/2}(\log n)^{3/2}) \\ & \leq (\|\hat{Z}_i - WZ_i\|_2 + \|WZ_i\|_2 + \|Z_j\|_2) O(n^{-1/2}(\log n)^{3/2}) \\ & = O(n^{-1}(\log n)^3) \end{aligned}$$

with high probability.

■

**Corollary 4.11** *For fixed  $m$ , the estimator based on ASE of MLE has the same entry-wise asymptotic bias as MLE, i.e.*

$$\lim_{n \rightarrow \infty} \text{Bias}(\tilde{P}_{ij}^{(1)}) = \lim_{n \rightarrow \infty} E[\tilde{P}_{ij}^{(1)}] - P_{ij} = \lim_{n \rightarrow \infty} E[\hat{P}_{ij}^{(1)}] - P_{ij} = \lim_{n \rightarrow \infty} \text{Bias}(\hat{P}_{ij}^{(1)}).$$

**Proof:** Direct result from Lemma 4.10 by noticing

$$\lim_{n \rightarrow \infty} E[\tilde{P}_{ij}^{(1)}] = \lim_{n \rightarrow \infty} E[\hat{P}_{ij}^{(1)}].$$

■

**Theorem 4.12**  $\text{Var}(\hat{Z}_i^T \hat{Z}_j) = O(n^{-2}(\log n)^6)$  with high probability.

**Proof:** By Lemma 4.10,

$$\begin{aligned} \text{Var}(\hat{Z}_i^T \hat{Z}_j) &= E[(\hat{Z}_i^T \hat{Z}_j - E[\hat{Z}_i^T \hat{Z}_j])^2] \\ &= E[(\hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j + Z_i^T Z_j - E[\hat{Z}_i^T \hat{Z}_j])^2] \\ &= E[(\hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j)^2] + E[(Z_i^T Z_j - E[\hat{Z}_i^T \hat{Z}_j])^2] \\ &\quad - 2E[(\hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j)(Z_i^T Z_j - E[\hat{Z}_i^T \hat{Z}_j])] \\ &\leq E[(\hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j)^2] + E[(Z_i^T Z_j - E[\hat{Z}_i^T \hat{Z}_j])^2] \\ &\quad + 2\sqrt{E[(\hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j)^2]E[(Z_i^T Z_j - E[\hat{Z}_i^T \hat{Z}_j])^2]} \\ &\leq 4E[(\hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j)^2] \\ &= O(n^{-2}(\log n)^6) \end{aligned}$$

with high probability.

■

**Corollary 4.13** *For fixed  $n$ ,  $1 \leq i, j \leq n$ ,  $\text{Var}(\hat{P}_{ij}^{(1)}) = \Theta(m^{-1})$ .*

**Proof:** Direct result from central limit theorem.

■

**Theorem 4.14** *For fixed  $m$ ,  $1 \leq i, j \leq n$  and  $i \neq j$ ,*

$$\frac{\text{Var}(\tilde{P}_{ij}^{(1)})}{\text{Var}(\hat{P}_{ij}^{(1)})} = O(mn^{-2}(\log n)^6).$$

Thus

$$\text{ARE}(\hat{P}_{ij}^{(1)}, \tilde{P}_{ij}^{(1)}) = 0.$$

Furthermore, as long as  $m$  goes to infinity of order  $o(n(\log n)^{-3})$ ,

$$\text{ARE}(\hat{P}_{ij}^{(1)}, \tilde{P}_{ij}^{(1)}) = 0.$$

**Proof:** The results are direct from Theorem 4.12 and Corollary 4.13.

■

RT: Expectation with high probability?

### 4.3 $\tilde{P}^{(q)}$ better than $\hat{P}^{(q)}$

**Lemma 4.15** Consider the model  $X_1, \dots, X_m \stackrel{iid}{\sim} \text{Exp}(\theta)$  with  $E[X_1] = \theta$ . Given any data  $x = (x_1, \dots, x_m)$  such that  $x_{(1)} > 0$  and not all  $x_i$ 's are the same, then  $\hat{\theta}_q(x) < \hat{\theta}_1(x)$  for  $0 < q < 1$ , i.e. MLqE [2, 5] is always less than MLE under exponential distribution no matter how the data is sampled.

**Proof:** The MLE is

$$\hat{\theta}_1(x) = \bar{x}.$$

And the MLqE  $\hat{\theta}_q(x)$  solves the equation

$$\sum_{i=1}^m e^{-\frac{(1-q)x_i}{\hat{\theta}_q(x)}} (x_i - \hat{\theta}_q(x)) = 0.$$

Consider the continuous function  $g(\theta, x) = \sum_{i=1}^m e^{-\frac{(1-q)x_i}{\theta}} (x_i - \theta)$ .

Let  $x_{(1)} \leq \dots \leq x_{(l)} \leq \bar{x} \leq x_{(l+1)} \leq \dots \leq x_{(m)}$ . Define  $s_i = \bar{x} - x_{(i)}$  for  $1 \leq i \leq l$ , and  $t_i = x_{(l+i)} - \bar{x}$  for  $1 \leq i \leq m-l$ . Note that  $\sum_{i=1}^l s_i = \sum_{i=1}^{m-l} t_i$ . Then we have

$$\begin{aligned} g(\hat{\theta}_1(x), x) &= g(\bar{x}, x) \\ &= \sum_{i=1}^m e^{-\frac{(1-q)x_{(i)}}{\bar{x}}} (x_{(i)} - \bar{x}) \\ &= -\sum_{i=1}^l e^{-\frac{(1-q)x_{(i)}}{\bar{x}}} s_i + \sum_{i=1}^{m-l} e^{-\frac{(1-q)x_{(l+i)}}{\bar{x}}} t_i \\ &\leq -e^{-(1-q)} \sum_{i=1}^l s_i + \sum_{i=1}^{m-l} e^{-\frac{(1-q)x_{(l+i)}}{\bar{x}}} t_i \\ &\leq -e^{-(1-q)} \sum_{i=1}^{m-l} t_i + \sum_{i=1}^{m-l} e^{-\frac{(1-q)x_{(l+i)}}{\bar{x}}} t_i \\ &\leq -\sum_{i=1}^{m-l} e^{-\frac{(1-q)x_{(l+i)}}{\bar{x}}} t_i + \sum_{i=1}^{m-l} e^{-\frac{(1-q)x_{(l+i)}}{\bar{x}}} t_i \\ &= 0, \end{aligned}$$

and equality holds if and only if all  $x_i$ 's are the same, which is excluded by the assumption. Thus  $g(\hat{\theta}_1(x), x) < 0$ .

Also we know:

- $g(\hat{\theta}_q(x), x) = 0$ ;
- $\lim_{\theta \rightarrow 0^+} g(\theta, x) = 0$ ;
- $g(\theta, x) > 0$  when  $\theta < x_{(1)}$ ;

Combined with  $g(\hat{\theta}_1(x), x) < 0$ , we have  $\hat{\theta}_q(x) < \hat{\theta}_1(x)$  for  $0 < q < 1$ . ■

**Theorem 4.16** Let  $P$  and  $C$  be two  $n$ -by- $n$  symmetric and hollow matrices satisfying element-wise conditions  $0 < P_{ij} \leq C_{ij} \leq R$  for some constant  $R > 0$ . For  $0 < \epsilon < 1$ , we define  $m$  symmetric and hollow matrices as

$$A^{(t)} \stackrel{iid}{\sim} (1 - \epsilon)\text{Exp}(P) + \epsilon\text{Exp}(C)$$

for  $1 \leq t \leq m$ . Let  $\hat{P}^{(q)}$  be the entry-wise MLqE based on exponential distribution with  $m$  observations. Define  $H^{(q)} = E[\hat{P}^{(q)}]$ , then for any constant  $c > 0$  there exists another constant  $n_0(c)$ , independent of  $n$ ,  $P$ ,  $C$  and  $\epsilon$ , such that if  $n > n_0$ , then for all  $\eta$  satisfying  $n^{-c} \leq \eta \leq 1/2$ ,

$$P \left( \|\hat{P}^{(q)} - H^{(q)}\|_2 \leq 8R\sqrt{2n \ln(n/\eta)} \right) \geq 1 - \eta.$$

**Remark:** This is the extended version of Theorem 3.1 in [4].

**Proof:** Let  $\{e_i\}_{i=1}^n$  be the canonical basis for  $\mathbb{R}^n$ . For each  $1 \leq i, j \leq n$ , define a corresponding matrix  $G_{ij}$ :

$$G_{ij} \equiv \begin{cases} e_i e_j^T + e_j e_i^T, & i \neq j; \\ e_i e_i^T, & i = j. \end{cases}$$

Thus  $\hat{P}^{(q)} = \sum_{1 \leq i < j \leq n} \hat{P}_{ij}^{(q)} G_{ij}$  and  $H^{(q)} = \sum_{1 \leq i < j \leq n} H_{ij}^{(q)} G_{ij}$ . Then we have  $\hat{P}^{(q)} - H^{(q)} = \sum_{1 \leq i < j \leq n} X_{ij}$ , where  $X_{ij} \equiv (\hat{P}_{ij}^{(q)} - H_{ij}^{(q)}) G_{ij}$ ,  $1 \leq i < j \leq n$ .

First consider the  $k$ -th moment of  $X_{ij}$  for  $1 \leq i < j \leq n$ . By Lemma 4.15 we have

$$\begin{aligned} \left| \hat{P}_{ij}^{(q)} - H_{ij}^{(q)} \right| &= \left| \hat{P}_{ij}^{(q)} - \hat{P}_{ij}^{(1)} + \hat{P}_{ij}^{(1)} - H_{ij}^{(1)} + H_{ij}^{(1)} - H_{ij}^{(q)} \right| \\ &\leq \left| \hat{P}_{ij}^{(q)} - \hat{P}_{ij}^{(1)} \right| + \left| \hat{P}_{ij}^{(1)} - H_{ij}^{(1)} \right| + \left| H_{ij}^{(1)} - H_{ij}^{(q)} \right| \\ &\leq \hat{P}_{ij}^{(1)} + \left| \hat{P}_{ij}^{(1)} - H_{ij}^{(1)} \right| + H_{ij}^{(1)} \\ &\leq 2 \left( \left| \hat{P}_{ij}^{(1)} - H_{ij}^{(1)} \right| + H_{ij}^{(1)} \right). \end{aligned}$$

Since

$$\begin{aligned} E[(\hat{P}_{ij}^{(1)} - H_{ij}^{(1)})^k] &\leq (1 - \epsilon) \exp(-H_{ij}/P_{ij}) P_{ij}^k \Gamma(1 + k, -H_{ij}/P_{ij}) \\ &\quad + \epsilon \exp(-H_{ij}/C_{ij}) C_{ij}^k \Gamma(1 + k, -H_{ij}/C_{ij}) \\ &\leq ((1 - \epsilon) \exp(-H_{ij}/P_{ij}) P_{ij}^k + \epsilon \exp(-H_{ij}/C_{ij}) C_{ij}^k) k! \\ &\leq ((1 - \epsilon) P_{ij}^k + \epsilon C_{ij}^k) k! \\ &\leq C_{ij}^k k!, \end{aligned}$$

Then

$$\begin{aligned}
E[(\hat{P}_{ij}^{(q)} - H_{ij}^{(q)})^k] &\leq E\left[\left|\hat{P}_{ij}^{(q)} - H_{ij}^{(q)}\right|^k\right] \\
&\leq 2^k E\left[\left(\left|\hat{P}_{ij}^{(1)} - H_{ij}^{(1)}\right| + H_{ij}^{(1)}\right)^k\right] \\
&\leq 2^k \sum_{s=0}^k \binom{k}{s} E\left[\left|\hat{P}_{ij}^{(1)} - H_{ij}^{(1)}\right|^s\right] \left(H_{ij}^{(1)}\right)^{k-s} \\
&\leq 2^k \sum_{s=0}^k \binom{k}{s} C_{ij}^s s! \left(H_{ij}^{(1)}\right)^{k-s} \\
&\leq 2^k k! \sum_{s=0}^k \binom{k}{s} C_{ij}^s \left(H_{ij}^{(1)}\right)^{k-s} \\
&= 2^k k! \left(C_{ij} + H_{ij}^{(1)}\right)^k.
\end{aligned} \tag{5}$$

Combined with for  $i \neq j$ ,

$$G_{ij}^k \equiv \begin{cases} e_i e_i^T + e_j e_j^T, & k \text{ is even;} \\ e_i e_j^T + e_j e_i^T, & k \text{ is odd,} \end{cases}$$

thus we have

1. When  $k$  is even,

$$E[X_{ij}^k] = E[(\hat{P}_{ij}^{(q)} - H_{ij}^{(q)})^k] G_{ij}^2 \preceq 2^{2k} k! R^k G_{ij}^2;$$

2. When  $k$  is odd,

$$E[X_{ij}^k] = E[(\hat{P}_{ij}^{(q)} - H_{ij}^{(q)})^k] G_{ij} \preceq 2^{2k} k! R^k G_{ij}^2.$$

So

$$E[X_{ij}^k] \preceq 2^{2k} k! R^k G_{ij}^2.$$

Let

$$\sigma^2 := \left\| \sum_{1 \leq i < j \leq n} (4\sqrt{2} R G_{ij})^2 \right\| = 32 R^2 \|(n-1)I\| = 32 R^2 (n-1),$$

notice that random matrices  $X_{ij}$  are independent, self-adjoint and have mean zero, apply Theorem 4.3 we have

$$\begin{aligned}
P\left(\lambda_{\max}(\hat{P}^{(q)} - H^{(q)}) \geq t\right) &\leq n \exp\left(-\frac{t^2/2}{\sigma^2 + 4Rt}\right) \\
&\leq n \exp\left(-\frac{t^2/2}{32R^2 n + Rt}\right).
\end{aligned}$$

Now consider  $Y_{ij} \equiv (H^{(q)} - \hat{P}^{(q)}) G_{ij}$ ,  $1 \leq i < j \leq n$ . Then we have  $H^{(q)} - \hat{P}^{(q)} = \sum_{1 \leq i < j \leq n} Y_{ij}$ . Since

$$E[(H^{(q)} - \hat{P}^{(q)})^k] = (-1)^k E[(\hat{P}^{(q)} - H^{(q)})^k],$$



1. When  $k$  is even,

$$E[Y_{ij}^k] = E[(\hat{P}^{(q)} - H^{(q)})^k] G_{ij}^2 \preceq 2^{2k} k! R^k G_{ij}^2;$$

2. When  $k$  is odd,

$$E[Y_{ij}^k] = -E[(\hat{P}^{(q)} - H^{(q)})^k] G_{ij}^2 \preceq 2^{2k} k! R^k G_{ij}^2.$$

Thus

$$\begin{aligned} P\left(\lambda_{\min}(\hat{P}^{(q)} - H^{(q)}) \leq -t\right) &= P\left(\lambda_{\max}(H^{(q)} - \hat{P}^{(q)}) \geq t\right) \\ &\leq n \exp\left(-\frac{t^2/2}{32R^2n + Rt}\right). \end{aligned}$$

Therefore we have

$$P\left(\|\hat{P}^{(q)} - H^{(q)}\| \geq t\right) \leq n \exp\left(-\frac{t^2/2}{32R^2n + Rt}\right).$$

Now let  $c > 0$  be given and assume  $n^{-c} \leq \eta \leq 1/2$ . Then there exists a  $n_0(c)$  independent of  $n$ ,  $P$ ,  $C$  and  $\epsilon$  such that whenever  $n > n_0(c)$ ,

$$t = 8R\sqrt{2n \ln(n/\eta)} \leq 32Rn.$$

Plugging this  $t$  into the equation above, we get

$$P(\|\hat{P}^{(q)} - H^{(q)}\| \geq 8R\sqrt{2n \ln(n/\eta)}) \leq n \exp\left(-\frac{t^2}{64R^2n}\right) = \eta.$$

■

As we define  $H^{(q)} = E[\hat{P}^{(q)}]$ , let  $d^{(q)} = \text{rank}(H^{(q)})$  be the dimension in which we are going to embed  $\hat{P}^{(q)}$ . Then we can define  $H^{(q)} = ZZ^T$  where  $Z \in \mathbb{R}^{n \times d^{(q)}}$ .

For simplicity, from now on, we will use  $\hat{P}$  to represent  $\hat{P}^{(q)}$ , use  $H$  to represent  $H^{(q)}$  and use  $k$  to represent the dimension  $d^{(q)}$  we are going to embed. Assume  $H = USU^T = ZZ^T$ , where  $Z$  is a  $n$ -by- $k$  matrix. Then our estimate for  $Z$  up to rotation is  $\hat{Z} = \hat{U}\hat{S}^{1/2}$ , where  $\hat{U}\hat{S}\hat{U}^T$  is the rank- $d$  spectral decomposition of  $|\hat{P}| = (\hat{P}^T \hat{P})^{1/2}$ .

RT: We don't have any bound on  $d^{(q)}$ ?

**Lemma 4.17** *Under the above assumptions,  $\lambda_i(H) = \Theta(n)$  with high probability when  $i \leq k$ , i.e. the largest  $k$  eigenvalues of  $H$  is of order  $n$ . Moreover, we have  $\|S\|_2 = \Theta(n)$  and  $\|\hat{S}\|_2 = \Theta(n)$  with high probability.*

**Remark:** This is an extended version of Proposition 4.3 in [7].

**Proof:** Exactly the same as proof for Lemma 4.5. ■

**Lemma 4.18** *Let  $W_1 \Sigma W_2^T$  be the singular value decomposition of  $U^T \hat{U}$ . Then for sufficiently large  $n$ ,*

$$\|U^T \hat{U} - W_1 W_2^T\|_F = O(n^{-1} \log n)$$

*with high probability.*

**Proof:** Exactly the same as proof for Lemma 4.6. ■

We will denote the orthogonal matrix  $W_1 W_2^T$  by  $W^*$ .

**Lemma 4.19** *For sufficiently large  $n$ ,*

$$\|W^* \hat{S} - S W^*\|_F = O(\log n),$$

$$\|W^* \hat{S}^{1/2} - S^{1/2} W^*\|_F = O(n^{-1/2} \log n)$$

and

$$\|W^* \hat{S}^{-1/2} - S^{-1/2} W^*\|_F = O(n^{-3/2} \log n)$$

with high probability.

**Proof:**

By Proposition 2.1 in [6] and Equation (2), we have for some orthogonal matrix  $W$ ,

$$\|\hat{U} - UW\|_F^2 \leq \frac{2\|\hat{U}\hat{U}^T - UU^T\|_F^2}{\delta^2} = O(n^{-1/2} \sqrt{\log n}).$$

Let  $Q = \hat{U} - UU^T \hat{U}$ . And  $Q$  is the residual after projecting  $\hat{U}$  orthogonally onto the column space of  $U$ , we have

$$\|Q\|_F = \|\hat{U} - UU^T \hat{U}\|_F \leq \|\hat{U} - UT\|_F = O(n^{-1/2} \sqrt{\log n}). \quad (6)$$

for all  $k \times k$  matrices  $T$ . Then

$$\begin{aligned} W^* \hat{S} &= (W^* - U^T \hat{U}) \hat{S} + U^T \hat{U} \hat{S} = (W^* - U^T \hat{U}) \hat{S} + U^T \hat{P} \hat{U} \\ &= (W^* - U^T \hat{U}) \hat{S} + U^T (\hat{P} - H) \hat{U} + U^T H \hat{U} \\ &= (W^* - U^T \hat{U}) \hat{S} + U^T (\hat{P} - H) Q + U^T (\hat{P} - H) U U^T \hat{U} + U^T H \hat{U} \\ &= (W^* - U^T \hat{U}) \hat{S} + U^T (\hat{P} - H) Q + U^T (\hat{P} - H) U U^T \hat{U} + S U^T \hat{U}. \end{aligned}$$

Combined with Theorem 4.16, Lemma 4.17, Lemma 4.18, we have

$$\begin{aligned} &\|W^* \hat{S} - S W^*\|_F \\ &= \|(W^* - U^T \hat{U}) \hat{S} + U^T (\hat{P} - H) Q + U^T (\hat{P} - H) U U^T \hat{U} + S(U^T \hat{U} - W^*)\|_F \\ &\leq \|W^* - U^T \hat{U}\|_F (\|\hat{S}\|_2 + \|S\|_2) + \|U^T\|_F \|\hat{P} - H\|_2 \|Q\|_F + \|U^T (\hat{P} - H) U\|_F \\ &\leq O(\log n) + O(\log n) + \|U^T (\hat{P} - H) U\|_F \end{aligned}$$

with high probability. And we know  $U^T (\hat{P} - H) U$  is a  $k \times k$  matrix with  $ij$ -th entry to be

$$u_i^T (\hat{P} - H) u_j = \sum_{s=1}^n \sum_{t=1}^n (\hat{P}_{st} - H_{st}) u_{is} u_{jt} = 2 \sum_{s < t} (\hat{P}_{st} - H_{st}) u_{is} u_{jt}$$

where  $u_i$  and  $u_j$  are the  $i$ -th and  $j$ -th columns of  $U$ . Thus, conditioned on  $H$ ,  $u_i^T (\hat{P} - H) u_j$  is a sum of independent mean 0 random variables.

RT: Not right here, I am using the result for spectral norm as frobenius norm.

By Equation (5), we have

$$\begin{aligned} & E \left[ \left( (\hat{P}_{st} - H_{st}) u_{is} u_{jt} \right)^k \right] \\ & \leq 2^k k! (C_{st} + H_{st}^{(1)})^k u_{is}^k u_{jt}^k \\ & \leq \frac{k!}{2} (4R)^{k-2} (4\sqrt{2}R u_{is} u_{jt})^2. \end{aligned}$$

Also we have

$$\sigma^2 := \left| \sum_{s < t} 32R^2 u_{is}^2 u_{jt}^2 \right| \leq 32R^2,$$

then by Theorem 4.3, we have

$$P \left( \left| 2 \sum_{s < t} (\hat{P}_{st} - H_{st}) u_{is} u_{jt} \right| \geq t \right) \leq \exp \left( \frac{-t^2/8}{32R^2 + 2Rt} \right),$$

thus each entry of  $U^T(\hat{P} - H)U$  is of order  $O(\log n)$  with high probability and thus

$$\|U^T(\hat{P} - H)U\|_F = O(\log n) \quad (7)$$

with high probability. Hence

$$\|W^* \hat{S} - SW^*\|_F = O(\log n)$$

with high probability. Also, since

$$W_{ij}^* (\lambda_j^{1/2}(\hat{P}) - \lambda_i^{1/2}(H)) = W_{ij}^* \frac{\lambda_j(\hat{P}) - \lambda_i(H)}{\lambda_j^{1/2}(\hat{P}) + \lambda_i^{1/2}(H)}$$

and the eigenvalues  $\lambda_j^{1/2}(\hat{P})$  and  $\lambda_i^{1/2}(H)$  are both of order  $\Theta(\sqrt{n})$ , we have

$$\|W^* \hat{S}^{1/2} - S^{1/2} W^*\|_F = O(n^{-1/2} \log n).$$

Similarly, since

$$W_{ij}^* (\lambda_j^{-1/2}(\hat{P}) - \lambda_i^{-1/2}(H)) = W_{ij}^* \frac{\lambda_i(H) - \lambda_j(\hat{P})}{(\lambda_j^{-1/2}(\hat{P}) + \lambda_i^{-1/2}(H)) \lambda_j(\hat{P}) \lambda_i(H)}$$

and the eigenvalues  $\lambda_j(\hat{P})$  and  $\lambda_i(H)$  are both of order  $\Theta(n)$ , we have

$$\|W^* \hat{S}^{-1/2} - S^{-1/2} W^*\|_F = O(n^{-3/2} \log n).$$

■

**Lemma 4.20** *There exists a rotation matrix  $W$  such that for sufficiently large  $n$ ,*

$$\|\hat{Z} - ZW\|_F = \|(\hat{P} - H)US^{-1/2}\|_F + O(n^{-1/2}(\log n)^{3/2})$$

*with high probability.*

**Proof:** Exactly the same as proof for Lemma 4.8.

■

**Theorem 4.21** *There exists a rotation matrix  $W$  such that for sufficiently large  $n$ ,*

$$\max_i \|\hat{Z}_i - WZ_i\|_2 = O(n^{-1/2}(\log n)^{3/2})$$

*with high probability.*

**Proof:** By Lemma 4.20, we have

$$\|\hat{Z} - ZW\|_F = \|(\hat{P} - H)US^{-1/2}\|_F + O(n^{-1/2}(\log n)^{3/2})$$

and similarly we could have the bound for each row vector

$$\begin{aligned} \max_i \|\hat{Z}_i - WZ_i\|_2 &\leq \frac{1}{\lambda_k^{1/2}(H)} \max_i \|((\hat{P} - H)U)_i\|_2 + O(n^{-1/2}(\log n)^{3/2}) \\ &\leq \frac{k^{1/2}}{\lambda_k^{1/2}(H)} \max_j \|(\hat{P} - H)u_j\|_\infty + O(n^{-1/2}(\log n)^{3/2}) \end{aligned}$$

where  $u_j$  denotes the  $j$ -th column of  $U$ . Now given  $i$  and  $j$ , the  $i$ -th element of the vector  $(\hat{P} - H)u_j$  is of the form

$$\sum_{s=1}^n (\hat{P}_{is} - H_{is})u_{js} = \sum_{s \neq i} (\hat{P}_{is} - H_{is})u_{js}.$$

Thus, conditioned on  $H$ , the  $i$ -th element of the vector  $(\hat{P} - H)u_j$  is a sum of independent mean 0 random variables. By Equation (5), we have

$$\begin{aligned} &E \left[ \left( (\hat{P}_{is} - H_{is})u_{js} \right)^k \right] \\ &\leq 2^k k! (C_{is} + H_{is}^{(1)})^k u_{js}^k \\ &\leq \frac{k!}{2} (4R)^{k-2} (4\sqrt{2}Ru_{js})^2. \end{aligned}$$

Also we have

$$\sigma^2 := \left| \sum_{s \neq i} 32R^2 u_{js}^2 \right| \leq 32R^2,$$

then by Theorem 4.3, we have

$$P \left( \left| \sum_{s \neq i} (\hat{P}_{is} - H_{is})u_{js} \right| \geq t \right) \leq \exp \left( \frac{-t^2/2}{32R^2 + Rt} \right),$$

i.e. it can be bounded by a constant with high probability. Taking the union bound over all  $i$  and  $j$ , we have

$$\max_i \|\hat{Z}_i - WZ_i\|_2 \leq \frac{Cd^{1/2}}{\lambda_d^{1/2}(H)} (\log n)^{3/2} + O(n^{-1/2}(\log n)^{3/2}) = O(n^{-1/2}(\log n)^{3/2}).$$

■

**Lemma 4.22**  $\left| \hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j \right| = O(n^{-1}(\log n)^3)$  *with high probability.*

**Proof:** Let  $W$  be the rotation matrix in Theorem 4.21, then

$$\begin{aligned} \left| \hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j \right| &= \left| \hat{Z}_i^T \hat{Z}_j - \hat{Z}_i^T W Z_j + \hat{Z}_i^T W Z_j - (W Z_i)^T W Z_j \right| \\ &\leq \left| \hat{Z}_i^T (\hat{Z}_j - W Z_j) + (\hat{Z}_i^T - (W Z_i)^T) W Z_j \right| \\ &\leq \|\hat{Z}_i\|_2 \|\hat{Z}_j - W Z_j\|_2 + \|Z_j\|_2 \|\hat{Z}_i^T - (W Z_i)^T\|_2. \end{aligned}$$

Since  $\|Z_i\|_2^2 = Z_i^T Z_i = H_{ii}^q = E[\hat{P}_{ii}^{(q)}] \leq E[\hat{P}_{ii}^{(1)}] = (1 - \epsilon)P_{ij} + \epsilon C_{ij} \leq R$ , we have  $\|Z_i\|_2 = O(1)$ . Combined with Theorem 4.21,

$$\begin{aligned} \left| \hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j \right| &= (\|\hat{Z}_i\|_2 + \|Z_j\|_2) O(n^{-1/2} (\log n)^{3/2}) \\ &\leq (\|\hat{Z}_i - W Z_i\|_2 + \|W Z_i\|_2 + \|Z_j\|_2) O(n^{-1/2} (\log n)^{3/2}) \\ &= O(n^{-1} (\log n)^3) \end{aligned}$$

with high probability. ■

**Corollary 4.23** *For fixed  $m$ , the estimator based on ASE of MLqE has the same entry-wise asymptotic bias as MLqE, i.e.*

$$\lim_{n \rightarrow \infty} \text{Bias}(\tilde{P}_{ij}^{(q)}) = \lim_{n \rightarrow \infty} E[\tilde{P}_{ij}^{(q)}] - P_{ij} = \lim_{n \rightarrow \infty} E[\hat{P}_{ij}^{(q)}] - P_{ij} = \lim_{n \rightarrow \infty} \text{Bias}(\hat{P}_{ij}^{(q)}).$$

**Proof:** Direct result from Lemma 4.22 by noticing

$$\lim_{n \rightarrow \infty} E[\tilde{P}_{ij}^{(q)}] = \lim_{n \rightarrow \infty} E[\hat{P}_{ij}^{(q)}].$$
■

**Theorem 4.24**  $\text{Var}(\hat{Z}_i^T \hat{Z}_j) = O(n^{-2} (\log n)^6)$  with high probability.

**Proof:**

$$\begin{aligned} \text{Var}(\hat{Z}_i^T \hat{Z}_j) &= E[(\hat{Z}_i^T \hat{Z}_j - E[\hat{Z}_i^T \hat{Z}_j])^2] \\ &= E[(\hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j + Z_i^T Z_j - E[\hat{Z}_i^T \hat{Z}_j])^2] \\ &= E[(\hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j)^2] + E[(Z_i^T Z_j - E[\hat{Z}_i^T \hat{Z}_j])^2] \\ &\quad - 2E[(\hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j)(Z_i^T Z_j - E[\hat{Z}_i^T \hat{Z}_j])] \\ &\leq E[(\hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j)^2] + E[(Z_i^T Z_j - E[\hat{Z}_i^T \hat{Z}_j])^2] \\ &\quad + 2\sqrt{E[(\hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j)^2] E[(Z_i^T Z_j - E[\hat{Z}_i^T \hat{Z}_j])^2]} \\ &\leq 4E[(\hat{Z}_i^T \hat{Z}_j - Z_i^T Z_j)^2] \\ &= O(n^{-2} (\log n)^6) \end{aligned}$$

with high probability. ■

**Theorem 4.25** *Let  $u_q(\theta) = E_\theta[\hat{\theta}_{q,n}]$ ,  $\phi_q(x; \theta) = \frac{\partial}{\partial \theta} L_q(f(x; \theta))$ , and  $\phi'_q(x; \theta) = \frac{\partial^2}{\partial \theta^2} L_q(f(x; \theta))$ . Then the asymptotic distribution of  $\hat{\theta}_{q,n}$  is  $\sqrt{n}(\hat{\theta}_{q,n} - u_q(\theta)) \sim \mathcal{N}(0, V_q(\theta))$ , where  $V_q(\theta) = E[\phi_q(X; \theta)^2] / E[\phi'_q(X; \theta)]^2$ .*

**Remark:** See Theorem 1 in <http://arxiv.org/pdf/1310.7278.pdf>.

**Corollary 4.26**  $\text{Var}(\hat{P}_{ij}^{(q)}) = \Theta(m^{-1})$ .

**Proof:** Direct result from Theorem 4.25. ■

**Theorem 4.27** For fixed  $m$ ,  $1 \leq i, j \leq n$ ,

$$\frac{\text{Var}(\tilde{P}_{ij}^{(q)})}{\text{Var}(\hat{P}_{ij}^{(q)})} = O(mn^{-2}(\log n)^6).$$

Thus

$$\text{ARE}(\hat{P}_{ij}^{(q)}, \tilde{P}_{ij}^{(q)}) = 0.$$

Furthermore, as long as  $m$  goes to infinity of order  $o(n(\log n)^{-3})$ ,

$$\text{ARE}(\hat{P}_{ij}^{(q)}, \tilde{P}_{ij}^{(q)}) = 0.$$

**Proof:** The results are direct from Theorem 4.24 and Corollary 4.26. ■

#### 4.4 $\tilde{P}^{(q)}$ better than $\tilde{P}^{(1)}$

**Theorem 4.28** For sufficiently large  $n$  and  $C$ , any  $1 \leq i, j \leq n$ ,

$$\lim_{m \rightarrow \infty} \text{Bias}(\tilde{P}_{ij}^{(1)}) > \lim_{m \rightarrow \infty} \text{Bias}(\tilde{P}_{ij}^{(q)})$$

**Proof:** Direct result from Lemma 3.1, Corollary 4.11 and Corollary 4.23. ■

**Theorem 4.29** For any fixed  $m$ , any  $1 \leq i, j \leq n$ ,

$$\lim_{n \rightarrow \infty} \text{Var}(\tilde{P}_{ij}^{(1)}) = \lim_{n \rightarrow \infty} \text{Var}(\tilde{P}_{ij}^{(q)}) = 0$$

**Proof:** Direct result from Theorem 4.12 and Theorem 4.24. ■

## 5 Generalization

We can generalize the exponential distribution to  $F$  and generalize MLqE to  $\hat{P}$  with the following assumptions:

- There exists  $C_0(P_{ij}, \epsilon) > 0$  such that under the contaminated model with  $C > C_0(P_{ij}, \epsilon)$ ,

$$\lim_{m \rightarrow \infty} \left| E[\hat{P}_{ij}] - P_{ij} \right| < \lim_{m \rightarrow \infty} \left| E[\hat{P}_{ij}^{(1)}] - P_{ij} \right|,$$

for  $1 \leq i, j \leq n$ .

- Let  $A^{(t)} \stackrel{iid}{\sim} (1 - \epsilon)F(P) + \epsilon F(C)$  and  $H_{ij}^{(1)} = E[\hat{P}_{ij}^{(1)}] = (1 - \epsilon)E_F(P_{ij}) + \epsilon E_F(C_{ij})$ , then  $E[(A_{ij}^{(t)} - H_{ij}^{(1)})^k] \leq \text{const} \cdot k!$ .
- $\hat{P}_{ij} \leq \text{const} \cdot \hat{P}_{ij}^{(1)}$ . This might be generalized to with high probability later.
- $\lim_{m \rightarrow \infty} \text{Var}(\hat{P}_{ij}) = 0$

## 6 Analysis of MLqE under Exponential Model

Let  $f(x, \theta) = \sum_{i=1}^m e^{-\frac{(1-q)x_i}{\theta}} (x - \theta)$ , then  $\frac{\partial f}{\partial x_i}(x, \theta) = e^{-\frac{(1-q)x_i}{\theta}} \left( -\frac{(1-q)x_i}{\theta} + 2 - q \right)$ .

Define  $c_q = 1 + \frac{1}{1-q}$ , then

$$\frac{\partial f}{\partial x_i}(x, \theta) = 0 \Leftrightarrow x_i \rightarrow \infty \text{ or } x_i = \left(1 + \frac{1}{1-q}\right)\theta$$

Define  $x = (x_1, \dots, x_m)$  and  $x^- = (x_1, \dots, x_m) \setminus x_i$ , then

1.  $\theta_{\max}, x_i = c_q \theta \leq c_q \bar{x}$   
 $\frac{1}{m} (c_q \bar{x} + (m-1)\bar{x}^-) = \bar{x}$   
Thus  $\theta_{\max} \leq \frac{m-1}{m-C_q} \bar{x}^-$
2.  $\theta_{\min}$   
2 candidates:  $x_i = 0$  or  $x_i = \infty$ .  
When  $x_i = 0$ , It's like  $m-1$  points  $x^-$ ;  
When  $x_i = \infty$ ,  $f(x, \theta) = f(x^-, \theta) - \theta$ .  
Thus  $\theta_{\min}$  solves  $f(x, \theta)|_{x_i=0} = 0$ .

## 7 Appendix

**Definition 7.1** A random variable  $Z$  is called central moment bounded with real parameter  $L > 0$ , if for any integer  $i \geq 1$  we have

$$E \left[ |Z - E[Z]|^i \right] \leq i \cdot L \cdot E \left[ |Z - E[Z]|^{i-1} \right].$$

## References

- [1] Chandler Davis and William Morton Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.
- [2] Davide Ferrari and Yuhong Yang. Maximum lq-likelihood estimation. *Ann. Statist.*, 38(2):753–783, 04 2010.
- [3] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.
- [4] Roberto Imbuzeiro Oliveira. Concentration of the adjacency matrix and of the laplacian in random graphs with independent edges. *arXiv preprint arXiv:0911.0600*, 2009.
- [5] Yichen Qin and Carey E Priebe. Maximum l q-likelihood estimation via the expectation-maximization algorithm: A robust estimation of mixture models. *Journal of the American Statistical Association*, 108(503):914–928, 2013.
- [6] Karl Rohe, Sourav Chatterjee, and Bin Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, pages 1878–1915, 2011.



- [7] Daniel L Sussman, Minh Tang, and Carey E Priebe. Consistent latent position estimation and vertex classification for random dot product graphs. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(1):48–57, 2014.
- [8] Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012.