READING GROUP: ASYMPTOTIC THEORY FOR STATIONARY PROCESSES (WU, 2011)

Heman Leung

Spring, 2020

Motivation (p.1)

- Finite-sample distributions can be impossible to derive in time series
- Usual tools like CLT and LLN are not tailored for time series
  - *Simple CLT relies on independence*
  - *Advance CLT imposes condition like α-mixing, which is hard to verify*
- Some asymptotic tools developed for linear time series before
  - *What if nonlinear?*
  - *Goal of this survey: general stationary process*

Stationary process

- Form: $X_i = H(\dots, \epsilon_{i-1}, \epsilon_i)$ where $\epsilon_i, i \in \mathbb{Z}$ are iid random variables
  - *Basis of Wu's 2005 PNAS paper*
  - *Avoid the use of strong mixing condition*
  - *Sometimes called nonlinear Wold representation*
- Casual interpretation: $\{\epsilon_t\}_{t \le i}$ "cause" X_i
  - X_i *is independent of future innovation* ϵ_j *where* $j > i$
  - *Reason will be argued in next section*

REPRESENTATION THEORY

Section 2

Wold representation (1938) (p.2)

- Any weakly stationary process can be decomposed into
  - *A regular process (a moving average sum of white noises)*
  - *And a singular process (a linearly deterministic component)*
  - *Form:* $X_t = \sum_{j=0}^{\infty} b_j \epsilon_{t-j} + \eta_t$
- Also gives causal interpretation
  - *Stronger in the sense that it is linear*
- However not much insight for asymptotic distribution
  - *Joint distribution of white noises can be too complicated*

Rosenblatt transformation (1952)

- Any finite dimensional random vector can be expressed in distribution as functions of iid uniforms
  - *Based on quantile transformation:* $X_n \stackrel{d}{=} \left(X_{n-1}, G_n(X_{n-1}, U_n)\right)$
- Not applicable on stationary ergodic process
  - *However suggest the usefulness of nonlinear Wold representation*
  - *Stationary: distribution of the random variables*
  - *Ergodic: statistical property can be deduced from sample paths*

Comparison

Representation	Form	Requirement	Asymptotic
Wold	$X_t = \sum_{j=0}^{\infty} b_j \epsilon_{t-j} + \eta_t$	Weakly stationary	No
Nonlinear Wold	$X_t = H(\dots, \epsilon_{t-1}, \epsilon_t)$	Strictly stationary	Yes

- Weakly stationary ⟺ Wold
- Strictly stationary ⟺ nonlinear Wold?
  - *Previous example suggest that nonlinear Wold can represent lots of process*
  - *Is there strictly stationary process that cannot be represented?*
    - Then we cannot use the asymptotic in this paper

DEPENDENCE MEASURES

Section 3

Physical dependence

- Physical dependence measure: $\delta_p(j) \stackrel{\text{def}}{=} \|X_j - X_j^*\|_p$ where $j \geq 0$
  - $\mathcal{L}^p$ *norm:* $\|X\|_p \stackrel{\text{def}}{=} (E|X|^p)^{\frac{1}{p}}$
  - *Shift process:* $\mathcal{F}_i \stackrel{\text{def}}{=} (\dots, \epsilon_{i-1}, \epsilon_i)$
    - A bit like filtration of innovation
  - *Coupled innovation:* $(\epsilon_i')_{i\in\mathbb{Z}}$ *is iid copy of* $(\epsilon_i)_{i\in\mathbb{Z}}$
  - *Coupled* X_j*:* $X_j^* = H(F_j^*)$ *where* $F_j^* = (\dots, \epsilon_{-1}, \epsilon_0', \epsilon_1, \dots, \epsilon_{j-1}, \epsilon_j)$
    - Exchangeable: $\left(X_j, X_j^*\right) \stackrel{d}{=} \left(X_j^*, X_j\right)$
  - *Idea: measure the causal effect of changing initial input* ϵ_0 *on output* X_j
    - Rubin causality?

Predictive dependence

- Predictive dependence measure: $\omega_p(j) \stackrel{\text{def}}{=} \|g_j(\mathcal{F}_0) - g_j(\mathcal{F}_0^*)\|_p$ where $j \geq 0$
  - $g_j(\mathcal{F}_0) \stackrel{\text{def}}{=} E\left(X_j \middle| \mathcal{F}_0\right)$
    - Nonlinear analogue of Kolmogorov's (1941, written in Russian) linear predictor
  - *Idea: measure the predictive effect of knowing initial input ϵ_0 on output X_j*
    - Granger causality?
- Lemma 1: $\theta_p(i) \leq \omega_p(i) \leq 2\theta_p(i)$
  - $\theta_p(i) \stackrel{\text{def}}{=} \|\mathcal{P}_0 X_i\|_p$
  - *Projection operator:* $\mathcal{P}_j \cdot \stackrel{\text{def}}{=} E\left(\cdot \middle| \mathcal{F}_j\right) - E\left(\cdot \middle| \mathcal{F}_{j-1}\right)$ *where* $j \in \mathbb{Z}$
    - This naturally leads to martingale differences
  - *Interchangeable use of* $\omega_p(i)$ *and* $\theta_p(i)$

Projection operator

- Definition: $\mathcal{P}_j \cdot \overset{\text{def}}{=} E\left(\cdot \,\middle|\mathcal{F}_j\right) - E\left(\cdot \,\middle|\mathcal{F}_{j-1}\right)$ where $j \in \mathbb{Z}$
  - *Help decompose the predictive effect of knowing ϵ_j on X_i*
  - $A = E(A) + \sum_{j=-\infty}^{\infty} \mathcal{P}_j A =$ *mean effect + contribution of ϵ_j on predicting A*
  - $\mathcal{P}_i \mathcal{P}_j A = \begin{cases} \mathcal{P}_j A, & i = j \\ 0, & i \neq j \end{cases}$
    - result of tower property
  - $\mathcal{P}_j X_i = 0$ *if* $j > i$
    - future ϵ_j cannot cause X_i
  - $X_i = E(X_i) + \sum_{j=-\infty}^{i} \mathcal{P}_j X_i$

Stability (p.3)

- Stability: p-stable if $\Delta_p \stackrel{\text{def}}{=} \sum_{j=0}^{\infty} \delta_p(j) < \infty$
  - *Interpretation: cumulative impact of ϵ_0 on $\{X_i\}_{i\geq 0}$ is finite*
  - *Short-range dependence condition*
- Weak stability: weakly p-stable if $\Omega_p \stackrel{\text{def}}{=} \sum_{j=0}^{\infty} \omega_p(j) < \infty$
  - *Interpretation: cumulative contribution of ϵ_0 in predicting $\{X_i\}_{i\geq 0}$ is finite*
  - *Weak stability with $p = 2$ guarantees an invariance principle for the partial sum process $S_n = \sum_{i=1}^{n} X_i$*
    - Invariance principle: functional extension of CLT
- Why weak? (Wu, 2005): $\delta_p(j) \geq \omega_p(j)$ where $p \geq 1,\ j \geq 0$

Relationship

- Note that definitions in Wu (2005) are used instead

Name	Definition	Sum
Physical dependence $\delta_p(j)$	$\|X_j - X_j^*\|_p$	$\Delta_p < \infty \Rightarrow p$-stable
Predictive dependence $\omega_p(j)$	$\left\|E(X_j\|\mathcal{F}_0) - E(X_j\|\mathcal{F}_0^*)\right\|_p$	$\Omega_p < \infty \Rightarrow$ weakly p-stable
Projection $\theta_p(j)$	$\left\|E(X_j\|\mathcal{F}_0) - E(X_j\|\mathcal{F}_{-1})\right\|_p$	$\Theta_p < \infty \Rightarrow$ weakly p-stable

- $\theta_p(j) \leq \omega_p(j) \leq \min\left[2\theta_p(j), \delta_p(j)\right]$
- $\Delta_p < \infty \Rightarrow \Theta_p \leq \Omega_p < \infty$ (stability implies weak stability)

Examples (p.4)

- Linear process: $X_t = \sum_{i=0}^{\infty} a_i \epsilon_{t-i}$
  - *Existence can be confirmed with Kolmogorov's Three Series Theorem*
  - $\delta_p(n) = \|a_n \epsilon_0 - a_n \epsilon_0'\|_p = |a_n| \times \|\epsilon_0 - \epsilon_0'\|_p = \omega_p(n)$
  - *Stable if* $\sum_{i=0}^{\infty} |a_i| < \infty$
- ARMA process: $X_t = \epsilon_t + \sum_{j=1}^{p} \phi_j X_{t-j} + \sum_{l=1}^{q} \theta_l \epsilon_{t-l}$
  - *Special class of linear process*
  - a_i *is the coefficient of* $\frac{1+\sum_{l=1}^{q} \theta_l z^l}{1-\sum_{j=1}^{p} \phi_j z^j}$

Volterra series

- Volterra expansion: functional extension of Taylor expansion
- Nonlinear Wold: $H(\ldots, \epsilon_{n-1}, \epsilon_n) = \sum_{k=1}^{\infty} \sum_{u_1,\ldots,u_k=0}^{\infty} g_k(u_1, \ldots, u_k)\epsilon_{n-u_1} \ldots \epsilon_{n-u_k}$
  - *g_k are called Volterra kernel*
  - *Under the assumptions below, $X_n \in \mathcal{L}^2$ and exists*
    - ϵ_t are iid with mean 0 and variance 1
    - $g_k(u_1, \ldots, u_k)$ symmetric and $= 0$ if $u_i = u_j$ for some $1 \leq i < j \leq k$
    - $\sum_{k=1}^{\infty} \sum_{u_1,\ldots,u_k=0}^{\infty} g_k^2(u_1, \ldots, u_k) < \infty$
- Physical dependence: $\delta_p^2(n) = 2 \sum_{k=1}^{\infty} k \sum_{u_2,\ldots,u_k=0}^{\infty} g_k^2(\mathrm{n}, \mathrm{u}_2, \ldots, u_k)$
- Predictive dependence: $\omega_p^2(n) = 2 \sum_{k=1}^{\infty} k \sum_{u_2,\ldots,u_k=n+1}^{\infty} g_k^2(\mathrm{n}, \mathrm{u}_2, \ldots, u_k)$
  - *Stable if $\sum_{i=1}^{\infty} \omega_p(i) < \infty$*

NONLINEAR TIME SERIES

Section 4

Summary (p.4-7)

- Nonlinear $AR(p)$ model: $X_n = G(X_{n-1}, \dots, X_{n-p}; \epsilon_n)$ where $p \geq 1$ and $n \in \mathbb{Z}$
- Present sufficient condition for the above to have
  - *Stationary representation in form of nonlinear Wold*
  - *Geometric-moment contracting (GMC) property*
- Implication of GMC: $\delta_p(n) = O(r^n)$ for some $r \in (0,1)$
  - *Thus p-stable?*
- The rest are special cases of the above model
  - *E.g. Threshold AR, GARCH*

CENTRAL LIMIT THEORY

Section 5

Invariance principle (p.7-8)

- For simplicity, assume $E(X_i) = 0$ and $Cov(X_0, X_k) = \gamma_k$
- Traditional CLT: $\frac{S_n}{\sqrt{n}} \xrightarrow{d} N(0, \sigma^2)$ where $S_n = \sum_{i=1}^{n} X_i$
  - *Problems: autocorrelation, heteroskedasticity, no representation of* σ^2
- Invariance principle: $\left\{\frac{S_{nu}}{\sqrt{n}}, 0 \leq u \leq 1\right\} \xrightarrow{d} \{\sigma W_u, 0 \leq u \leq 1\}$
  - $S_t = S_{\lfloor t \rfloor} + (t - \lfloor t \rfloor) X_{\lfloor t \rfloor + 1}$ *(extend* S_n *to a stochastic process)*
  - W_u *is a standard Brownian motion*
  - *Entails CLT on* S_n *if holds*

Weak stability (p.8)

- If a time series is weakly p-stable, then $\Theta_{\mathrm{p}} \overset{\text{def}}{=} \sum_{i=0}^{\infty} \theta_p(i) < \infty$
  - *Weakly p-stable:* $\Omega_p = \sum_{j=0}^{\infty} \omega_p(j) < \infty$
  - *So this follows from lemma 1 that* $\Theta_{\mathrm{p}} \leq \Omega_p < \infty$
- Assume $E(X_i) = 0$ and $\Theta_{\mathrm{p}} < \infty$, then we have
  - *Moment inequality:* $\|S_n\|_p \leq \begin{cases} (p-1)^{\frac{1}{2}} n^{\frac{1}{2}} \Theta_p, \; p > 2 \\ (p-1)^{-1} n^{\frac{1}{p}} \Theta_p, \; 1 < p \leq 2 \end{cases}$
    - Help to bound the order of remainder term related to moment of S_n
    - Example: $\hat{\theta}(\bar{X}) = \hat{\theta}(\mu) + \left[\hat{\theta}(\bar{X}) - \hat{\theta}(\mu)\right]$ (potentially easier to deal with $\hat{\theta}(\mu)$)
  - *If* $\Theta_2 < \infty$, $\left\{\frac{S_{nu}}{\sqrt{n}}, 0 \leq u \leq 1\right\} \overset{d}{\rightarrow} \{\sigma W_u, 0 \leq u \leq 1\}$
    - Where $\sigma^2 = \|\sum_{i=0}^{\infty} \mathcal{P}_0 X_i\|^2 = \sum_{k \in \mathbb{Z}} \gamma_k$
    - Martingale approximation: Volný (1993) Theorem B and Theorem 6

Martingale approximation

- Martingale property: X_t satisfying $E\left(X_t \middle| \mathcal{F}_j\right) = X_j$ (best guess of future is present)
- Martingale difference sequence: $D_t = X_t - X_{t-1}$ satisfying $E(D_t|\mathcal{F}_{t-1}) = 0$
  - *Projection $\mathcal{P}_{i-l}X_i$ is MDS by tower property since $\mathcal{F}_{i-l} \subset \mathcal{F}_{i-1}$*
- Martingale approximation: $X_i = \sum_{l\in\mathbb{Z}} \mathcal{P}_{i-l}X_i$ ($E(X_i) = 0$ by assumption)
- Minkowski inequality: $\|v + w\|_p \leq \|v\|_p + \|w\|_p$
  - $\|S_n\|_p = \|\sum_{i=1}^n \sum_{l\in\mathbb{Z}} \mathcal{P}_{i-l}X_i\|_p \leq \sum_{l\in\mathbb{Z}}\|\sum_{i=1}^n \mathcal{P}_{i-l}X_i\|_p$
- Burkholder's inequality (1988): $c_p \left\|\sqrt{\sum_{i=1}^n D_i^2}\right\|_p \leq \|X_n\|_p \leq C_p \left\|\sqrt{\sum_{i=1}^n D_i^2}\right\|_p$
  - *X_i is a martingale, $D_i = X_i - X_{i-1}$ is a MDS*
  - *$c_p < C_p$ are positive constants depends on p where $1 < p < \infty$*
  - *Idea: relate maximum of a martingale with its quadratic variation*

Proof of moment inequality

- Apply Burkholder's inequality to $\|\sum_{i=1}^{n} \mathcal{P}_{i-l} X_i\|_p^p$
  - $\|\sum_{i=1}^{n} \mathcal{P}_{i-l} X_i\|_p^p \leq C_p' \left\| \sqrt{\sum_{i=1}^{n} (\mathcal{P}_{i-l} X_i)^2} \right\|_p^p = C_p' E \left|\sum_{i=1}^{n} (\mathcal{P}_{i-l} X_i)^2\right|^{\frac{p}{2}}$
  - $\leq C_p' \sum_{i=1}^{n} E|\mathcal{P}_{i-l} X_i|^p = C_p' \sum_{i=1}^{n} \|\mathcal{P}_{i-l} X_i\|_p^p$ *(by power of sum inequality)*
  - $= C_p' n \|\mathcal{P}_0 X_l\|_p^p$ *(by $\mathcal{L}^p$ stationarity)*
- By summing $C_p n^{\frac{1}{p}} \|\mathcal{P}_0 X_l\|_p$, we have $\|S_n\|_p \leq (p-1)^{-1} n^{\frac{1}{p}} \Theta_p, 1 < p \leq 2$
  - *The other case uses result of Rio (2009)*
  - *Tighter bound may have appeared in recent years but Wu's framework provides a uniform way to describe the condition*

Proof of invariance principle

- Doob's martingale inequality: to be discussed in the next paper

GAUSSIAN APPROXIMATION

Section 6

Strong invariance principle (p.8-9)

- Invariance principle does not specify rate of convergence
- Komlós–Major–Tusnády approximation (1975, 1976) for iid r.v.
  - *Assume* $X_i \in \mathcal{L}^p$ *where* $p > 2$, $E(X_i) = 0$
  - *On a richer probability space, there exists* $\{X_i'\}_{i\in\mathbb{Z}} \overset{d}{=} \{X_i\}_{i\in\mathbb{Z}}$ *and* $S_n' = \sum_{i=1}^{n} X_i'$
  - *We have* $\max_{0\le t\le n} |S_t' - \sigma W_t| = o_{a.s.}\left(n^{\frac{1}{p}}\right)$
    - Where $\sigma = \|X_i\|$
  - *Further assume* $E\left(e^{t|X_1|}\right) < \infty$, *we have* $\max_{0\le t\le n} |S_t' - \sigma W_t| = o_{a.s.}(\log n)$
    - A bit like MGF exists

KMT approximation under dependence

- Theorem (Wu, 2007):
  - *Assume* $X_i \in \mathcal{L}^p$ *where* $2 < p \leq 4$ *has nonlinear Wold representation*
  - *Assume* $E(X_i) = 0$ *and* $\sum_{i=1}^{\infty}\left[\delta_p(i) + i\omega_p(i)\right] < \infty$
    - $\sum_{i=1}^{\infty} i\delta_p(i) < \infty$ is sufficient
  - *Then on a richer probability space, there exists* $\{X_i'\}_{i\geq 0} \overset{d}{=} \{X_i\}_{i\geq 0}$
  - *We have* $\max_{0\leq t\leq n} |S_t' - \sigma W_t| = o_{a.s.}\left[n^{\frac{1}{p}}(\log n)^{\frac{1}{2}+\frac{1}{p}}(\log\log n)^{\frac{2}{p}}\right]$
    - Where $\sigma = \|\sum_{i=0}^{\infty} \mathcal{P}_0 X_i\|$
- Theorem (Berkes, Liu & Wu, 2014):
  - *Under some other mild assumptions, we have* $S_n' - \sigma W_n = o_{a.s.}\left(n^{\frac{1}{p}}\right)$
  - *Read the AoP paper if you are interested*

SAMPLE COVARIANCE FUNCTION

Section 7

CLT with bounded k

- Sample autocovariance: $\hat{\gamma}_k \overset{\text{def}}{=} \frac{1}{n}\sum_{i=k+1}^{n}(X_i - \bar{X})(X_{i-k} - \bar{X})$ where $0 \le k < n$
  - *If* $\mu = 0$*, then* $\hat{\gamma}_k = \frac{1}{n}\sum_{i=k+1}^{n} X_i X_{i-k}$
- Assume $E(X_i) = 0, X_i \in \mathcal{L}^p$ where $2 < p \le 4$ and $\Delta_p < \infty$
  - *Let* $Y_i = (X_i, \dots, X_{i-k})^T$ *and* $\Gamma_k = (\gamma_0, \dots, \gamma_k)^T$ *where* $k \in \mathbb{N}$ *is fixed, we have*
  - *Moment inequality:* $\left\| \hat{\gamma}_k - \left(1 - \frac{k}{n}\right)\gamma_k \right\|_{\frac{p}{2}} \le \frac{3p-3}{n}\Theta_p^2 + \frac{4n^{\frac{2}{p}-1}}{p-2}\|X_1\|_p \Delta_p$
  - *If* $X_i \in \mathcal{L}^4$ *and* $\Delta_4 < \infty$, $\sqrt{n}(\hat{\gamma}_0 - \gamma_0, \dots, \hat{\gamma}_k - \gamma_k) \overset{d}{\to} N[0, E(D_0 D_0^T)]$
    - Where $D_0 = \sum_{i=0}^{\infty} \mathcal{P}_0(X_i Y_i) \in \mathcal{L}^2$

Proof of bounded k (p.10)

- Jensen's inequality: $\phi[E(X)] \leq E[\phi(X)]$ where $\phi(\cdot)$ is convex
  - *Every $\mathcal{L}^p$ norm is convex by Minkowski inequality*
- Product identity: $ab - \hat{a}\hat{b} \equiv a\left(b - \hat{b}\right) + \hat{b}(a - \hat{a})$
  - *As we usually have good knowledge of $a - \hat{a}$ and $b - \hat{b}$*
- Cauchy–Schwarz inequality: $\left|\sum_{i=1}^{n} a_i b_i\right| \leq \sqrt{\left(\sum_{i=1}^{n} a_i\right)^2 \left(\sum_{i=1}^{n} b_i\right)^2}$
  - *Probabilistic version:* $|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$
- Cramér–Wold device: $\vec{X}^{(n)} \xrightarrow{d} \vec{X} \Leftrightarrow \sum_{i=1}^{k} t_i X_i^{(n)} \xrightarrow{d} \sum_{i=1}^{k} t_i X_i \;\; \forall \vec{t} \in \mathbb{R}^k$
  - *Help establish multivariate convergence using univariate result*

CLT with unbounded k

- Asymptotic distribution does not depend on the speed of $k_n \to \infty$
- Assume $E(X_i) = 0, k_n \to \infty$ and $\Delta_p < \infty$
  - *Let* $Z_i = (X_i, \dots, X_{i-h+1})^T$ *where* $h \in \mathbb{N}$ *is fixed, we have*
  - $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[X_i Z_{i-k_n} - E\left(X_{k_n} Z_0\right)\right] \xrightarrow{d} N(0, \Sigma_h)$
    - Where Σ_h has entries $\sigma_{ab} = \sum_{j \in \mathbb{Z}} \gamma_{j+a} \gamma_{j+b} = \sigma_{0,a-b}$ where $1 \le a, b \le h$
  - *If* $\frac{k_n}{n} \to 0$, $\sqrt{n}\left[\left(\hat{\gamma}_{k_n}, \dots, \hat{\gamma}_{k_n-h+1}\right)^T - \left(\gamma_{k_n}, \dots, \gamma_{k_n-h+1}\right)^T\right] \xrightarrow{d} N(0, \Sigma_h)$
- The above results come from Wu (2008) for short-range dependence
  - *Can be extended to long-range linear process*

Estimation problem of γ_k

- $\hat{\gamma}_k$ is not a good estimator of γ_k when k is large
  - *Example:* $k \to \infty$ *with* $\frac{k}{n} \to 0$ *satisfies* $\sqrt{n}\gamma_k \to 0$
  - *MSE of* $\hat{\gamma}_k$*:* $E[(\hat{\gamma}_k - \gamma_k)^2] \sim \frac{\sigma_{00}}{n}$
  - *MSE of* $\tilde{\gamma}_k = 0$*:* $E[(\tilde{\gamma}_k - \gamma_k)^2] = o\left(\frac{1}{n}\right) \ll O\left(\frac{1}{n}\right)$
    - Little o: $f(x) = o\big(g(x)\big) \Leftrightarrow \lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$
- Shrinkage estimate $\hat{\gamma}_k \mathbb{I}_{|\hat{\gamma}_k| \geq c_n}$ with carefully chosen $c_n \to 0$ can reduce MSE
  - *Similar to Stein's phenomenon (discussed last semester)*
  - *Details to be discussed in next section*

ESTIMATION OF COVARIANCE MATRIX

Section 8

Convergence problem of $\hat{\Sigma}_n$ (p.11)

- Operator norm: $\rho(A) \stackrel{\text{def}}{=} \max_{x \in \mathbb{R}^n : |x|=1} |Ax|$
  - $|x| \stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^{n} x_i^2}$
  - *Hence $\rho^2(A)$ is the largest eigenvalue of $A^T A$*
- Previous entry-wise convergence result does not imply matrix convergence of $\hat{\Sigma}_n$
  - *Inconsistency is due to previous estimation problem*
  - *Conjecture (Wu & Pourahmadi, 2009):* $\rho\left(\hat{\Sigma}_n - \Sigma_n\right) \xrightarrow{d} Gumbel(0,1)$
    - With proper centering and scaling
    - Gumbel distribution is usually used in extreme value theory

Truncation technique

- Banded covariance matrix estimator: $\hat{\Sigma}_{n,l_n} = \left(\hat{\gamma}_{i-j}\mathbb{I}_{|i-j|\le l_n}\right)_{1\le i,j\le n}$
  - *Under suitable conditions on banding parameter l_n, $\hat{\Sigma}_{n,l_n}$ is consistent*
  - *However $\hat{\Sigma}_{n,l_n}$ may not be non-negative definite*
- Tapered version: $\tilde{\Sigma}_{n,l_n} = \left(\hat{\gamma}_{i-j} w\left(\frac{|i-j|}{l_n}\right)\right)_{1\le i,j\le n} = \hat{\Sigma}_n \star W_n$
  - *$\star$ is the element-wise product*
  - *$w(\cdot)$ is a lag window function (aka kernel) satisfying some conditions*
    - Such that W_n is non-negative definite
    - Example (Bartlett kernel/triangular window): $w_B(u) = \max(0,1-|u|)$
- Schur Product Theorem: $A \star B$ is non-negative definite if A, B are non-negative definite

Result of $\hat{\Sigma}_n$

- Theorem (Wu & Pourahmadi, 2009):
  - *Assume* X_i *has nonlinear Wold representation and* $\Theta_2 < \infty$
  - *If* $\sigma = \|\sum_{i=0}^{\infty} \mathcal{P}_0 X_i\| > 0$*, then* $\rho(\hat{\Sigma}_n - \Sigma_n) \overset{p}{\nrightarrow} 0$
    - i.e. inconsistency when long-run variance is non-zero
- Theorem (Xiao & Wu, 2010):
  - *Assume* X_i *has nonlinear Wold representation and* $E(X_i) = 0$
  - *Assume* $X_i \in \mathcal{L}^p$ *where* $p > 2$ *and* $\sum_{i=j}^{\infty} \delta_p(i) = o\left(\frac{1}{\log j}\right)$ *as* $j \to \infty$
  - *Assume* $\min_{\vartheta} f(\vartheta) > 0$
    - f is the spectral density function which will be discussed in next section
  - *Then* $\exists c > 0\ s.t. \lim_{n\to\infty} P\left[c^{-1} \log n \le \rho(\hat{\Sigma}_n - \Sigma_n) \le c \log n\right] = 1$
    - i.e. $\rho(\hat{\Sigma}_n - \Sigma_n) = o(\log n)$ as $n \to \infty$
- These two theorems implies $\hat{\Sigma}_n$ usually does not converge

Result of $\tilde{\Sigma}_{n,l_n}$

- Upper bound: assume $E(X_i) = 0$ and $\Delta_p < \infty$ for $2 < p \leq 4$
  - *Let* $b_n = \sum_{k=1}^{l} \left|1 - w\left(\frac{k}{l}\right) + \frac{k}{n} w\left(\frac{k}{l}\right)\right| |\gamma_k| + \sum_{j=l+1}^{n} |\gamma_j|$
  - *Then* $\left\|\rho\left(\tilde{\Sigma}_{n,l} - \Sigma_n\right)\right\|_q \leq 2b_n + (l+1)\frac{4\|X_1\|_p \Delta_p}{n^{1-\frac{1}{q}(p-2)}}$ *where* $q = \frac{p}{2}, 0 \leq l < n$
    - Note that this bound is non-asymptotic
  - *Hence if* $l = l_n \to \infty$ *and* $\frac{l_n}{n^{1-\frac{1}{q}}} \to 0$, $\left\|\rho\left(\tilde{\Sigma}_{n,l} - \Sigma_n\right)\right\|_q \to 0$
- Theorem (Xiao & Wu, 2010):
  - *Assume* $X_i \in \mathcal{L}^p$ *where* $p > 4$ *and* $\Theta_p(m) = O(m^{-\alpha})$ *where* $\alpha > 0$
  - *Let* $l_n \asymp n^{\lambda}$ *where* $\lambda \in (0,1)$ *satisfy* $\lambda < \frac{p\alpha}{2}$ *and* $(1-2\alpha)\lambda < 1 - \frac{4}{p}$
  - *Then* $\rho\left(\tilde{\Sigma}_{n,l} - \Sigma_n\right) = O(b_n) + O_{\mathbb{P}}\left[n^{-\frac{1}{2}}(l_n \log l_n)^{\frac{1}{2}}\right]$
  - *Adding additional assumptions gives a lower bound*

Special cases (p.12)

- Assume $p = 4$ and $\gamma_k = O(\rho^k)$ for some $0 < \rho < 1$
- Rectangular window: $w(k) = 1$ for $|k| \le l$
  - *Choose* $l = l_n = \left\lfloor \frac{\log n}{-2 \log \rho} \right\rfloor$, *then* $\left\| \rho\left(\tilde{\Sigma}_{n,l} - \Sigma_n\right) \right\| = O\left(n^{-\frac{1}{2}} \log n\right)$
    - Almost optimal but may not be non-negative definite
- Bartlett window: $w(k) = 1 - \frac{|k|}{l}$ for $|k| \le l$
  - *Choose* $l \asymp n^{\frac{1}{4}}$, *then*
  - $\left\| \rho\left(\tilde{\Sigma}_{n,l} - \Sigma_n\right) \right\| = O(1) \sum_{k=1}^{l} [1 - w(k)] |\gamma_k| + O\left(l n^{-\frac{1}{2}} + \rho^l\right) = O\left(n^{-\frac{1}{4}}\right)$
- Parzen window: $1 - w_P(u) = O(u^2)$
  - *Choose* $l \asymp n^{\frac{1}{6}}$, *then*
  - $\left\| \rho\left(\tilde{\Sigma}_{n,l} - \Sigma_n\right) \right\| = O\left(l^{-2} + l n^{-\frac{1}{2}} + \rho^l\right) = O\left(n^{-\frac{1}{3}}\right)$

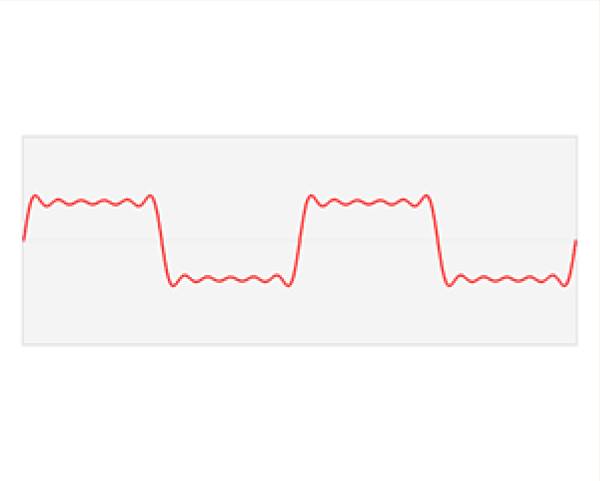
Application

- The upper bound can be applied to Wiener-Kolmogorov prediction
  - *Since γ_k can only be estimated with finite sample in practice*
  - *Probably referring to Wiener filter*
  - *Kalman filter is nonstationary extension of Wiener filter*
    - More popular in practice
- Help establish asymptotic theory of estimates of coefficients in
  - *Wold decomposition theorem*
  - *Discrete Wiener-Hopf equations*
    - A method to solve systems of integral equations

PERIODOGRAMS

Section 9

Frequency domain



- Time domain: changes of a signal with respect to time
- Frequency domain: changes of a signal with respect to frequency
  - *How much of a signal lies within each given band over a range of frequencies*
- Why frequency domain?
  - *Simplify the mathematical analysis*
  - *Give an intuitive understanding of the qualitative behavior of the system*
    - E.g. periodicity, power

Tools (p.12-13)

- Periodogram: $I_n(\phi) \stackrel{\text{def}}{=} \frac{|S_n(\phi)|^2}{n}\ \forall \phi \in \mathbb{R}$
  - *Discrete Fourier transform (DFT):* $S_n(\phi) = \sum_{t=1}^{n} x_t e^{it\phi}$ *where* $i = \sqrt{-1}$
- Spectral distribution function: right-continuous, non-decreasing F that satisfy
  - $\gamma_k = \int_0^{2\pi} e^{ik\phi} dF(\phi)$ *and is bounded on* $[0,2\pi]$
- Spectral density function: $f = F'$ if F is absolutely continuous
- Theorem (Peligrad & Wu, 2010): for regular nonlinear Wold process,
  - *If* $\sum_{k\in\mathbb{Z}} |\gamma_k| < \infty$, *then* $f(\phi) = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} \gamma_k e^{ik\phi} = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} \gamma_k \cos(k\phi)$
    - Euler's formula: $e^{i\phi} = \cos\phi + i\sin\phi$
  - *Continuity property: if* $u_p = \sum_{k=1}^{\infty} k^p |\gamma_k| < \infty$, *then* $f \in \mathcal{C}^p(\mathbb{R})$
    - Larger p such that $u_p < \infty$ means weaker serial dependence

Central limit problem of $S_n(\phi)$

- Assume X_t is second order stationary, $E(X_t) = 0$ and $\sum_{k\in\mathbb{Z}}|\gamma_k| < \infty$
  - *Then* $E[I_n(\phi)] = \sum_{k=1-n}^{n-1}\left(1 - \frac{|k|}{n}\right)\gamma_k \cos(k\phi) \to 2\pi f(\phi)$ *as* $n \to \infty$
  - *Hence* $\frac{|S_n(\phi)|^2}{2\pi n}$ *is asymptotically unbiased for* $f(\phi)$
    - Note that $S_n(\phi)$ comes from discrete transform but $f(\phi)$ comes from continuous
  - *However it is inconsistent by the following theorem*
- Theorem: assume $E(X_t^2) < \infty$
  - *For almost all* $\vartheta \in \mathbb{R}$ *(Lebesgue), we have* $\binom{\Re}{\Im}\frac{S_n(\vartheta)}{\sqrt{n}} \xrightarrow{d} N[0, \pi f(\vartheta) Id_2]$
    - Consequently, $\frac{I_n(\vartheta)}{2\pi f(\vartheta)} \xrightarrow{d} Exp(1)$
    - Proposition: this holds $\forall\vartheta \in (0, 2\pi)$ if $\sum_{i=0}^{\infty}\|\mathcal{P}_0 X_i - \mathcal{P}_0 X_{i+1}\| < \infty$
      - *A sufficient condition is* $\Theta_p < \infty$
  - *For almost all pairs* (ϑ, φ) *(Lebesgue),* $\frac{S_n(\vartheta)}{\sqrt{n}} \perp \frac{S_n(\varphi)}{\sqrt{n}}$ *asymptotically*

Fast Fourier transform

- Discrete Fourier transform: $O(n^2)$ time complexity to obtain $S_n(\vartheta_j), j = 1, \ldots, n$
- Fast Fourier transform: $O(n \log_2 n)$ time complexity to obtain same estimate
  - *A size-N ($N = N_1 N_2$) DFT can be expressed as two DFTs with size N_1, N_2*
  - *This is possible by complex root of unity (aka twiddle factors)*
  - *Cooley-Tukey algorithm*
- Faster Fourier transform?
  - *Lower bound on time complexity of FFT is an open problem*
  - *Some results under sparsity (Hassanieh et. al., 2012)*

Central limit problem of $S_n(\vartheta_j)$

- Theorem: assume X_i is nonlinear Wold, $\Theta_p < \infty$ and $\min_\vartheta f(\vartheta) > 0$,
  - *Let* $q \in \mathbb{N}$, $m = \left\lfloor \frac{n-1}{2} \right\rfloor$ *and* $Y_k \overset{iid}{\sim} N(0,1)$ *for* $1 \le k \le 2q$
  - *Then* $\left\{ \frac{S_n\left(\vartheta_{l_j}\right)}{\sqrt{n\pi f(\vartheta_{l_j}}}, 1 \le j \le q \right\} \overset{d}{\to} \left\{ Y_{2j-1} + iY_{2j}, 1 \le j \le q \right\}$
    - Where $1 \le l_1 < \cdots < l_q \le m$ may depend on n
    - Consequently, $\left\{ \frac{I_n\left(\vartheta_{l_j}\right)}{f\left(\vartheta_{l_j}\right)}, 1 \le j \le q \right\} \overset{d}{\to} \left\{ E_j, 1 \le j \le q \right\}$ where $E_j \overset{iid}{\sim} Exp(1)$
  - *The remaining part discuss the maximum error of approximation (?)*
    - Continuous mapping theorem: $X_n \overset{d}{\to} X \Rightarrow g(X_n) \overset{d}{\to} g(X)$ if $g(\cdot)$ is continuous
      - *This is only the part used. Check standard reference for full theorem*
    - Theorem (Lin & Liu, 2009) states convergence to standard Gumbel

ESTIMATION OF SPECTRAL DENSITIES

Section 10

Estimation problem of $f(\theta)$ (p.14)

- Inconsistency of $I_n(\vartheta)$ (though unbiased) proved in last section
- Lag window estimator: $f_n(\theta) = \frac{1}{2\pi}\sum_{k=1-n}^{n-1} K\left(\frac{k}{B_n}\right)\hat{\gamma}_k e^{ik\theta}$ where
  - *Bandwidth B_n satisfies $B_n \to \infty$ and $\frac{B_n}{n} \to 0$*
  - *Window K is symmetric, bounded, continuous at 0 and $K(0) = 1$*
  - *This estimator is consistent but its limiting distribution is highly nontrivial*
- Theorem (Liu & Wu, 2010): assume $E(X_t) = 0, E(X_t^4) < \infty$ and $\Delta_4 < \infty$
  - *Let $B_n \to \infty$ and $B_n = o(n)$ as $n \to \infty$*
  - *Assume K is symmetric, bounded, $\lim_{u\to 0} K(u) = K(0) = 1$*
  - *Assume $\kappa \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} K^2(u)du < \infty$ and K is continuous at all but finite points*
  - *Assume $\sup_{0<w\leq 1} \sum_{j\geq \frac{c}{w}} K^2(jw) \to 0$ as $c \to \infty$*
  - *Then $\sqrt{\frac{n}{B_n}}\{f_n(\theta) - E[f_n(\theta)]\} \xrightarrow{d} N\left[0, \left(1 + \mathbb{I}_{\frac{\theta}{\pi}\in\mathbb{Z}}\right) f^2(\theta)\kappa\right]$ for $0 \leq \theta < 2\pi$*

Long-run variance

- Long-run variance: $2\pi f(0) = \frac{2\pi}{2\pi}\sum_{k\in\mathbb{Z}}\gamma_k \cos(0) = \sum_{k\in\mathbb{Z}}\gamma_k = \sigma^2$
  - *First equality is due to theorem in Peligrad & Wu (2010) (slide p.41)*
  - *Last equality is due to probabilistic representation of σ^2 (slide p.20)*
- Put $\theta = 0$ in the previous CLT, we have $\sqrt{\frac{n}{B_n}}\{f_n(0) - f(0)\} \xrightarrow{d} N[0, 2f^2(0)\kappa]$
  - *If the bandwidth $b_n = B_n^{-1}$ satisfy*
  - $2\pi\{E[f_n(0)] - f(0)\} = \sum_{k=1-n}^{n-1} K(kb_n)\left(1 - \frac{|k|}{n}\right)\gamma_k - \sum_{k\in\mathbb{Z}}\gamma_k = O\left[(nb_n)^{-\frac{1}{2}}\right]$
- Log transformation can stabilize the variance (ease CI but may lose good property)
  - $\sqrt{\frac{n}{B_n}}[\log f_n(0) - \log f(0)] \xrightarrow{d} N(0, 2^2)$

Recursive estimation

- Lag window estimator is non-recursive as bandwidth depends on n
  - *When bandwidth changes, all blocks need to be updated (time complexity)*
  - *If all blocks need to be updated, we need store all data (space complexity)*
- Recursive estimation is possible by letting bandwidth depends on i
  - *When bandwidth changes, only the new block need to be updated*
  - *If bandwidth is increasing in a block, we only need the new data*
- Xiao and Wu (2010) provides algorithm for spectral density
- Wu (2009), Chan and Yau (2017) provides algorithms for long-run variance

KERNEL ESTIMATION

Section 11

Kernel regression

- Model: $Y_i = G(X_i, \eta_i), X_i = (\dots, \epsilon_{i-1}, \epsilon_i)$
  - *Important example: the autoregressive model* $X_{i+1} = R(X_i, \epsilon_{i+1})$
- Nadaraya-Watson estimator of $g(x_0) = E(Y_n | X_n = x_0)$: $g_n(x_0) = \frac{T_n(x_0)}{f_n(x_0)}$
  - $T_n(x) = \frac{1}{n}\sum_{t=1}^{n} Y_t K_{b_n}(x - X_t)$ *where* $K_{b_n}(x) = \frac{1}{b_n} K\left(\frac{x}{b_n}\right)$
    - Kernel K is symmetric, bounded on $\mathbb{R}$, has bounded support and $\int_{\mathbb{R}} K(u)du = 1$
    - Bandwidth $b_n \to 0$ and $nb_n \to \infty$
  - $f_n(x_0) = \frac{1}{n}\sum_{t=1}^{n} K_{b_n}(x_0 = X_t)$
    - Rosenblatt’s (1956) kernel density estimate

CLT for $g_n(x_0)$

- l-step ahead conditional densities: $F_l(X_{i+l} \le x|\mathcal{F}_i)$, $f_l(x|\mathcal{F}_i) = \frac{d}{dx}F_l(x|\mathcal{F}_i)$
- Theorem (Wu, 2005; Wu, Huang & Huang, 2010)
  - *Assume* $\exists c_0 < \infty\ s.t. \sup_{x\in\mathbb{R}} f_1(x|\mathcal{F}_i) \le c_0\ a.s.$ *and* $\sum_{i=1}^{\infty} \sup_x \|\mathcal{P}_0 f_1(x|\mathcal{F}_i)\| < \infty$
  - *Assume* $b_n \to 0, nb_n \to \infty$ *and let* $\kappa = \int_{\mathbb{R}} K^2(u)du$
  - *Then* $\sqrt{nb_n}\{f_n(x_0) - E[f_n(x_0)]\} \xrightarrow{d} N[0, f(x_0)\kappa]$
  - *Let* $V_p(x) = E|G(x,\eta_n)|^p$ *and* $\sigma^2(x) = V_2(x) - g^2(x)$
  - *If* $f(x_0) > 0, V_2, g \in \mathcal{C}(\mathbb{R})$ *and* $V_p(x)$ *is bounded on a neighborhood of* x_0
  - *Then* $\sqrt{nb_n}\left\{g_n(x_0) - \frac{E[T_n(x_0)]}{E[f_n(x_0)]}\right\} \xrightarrow{d} N\left[0, \frac{\sigma^2(x_0)\kappa}{f(x_0)}\right]$
    - Note that $E\left[\frac{T_n(x_0)}{f_n(x_0)}\right]$ does not necessary equal to $\frac{E[T_n(x_0)]}{E[f_n(x_0)]}$
    - This implies they are probably independent

Maximum deviation

- Maximum deviation: $\Lambda_n \stackrel{\text{def}}{=} \sup_{l \le x \le u} \sqrt{\frac{nb}{\kappa f(x)}} |f_n(x) - E[f_n(x)]|$
- Theorem (Liu & Wu, 2010)
  - *Assume* $X_n = a_0\epsilon_n + g(\dots, \epsilon_{i-2}, \epsilon_{i-1}) \in \mathcal{L}^p$ *for some* $p > 0$ *where* $a_0 \neq 0$
  - *Assume pdf* f_ϵ *of* ϵ_1 *is positive and* $\sup_{x \in \mathbb{R}} [f_\epsilon(x) + |f'_\epsilon(x)| + |f''_\epsilon(x)|] < \infty$
  - *Assume* $\exists 0 < \delta_2 \le \delta_1 < 1$ *such that* $n^{-\delta_1} = O(b_n)$ *and* $b_n = O\left(n^{-\delta_2}\right)$
  - *Let* $p' = \min(p, 2)$ *and* $\Theta_n = \sum_{i=0}^{n} \delta_{p'}(i)^{\frac{p'}{2}}$
  - *Assume* $\Psi_{n,p'} = O(n^{-\gamma})$ *for some* $\gamma > \frac{\delta_1}{1-\delta_1}$
  - *Assume* $\sum_{k=-n}^{\infty} (\Theta_{n+k} - \Theta_k)^2 = o(b_n^{-1} n \log n)$
  - *Let the kernel* $K \in \mathcal{C}^1[-1,1]$ *with* $K(\pm 1) = 0, l = 0$ *and* $u = 1$
  - *Then* $P\left[(2 \log b^{-1})^{-\frac{1}{2}} \Lambda_n - 2 \log b^{-1} - \frac{1}{2} \log K_3 \le z\right] \to e^{-2e^{-z}}\ \forall z \in \mathbb{R}$
    - Where $K_3 = \frac{\int_{-1}^{1} [K'(t)]^2 dt}{4\pi^2 \int_{-1}^{1} K^2(t) dt}$

U-STATISTICS

Section 12

U-statistic

- Weighted U-statistic: $U_n = \sum_{1\le i,j\le n} w_{i-j} K(X_i, X_j)$
  – *Where $w_i = w_{-i}$ are weights and K is symmetric measurable function*
  – *Predictive dependence: $\theta_{i,j} = \left\|\mathcal{P}_0 K(X_i, X_j)\right\|$*
- Theorem (Hsing and Wu, 2004)
  – *Assume $\sum_{k=0}^{\infty}\sum_{i=0}^{\infty}|w_k|\theta_{i,i-k} < \infty$ (summable weights)*
  – *Then $\exists\sigma^2 < \infty$ such that $\frac{1}{\sqrt{n}}[U_n - E(U_n)] \xrightarrow{d} N(0,\sigma^2)$*
  – *Let $W_n(i) = \sum_{j=1}^{n} w_{i-j}$ and $W_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n} W_n^2(i)}$*
  – *Assume $\sum_{i=1}^{\infty}|w_i| = \infty, \sum_{k=0}^{n}(n-k)w_k^2 = o(nW_n^2)\,, \liminf_{n\to\infty}\frac{W_n}{\sum_{i=0}^{\infty}|w_i|} > 0$*
  – *Assume $\sum_{l=0}^{\infty}\sup_{j\in\mathbb{Z}}\left\|K(X_0, X_j) - K(\tilde{X}_0, \tilde{X}_j)\right\| < \infty$*
    - Where $\tilde{X}_j = E(X_j | \epsilon_{j-l}, \dots, \epsilon_j)$
  – *Then $\exists\sigma_U^2 < \infty$ such that $\frac{1}{W_n\sqrt{n}}[U_n - E(U_n)] \xrightarrow{d} N(0,\sigma_U^2)$*