Optimal Estimation of Large Toeplitz Covariance Matrices

Harrison H. Zhou Department of Statistics Yale University

Joint work with T. Tony Cai and Zhao Ren

<u>Outline</u>

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

Introduction

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



Introduction – Spectral Density Estimation

The model given by observing

 $\mathbf{X}_{1} \quad N\left(0, \Sigma_{\boldsymbol{p} \times \boldsymbol{p}}\right)$

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

Spectral Density Estimation

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

where

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

Here we have $\sigma_{-m} = \sigma_m$.

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:

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

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_i is asymptotically equivalent to the **Gaussian white noise**

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

under some assumptions on the unknown f.

For example,

 $(M, \epsilon) = |f: |f(t_1) - f(t_2)| \le M |t_1 - t_2|$ and $f(t) = \epsilon$.

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

Intuitively, the model

$$\mathbf{X}_{i} \quad N\left(0, \Sigma_{\boldsymbol{p} \times \boldsymbol{p}}\right), \ i = 1, 2, \dots, n$$

is asymptotically equivalent to

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

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

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

where

$$\left\|\Sigma_{\infty\times\infty}\right\| = \sup_{\|\mathbf{v}\|_{2}=1} \left\|\Sigma_{\infty\times\infty}v\right\|_{2}, \text{ and } \left\|f\right\|_{\infty} = \sup_{\mathbf{x}} \left|f\left(\mathbf{x}\right)\right|.$$

Intuitively

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

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



$$\inf_{\substack{\stackrel{}{}_{p\times p} \; \mathsf{F}_{\alpha}}} \sup_{\mathsf{F}_{\alpha}} \mathbb{E} \left\| \begin{array}{c} \mathsf{f}^{\mathsf{A}} & \mathsf{f} \end{array} \right\|_{1}^{2} \qquad \left(\frac{\mathsf{np}}{\mathsf{log}\,(\mathsf{pn})} \right)^{\frac{2\alpha}{2\alpha+1}}.$$

Again,

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

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



A more informative model

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

Define

$$\omega_j = \frac{2\pi j}{2p-1}, \ |j| \le p-1$$

and where

$$f_p(t) = \frac{1}{2\pi} \left(\sigma_0 + 2 \sum_{m=1}^{p-1} \sigma_m \cos\left(mt\right) \right).$$

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

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

where

$$\lambda_j = f_p(\omega_j), \ |j| \le p - 1$$

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

The more informative model is *exactly* equivalent to

$$Z_j = f_p(\omega_j) \xi_j, \ |j| \leq p-1, Var(\xi_j) \asymp 1/n.$$

For this model it is easy to show

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

We have

$$\begin{aligned} \left| \hat{\Sigma}_{p \times p} - \Sigma_{p \times p} \right\| & \sup_{t \in [-;]} \left| (\sigma_0 - \hat{\sigma}_0) + 2 \sum_{m=1}^p (1 - \frac{m}{p}) \left(\hat{\sigma}_m - \sigma_m \right) e^{imt} \right| \\ &= \sup_{t \in [-;]} \left| \hat{f}(t) - f(t) \right| + \text{negligible term} \end{aligned}$$

based on a fact

$$\begin{split} \|\Sigma_{p\times p}\| &\geqslant \sup_{t\in[-;]} \frac{1}{p} \ \Sigma_{p\times p} v_t, v_t \ = \sup_{t\in[-;]} \left| \sigma_0 + 2\sum_{m=1}^p (1-\frac{m}{p}) \sigma_m e^{imt} \right| \\ \text{where } v_t &= (e^{it}, e^{i2t}, \cdots, e^{ipt}). \text{ Thus} \\ &\sup_{\mathcal{F}_{\alpha}} \mathbb{E} \left\| \hat{\Sigma}_{p\times p} - \Sigma_{p\times p} \right\|^2 \quad c \left(\frac{np}{\log(pn)} \right)^{-\frac{2\alpha}{2\alpha+1}}. \end{split}$$

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 $\hat{\Sigma}_{p \times p}$ such that

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

for some C > 0.

Main Results – Upper bound

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

We have

$$\left\|\hat{\Sigma}_{k} - \Sigma_{p}\right\|^{2} \leq 2\left\|\hat{\Sigma}_{k} - \Sigma_{k}\right\|^{2} + 2\left\|\Sigma_{k} - \Sigma_{p}\right\|^{2} \leq 8\pi^{2}\left(\left\|\hat{f}_{k} - f_{k}\right\|_{\infty}^{2} + \left\|f_{k} - f_{p}\right\|_{\infty}^{2}\right)\right\|^{2}\right\|$$

since

$$\|\Sigma_{k}\| \leq 2\pi \| f_{k} \|_{\infty} = \sup_{[-;]} |\sigma_{0} + 2\sum_{m=1}^{k-1} \sigma_{m} \cos(mt)|.$$

Main Results – Upper bound

Variance-bias trade-off

Variance part:

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

Bias part:

$$\parallel f_k - f_p \parallel_{\infty}^2 \leq Ck^{-2} \ .$$

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

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

Remark: For simplicity we consider only the case $k_{optimal} \leq p$.

Main Result

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

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

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.