Optimal Estimation of Large Toeplitz Covariance Matrices

Harrison H. Zhou
Department of Statistics
Yale University

Joint work with T. Tony Cai and Zhao Ren

Outline

- **Introduction**
- **Motivation from Asymptotic Equivalence Theory**
- **Main Results**
- **Summary**

Introduction

Let $\mathbf{X}_1, \ldots, \mathbf{X}$ be i.i.d. p-variate Gaussian with an unkown Toeplitz covariance matrix $\Sigma_{\ \times\ }$,

$$\begin{pmatrix} \sigma_0 & \sigma_1 & \cdots & \sigma_{\ -2} & \sigma_{\ -1} \\ \sigma_1 & \sigma_0 & & & \sigma_{\ -2} \\ \vdots & & \ddots & & \vdots \\ \sigma_{\ -2} & & & \sigma_0 & \sigma_1 \\ \sigma_{\ -1} & \sigma_{\ -2} & \cdots & \sigma_1 & \sigma_0 \end{pmatrix}.$$

Goal: Estimate $\Sigma_{\ \times\ }$ based on the sample $\ \mathbf{X}\ : 1 \leq i \leq n\ $.

Introduction – Spectral Density Estimation

The model given by observing

$$\mathbf{X}_1 \quad N(0, \Sigma_{\times})$$

with $\Sigma_{\times}$ Toeplitz is commonly called

Spectral Density Estimation

$\mathbf{X}_1$, a stationary centered Gaussian sequence with spectral density f

where

$$f(t) = \frac{1}{2\pi} \sum_{=-\infty}^{\infty} \sigma \; \exp(imt) = \frac{1}{2\pi}[\sigma_0 + 2 \sum_{=1} \sigma \; \cos(mt)], \; t \quad [-\pi, \pi].$$

Here we have $\sigma_{-} = \sigma$.

Remark: there is a one-to-one correspondence between f and $\Sigma_{\infty\times\infty}$.

Introduction – Problem of Interest

We want to understand the minimax risk:

$$\inf_{\hat{\Sigma}} \sup_{\mathcal{F}} \mathbb{E} \quad \hat{\Sigma} - \Sigma \quad ^2$$

where $\cdot$ denotes the spectral norm and is some parameter space for f.

Motivation from Asymptotic Equivalence Theory

Golubev, Nussbaum and Z. (2010, AoS)

The **Spectral Density Estimation** given by observing each **X** is asymptotically equivalent to the **Gaussian white noise**

$$dy\ (t) = \log f(t)dt + 2\pi^{1\ 2} p^{-1\ 2} dW\ (t)\ ,\ t \quad [-\pi, \pi]$$

under some assumptions on the unknown f.

For example,

$$(M, \epsilon) = \quad f :\ f(t_1) - f(t_2) \ \leq M\ t_1 - t_2 \quad \text{and } f(t) \geq \epsilon \quad .$$

We need $\alpha > 1/2$ to establish the asymptotic equivalence.

Intuitively, the model

$$\mathbf{X} \sim N\left(0, \Sigma_{\times}\right), \; i = 1, 2, \ldots, n$$

is asymptotically equivalent to

$$dy\left(t\right) = \log f(t)dt + 2\pi^{1\ 2}\left(np\right)^{-1\ 2} dW\left(t\right), \; t \quad [-\pi, \pi]$$

possibly under some strong assumptions on the unknown f .

"Equivalent" Losses

Let $\hat{\Sigma}_{\infty\times\infty}$ be a Toeplitz matrix and $\hat{f}$ be the corresponding spectral density.

We know

$$\left\|\hat{\Sigma}_{\infty\times\infty} - \Sigma_{\infty\times\infty}\right\| = 2\pi \left\|\hat{f} - f\right\|_{\infty}$$

based on a well known result

$$\Sigma_{\infty\times\infty} \quad = 2\pi \quad f \quad _{\infty}$$

where

$$\Sigma_{\infty\times\infty} \quad = \sup_{\| \quad \|_2=1} \Sigma_{\infty\times\infty} v \quad _2, \text{ and } \quad f \quad _{\infty} = \sup \; f(x) \quad .$$

Intuitively

$$\left\|\hat{\Sigma} \quad _{\times} \quad - \Sigma \quad _{\times} \quad \right\| \quad \left\|\hat{\Sigma}_{\infty\times\infty} - \Sigma_{\infty\times\infty}\right\| ?$$

Thus optimal estimation on f may imply optimal estimation on Σ.

Question

Can we show

$$\inf_{\hat{\Sigma}_{p\times p}} \sup_{\mathcal{F}_\alpha} \mathbb{E} \left\| \hat{\Sigma}_{p\,p} - \Sigma_{p\,p} \right\|^2 \quad \left(\frac{np}{\log(pn)}\right)^{\frac{2\alpha}{2\alpha+1}} ?$$

Remark : Classical result on nonparametric function estimation under the sup norm:

$$\inf_{\hat{\Sigma}_{p\times p}} \sup_{\mathcal{F}_\alpha} \mathbb{E} \left\| \hat{f} - f \right\|_{\infty}^2 \quad \left(\frac{np}{\log(pn)}\right)^{\frac{2\alpha}{2\alpha+1}} .$$

Again,

- We don't really have the asymptotic equivalence.
- The following claim is very intuitive

$$\left\|\hat{\Sigma}_{\ \times\ } - \Sigma_{\ \times\ }\right\| \text{[illegible]} \left\|\hat{\Sigma}_{\infty\times\infty} - \Sigma_{\infty\times\infty}\right\|.$$

Main Results –Lower bound

Show that

$$\inf_{\hat{\Sigma}_{p\times p}} \sup_{\mathcal{F}_\alpha} \mathbb{E} \left\| \hat{\Sigma}_{\times} - \Sigma_{\times} \right\|^2 \geq c \left(\frac{np}{\log(pn)} \right)^{-\frac{2\alpha}{2\alpha+1}}$$

for some $c > 0$.

Main Results –Lower bound

A more informative model

Observe $\mathbf{Y}_1 = (\mathbf{X}_1, \mathbf{W}_1)$ with a circulant covariance matrix $\tilde{\Sigma}_{(2\ \ -1)\times(2\ \ -1)}$

$$\begin{pmatrix} \sigma_0 & \sigma_1 & \cdots & \sigma_{\ \ -2} & \sigma_{\ \ -1} & \sigma_{\ \ -2} & \cdots & \sigma_2 & \sigma_1 \\ \sigma_1 & \sigma_0 & & & \sigma_{\ \ -2} & \sigma_{\ \ -1} & & & \sigma_2 \\ \vdots & & \ddots & & \vdots & & \ddots & & \vdots \\ \sigma_{\ \ -2} & & & \sigma_0 & \sigma_1 & & & \sigma_{\ \ -1} & \sigma_{\ \ -2} \\ \sigma_{\ \ -1} & \sigma_{\ \ -2} & \cdots & \sigma_1 & \sigma_0 & \sigma_2 & \cdots & \sigma_{\ \ -2} & \sigma_{\ \ -1} \\ & & & & \cdots\cdots\cdots\cdots & & & & \end{pmatrix}.$$

Define

$$\omega \ = \frac{2\pi j}{2p-1}, \quad j \ \le p-1$$

and where

$$f \ (t) = \frac{1}{2\pi}\left(\sigma_0 + 2\sum_{=1}^{-1} \sigma \ \ \cos\left(mt\right)\right).$$

It is well known that the spectral decomposition of $\tilde{\Sigma}_{(2\ \ -1)\times(2\ \ -1)}$ can be described as follows:

$$\tilde{\Sigma}_{(2\ \ -1)\times(2\ \ -1)} = \sum_{|\ \ |\le\ \ -1} \lambda \ \mathbf{u} \ \mathbf{u}'$$

where

$$\lambda \ = f \ (\omega \ ), \quad j \ \le p-1$$

and the eigenvector $\mathbf{u}$ doesn't depend on $\sigma \ \ : 0 \le m \le p-1$.

Main Results –Lower bound

The more informative model is *exactly* equivalent to

$$Z \ = f \ (\omega \ ) \, \xi \ , \quad j \ \leqslant p - 1, Var\,(\xi \ ) \asymp 1/n.$$

For this model it is easy to show

$$\sup_{\mathcal{F}_\alpha} \mathbb{E} \left\| \hat{f} - f \right\|_\infty^2 \geq c \left(\frac{np}{\log{(pn)}} \right)^{-\frac{2\alpha}{2\alpha+1}}.$$

Main Results –Lower bound

We have

$$\left\|\hat{\Sigma}_{\ \times\ } - \Sigma_{\ \times\ }\right\| \geq \sup_{\in[-\ \ ]} \left|(\sigma_0 - \hat{\sigma}_0) + 2\sum_{=1}(1 - \frac{m}{p})\,(\hat{\sigma}\ \ - \sigma\ \ )\,e\ \ \right|$$

$$= \sup_{\in[-\ \ ]} \left|\hat{f}\,(t) - f\,(t)\right| + \text{negligible term}$$

based on a fact

$$\Sigma_{\ \times\ } \geqslant \sup_{\in[-\ \ ]} \frac{1}{p}\ \Sigma_{\ \times\ }\ v\ ,v\ = \sup_{\in[-\ \ ]} \left|\sigma_0 + 2\sum_{=1}(1 - \frac{m}{p})\sigma\ \ e\ \ \right|$$

where $v\ = (e\ ,e^{\ 2}, \cdots, e\ )$. Thus

$$\sup_{\mathcal{F}_\alpha} \mathbb{E} \left\|\hat{\Sigma}_{\ \times\ } - \Sigma_{\ \times\ }\right\|^2 \geq c \left(\frac{np}{\log{(pn)}}\right)^{-\frac{2\alpha}{2\alpha+1}}.$$

Remark: Need to have some assumptions on (n, p, α) such that the "negligible term" is truly negligible.

Main Results – Upper bound

Show that there is a $\hat{\Sigma}_{\times}$ such that

$$\sup_{\mathcal{F}_\alpha} \mathbb{E} \left\| \hat{\Sigma}_{\times} - \Sigma_{\times} \right\|^2 \leq C \left(\frac{np}{\log(pn)} \right)^{-\frac{2\alpha}{2\alpha+1}}$$

for some $C > 0$.

Main Results – Upper bound

Let $\Sigma\ = [\sigma\ 1_{\{\ \leqslant\ -1\}}]$ be a banding approximation of $\Sigma_{\ \times\ }$,, and $\tilde{\Sigma}$ be a banding approximation of the sample covariance matrix $\hat{\Sigma}_{\ \times\ }$. Note that $\mathbb{E}\tilde{\Sigma}\ = \Sigma$. Let $\hat{\Sigma}$ be a Toeplitz version of $\tilde{\Sigma}$ by taking the average of elements along the diagonal.

We have

$$\left\|\hat{\Sigma}\ - \Sigma\ \right\|^2 \le 2\left\|\hat{\Sigma}\ - \Sigma\ \right\|^2 + 2\ \ \Sigma\ - \Sigma\ \ ^2 \le 8\pi^2\left(\ \hat{f}\ - f\ \ ^2_\infty + \ f\ - f\ \ ^2_\infty\right)$$

since

$$\Sigma\ \ \leqslant 2\pi\ \ f\ \ _\infty = \sup_{[-\ \ ]} \sigma_0 + 2\sum_{=1}^{-1} \sigma\ \cos(mt)\ .$$

Main Results – Upper bound

Variance-bias trade-off

Variance part:

$$\mathbb{E} \quad \hat{f} - f \quad {}_{\infty}^{2} \leq C\frac{k}{np}\log(np)\,.$$

Bias part:

$$f \quad - f \quad {}_{\infty}^{2} \leq Ck^{-2} \quad .$$

Set the optimal $k : k \qquad \asymp \left(\frac{}{\log}\right)^{\frac{1}{2\alpha+1}}$ which gives

$$\sup_{\mathcal{F}_\alpha} \mathbb{E} \left\| \hat{\Sigma} \ \times \ - \Sigma \ \times \ \right\|^2 \leq C\left(\frac{np}{\log(pn)}\right)^{-\frac{2\alpha}{2\alpha+1}}$$

Remark: For simplicity we consider only the case $k \qquad \leq p$.

Main Result

Theorem. The minimax risk of estimating the covariance matrix $\Sigma_{\times}$ over the class satisfies

$$\inf_{\hat{\Sigma}_{p\times p}} \sup_{\mathcal{F}_\alpha} \mathbb{E} \left\| \hat{\Sigma}_{\times} - \Sigma_{\times} \right\|^2 \asymp \left(\frac{np}{\log(pn)} \right)^{-\frac{2\alpha}{2\alpha+1}} ?$$

under some assumptions on (n, p, α).

Remarks

- Full asymptotic equivalence?
- Sharp asymptotic minimaxity?

Summary

- We studied rate-optimality of Toeplitz matrices estimation.
- Le Cam's theory plays important roles.
- Full asymptotic equivalence and sharp asymptotics remain unknown.