STAT665: ADVANCED TIME SERIES ANALYSIS `16:960:665:01`

FALL 2021, TUESDAY/THURSDAY 5:00–6:20PM

Course Information

- Instructor: Han Xiao
- Office: Hill Center 451
- Office Hours: Thursday 1:30-2:30 on Zoom
  `https://rutgers.zoom.us/j/98845998357?pwd=dzdNYWlqMmVqd2laVXh5UkJMUWhpQT09`
- Email: `hxiao``atsign''stat.rutgers.edu` (**I only check this email account regularly!**)
- Texts.
  - *Time Series: Theory and Methods*, by Peter J. Brockwell and Richard A. Davis. Springer, 1991, 2ed. (TSTM)
  - *Asymptotic Theory of Weakly Dependent Random Processes*, by Emmanuel Rio. Springer, 2017.
- Course work: (almost) weekly homework.

Outline

1. Foundations. (4 weeks)
   - Stochastic processes, stationarity, autocovariance fuctions.
   - Spectral representation, spectral decomposition.
   - Linear prediction.
2. ARMA Models. (4 weeks)
   - ARMA processes, linear prediction, estimation.
   - Asymptotic theory for the MLE.
3. Ergodic theorem and CLT for stationary processes. (4 weeks)
   - Strict stationarity, measure preserving transformation, ergodic theorem.
   - CLT for strictly stationary processes, mixing conditions, dependence measures.
4. Analysis of complex time series data. (2 weeks)
   - VAR models.
   - Dynamic factor models.
   - Matrix and tensor-valued time series.

Last updated on May 30, 2022.

1. Reading Assignments

Week 01. Chapter 1, 2 of TSTM.

Week 02. Section 3.1∼3.4 of TSTM, please DO read Section 3.3 by yourself, since I did not cover enough details in the lecture.

Week 03. Section 3.4, 4.1, 4.2, 4.3 of TSTM. Please also read Section 2.1 and 2.2 of this document.

Week 04. Section 4.3, 4.4 of TSTM.

Week 05. Section 5.1∼5.5, 5.7, 5.8 of TSTM.

Week 06. Section 5.7, 5.8, 8.1, 8.2, 8.10 of TSTM.

Week 07. Section 8.1, 8.2, 8.6∼8.11 of TSTM.

Week 08. Section 8.8, 10.1, 10.8 of TSTM.

Week 09. Section 10.8 of TSTM.

Week 10. Section 6.3, 10.8 of TSTM.

Week 11. Section 24 and 36 of *Probability and Measure* by Patrick Billingsley.

Week 12. Various places of Rio (2017), Wu (2005).

Week 13. Lam et al. (2011), Chen et al. (2021).

References

Chen, R., Yang, D., and Zhang, C.-H. (2021). Factor models for high-dimensional tensor time series. *Journal of the American Statistical Association*, pages 1–23.

Lam, C., Yao, Q., and Bathia, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika*, 98(4):901–918.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences*, 102(40):14150–14154.

2. Some Additional Notes

2.1. **On the proof of the Herglotz Theorem (Theorem 4.3.1).** The proof relies on the Helly's Theorem, and here are the precise statements. For the proofs, see Section 25 of Patrick Billingsley's *Probability and Measure*.

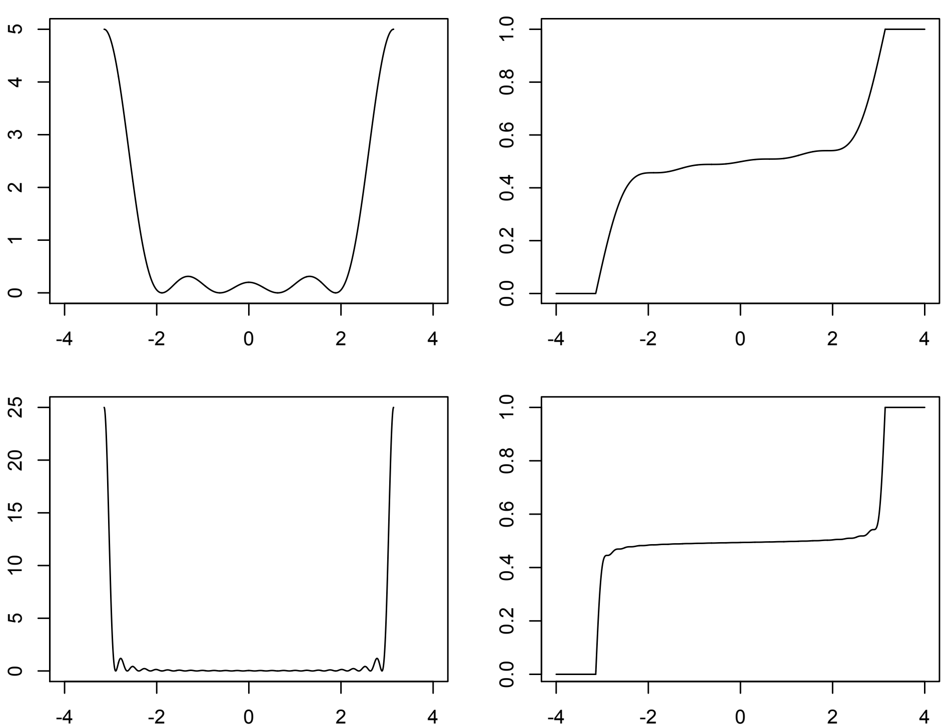
First recall the concept of *tightness*. A sequence of distribution functions $\{F_n\}$ on $\mathbb{R}$ is said to be *tight* if for any $\epsilon > 0$, there exist x and y such that $\sup_n F_n(x) \leq \epsilon$ and $\sup_n[1-F_n(y)] \leq \epsilon$. *Tightness* for a sequence of random variables $\{X_n\}$, denoted by the commonly seen notation $X_n = O_p(1)$, is defined through the corresponding distribution functions.

Theorem 1 (Helly's Theorem)**.**
A. For every sequence $\{F_n\}$ of distribution functions there exists a subsequence $\{F_{n_k}\}$ and a nonnegative, nondecreasing, right-continuous function F such that $\lim_k F_{n_k}(x) = F(x)$ at all continuity points x of F.
B. Tightness is a necessary and sufficient condition that for every subsequence $\{F_{n_k}\}$ there exists a further subsequence $\{F_{n_{k(j)}}\}$ and a distribution function F such that $F_{n_{k(j)}}$ converges to F in distribution as $j \to \infty$.

Note that A does not guarantee that F is a distribution function, while B does.

The proof of Theorem 4.3.1 seems an immediate application of the Helly's Theorem, but there is actually some subtlety, which we will clarify in this remark. First of all, let us emphasize that the spectral distribution is only supported on the interval $(-\pi, \pi]$. While there can be a point mass at π, the openness of the left end precludes such a possibility at $-\pi$.

We start with an example. Consider the autocovariance function $\gamma(h) = (-1)^h$ of the stochastic process $X_t = Xe^{i\pi t}$, where X has mean zero and variance 1. The f_N and F_N constructed in the proof are plotted below for $N = 5$, and 25. You can imagine that the limit of F_N is a distribution function $\tilde{F}$ with two jumps at $\pm\pi$, of the same size $\frac{1}{2}$. (What's the problem? The spectral distribution does not allow a point mass at$= -\pi$!)



Now let's return to the proof of Theorem 4.3.1, i.e. the following arguments will cover not only the preceding example, but all non-negative definite functions $\gamma(\cdot)$. Let $\tilde{F}$ be the limit of the subsequence $\{F_{N_k}\}$, and keep in mind that $\tilde{F}$ may have two point masses of the same size at $\pm\pi$. Any continuous function $g(\cdot)$ on the interval $[-\pi,\pi]$ such that $g(\pi)=g(-\pi)$ can be extended as a periodic function on $\mathbb{R}$. Since $g(\cdot)$ as a periodic function on $\mathbb{R}$ is bounded and continuous, it holds that

$$\lim_{N_k}\int g\,dF_{N_k}=\int g\,d\tilde{F}.$$

Obviously the integral on the left hand side equals to $\int_{-\pi}^{\pi} g(x)f_{N_k}(x)\,dx=\int_{(-\pi,\pi]} g\,dF_{N_k}$. Since $\tilde{F}$ is supported on $[-\pi,\pi]$, the integral on the right hand side equals to

$$\int_{[-\pi,\pi]} g\,d\tilde{F}=g(-\pi)*\tilde{F}(\{-\pi\})+g(\pi)*\tilde{F}(\{\pi\})+\int_{(-\pi,\pi)} g\,d\tilde{F}. \tag{1}$$

Now define a new distribution function F, supported on $(-\pi,\pi]$, by $F(x)=\tilde{F}(x)-\tilde{F}(-\pi)$ for $-\pi<x<\pi$, and $F(\pi)=\tilde{F}(\pi)$. Intuitively speaking, F is obtained from $\tilde{F}$ by moving its point mass on $-\pi$ to π. Therefore,

$$\int_{(-\pi,\pi]} g\,dF=g(\pi)*\Big[\tilde{F}(\{\pi\})+\tilde{F}(\{-\pi\})\Big]+\int_{(-\pi,\pi)} g\,d\tilde{F}. \tag{2}$$

Since $g(\pi)=g(-\pi)$, the integrals in (1) and (2) are equal, and consequently

$$\lim_{N_k}\int_{-\pi}^{\pi} g(x)f_{N_k}(x)\,dx=\int_{(-\pi,\pi]} g\,dF.$$

2.2. **On the uniqueness of the spectral distribution.** To show that the spectral distribution is uniquely determined by the autocovariance function $\gamma(\cdot)$, the arguments at the bottom of Page 119 involves an application of Theorem 2.11.1, in order to show that

$$\int_{(-\pi,\pi]}\phi(\nu)\,dF(\nu)=\int_{(-\pi,\pi]}\phi(\nu)\,dG(\nu)\quad\text{if }\phi\text{ is continuous with }\phi(-\pi)=\phi(\pi). \tag{3}$$

Showing that (3) implies F and G are the same is a good exercise for measure theory. Please try it if you have not done anything similar before (Hint: approximate an indicator function by continuous functions, and then apply the dominated convergence theorem). In this remark we provide more details on how Theorem 2.11.1 leads to (3).

Adopt the notations of Section 2.11 of the Fourier approximations. Recall that $\phi(\cdot)$ is a continuous function on $[-\pi,\pi]$ such that $\phi(\pi)=\phi(-\pi)$. If $\gamma(h)=\int_{(-\pi,\pi]} e^{ih\nu}\,dF(\nu)=\int_{(-\pi,\pi]} e^{ih\nu}\,dG(\nu)$ for all $h\in\mathbb{Z}$, then for any $j\in\mathbb{Z}$

$$\int_{(-\pi,\pi]} S_n\phi(\nu)\,dF(\nu)=\int_{(-\pi,\pi]} S_n\phi(\nu)\,dG(\nu)=\sum_{|j|\le n}\langle f,e_j\rangle\gamma(j).$$

Let $C_n\phi:=n^{-1}(S_0\phi+S_1\phi+\cdots+S_{n-1}\phi)$, it follows that

$$\int_{(-\pi,\pi]} C_n\phi(\nu)\,dF(\nu)=\int_{(-\pi,\pi]} C_n\phi(\nu)\,dG(\nu).$$

By Theorem 2.11.1, $C_n\phi\to\phi$ uniformly on $[-\pi,\pi]$. So there exists a constant $K>0$ such that $|C_n\phi(\nu)|<K$ for all $n\ge 1$ and $\nu\in[-\pi,\pi]$. By the dominated convergence theorem,

$$\int_{(-\pi,\pi]}\phi(\nu)\,dF(\nu)=\lim_n\int_{(-\pi,\pi]} C_n\phi(\nu)\,dF(\nu)=\lim_n\int_{(-\pi,\pi]} C_n\phi(\nu)\,dG(\nu)=\int_{(-\pi,\pi]}\phi(\nu)\,dG(\nu).$$

2.3. **Statement and Proof of Kolmogorov's Formula.** First of all, it is equivalent and more convenient to view a spectral distribution as a distribution over the unit circle $\mathbb{T}$ on the complex plane, indexed by $e^{i\lambda}$, with $\lambda \in (-\pi, \pi]$, so a density $f(\lambda)$ over $(\pi, \pi]$ is also equivalently written as $f(e^{i\lambda})$. Denote $\mathbb{U}$ the open disc with radius 1 on the complex plane. Suppose F is a spectral distribution, and $\{X_t,\, t \in \mathbb{Z}\}$ is a stationary process which has F as its spectral distribution. Define $\mathscr{M}_t = \overline{\mathrm{sp}}\{X_k,\, k \le t\}$, and

$$\sigma^2 = \|X_{t+1} - \mathcal{P}_{\mathscr{M}_t} X_{t+1}\|^2.$$

Theorem 2 (Kolmogorov's Formula)**.** *Let F be a spectral distribution, and f be its derivative (which is defined almost everywhere). Then $\sigma^2 > 0$ if and only if $\int_{-\pi}^{\pi} \log f(\lambda)\, d\lambda > -\infty$. Furthermore, the following identity always holds (even when $\sigma^2 = 0$).*

$$\sigma^2 = 2\pi \exp\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\lambda)\, d\lambda \right\}.$$

The proof uses the theory of H^p spaces, and involves the concepts of Poisson integral, radial limit, outer function, inner function, Blaschke product etc, all of which can be found in Walter Rudin's *Real and Complex Analysis* (McGraw-Hill, 3ed, 1986). All the theorems cited in the proof refer to the same book.

Proof. Suppose $\{X_t,\, t \in \mathbb{Z}\}$ is a stationary process which has F as its spectral distribution. If $\sigma^2 > 0$, let

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t =: U_t + V_t$$

be the Wold Decomposition of $\{X_t\}$, where $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$. The function $\psi(z) := \sum_{j=0}^{\infty} \psi_j z^j \in H^2$ (Theorem 17.12). Let ψ^* be the radial limit of $\psi(z)$, then $\log|\psi^*| \in L^1(\mathbb{T})$ (Theorem 17.17). Define the outer function

$$Q_\psi(z) = \exp\left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{i\lambda} + z}{e^{i\lambda} - z} \log|\psi^*(e^{i\lambda})|\, d\lambda \right].$$

It holds that $Q_\psi \in H^2$ (Theorem 17.16), and $|Q_\psi^*(e^{i\lambda})| = |\psi^*(e^{i\lambda})|$ a.e. on $\mathbb{T}$. According to Theorem 17.17, there is an inner function M_ψ such that $\psi = M_\psi Q_\psi$. We shall prove that M_f is a constant inner function, which implies that $M_f = 1$, and

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \log|\psi^*(e^{i\lambda})|\, d\lambda = 0. \tag{4}$$

Let's do proof by contradiction. If M_f is not a constant, then by Theorem 17.17,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \log|\psi^*(e^{i\lambda})|\, d\lambda > 0.$$

Let $Q_\psi(z) = \sum_{n=0}^{\infty} a_n z^n$ be the power series representation of Q_ψ on $\mathbb{U}$, then $a_0 = Q_\psi(0) > 1$. Note that $Q_\psi \in H^2$ implies that $\sum_{n=0}^{\infty} |a_n|^2 < \infty$. Consider the process

$$Y_t = \sum_{j=0}^{\infty} a_j Z_{t-j} + V_t. \tag{5}$$

By construction $\{Y_t\}$ and $\{X_t\}$ have the same spectral distribution, since $|Q_\psi^*(e^{i\lambda})| = |\psi^*(e^{i\lambda})|$ a.e. However, the representation (5) (this is actually a Wold Decomposition, but we don't need this fact) implies that

$$a_0^2 \sigma^2 \le \|Y_{t+1} - \mathcal{P}_{\overline{\mathrm{sp}}\{Y_k,\, k \le t\}} Y_{t+1}\| = \|X_{t+1} - \mathcal{P}_{\mathcal{M}_t} X_{t+1}\|^2 = \sigma^2,$$

which is impossible!

It follows from (4) that

$$\sigma^2 = 2\pi \exp\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log\left[|\psi^*(e^{i\lambda})|^2 \cdot \frac{\sigma^2}{2\pi} \right] d\lambda \right\}.$$

From here, the proof proceeds as (i) prove that F_U and F_V are singular with each other relative to F (see Doob (1953) or Kolmogorov's seminal paper *Stationary Sequences in Hilbert Space*), and (ii) since the spectral density density of F_U is positive a.e., it follows that F_V must be singular to the Lebesgue measure. Therefore, $F = F_U + F_V$ is precisely the Lebesgue decomposition of F, and hence $f = \psi^*$ a.e. and the proof of the "$\Rightarrow$" direction is complete.

Now we prove the other direction. Let $dF = f + d\nu$ be the Lebesgue decomposition of F. If $\log f \in L^1(\mathbb{T})$, define the function

$$Q_f(z) = \exp\left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{i\lambda} + z}{e^{i\lambda} - z} \log \sqrt{f(e^{i\lambda})}\, d\lambda \right].$$

By Theorem 17.16, $Q_f \in H^2$ and $Q_f^*(e^{i\lambda}) = \sqrt{f(e^{i\lambda})}$ a.e. Let $Q_f = \sum_{n=0}^{\infty} b_n z^n$ be the power series representation of Q_f over $\mathbb{U}$, it holds that $b_0 = Q_f(0) > 0$. Suppose $\{G_t,\, t \in \mathbb{Z}\} \sim \mathrm{WN}(0, 2\pi)$, and $\{H_t,\, t \in \mathbb{Z}\}$ has ν as its spectral distribution, and $\{G_t,\, t \in \mathbb{Z}\}$ and $\{H_t,\, t \in \mathbb{Z}\}$ are uncorrelated with each other. Define

$$W_t = \sum_{j=0}^{\infty} b_j G_{t-j} + H_t.$$

It is straightforward to verify that $\{W_t,\, t \in \mathbb{Z}\}$ has the same spectral distribution as $\{X_t\}$. But

$$\|W_{t+1} - \mathcal{P}_{\overline{\mathrm{sp}}\{W_k,\, k \le t\}} W_{t+1}\|^2 \ge 2\pi b_0^2 > 0,$$

which implies $\sigma^2 > 0$, and the proof is complete. □

2.4. **Martingale central limit theorems.** I am copying some martingale central limit theorems from Hall and Heyde's classic *Martingale Limit Theory and Its Application.* Although the versions presented here are simplified (see the book for the more general results), they are already very powerful.

Theorem 3 (Theorem 3.2.)**.** *Let $\{k_n\}$ be a non-decreasing sequence of positive integers which tends to infinity. Let $\{S_{ni}, \mathcal{F}_{ni},\, 1 \le i \le k_n,\, n \ge 1\}$ be a zero mean, square-integrable martingale array with differences X_{ni}, and let $\eta^2 > 0$ be a positive constant. Suppose that*

$$\max_{1 \le i \le k_n} |X_{ni}| \xrightarrow{p} 0, \tag{6}$$

$$\sum_{i=1}^{k_n} X_{ni}^2 \xrightarrow{p} \eta^2, \tag{7}$$

$$\mathbb{E}(\max_{1 \le i \le k_n} X_{ni}^2) \;\; \textit{is bounded in } n, \tag{8}$$

$$\mathcal{F}_{n,i} \subset \mathcal{F}_{n+1,i} \quad \textit{for } 1 \le i \le k_n, \;\; n \ge 1. \tag{9}$$

Then $S_n := \sum_{i=1}^{k_n} X_{ni}$ converges in distribution to $N(0, \eta^2)$.

Theorem 4 (Corollary 3.1.)**.** *If (6) and (8) are replaced by the conditional Lindeberg condition:*

$$\textit{for all } \epsilon > 0, \quad \sum_{i=1}^{k_n} \mathbb{E}\left[X_{ni}^2 I(|X_{ni}| > \epsilon) | \mathcal{F}_{n,i-1} \right] \xrightarrow{p} 0,$$

and if (7) *is replaced by an analogous condition on the conditional variance:*

$$V_n^2 := \sum_{i=1}^{k_n} \mathbb{E}(X_{ni}^2 | \mathcal{F}_{n,i-1}) \xrightarrow{p} \eta^2,$$

and if (9) *holds, then* S_n *converges in distribution to* $N(0, \eta^2)$.

Theorem 5 (Theorem 5.2.). *Let* $\{X_n, n \in \mathbb{Z}\}$ *be a strictly stationary and ergodic sequence of real random variables. Let*

$$S_n = \sum_{i=1}^{n} (X_i - \mathbb{E}X_i), \qquad \text{and } \mathcal{F}_0 = \sigma\{X_k,\, k \le 0\}.$$

If

$$\sum_{k=1}^{\infty} \mathrm{Cov}[\mathbb{E}(X_n|\mathcal{F}_0), X_k] \text{ converges for every } n \ge 0,$$

and

$$\lim_{n\to\infty} \sup_{K \ge 1} \left| \sum_{k=K}^{\infty} \mathrm{Cov}[\mathbb{E}(X_n|\mathcal{F}_0), X_k] \right| = 0,$$

then $n^{-1}\mathrm{Var}(S_n)$ *converges to* σ^2 *with* $0 \le \sigma^2 < \infty$, *and* $n^{-1/2}S_n$ *converges in distribution to* $N(0, \sigma^2)$.

2.5. **A few lemmas for the proof of Theorem 10.8.1.** Throughout this subsection, assume $\{X_t\}$ is a strictly stationary mean zero process with absolutely summable autocovariances $\gamma(n)$, so that its spectral density $f(\lambda) = (2\pi)^{-1} \sum_{n\in\mathbb{Z}} \gamma(n) e^{-in\lambda}$ is a continuous function. Define the event A as the set on which

$$\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n) \to 0, \quad \text{and} \quad \tilde{\gamma}(k) = \frac{1}{n} \sum_{t=k+1}^{n} X_t X_{t-k} \to \gamma(k),\ \forall\, k \ge 0.$$

Note that if $\{X_t\}$ is a causal and invertible ARMA(p, q) process $\phi_0(B)X_t = \theta_0(B)Z_t$ with $\{Z_t\} \sim$ IID$(0, \sigma^2)$, then $P[A] = 1$.

For any vector $\boldsymbol{\varphi} = (\varphi_0, \varphi_1, \ldots, \varphi_m)' \in \mathbb{R}^{m+1}$, define $\varphi(z) = \varphi_0 + \varphi_1 z + \cdots + \varphi_m z^m$. Note that $\varphi(z)$ differs from $\phi(z)$ and $\theta(z)$ in that it can have a constant term different from 1.

Lemma 6. *On the event* A, *it holds that*

(i) *For any* $\boldsymbol{\varphi} = (\varphi_0, \varphi_1, \ldots, \varphi_m)'$,

$$\frac{1}{n} \sum_j I_n(\omega_j) |\varphi(e^{-i\omega_j})|^2 \to \int_{-\pi}^{\pi} f(\lambda) |\varphi(e^{-i\lambda})|^2 \, d\lambda.$$

(ii) *For any sequence* $\{\boldsymbol{\beta}_n\} \in C_{p,q}$ *such that* $\boldsymbol{\beta}_n \to \boldsymbol{\beta} \in C_{p,q}$,

$$\frac{1}{n} \sum_j \frac{I_n(\omega_j)}{g(\omega_j; \boldsymbol{\beta}_n)} \to \int_{-\pi}^{\pi} \frac{f(\lambda)}{g(\lambda; \boldsymbol{\beta})} \, d\lambda.$$

(iii) *For any sequence* $\{\boldsymbol{\beta}_n\} \in C_{p,q}$ *such that* $\boldsymbol{\beta}_n \to \boldsymbol{\beta} \in \partial C_{p,q}$,

$$\liminf_{n\to\infty} \frac{1}{n} \sum_j \frac{I_n(\omega_j)}{g(\omega_j; \boldsymbol{\beta}_n)} \ge \int_{-\pi}^{\pi} \frac{f(\lambda)}{g(\lambda; \boldsymbol{\beta})} \, d\lambda.$$

Proof of (iii). First of all, the discussion is always on the event A. According to Corollary 4.4.2, for a fixed $\delta > 0$, there exists a φ such that

$$|\theta(e^{-i\lambda})|^2 + \delta \le |\varphi(e^{-i\lambda})|^{-2} \le |\theta(e^{-i\lambda})|^2 + 2\delta.$$

Let $(\phi_n\varphi)(e^{-i\lambda})|^2 = \sum_{k=-m}^{m} b_{n,k}e^{-ik\lambda}$ and $(\phi\varphi)(e^{-i\lambda})|^2 = \sum_{k=-(m)}^{m} b_k e^{-ik\lambda}$. Note that $b_{n,k} \to b_k$ for each $|k| \le m$ (as $\boldsymbol{\phi}_n \to \boldsymbol{\phi}$). Since

$$\frac{1}{n}\sum_j \left[I_n(\omega_j)|(\phi_n\varphi)(e^{-i\omega_j})|^2\right] = \sum_{k=-(m)}^{m} b_{n,k}\tilde{\gamma}(k) + 2\sum_{k=1}^{m} b_{n,k}\tilde{\gamma}(n-k),$$

it holds that (on the event A)

$$\frac{1}{n}\sum_j \left[I_n(\omega_j)|(\phi_n\varphi)(e^{-i\omega_j})|^2\right] \to \sum_{k=-(m)}^{m} b_k\gamma(k) = \int_{-\pi}^{\pi} f(\lambda)|(\phi\varphi)(e^{-i\lambda})|^2\,d\lambda.$$

Therefore,

$$\begin{aligned}
\liminf_{n\to\infty}\frac{1}{n}\sum_j \frac{I_n(\omega_j)}{g(\omega_j;\boldsymbol{\beta}_n)} &\ge \liminf_{n\to\infty}\frac{1}{n}\sum_j \left[I_n(\omega_j)|(\phi_n\varphi)(e^{-i\omega_j})|^2\right] \\
&= \int_{-\pi}^{\pi} f(\lambda)|(\phi\varphi)(e^{-i\lambda})|^2\,d\lambda \\
&\ge \int_{-\pi}^{\pi} f(\lambda)\frac{|\phi(e^{-i\lambda})|^2}{|\theta(e^{-i\lambda})|^2 + 2\delta}\,d\lambda,
\end{aligned}$$

where the first inequality is due to $\boldsymbol{\theta}_n \to \boldsymbol{\theta}$ and hence $|\theta_n(e^{-i\lambda})|^2 \le |\varphi(e^{-i\lambda})|^{-2}$ for all λ when n is large enough. Letting $\delta \downarrow 0$ and applying the MCT leads to the conclusion of (iii). $\square$

For any rational function $h(z) = h_1(z)/h_2(z)$ such that $h_1(z)h_2(z) \neq 0$ on $|z| \le 1$, let $G_n(h)$ be the $n \times n$ autocovariance matrix corresponding to the spectral density $(2\pi)^{-1}|h(e^{-i\lambda})|^2$. Following this definition, the matrix $G_n(\boldsymbol{\beta})$ with $\boldsymbol{\beta} \in C_{p,q}$ can also be expressed as $G_n(\boldsymbol{\beta}) = G_n(\theta/\phi)$.

Lemma 7. *On the event A, it holds that*

(i) For any $\varphi(z) := \varphi_0 + \varphi_1 z + \cdots + \varphi_m z^m$ with real coefficients,

$$\frac{1}{n}\boldsymbol{X}_n' G_n(\varphi)\boldsymbol{X}_n \to \int_{-\pi}^{\pi} f(\lambda)|\varphi(e^{-i\lambda})|^2\,d\lambda.$$

(ii) For any sequence $\{\boldsymbol{\beta}_n\} \in C_{p,q}$ such that $\boldsymbol{\beta}_n \to \boldsymbol{\beta} \in C_{p,q}$,

$$\frac{1}{n}\boldsymbol{X}_n' G_n^{-1}(\boldsymbol{\beta}_n)\boldsymbol{X}_n \to \int_{-\pi}^{\pi} \frac{f(\lambda)}{g(\lambda;\boldsymbol{\beta})}\,d\lambda.$$

(iii) For any sequence $\{\boldsymbol{\beta}_n\} \in C_{p,q}$ such that $\boldsymbol{\beta}_n \to \boldsymbol{\beta} \in \partial C_{p,q}$,

$$\liminf_{n\to\infty}\frac{1}{n}\boldsymbol{X}_n' G_n^{-1}(\boldsymbol{\beta}_n)\boldsymbol{X}_n \ge \int_{-\pi}^{\pi} \frac{f(\lambda)}{g(\lambda;\boldsymbol{\beta})}\,d\lambda.$$

Proof of (iii), sketch. Similar to the proof of Lemma 6 (iii), for each $\delta > 0$, find the φ such that

$$|\theta(e^{-i\lambda})|^2 + \delta \le |\varphi(e^{-i\lambda})|^{-2} \le |\theta(e^{-i\lambda})|^2 + 2\delta.$$

Note that $\varphi(z) \neq 0$ on $|z| \le 1$. Let $\alpha_n(z) := \varphi_0^{-1}(\phi_n\varphi)(z) = \sum_{k=0}^{m} a_{n,k}z^k$ and $\alpha(z) := \varphi_0^{-1}(\phi\varphi)(z) = \sum_{k=0}^{m} a_n z^k$. Note that $a_{n,0} = a_0 = 1$. Recall that $G_n(1/\alpha_n)$ is the $n \times n$ autoconvariance matrix corresponding to the spectral density $(2\pi)^{-1}|\alpha_n(e^{-i\lambda})|^{-2}$. Since $\theta_n \to \theta$, when n is large enough, $|\theta_n(e^{-i\lambda})|^2 \le |\varphi(e^{-i\lambda})|^{-2}$ for all λ, which implies that

$$\boldsymbol{X}_n' G_n^{-1}(\boldsymbol{\beta}_n)\boldsymbol{X}_n \ge \varphi_0^2 \cdot \boldsymbol{X}_n' G_n^{-1}(1/\alpha_n)\boldsymbol{X}_n.$$

Let $\{Y_{nt}\}$ be the autoregressive process $\alpha_n(B)Y_{nt} = W_t$, where $\{W_t\} \sim \mathrm{WN}(0,1)$. Now perform the Gram-Schmidt procedure:

$$\begin{aligned}
W_{n1} &= \delta_{n11}Y_{n1} \\
W_{n2} &= \delta_{n21}Y_{n1} + \delta_{n22}Y_{n2} \\
&\cdots \\
W_{nm} &= \delta_{nm1}Y_{n1} + \cdots + \delta_{nmm}Y_{nm} \\
W_{n,m+1} &= \alpha_n(B)Y_{n,m+1} \\
&\cdots \\
W_{nn} &= \alpha_n(B)Y_{nn},
\end{aligned} \tag{10}$$

such that $\{W_{n1}, \ldots, W_{nn}\}$ are uncorrelated with variance 1. Denote the matrix on the RHS of (10) by $\boldsymbol{T}_n$, and note that $G_n^{-1}(1/\alpha_n) = \boldsymbol{T}_n'\boldsymbol{T}_n$. Note that $\boldsymbol{T}_n'\boldsymbol{T}_n$ is the same as the matrix $G(\alpha_n)$ except for the upper-left and bottom-right $m \times m$ blocks. It follows that

$$\begin{aligned}
&\boldsymbol{X}_n' G_n^{-1}(1/\alpha_n)\boldsymbol{X}_n/n - \boldsymbol{X}_n' G(\alpha_n)\boldsymbol{X}_n/n \\
=&\frac{1}{n}\sum_{1\le j,k\le m} X_j X_k \left\{(\boldsymbol{T}_n'\boldsymbol{T}_n)[j,k] - G(\alpha_n)[j,k]\right\} \\
&+ \frac{1}{n}\sum_{0\le j,k\le m-1} X_{n-j}X_{n-k}\left\{(\boldsymbol{T}_n'\boldsymbol{T}_n)[n-j,n-k] - G(\alpha_n)[n-j,n-k]\right\},
\end{aligned}$$

which converges to zero on the event A. Here we also use the fact that $\sup_n \max_{0\le k\le m} |a_{nk}| < \infty$ and (HW Problem)

$$\sup_n \max_{1\le k\le j\le m} |\delta_{njk}| < \infty.$$

Let $|\alpha_n(e^{-i\lambda})|^2 = \sum_{k=-m}^m b_{n,k}e^{-ik\lambda}$ and $|\alpha(e^{-i\lambda})|^2 = \sum_{k=-(m)}^m b_k e^{-ik\lambda}$. Note that $b_{n,k} \to b_k$ for each $|k| \le m$. Similar to (i), it holds that (when n is large enough)

$$\boldsymbol{X}_n' G(\alpha_n)\boldsymbol{X}_n/n = \int_{-\pi}^{\pi} \frac{1}{2\pi}|\alpha_n(e^{-i\lambda})|^2 \sum_{|k|<n}\tilde{\gamma}(k)e^{-ik\lambda}\,d\lambda = \sum_{k=-m}^{m} b_{n,k}\tilde{\gamma}(k),$$

which on event A converges to

$$\sum_{k=-m}^{m} b_k\gamma(k) = \int_{-\pi}^{\pi} f(\lambda)|\alpha(e^{i\lambda})|^{-2}\,d\lambda.$$

To summarize, we have shown that on event A,

$$\begin{aligned}
\liminf_{n\to\infty} \boldsymbol{X}_n' G^{-1}(\boldsymbol{\beta}_n)\boldsymbol{X}_n/n &\ge \lim_{n\to\infty} \varphi_0^2 \cdot \boldsymbol{X}_n' G_n^{-1}(1/\alpha_n)\boldsymbol{X}_n/n = \varphi_0^2 \int_{-\pi}^{\pi} f(\lambda)|\alpha(e^{i\lambda})|^2\,d\lambda \\
&= \int_{-\pi}^{\pi} f(\lambda)|(\varphi\phi)(e^{i\lambda})|^2\,d\lambda \ge \int_{-\pi}^{\pi} f(\lambda)\frac{|\phi(e^{-i\lambda})|^2}{|\theta(e^{i\lambda})|^2 + 2\delta}\,d\lambda.
\end{aligned}$$

The proof is completed by letting $\delta \downarrow 0$ and applying the MCT. □

2.6. **On the Proof of Theorem 10.8.2.** The relationship between $\partial\bar{\sigma}^2(\boldsymbol{\beta}_0)/\partial\boldsymbol{\beta}$ and $\partial\bar{\sigma}^2(\bar{\boldsymbol{\beta}}_n)/\partial\boldsymbol{\beta}$ should be written as

$$\frac{\partial\bar{\sigma}^2(\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}} = \frac{\partial\bar{\sigma}^2(\bar{\boldsymbol{\beta}}_n)}{\partial\boldsymbol{\beta}} - \begin{pmatrix} \frac{\partial\bar{\sigma}^2\left(\boldsymbol{\beta}_n^{(1)}\right)}{\partial\beta_1\partial\beta_1} & \frac{\partial\bar{\sigma}^2\left(\boldsymbol{\beta}_n^{(1)}\right)}{\partial\beta_1\partial\beta_2} & \cdots & \frac{\partial\bar{\sigma}^2\left(\boldsymbol{\beta}_n^{(1)}\right)}{\partial\beta_1\partial\beta_{p+q}} \\ \frac{\partial\bar{\sigma}^2\left(\boldsymbol{\beta}_n^{(2)}\right)}{\partial\beta_2\partial\beta_1} & \frac{\partial\bar{\sigma}^2\left(\boldsymbol{\beta}_n^{(2)}\right)}{\partial\beta_2\partial\beta_2} & \cdots & \frac{\partial\bar{\sigma}^2\left(\boldsymbol{\beta}_n^{(2)}\right)}{\partial\beta_2\partial\beta_{p+q}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial\bar{\sigma}^2\left(\boldsymbol{\beta}_n^{(p+q)}\right)}{\partial\beta_{p+q}\partial\beta_1} & \frac{\partial\bar{\sigma}^2\left(\boldsymbol{\beta}_n^{(p+q)}\right)}{\partial\beta_{p+q}\partial\beta_2} & \cdots & \frac{\partial\bar{\sigma}^2\left(\boldsymbol{\beta}_n^{(p+q)}\right)}{\partial\beta_{p+q}\partial\beta_{p+q}} \end{pmatrix} (\bar{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0),$$

where each $\boldsymbol{\beta}_n^{(k)}$ lies on the segment connecting $\boldsymbol{\beta}_0$ and $\bar{\boldsymbol{\beta}}_n$, $1 \le k \le p+q$. The formula for $\tilde{\sigma}^2(\boldsymbol{\beta})$ is similar.