

Proof for ARE Theory

Runze

June 14, 2015

1 Problem Description

Under the stochastic block model with parameters (B, ρ) , we have $X_i \stackrel{iid}{\sim} \sum_{k=1}^K \rho_k \delta_{\nu_k}$, where $\nu = [\nu_1, \dots, \nu_K]^T$ satisfies $B = \nu^T \nu$. Define the block assignment τ such that $\tau_i = k$ if and only if $X_i = \nu_k$. Let $P = XX^T$ where $X = [X_1, \dots, X_n]^T$.

First draw τ from the multinomial distribution with parameter ρ . Then we are going to sample m i.i.d. graphs $A^{(1)}, \dots, A^{(m)}$ such that $A_{ij}^{(k)} \stackrel{iid}{\sim} \text{Bern}(P_{ij})$, $1 \leq k \leq m, 1 \leq i, j \leq n$.

Define $\bar{A} = \frac{1}{m} \sum_{k=1}^m A^{(k)}$. Let USU^T be the rank- d decomposition of \bar{A} , then we define $\hat{X} = US^{1/2}$.

We would like to know either \bar{A} or $\hat{P} = \hat{X}\hat{X}^T$ is a better estimation of P .

2 Asymptotic Relative Efficiency

When two reasonable estimators S and T of the parameter θ are considered, we always want to know which one is preferred. When both of them are unbiased, the one with a smaller variance would be more efficient. So if $\{S_n\}$ and $\{T_n\}$ are asymptotic unbiased for θ , then define the asymptotic relative efficiency of $\{S_n\}$ with respect to $\{T_n\}$ to be

$$ARE(S_n, T_n) = \lim_{n \rightarrow \infty} \frac{\text{Var}(T_n)}{\text{Var}(S_n)}.$$

An extended version of ARE is that, when $\{S_n\}$ and $\{T_n\}$ are sequences of estimators for θ (not necessarily to be asymptotic unbiased), then define the asymptotic relative efficiency of $\{S_n\}$ with respect to $\{T_n\}$ to be

$$ARE(S_n, T_n) = \lim_{n \rightarrow \infty} \frac{\text{Var}(T_n)/E[T_n]^2}{\text{Var}(S_n)/E[S_n]^2}.$$

3 Proofs

3.1 \bar{A}

Since \bar{A}_{ij} is the mean of m i.i.d. Bernoulli random variables with parameter P_{ij} , we have $E[\bar{A}_{ij}] = P_{ij}$ and $\text{Var}(\bar{A}_{ij}) = \frac{1}{m} P_{ij}(1 - P_{ij})$.

3.2 \hat{P}

First, we need to prove $E[\hat{X}_i] = X_i$.

$$ARE = \lim_{n \rightarrow \infty} \frac{Var(\hat{P}_{ij})}{Var(\hat{A}_{ij})} = \lim_{n \rightarrow \infty} \frac{E[(\hat{P}_{ij} - P_{ij})^2]}{E[(\hat{A}_{ij} - P_{ij})^2]}$$

In Athreya et al. (2013), Theorem 4.8 suggests that conditioned on $X_i = \nu_k$, $P\left(\sqrt{n}(\hat{X}_i - \nu_k) \leq z | X_i = \nu_k\right) \rightarrow \Phi(z, \Sigma(x_i))$ as $n \rightarrow \infty$, where $\Sigma(x) = \Delta^{-1}E[X_j X_j^T (x^T X_j - (x^T X_j)^2)]\Delta^{-1}$, $\Delta = E[X_1 X_1^T]$ and $\Phi(z, \Sigma)$ denotes the cumulative distribution function for the multivariate normal, with mean zero and covariance matrix Σ , evaluated at z . So the sequence of random variables $\sqrt{n}(\hat{X}_i - \nu_k)$ converges in distribution to a normal distribution. So conditioned on $X_i = \nu_k$, we have

- $\lim_{n \rightarrow \infty} E[\hat{X}_i] = \nu_k$;
- $\lim_{n \rightarrow \infty} nCov(\hat{X}_i, \hat{X}_i) = \Sigma(\nu_k)$, where $\Sigma(x) = \Delta^{-1}E[X_j X_j^T (x^T X_j)(1 - x^T X_j)]\Delta^{-1}$ and $\Delta = E[X_1 X_1^T]$.

Here the setting is a little different from Avanti's CLT theorem. We have m i.i.d. graph, so \bar{A} is closer to P than A_i with a scale m . Thus the new version is: conditioned on $X_i = \nu_k$, we have

- $\lim_{n \rightarrow \infty} E[\hat{X}_i] = \nu_k$;
- $\lim_{n \rightarrow \infty} nCov(\hat{X}_i, \hat{X}_i) = \Sigma(\nu_k)/m$.

I will update more details later.

In Athreya et al. (2013), Corollary 4.11 says \hat{X}_i and \hat{X}_j are asymptotically independent. Since $\hat{P}_{ij} = \hat{X}_i^T \hat{X}_j$ is a noisy dot product of $\nu_s^T \nu_t$, by Equation 5 in Brown and Rutemiller (1977), combined with asymptotic independence between \hat{X}_i and \hat{X}_j and $E[\hat{X}_i] = \nu_s$ when conditioning on $X_i = \nu_s$ and $X_j = \nu_t$, when n is large enough, we have

$$E[(\hat{P}_{ij} - P_{ij})^2] \approx \frac{1}{mn} (\nu_s^T \Sigma(\nu_t) \nu_s + \nu_t^T \Sigma(\nu_s) \nu_t^T) + \frac{1}{mn^2} (tr(\Sigma(\nu_s) \Sigma(\nu_t))). \quad (1)$$

Lemma 3.1 $\nu_s^T \Sigma(\nu_t) \nu_s = \frac{1}{\rho_s} \nu_s^T \nu_t (1 - \nu_s^T \nu_t)$.

Proof: Under the stochastic block model with parameters (B, ρ) , we have $X_i \stackrel{iid}{\sim} \sum_{k=1}^K \rho_k \delta_{\nu_k}$, where $\nu = [\nu_1, \dots, \nu_K]^T$ satisfies $B = \nu^T \nu$. Without loss of generality, we could assume that $\nu = US$ where $U = [u_1, \dots, u_K]^T$ is orthonormal in columns and S is a diagonal matrix. Here we can conclude that $\nu_s^T = u_s^T S$. Also define $R = \text{diag}(\rho_1, \dots, \rho_K)$, then we have

$$\Delta = E[X_1 X_1^T] = \sum_{k=1}^K \rho_k \nu_k \nu_k^T = \nu^T R \nu = S U^T R U S.$$

Thus

$$\begin{aligned}
\nu_s^T \Sigma(\nu_t) \nu_s &= \nu_s^T \Delta^{-1} \sum_{k=1}^K \rho_k \nu_k \nu_k^T (\nu_t^T \nu_k) (1 - \nu_t^T \nu_k) \Delta^{-1} \nu_s \\
&= \sum_{k=1}^K \rho_k (\nu_s^T \Delta^{-1} \nu_k) (\nu_k^T \Delta^{-1} \nu_s) (\nu_t^T \nu_k) (1 - \nu_t^T \nu_k) \\
&= \sum_{k=1}^K \rho_k (u_s^T U^T R^{-1} U u_k)^2 (\nu_t^T \nu_k) (1 - \nu_t^T \nu_k) \\
&= \sum_{k=1}^K \rho_k (e_s^T R^{-1} e_k)^2 (\nu_t^T \nu_k) (1 - \nu_t^T \nu_k) \\
&= \sum_{k=1}^K \rho_k \delta_{sk} \rho_s^{-2} (\nu_t^T \nu_k) (1 - \nu_t^T \nu_k) \\
&= \frac{1}{\rho_s} \nu_t^T \nu_s (1 - \nu_t^T \nu_s)
\end{aligned}$$

■

Lemma 3.2 $tr(\Sigma(\nu_s)\Sigma(\nu_t)) = \sum_{k,l=1}^K u_k^T S^{-2} U_l \nu_l^T \nu_k (1 - \nu_t^T \nu_k) \nu_s^T \nu_l (1 - \nu_s^T \nu_l)$.

Proof: Same as proof for Lemma ??.

■

This lemma is not used here, just a draft for possible further exploration.

By Lemma ?? and Lemma ??, as $n \rightarrow \infty$, Equation ?? can be written as:

$$E[(\hat{P}_{ij} - P_{ij})^2] \approx \frac{1}{mn} \left(\nu_{\tau_i}^T \Sigma(\nu_{\tau_j}) \nu_{\tau_i} + \nu_{\tau_j}^T \Sigma(\nu_{\tau_i}) \nu_{\tau_j} \right) + o\left(\frac{1}{mn}\right) \quad (2)$$

$$\approx \frac{1}{mn} \left(\frac{1}{\rho_{\tau_i}} + \frac{1}{\rho_{\tau_j}} \right) \nu_{\tau_i}^T \nu_{\tau_j} (1 - \nu_{\tau_i}^T \nu_{\tau_j}) \quad (3)$$

$$= \frac{1}{mn} \left(\frac{1}{\rho_{\tau_i}} + \frac{1}{\rho_{\tau_j}} \right) P_{ij} (1 - P_{ij}) \quad (4)$$

3.3 ARE

So the asymptotic relative efficiency is

$$\begin{aligned}
nARE(\bar{A}_{ij}, \hat{P}_{ij}) &= \lim_{n \rightarrow \infty} n \frac{Var(\hat{P}_{ij})}{Var(\bar{A}_{ij})} = \lim_{n \rightarrow \infty} n \frac{E[(\hat{P}_{ij} - P_{ij})^2]}{Var(\bar{A}_{ij})} \\
&= \lim_{n \rightarrow \infty} n \frac{(1/\rho_{\tau_i} + 1/\rho_{\tau_j}) P_{ij} (1 - P_{ij}) / mn}{P_{ij} (1 - P_{ij}) / m} \\
&= \rho_{\tau_i}^{-1} + \rho_{\tau_j}^{-1}
\end{aligned}$$

And the relative efficiency could be approximated by $(\rho_{\tau_i}^{-1} + \rho_{\tau_j}^{-1}) / n$ when n is large enough.