Estimation and Inference in Sparse Autoregressive Networks

Binyan Jiang, Yutong Wang, Qiwei Yao
by.jiang@polyu.edu.hk, y.wang487@lse.ac.uk, q.yao@lse.ac.uk

Problem

To predict the evolution of dynamic networks, we model it by a network AR(1) process.

Given a sample of adjacency matrix $\{X_1, \cdots, X_n\}$, our first purpose is to estimate the parameters $(\alpha_{i,j})_{p\times p}$, $(\beta_{i,j})_{p\times p}$, and find a proper embedding into a space with lower dimension (find a simpler representation for parameters). Thus, the second purpose is to estimate $(\theta_i, \eta_i)_{i=1}^p$.

Concepts

The **adjacency matrix** is one way preferred by mathematicians to represent networks. A network with n nodes can be represented by an n-by-n matrix X, where node i and j are connected once $X_{i,j} = 1$.

α-mixing coefficient is firstly defined for two σ-algebra $\mathcal{A}$ and $\mathcal{B}$:

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A\in\mathcal{A}, B\in\mathcal{B}} |P(A\cap B) - P(A)P(B)|.$$

For time series $\{X_t\}_{t=0}^\infty$, it is defined as:

$$\alpha_{X_t}(n) = \sup_{k\geq 1} \alpha(\mathcal{M}_k, \mathcal{G}_{k+n}),$$

where $\mathcal{M}_j = \sigma(\{X_i, i \leq j\})$, $\mathcal{G}_j = \sigma(\{X_i, i \geq j\})$

Models

We consider an AR(1) dynamic network defined on p fixed nodes, denoted by $\{1, \cdots, p\}$, with the $p\times p$ adjacency matrix $X_t = (X_{i,j}^t)$ at time t defined by

$$X_{i,j}^t = X_{i,j}^{t-1} I(\varepsilon_{i,j}^t = 0) + I(\varepsilon_{i,j}^t = 1), t\geq 1, \quad (1)$$

innovations $\varepsilon_{i,j}^t$, $1\leq i < j \leq p$, are independent, and

$P(\varepsilon_{i,j}^t = 1) = \alpha_{i,j}, P(\varepsilon_{i,j}^t = -1) = \beta_{i,j}$,

$P(\varepsilon_{i,j}^t = 0) = 1 - \alpha_{i,j} - \beta_{i,j}$.

Thus, $\{X_t\}$ is a Markov process, with

$$P(X_{i,j}^t = 1 | X_{i,j}^{t-1} = 0) = \alpha_{i,j},$$

$$P(X_{i,j}^t = 0 | X_{i,j}^{t-1} = 1) = \beta_{i,j}.$$

In addition, assume parameters $\alpha_{i,j}$ and $\beta_{i,j}$ is generated from $\{\theta_i, \eta_i\}_{i=1}^p$ by:

$$\alpha_{i,j} = \theta_i\theta_j, \quad \beta_{i,j} = \eta_i\eta_j.$$

This setting comes from the insights that connecting and breaking probability $\alpha_{i,j}, \beta_{i,j}$ should be explained by node i and node j's node-specific property: θ_i, η_i and θ_j, η_j. Here the property is in dimension 1, and the dimension could be higher, the corresponding model is called dot-product random graph.

References

[1] Merlevede, F., Peligrad, M., Rio, E., et al. (2009). Bernstein inequality and moderate deviations under strong mixing conditions. In *High dimensional probability V: the Luminy volume, pages 273–292. Institute of Mathematical Statistics.*

[2] Jiang, B., Li, J., and Yao, Q. (2020). Autoregressive networks. *arXiv preprint arXiv:2010.04492.*

Sparsity is an Issue in Network Parameter Estimation

The sparsity in networks is not like in the linear model, where we assume only a small number of all features are strong features that actually affect the response variable. Here, sparsity means the expected number of edges divided by the number of all possible edges $\rho_p = \frac{\sum_{i,j=1}^p E[X_{i,j}]}{p(p-1)/2}$ goes to zero as p goes to infinity.

Under our setting, $\rho_p = \frac{\sum_{i,j=1}^p E[X_{i,j}]}{p(p-1)/2} = \frac{2}{p(p-1)}\sum_{1\leq i<j\leq p} \frac{\alpha_{i,j}}{\alpha_{i,j}+\beta_{i,j}} = \frac{2}{p(p-1)}\sum_{1\leq i<j\leq p} \frac{1}{1+\beta_{i,j}/\alpha_{i,j}}$. It is clear that if $\alpha_{i,j}$ and $\beta_{i,j}$ are bounded away from 0 ($\liminf_{p\to\infty} \alpha_{i,j} > 0$, $\liminf_{p\to\infty} \beta_{i,j} > 0$), then the network is not sparse. Asymptotic results under non-sparse settings have been thoroughly investigated.

One way to understand why sparse is an issue is to treat it as signal processing:

$$X = E[X] + P$$

where $E[X] = (\frac{\alpha_{i,j}}{\alpha_{i,j}+\beta_{i,j}})_{p\times p}$ is the expected value of the adjacency matrix, while X is the realisation (of $p(p-1)/2$ number of Bernoulli distribution), while P satisfying $E[P] = 0$ is the noise (or error term).

Methods

We estimate α and β by conditional Maximum Likelihood Estimation.

$$\widehat{\alpha}_{i,j} = \frac{\sum_{t=1}^n X_{i,j}^t(1 - X_{i,j}^{t-1})}{\sum_{t=1}^n (1 - X_{i,j}^{t-1})}, \quad \widehat{\beta}_{i,j} = \frac{\sum_{t=1}^n (1 - X_{i,j}^t)X_{i,j}^{t-1}}{\sum_{t=1}^n X_{i,j}^{t-1}}, \quad \widehat{\pi}_{i,j} = \frac{\widehat{\alpha}_{i,j}}{\widehat{\alpha}_{i,j} + \widehat{\beta}_{i,j}}. \quad (2)$$

Next, by using $\widehat{\alpha}$, we aim to estimate θ. Directly listing all equations $\widehat{\alpha}_{i,j} = \widehat{\theta}_i\widehat{\theta}_j$ for $1\leq i < j \leq p$ does not necessarily yield solutions, since $\widehat{\alpha}_{i,j}$ are noise versions of $\alpha_{i,j}$, therefore the matrix $(\widehat{\alpha}_{i,j})_{p\times p}$ are very likely to not be in 1 dimension. (there are $p(p-1)/2$ number of equations, and only p number of variables)

We propose a method to solve for $\widehat{\theta}_i$: consider summation for p number of rows:

$$\sum_{j=1, j\neq i}^p \widehat{\theta}_i\widehat{\theta}_j = \sum_{j=1, j\neq i}^p \widehat{\alpha}_{i,j}, \quad \forall i = 1, \cdots, p.$$

Now there are p number of equations and p number of variables. Although it is in the quadratic form, we prove this is a convex problem, thus having a unique solution.

Given the estimation strategy, we could derive the probability bound for these estimators.

Lemma 1. *For $t\geq 1$, Define $Y_{i,j}^t = X_{i,j}^t(1 - X_{i,j}^t)$. Set $c_Y = \frac{1}{4}(\alpha_{i,j} + \beta_{i,j})$, then, for any $n\in\mathbb{N}$,*

$$\boldsymbol{\alpha}_{Y_{i,j}}(n) \leq \exp\{-2c_Y n\}. \quad (3)$$

Theorem 1. *Let $n\geq 4$. For any t such that*

$$0 < t \leq \frac{\alpha_{i,j}\beta_{i,j}}{8\left[\log_2\left(\frac{n}{\alpha_{i,j}+\beta_{i,j}}\right)\right]^2 (\alpha_{i,j} + \beta_{i,j})},$$

we have the non-asymptotic bound for the Moment Generating Function of $S_{(0,n]}$:

$$\log \mathbb{E}\exp\{tS_{(0,n]}\} \leq 15.5t^2v^2n + 1.4n\exp\left\{-\frac{\alpha_{i,j}+\beta_{i,j}}{24t}\right\} + \frac{8t^2n\alpha_{i,j}\beta_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^3} + \frac{1}{8}. \quad (4)$$

Furthermore, for any $\varepsilon_{n,p} > 0$ and $\varepsilon_{n,p} = o\left(\frac{(\alpha_{i,j}\beta_{i,j})^2}{(\alpha_{i,j}+\beta_{i,j})^4\left[\log_2\left(\frac{n}{\alpha_{i,j}+\beta_{i,j}}\right)\right]^2}\right)$, there exists a constant $C > 0$ only depends on a upper bound of α-mixing coefficient of $\{X_{i,j}^t\}_{t=0}^\infty$, such that the inequality below holds for all sufficiently large n.

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{t=1}^n (X_{i,j}^t - \pi_{i,j})\right| \geq \varepsilon_{n,p}\right) \leq 10\exp\left\{-\frac{Cn(\alpha_{i,j}+\beta_{i,j})^3\varepsilon_{n,p}^2}{\alpha_{i,j}\beta_{i,j}}\right\}. \quad (5)$$

(C1) As $n, p\to\infty$, it holds that $\frac{(\alpha_{i,j}+\beta_{i,j})^3}{(\alpha_{i,j}\beta_{i,j})^{3/2}}\left(\log\frac{n}{\alpha_{i,j}+\beta_{i,j}}\right)^2\sqrt{\frac{\log p}{n(\alpha_{i,j}+\beta_{i,j})}} \to 0$.

Corollary 1. *Let condition (C1) hold. For any $\kappa > 0$, and any p there exists a constant C_κ only depends on κ, such that for all sufficiently large n,*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{t=1}^n X_{i,j}^t - \pi_{i,j}\right| \geq C_\kappa\sqrt{\frac{\alpha_{i,j}\beta_{i,j}\log p}{n(\alpha_{i,j}+\beta_{i,j})^3}}\right) \leq p^{-\kappa}. \quad (6)$$