'Oracle inequalites' für Netzwerkmodelle

Saskia F. Glaffig 06.12.17

Modell und Bezeichnungen

Das Modell, Schätzung der Verbindungswahrscheinlichkeiten

Wir betrachten ein Netzwerk, welches als ungerichteter Graph mit n Knoten definiert wird. A bezeichne die (symmetrische) Adjazenzmatrix des Graphen, wobei wir $A_{i,i} = 0$ für alle $1 \leq i \leq n$ setzen und $A_{i,j}$ für alle $1 \leq j < i \leq n$ unabhängige, Bernoulli-verteilte Zufallsvariablen mit Parameter $(\Theta_0)_{i,j}$ seien. Die symmetrische $(n \times n)$-Matrix der Verbindungswahrscheinlichkeiten mit den Einträgen $(\Theta_0)_{i,j}$ und $(\Theta_0)_{i,i} = 0$ für alle $1 \leq i \leq n$ bezeichnen wir mit Θ_0.
Der Graph wird durch das 'stochastic block model' mit $k \leq n$ Klassen approximiert. Jede dieser Klassen habe mindestens n_0 Elemente.

Bezeichnungen

- Für $m \in \mathbb{N}$ sei $[m] = \{1, ..., m\}$.
- Seien $n, k, n_0 \in \mathbb{N}$, so dass $2 \leq k \leq n$ und $n_0 \leq n$. Weiter sei

$$\mathcal{Z}_{n,k,n_0} = \{z : [n] \to [k] \quad | \min_{a \in [k]} |z^{-1}(a)| \geq n_0\}.$$

- $||B||_F$ bezeichne die Frobenius-Norm und $||B||_\infty$ die Supremumsnorm einer Matrix $B \in \mathbb{R}^{n \times n}$.

1. 'Oracle inequalities'

Im Folgenden sei $k \in \mathbb{N}$ fest.

Kleinster-Quadrate-Schätzer (LSE)

Für $z \in \mathcal{Z}_{n,k,n_0}$ und $Q \in \mathbb{R}^{k \times k}_{sym}$ sei

$$L(Q, z) = \sum_{(a,b) \in [k] \times [k]} \quad \sum_{(i,j) \in z^{-1}(a) \times z^{-1}(b), j < i} (A_{i,j} - Q_{a,b})^2.$$

Der LSE von (Q, z) ist dann

$$(\hat{Q}, \hat{z}) \quad \in \underset{Q \in \mathbb{R}^{k\times k}_{sym}, z \in \mathcal{Z}_{n,k,n_0}}{\arg\min} \quad L(Q, z),$$

und der LSE von Θ_0 sei definiert als symmetrische $(n \times n)$-Matrix mit den Einträgen

$$\begin{aligned} \hat{\Theta}_{i,j} &= \hat{Q}_{\hat{z}(i),\hat{z}(j)} \qquad && \forall i > j, \\ \hat{\Theta}_{i,i} &= 0 && \forall i. \end{aligned}$$

Eingeschränkter kleinste-Quadrate-Schätzer

Sei $r \in (0, 1]$ gegeben und $\mathcal{B}_r = \{Q \in \mathbb{R}^{k\times k}_{sym} : ||Q||_\infty \le r\}$. Definiere

$$(\hat{Q}_r, \hat{z}_r) \quad \in \underset{Q \in \mathcal{B}_r, z \in \mathcal{Z}_{n,k,1}}{\arg\min} \; L(Q, z).$$

Der eingeschränkte LSE $\hat{\Theta}^r$ von Θ_0 sei die symmetrische $(n\times n)$-Matrix mit den Einträgen

$$\begin{aligned} \hat{\Theta}^r_{i,j} &= (\hat{Q}_r)_{\hat{z}_r(i),\hat{z}_r(j)} \qquad && \forall i > j \\ \hat{\Theta}^r_{i,i} &= 0 && \forall i. \end{aligned}$$

'Oracle inequalities'

Sei Θ_{*,n_0} die optimale Approximation von Θ_0 in der Menge

$$\mathcal{T}_{n_0}[k] = \{\Theta \in \mathbb{R}^{n\times n} : \exists z \in \mathcal{Z}_{n,k,n_0}, Q \in \mathbb{R}^{k\times k}_{sym}, \text{ so dass } \Theta_{i,j} = Q_{z(i),z(j)}, i \ne j \text{ und } \Theta_{i,i} = 0\, \forall i\}$$

bezüglich der Frobenius-Norm.

Theorem 1. *Es gibt positive absolute Konstanten C_1, C_2, C_3, so dass die Ungleichung*

$$\mathbb{E}\left[\frac{1}{n^2}||\hat{\Theta} - \Theta_0||_F^2\right] \le \frac{C_1}{n^2}||\Theta_0 - \Theta_{*,n_0}||_F^2 + C_2||\Theta_0||_\infty \left(\frac{\log(k)}{n} + \frac{k^2}{n^2}\right) + C_3 \frac{\log(n/n_0)}{n_0}\left(\frac{\log(k)}{n} + \frac{k^2}{n^2}\right) \tag{1}$$

für alle $n_0 \ge 2$ erfüllt ist.

Im Fall von gleichmäßigen Klassen ($n_0 = \mathcal{O}(n/k)$) ergibt sich daraus das folgende Resultat.

Korollar 2. *Es gelte $n_0 \ge Cn/k$ für ein $C > 0$ und $||\Theta_0||_\infty \ge \frac{k\log(k)}{n}$. Dann gibt es positive Konstanten C_1, C_2, so dass die Ungleichung*

$$\mathbb{E}\left[\frac{1}{n^2}||\hat{\Theta} - \Theta_0||_F^2\right] \le \frac{C_1}{n^2}||\Theta_0 - \Theta_{*,n_0}||_F^2 + C_2||\Theta_0||_\infty \left(\frac{\log(k)}{n} + \frac{k^2}{n^2}\right) \tag{2}$$

für $n_0 \ge 2$ erfüllt ist.

Theorem 3. *Es gibt positive Konstanten C_1, C_2, so dass die folgende Aussage gilt: Ist $||\Theta_0||_\infty \leq r$, so folgt*

$$\mathbb{E}\left[\frac{1}{n^2}||\hat{\Theta}^r - \Theta_0||_F^2\right] \leq \frac{C_1}{n^2}||\Theta_0 - \Theta_{*,1}||_F^2 + C_2 r\left(\frac{\log(k)}{n} + \frac{k^2}{n^2}\right). \tag{3}$$

Literatur

[1] O. Klopp, A. B. Tsybakov und N. Verzelen. «Oracle Inequalities for Network Models and Sparse Graphon Estimation». In: *The Annals of Statistics* 45.1 (2017), S. 316–354.