16:960:665 - Syllabus

Pranav Tikkawar

September 30, 2025

Contents

Syllabus

Time Series: Theory and Methods. Brockwell and Davis
Asymtotic Theory of Weakly dependent Random Process
Martingale Limit Theory

Durret - Probability Theory and Examples

Questions

Ask what I need to get and review before classes start
Measure Theory: not hardcore
look into textbooks and ergodic theory
Ask the professors of the classes to audit
What is the $X.(\omega)$ noations

Acronyms

R.V. - Random Variable
S.P. - Stochastic Process
fn - Function
dist - Distribution
G.P. - Gaussian Process
iid - independent and identically distributed
a.s. - Almost Surely
w.p 1 - with probability 1

1 Notes

1.1 9/2/2025 Lecture 1

We use Stochastic Process to model time series data

Definition (Stochastic Process)**.** A stochastic process is a family of random variables $\{X_t : t \in \mathcal{T}\}$ defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$.
$\mathcal{T} = \mathbb{N}, \mathbb{Z}$ Discrete Time
$\mathcal{T} = \mathbb{R}$ Continuous Time (not focusing on this)
$\mathcal{T} \subseteq \mathbb{R}^n$ Geospatial, with location and time, (not focusing on this)
$\mathcal{T} \subseteq \mathbb{S}^3$ Unit Sphere w/Geophysics.

Definition (Realization of a S.P.)**.** The functions $\{X.(\omega), \omega \in \Omega\}(\mathcal{T} \to \mathbb{R})$ are realizations or sample passes of the process.

- Fix t, X_t is a fn of Ω
- Fix an outcome $\omega \in \Omega$, $X.(\omega)$ is a fn on $\mathcal{T}$
- The time series we observe is a realization of the S.P.
- Conventionally the observed time series is indexed by $\{1, 2, \ldots, n\}$ ie $\{X_1, X_2, \ldots, X_n\}$ (known as the lens/sample size)

Example (1.2.1 from book)**.** Suppose $A \geq 0$ is a R.V and given by $\Theta \sim Uniform(0, 2\pi)$. and they are independent. and $v > 0$ is a known constant
Then $X_t = A\cos(vt + \Theta), t \in \mathbb{Z}$
Fore every $\omega \in \Omega$, $A(\omega), \Theta(\omega)$ are fixed
$X_t(\omega) = A(\omega)\cos(vt + \Theta(\omega))$
A determines the amplitude and Θ determines the phase.
What we do is we take a model, and have the data as a realization, and solve the inverse problem of determining the parameters of the model.

Example (1.2.2 from the book)**.** Consider $X_1, X_2, X_3, \ldots$ are IID and take value $1, -1$ with probability $1/2$
Im considering to use some binomal theorem thing...

Example (1.2.3 from the book)**.** Suppose X_t coming from prior question.
$S_t = \sum_{i=1}^{t} X_i = X_1 + X_2 + \ldots + X_t$ $S_t : t \in \mathbb{N}$ is a S.P. called a simple symmetric random walk
Consider a man in 1D who starts at 0, and takes a random draw to walk left or right. The path of this miserable guys is S_t
The realization is a plot of $S_t(\omega)$ against t.

Definition (The Distribution of a Stochastic Process)**.** Let $\mathcal{I}$ be the collection of all tuples $\{\mathbf{t} = (t_1, t_2, \ldots, t_n)' \in \mathcal{T}, t_1 < t_2 < \ldots < t_n\}$ The finite dimensional dist. fns of $\{X_t, t \in \mathcal{T}\}$ are the collection of fns $\{F_t(\cdot)\mathbf{t} \in \mathcal{I}\}$ where

$$F_{\mathbf{t}}(\mathbf{x}) = P(X_{\mathbf{t_1}} = x_1, X_{\mathbf{t_2}} = x_2, \ldots, X_{\mathbf{t_n}} = x_n)$$

$$\mathbf{x} = (x_1, x_2, \ldots, x_n)' \in \mathbb{R}^n$$

ie $F_{\mathbf{t}}(\mathbf{x})$ is the joint distribution of the process of the R.V. $\mathbf{x}$.

Theorem 1 (Kolmogrov (consistincy) Theorem)**.** *The prob. distribution fns $\{F_{\mathbf{t}}(\cdot) : \mathbf{t} \in \mathcal{I}\}$ are the distribution functions of some S.P. $\iff$ for any $n \in \mathbb{N}$, $\mathbf{t} = (t_1, t_2, \ldots, t_n)' \in \mathcal{T}$ and $1 \leq i \leq n$*

$$\lim_{x_i \to \infty} F_{\mathbf{t}}(\mathbf{x}) = F_{\mathbf{t_i}}(\mathbf{x_i})$$

Where $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$,
$\mathbf{t_i} = (t_1, t_2, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n)'$ and $\mathbf{x_i} = (x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)'$

essentially the i are the missing ones

$$F(x_1, x_2) = \mathbb{P}(X_1 \le x_1, X_2 \le x_2)$$
$$\lim_{x_2 \to \infty} F(x_1, x_2) = \mathbb{P}(X_1 \le x_1)$$

"https://en.wikipedia.org/wiki/Kolmogorov_extension_theorem"
We essentially only need to specify the consistency of the finite dimensional distributions to define a S.P.

1.2 9/9/2025 Lecture 2

Definition (Autocovariance function)**.** If $X_t, t \in \mathcal{T}$ is a S.P. s.t $E(X_t^2) < \infty$, then for every $t \in \mathcal{T}$ the the autocovariance function is defined as

$$\gamma_x(r, s) = Cov(X_r, X_s), r, s \in \mathcal{T}$$

Definition (Autocorrelation function)**.** If $X_t, t \in \mathcal{T}$ is a S.P. s.t $E(X_t^2) < \infty$, then for every $t \in \mathcal{T}$ the the autocorrelation function is defined as

$$\rho_x(r, s) = Corr(X_r, X_s) = \frac{\gamma_x(r, s)}{\sqrt{\gamma_x(r, r)\gamma_x(s, s)}}, r, s \in \mathcal{T}$$

Definition (Stationary S.P)**.** A stochastic process $X_t, t \in \mathcal{T}$ is said to be stationary

- $E(X_t^2) < \infty$ for all $t \in \mathcal{T}$
- $E(X_t) = \mu$ for all $t \in \mathcal{T}$
- $\gamma_x(r, s) = \gamma_x(r + h, s + h)$ for all $r, s, h \in \mathcal{T}$

Weakly Stationary/Covariance Stationary/Wide Sense Stationary/Second Order Stationary
ASK: If our $\mathcal{T}$ is a non convex set, does this still hold?
Also if X_t is stationary, then $\gamma_x(r, s) = \gamma_x(0, s - r) = \gamma_x(s - r)$ ie we can define the autocovariance as a fn of the one variable: the lag $h = s - r$
Similarly $\rho_x(h) = \frac{\gamma_x(h)}{\gamma_x(0)}$

Definition (Strict Stationarity)**.** A stochastic process $X_t, t \in \mathcal{T}$ is said to be strictly stationary if for every $n \in \mathbb{N}$, $t_1, t_2, \ldots, t_n \in \mathcal{T}$ and $h \in \mathcal{T}$ the random vectors $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})'$ and $(X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_n+h})'$ have the same distribution.
ie the finite dimensional distributions are shift invariant.
If Strict Stationarity with finite second moments $\implies$ Weak Stationarity.

Definition (Gaussian Time Series (S.P))**.** A Gaussian S.P. is a S.P. $X_t, t \in \mathcal{T}$ if all the finite dimensional distributions fns of $\{X_t\}$ are multivariate normal.
ie for every $n \in \mathbb{N}$ and $t_1, t_2, \ldots, t_n \in \mathcal{T}$ the random vector $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})'$ has a multivariate normal distribution. - IF a G.P. is stationary, then it is strictly stationary.

Definition (Stationarity of IID)**.** IID variables are strictly stationary.

Definition (White Noise)**.** A S.P. X_t is said to be white noise if can also be written as $WN(0, \sigma^2)$

- $E(X_t) = 0$ for all t
- $Var(X_t) = \sigma^2 < \infty$ for all t
- $Cov(X_t, X_s) = 0$ for all $t \neq s$

It is a weakly stationary S.P.

Example (Example of White Noise not Strictly Stationary)**.** Let X_t with $t = even$ be $N(0, 1)$ and X_t with $t = odd$ be $Rademacher(0, 1)$
Then X_t is white noise but not strictly stationary.

Example (1.3.1)**.** $X_t = A\cos(\Theta t) + B\sin(\Theta t)$ where $E(A) = E(B) = 0$, $Var(A) = Var(B) = 1$, $Cov(A, B) = 0$

- $E(X_t) = 0$
- $Var(X_t) = E(A^2\cos^2(\Theta t) + B^2\sin^2(\Theta t)) = \cos^2(\Theta t) + \sin^2(\Theta t) = 1$
- $Cov(X_t, X_s) = E(X_tX_s) = E[(A\cos(\Theta t) + B\sin(\Theta t))(A\cos(\Theta s) + B\sin(\Theta s))] = E[A^2]\cos(\Theta t)\cos(\Theta s) + E[B^2]\sin(\Theta t)\sin(\Theta s) = \cos(\Theta t)\cos(\Theta s) + \sin(\Theta t)\sin(\Theta s) = \cos(\Theta(t - s))$

Note that the $Cov(X_t, X_s)$ is only a fn of $t - s$
Thus X_t is weakly stationary.

Example (1.3.2)**.** Let $Z_t, t \in \mathbb{Z}$ be IID(0, σ^2)
$X_t = Z_t + \Theta Z_{t-1}$

- $E(X_t) = 0$
- $Var(X_t) = Var(Z_t) + \Theta^2 Var(Z_{t-1}) = (1 + \Theta^2)\sigma^2$
- $Cov(X_t, X_s) = E(X_tX_s) = E[(Z_t + \Theta Z_{t-1})(Z_s + \Theta Z_{s-1})] = \Theta\sigma^2$ if $|t - s| = 1$, $(1 + \Theta^2)\sigma^2$ if $t = s$, 0 otherwise

Thus X_t is weakly stationary.

Example (1.3.4)**.** Assume X_t is IID(0, σ^2)
$S_t = X_1 + X_2 + \ldots + X_t$ $t \geq 1$

- $E(S_t) = 0$
- $Var(S_t) = t\sigma^2$ Not constant

- $Cov(S_r, S_t) = E(S_r S_t) = r\sigma^2$ WLOG $r \leq t$
- $Cov(S_r, S_t) = (r \wedge t)\sigma^2$

Proposition 1 (1.5.1)**.** *Suppose X_t is weakly stationary with $\gamma_x(h), \rho_x(h)$ as the autocovariance and autocorrelation fns. Then*

- $\gamma_x(0) \geq 0$
- $|\gamma_x(h)| \leq \gamma_x(0)$ *for all* $h \in \mathcal{T}$
- $\gamma_x(h) = \gamma_x(-h)$ *for all* $h \in \mathcal{T}$

Remark (Some Statistics...)**.** Observe $\{X_t\}, t = 1, 2, \ldots, n$ Want to estimate $\mu, \gamma(0), \gamma(1), \ldots, \gamma(n-1)$

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i := \bar{X}$$

$$\hat{\gamma}(0) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\hat{\gamma}(1) = \frac{1}{n}\sum_{i=1}^{n-1}(X_i - \bar{X})(X_{i+1} - \bar{X})$$

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{i=1}^{n-h}(X_i - \bar{X})(X_{i+h} - \bar{X})$$

The reason why we divide by n we want to shrink it. intuition is that we want to make autocorrelation smaller as n increases.

1.3 9/11/2025 Lecture 3

Remark (Matrix Form of Autocovariance)**.** Observe $X_1, X_2, \ldots, X_n$
$\hat{\gamma}(h) = \frac{1}{n}\sum_{i=1}^{n-h}(X_i - \bar{X})(X_{i+h} - \bar{X})$.

$$\Gamma_n = \text{Cov}\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} \gamma_x(0) & \gamma_x(1) & \gamma_x(2) & \ldots & \gamma_x(n-1) \\ \gamma_x(1) & \gamma_x(0) & \gamma_x(1) & \ldots & \gamma_x(n-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_x(n-1) & \gamma_x(n-2) & \gamma_x(n-3) & \ldots & \gamma_x(0) \end{bmatrix}$$

This is a Toeplitz matrix. ie constant along the diagonals. It is also positive semidefinite. ie $a'\Gamma_n a \geq 0$ for all $a \in \mathbb{R}^n$.
For the Sample version, we have

$$\hat{\Gamma}_n = \begin{bmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \hat{\gamma}(2) & \ldots & \hat{\gamma}(n-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \hat{\gamma}(1) & \ldots & \hat{\gamma}(n-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}(n-1) & \hat{\gamma}(n-2) & \hat{\gamma}(n-3) & \ldots & \hat{\gamma}(0) \end{bmatrix}$$

We use n as a common denominator to ensure that $\hat{\Gamma}_n$ is positive semidefinite.
Γ_n is called the order-n autocovariance matrix of the process.
$\hat{\Gamma}_n$ is called the order-n sample autocovariance

Theorem 2. *A real valued fn defined on the integers is the autocovariance fn of a weakly stationary Time Series iff*

- *It is even. ie $\gamma(h) = \gamma(-h)$ for all $h \in \mathcal{T}$*
- *It is non-negative definite. ie for every $n \in \mathbb{N}$ and $a_1, a_2, \ldots, a_n \in \mathbb{R}$*

IE $\sum_{i,j}^n a_i k(t_i - t_j) a_j \geq 0$ for all $n \geq 1, \mathbf{t} = (t_1, t_2, \ldots, t_n)' \in \mathbb{N}^n$, and $\mathbf{a} = (a_1, a_2, \ldots, a_n)' \in \mathbb{R}^n$

Proof. **LOOK MORE INTO THIS THEOREM**
$\Longrightarrow$
It is straightforward to see that $\gamma_x(h)$ is even.
Let $n \in \mathbb{N}$, $\mathbf{t} = (t_1, t_2, \ldots, t_n)' \in \mathbb{N}^n$, and $\mathbf{a} = (a_1, a_2, \ldots, a_n)' \in \mathbb{R}^n$

$$\sum_{i,j}^n a_i \gamma_x(t_i - t_j) a_j = \sum_{i,j}^n a_i Cov(X_{t_i}, X_{t_j}) a_j = Cov(\sum_{i=1}^n a_i X_{t_i}, \sum_{j=1}^n a_j X_{t_j}) = Var(\sum_{i=1}^n a_i X_{t_i}) \geq 0$$

$\Longleftarrow$
Let $k(h)$ be a real valued fn defined on the integers which is even and non-negative definite.
Let $n \in \mathbb{N}$, $\mathbf{t} = (t_1, t_2, \ldots, t_n)' \in \mathbb{N}^n$, and $\mathbf{a} = (a_1, a_2, \ldots, a_n)' \in \mathbb{R}^n$
Define $\Gamma_n = [k(t_i - t_j)]_{i,j=1}^n$
Then Γ_n is a non-negative definite matrix. ie $a'\Gamma_n a \geq 0$ for all $a \in \mathbb{R}^n$.
Thus by the spectral theorem, there exists a random vector $\mathbf{X} = (X_{t_1}, X_{t_2}, \ldots, X_{t_n})'$ with mean 0 and covariance matrix Γ_n. ie $E(\mathbf{X}) = 0$ and $Cov(\mathbf{X}) = \Gamma_n$.
ie $Cov(X_{t_i}, X_{t_j}) = k(t_i - t_j)$ for all $1 \leq i, j \leq n$
By Kolmogorov's theorem, there exists a S.P. $X_t, t \in \mathbb{Z}$ with autocovariance fn $k(h)$.

$\square$

Example. Suppose $k(h) = \begin{cases} 1 & h = 0 \\ \rho & h = \pm 1 \\ 0 & otherwise \end{cases}$
When is k an autocovariance fn of a weakly stationary S.P.?

- $|\rho| \leq .5$ then

Remember Z_t is IID(0,σ^2), $X_t = Z_t + \Theta Z_{t-1}$ with acovf $\gamma_x(h) = \begin{cases} (1+\Theta^2)\sigma^2 & h = 0 \\ \Theta\sigma^2 & h = \pm 1 \\ 0 & otherwise \end{cases}$

$\rho(1) = \frac{\Theta}{1+\Theta^2}$ then $1 + \Theta^2 \leq 2\theta$ ie $|\rho| \leq .5$

- If $.5 < \rho \leq 1$ then $k(h)$ is not an acovf.

Then you can find a n s.t.

$$\sum_{i,j}^{2n} a_i a_j k(i-j) = 2n - 2(n-1)\rho < 0$$

Where does this formula on the RHS come from?

- If $-1 \leq \rho < -.5$ then $k(h)$ is not an acovf.

Definition (Mixing Conditions)**.** Suppose $\mathcal{G}$ and $\mathcal{H}$ are two sub σ-fields on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. $\mathcal{G} \subset \mathcal{F}$ and $\mathcal{H} \subset \mathcal{F}$.

Definition (α-mixing:)**.** α-mixing: $\alpha(\mathcal{G}, \mathcal{H}) = \sup_{G \in \mathcal{G}, H \in \mathcal{H}} \left|\mathbb{P}(G \cap H) - \mathbb{P}(G)\mathbb{P}(H)\right|$
X_1 and X_2 are independent $\mathcal{G} = \sigma(X_1) = \sigma\langle [X_1 \leq c], c \in \mathbb{R}\rangle$ and $\mathcal{H} = \sigma(X_2)$

- $\alpha(\mathcal{G}, \mathcal{H}) = 0 \; iff \; \mathcal{G}$ and $\mathcal{H}$ are independent
- $0 \leq \alpha(\mathcal{G}, \mathcal{H}) \leq .25$ for all $\mathcal{G}, \mathcal{H}$ sub σ-fields of $\mathcal{F}$
- $\mathbb{P}(G \cap H) - \mathbb{P}(G)\mathbb{P}(H) = \mathbb{E}[I_G I_H] - \mathbb{E}[I_G]\mathbb{E}[I_H] = Cov(I_G, I_H)$

$|Cov(c_1 I_G, c_2 I_H)| \leq c_1 c_2 \alpha(\mathcal{G}, \mathcal{H})$

Definition (ϕ-mixing:)**.** ϕ-mixing: $\phi(\mathcal{G}, \mathcal{H}) = \sup_{G \in \mathcal{G}, H \in \mathcal{H}, \mathbb{P}(G) > 0} |\mathbb{P}(H|G) - \mathbb{P}(H)|$

- $\phi(\mathcal{G}, \mathcal{H}) = 0 \; iff \; \mathcal{G}$ and $\mathcal{H}$ are independent
- $0 \leq \phi(\mathcal{G}, \mathcal{H}) \leq 1$ for all $\mathcal{G}, \mathcal{H}$ sub σ-fields of $\mathcal{F}$
- $\alpha(\mathcal{G}, \mathcal{H}) \leq \frac{1}{2}\phi(\mathcal{G}, \mathcal{H})$

Example. X is G-measureable and Y is H-measureable, $|X| \leq C_1$ and $|Y| \leq C_2$ a.s.
Then $|\text{Cov}(X, Y)| \leq 4C_1 C_2 \alpha(\mathcal{G}, \mathcal{H})$

1.4 9/16/2025 Lecture 4

Remark (Last Class Review)**.** Mixing Conditions:
Suppose $\mathcal{G}$ and $\mathcal{H}$ are two sub σ-fields on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. $\mathcal{G} \subset \mathcal{F}$ and $\mathcal{H} \subset \mathcal{F}$. LOOK INTO TEXTBOOK ASSIGNMENTS

- α-mixing: $\alpha(\mathcal{G}, \mathcal{H}) = \sup_{G \in \mathcal{G}, H \in \mathcal{H}} \left|\mathbb{P}(G \cap H) - \mathbb{P}(G)\mathbb{P}(H)\right|$
- ϕ-mixing: $\phi(\mathcal{G}, \mathcal{H}) = \sup_{G \in \mathcal{G}, H \in \mathcal{H}, \mathbb{P}(G) > 0} |\mathbb{P}(H|G) - \mathbb{P}(H)|$

1. $\alpha(\mathcal{G}, \mathcal{H}) = 0 \; iff \; \mathcal{G}$ and $\mathcal{H}$ are independent

2. $0 \le \alpha(\mathcal{G},\mathcal{H}) \le .25$ for all $\mathcal{G},\mathcal{H}$ sub σ-fields of $\mathcal{F}$, $0 \le \phi(\mathcal{G},\mathcal{H}) \le 1$ for all $\mathcal{G},\mathcal{H}$ sub σ-fields of $\mathcal{F}$

3. $\alpha(\mathcal{G},\mathcal{H}) \le \frac{1}{2}\phi(\mathcal{G},\mathcal{H})$

Equal definition: $\alpha(X,Y) = \sup_{c_1,c_2\in\mathbb{R}} |\mathbb{P}(X \le c_1, Y \le c_2) - \mathbb{P}(X \le c_1)\mathbb{P}(Y \le c_2)|$
Equal definition: $\phi(X,Y) = \sup_{c_1,c_2\in\mathbb{R},\mathbb{P}(X\le c_1)>0} |\mathbb{P}(Y \le c_2|X \le c_1) - \mathbb{P}(Y \le c_2)|$

Theorem 3 (Ibragimov 1962)**.** $\mathbb{P}(G\cap H) - \mathbb{P}(G)\mathbb{P}(H) = Cov(I_G, I_H)$
$|Cov(c_1 I_G, c_2 I_H)| \le c_1c_2\alpha(\mathcal{G},\mathcal{H})$
Sup. $|X| \le C_1$ *and* $|Y| \le C_2$ *a.s.*
Then $|E(XY) - E(X)E(Y)| \le 4c_1c_2\alpha(\mathcal{G},\mathcal{H})$

Proof.

$$\begin{aligned}
E(XY) - E(X)E(Y) &= E[X(Y - E(Y))] \\
&= E[X(E(Y|X) - E(Y))] \\
&= E[E(XY|X) - E(Y)] \\
|E(XY) - E(X)E(Y)| &= |E[X(E(Y|X) - E(Y))]| \\
&\le c_1E|E(Y|X) - E(Y)| \\
\text{Define } \eta = \text{sign}(E(Y|X) - E(Y)) & \\
&= c_1E[\eta(E(Y|X) - E(Y))] \\
\eta E(Y|X) &= E(\eta Y|X) \\
c_1E[E(\eta Y|X) - \eta E(Y)] &= c_1[E(\eta Y) - E(\eta)E(Y)] \\
E(\eta Y) - E(\eta)E(Y) &\le E[Y[E(\eta|Y) - E(\eta)]] \\
\text{Let } \xi = \text{sign}(E(\eta|Y) - E(\eta)) & \\
E(\eta Y) - E(\eta)E(Y) &\le c_2(E[\xi\eta] - E(\xi)E(\eta)) \\
E(XY) - E(X)E(Y) &\le c_1c_2(E[\xi\eta] - E(\xi)E(\eta)) \\
\eta = I_{\eta=1} - I_{\eta=-1}, &\xi = I_{\xi=1} - I_{\xi=-1} \\
\text{Cov}(\xi,\eta) &= \text{Cov}(I_{\xi=1} - I_{\xi=-1}, I_{\eta=1} - I_{\eta=-1}) \\
&= \text{Cov}(I_{\xi=1}, I_{\eta=1}) + \text{Cov}(I_{\xi=-1}, I_{\eta=-1}) \\
&\quad - \text{Cov}(I_{\xi=1}, I_{\eta=-1}) - \text{Cov}(I_{\xi=-1}, I_{\eta=1}) \\
\implies |\text{Cov}(\xi,\eta)| &\le 4\alpha(\mathcal{G},\mathcal{H}) \\
|E(XY) - E(X)E(Y)| &\le 4c_1c_2\alpha(\mathcal{G},\mathcal{H})
\end{aligned}$$

∎

Why are we doing this?
Consider $X_1, X_2, \ldots$ *IID(0,* σ^2*)*
$\frac{X_1+X_2+\ldots+X_n}{n} \xrightarrow{d} N(0,\sigma^2)$
Now how do we get CLT?

Consider $X_1, X_2, \dots$ *is a weakly stationary S.P, with* $E(X_t) = 0$
$\frac{X_1+X_2+\dots+X_n}{n} \xrightarrow{d} N(0,\sigma^2)$
We can see this is the variance $S_n = X_1 + X_2 + \dots + X_n$
$Var(S_n) = \sum_{i=1}^n Var(X_i) + 2\sum_{1\le i<j\le n} Cov(X_i, X_j)$
$= n\gamma_x(0) + 2\sum_{1\le i<j\le n}(\gamma_x(j-i))$
$= n\gamma_x(0) + 2\sum_{h=1}^{n-1}(n-h)\gamma_x(h)$
$Var(\frac{S_n}{\sqrt{n}}) = \gamma_x(0) + 2\sum_{h=1}^{n-1}(1-\frac{h}{n})\gamma_x(h)$
$\lim_{n\to\infty} Var(\frac{S_n}{\sqrt{n}}) = \gamma_x(0) + 2\sum_{h=1}^{\infty}\gamma_x(h)$
We want this infinite series to converge. ie $\sum_{h=1}^{\infty}|\gamma_x(h)| < \infty$.

Consider $X_1, X_2, \dots$ *is a strictly stationary S.P.*
Define $\alpha_0 = \frac{1}{2}$, *and* $\alpha_n = \alpha(X_0, X_n)$ *for* $n \ge 1$
Assume $E|X_0|^p < \infty$ *for some* $p > 2$
Then $|\gamma_x(k)| = |Cov(X_0, X_k)| \le 8||X_0||_p^2\alpha_k^{1-\frac{2}{p}}$

Corollary (Only Y is bounded)**.** Suppose $E[X^2] < \infty$ for some $p > 1$ and $|Y| \le C$ a.s.
Then $E(XY) - E(X)E(Y) \le 6C||X||_p[\alpha(X,Y)]^{1-\frac{1}{p}}$ where $||X||_p = (E|X|^p)^{\frac{1}{p}}$

Proof. Through Truncation:
$X_1 = XI_{|X|\le C_1}$ and $X_2 = X - X_1$

$$\begin{aligned}
|E(XY) - E(X)E(Y)| &\le |E(X_1Y) - E(X_1)E(Y)| + |E(X_2Y) - E(X_2)E(Y)| \\
&\le 4CC_1\alpha(X,Y) + 2CE|X_2| \\
E|X_2| &= E|XI_{|X|>C_1}| \le \frac{E|X|^p}{C_1^{p-1}} \\
I_{|X|>C_1} &< \frac{|X|^p}{C_1^{p-1}} \\
&= \frac{||X||_p^p}{C_1^{p-1}}
\end{aligned}$$

Thus $|E(XY) - E(X)E(Y)| \le 4CC_1\alpha(X,Y) + \frac{||X||_p^p}{C_1^{p-1}}$.
Take $C_1 = \alpha^{-\frac{1}{p}}||X||_p$ to get best bound.
Then the corollary follows.
$\boxed{\text{Look into bernstein inequality}}$ □

Corollary (No bounded (Davydov 1968))**.** Suppose $E|X|^p < \infty$ and $E|Y|^q < \infty$ for some $p, q > 1$ and $\frac{1}{p} + \frac{1}{q} < 1$ then
$|E(XY) - E(X)E(Y)| \le 8||X||_p||Y||_q[\alpha(X,Y)]^{1-\frac{1}{p}-\frac{1}{q}}$

Review of Hilbert Spaces

Definition (Inner Product Space)**.** A vector space $\mathcal{V}$ over the field $\mathbb{F}$ is called an inner product space if there exists a fn $\langle \cdot, \cdot \rangle$

- $\langle u, v \rangle = \overline{\langle v, u \rangle}$ for all $u, v \in \mathcal{V}$
- $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ for all $u, v, w \in \mathcal{V}$
- $\langle cu, v \rangle = c\langle u, v \rangle$ for all $u, v \in \mathcal{V}$ and $c \in \mathbb{F}$
- $\langle u, u \rangle \geq 0$ for all $u \in \mathcal{V}$
- $\langle u, u \rangle = 0$ iff $u = 0$

We will see that for the prob space $\langle X, Y \rangle = E[XY]$ but this only holds a.s.

1.5 9/18/2025 Lecture 5

Definition (Inner Product Space)**.** $\mathcal{H}$ is an inner product space with inner product $\langle \cdot, \cdot \rangle$

Example (2.2.2)**.** $L^2(\Omega, \mathcal{F}, \mathbb{P}) = \{X : \Omega \to \mathbb{R} | X$ is measurable and $E(X^2) < \infty\}$
$\langle X, Y \rangle = E(XY) = \int_\Omega X(\omega)Y(\omega)d\mathbb{P}(\omega)$
$\langle X, X \rangle = E(X^2) = 0 \implies X = 0$ a.s.
Define an equivalence relation $X \sim Y$ if $X = Y$ a.s.

- The elements of L^2 are equivalence classes
- $||X|| = \sqrt{\langle X, X \rangle} = \sqrt{E(X^2)}$ is a norm on L^2

Remark. IP Properties:

- $|\langle x, y \rangle| \leq ||x||||y||$ (Cauchy-Schwarz Inequality)
- $||x + y|| \leq ||x|| + ||y||$ (Triangle Inequality)
- If $||x_n - x|| \to 0$ and $||y_n - y|| \to 0$ then $\langle x_n, y_n \rangle \to \langle x, y \rangle$ (Continuity of Inner Product)

Definition (Limit of a Sequence in Hilbert Space)**.** Let $\mathcal{H}$ be a Hilbert Space and $\{x_n\}$ be a sequence in $\mathcal{H}$
We say that $x_n \to x$ if $||x_n - x|| \to 0$ as $n \to \infty$.

Suppose $\{X_n\}$ is a sequence of random variables in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ which converges to X. Then consider the RV 1 (constant)
Consider $\langle X_n, 1 \rangle \to \langle X, 1 \rangle$
ie $E(X_n) \to E(X)$

$X_n \to X$

$\langle X_n, X_n\rangle \to \langle X, X\rangle$ ie $E(X_n^2) \to E(X^2)$

$X_n \to X$, $Y_n \to Y$
$\langle X_n, Y_n\rangle \to \langle X, Y\rangle$
ie $E(X_nY_n) \to E(XY)$

Definition (Cauchy Sequence)**.** A sequence of elemnts $\{x_n\}$ in an inner product space $\mathcal{H}$ is called a Cauchy sequence if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $||x_n - x_m|| < \epsilon$ for all $n, m \geq N$.

Definition (Hilbert Space)**.** An inner product space $\mathcal{H}$ is called a Hilbert Space if every Cauchy sequence in $\mathcal{H}$ converges to an element in $\mathcal{H}$.

Example. Consider $\mathcal{M}(\Omega, \mathcal{F}, \mathbb{P}) = \{X : |X| \leq C, C > 0\}$
$\langle X, Y\rangle = E(XY)$
$X \sim N(0, 1)$
$X_n = XI_{|X|\leq n}$
$E|X - X_n|^2 = E[X^2 I_{|X|>n}] \to 0$ by DCT (Dominated Convergence Theorem)
So $X_n \to X$ in L^2 but $X \notin \mathcal{M}$
Thus $\mathcal{M}$ is not a Hilbert Space.

Definition (Complex Random Variable)**.** A complex random variable is a fn $Z : \Omega \to \mathbb{C}$ such that $Z = X + iY$ where X, Y are real random variables.

Definition (Closed Subspace)**.** A linear subspace of a Hilbert Space $\mathcal{H}$ is called a closed subspace if $\mathcal{M}$ contains its limit points. ie if $\{x_n\} \subset \mathcal{M}$ and $x_n \to x$ in $\mathcal{H}$ then $x \in \mathcal{M}$.

Proposition 2 (2.3.1)**.** $\boxed{\textit{Review the definition}}$ *If $\mathcal{M}$ is a closed subset of a H.S $\mathcal{H}$ then the orthogonal compliment $\mathcal{M}^\perp = \{x \in \mathcal{H} : x \perp y, \forall y \in \mathcal{M}\}$ closed linear subspace of $\mathcal{H}$.*

Theorem 4 (2.3.1 Projection Theorm)**.** *If $\mathcal{M}$ is a closed linear subspace of a H.S $\mathcal{H}$ and $x \in \mathcal{H}$ then*
(i) there is a unique element $\hat{x} \in \mathcal{M}$ such that $||x - \hat{x}|| = \inf_{y\in\mathcal{M}} ||x - y||$
(ii) $\hat{x} \in \mathcal{M}$ and $||x - \hat{x}|| = \inf_{y\in\mathcal{M}} ||x - y||$ iff $\hat{x} \in \mathcal{M}$ and $x - \hat{x} \in \mathcal{M}^\perp$

Definition (2.4.1 Closed Span)**.** The closed span $\overline{sp}\{X_t, t \in \mathcal{T}\}$ of any subset $\{X_t, t \in \mathcal{T}\}$ of a H.S $\mathcal{H}$ is the smallest closed linear subspace of $\mathcal{H}$ containing $\{X_t, t \in \mathcal{T}\}$.

Definition (Orthonormal Set)**.** A set $\{e_t : t \in \mathcal{T}\}$ of element of an IP space is said to be orthonormal if $\langle e_s, e_t\rangle = \begin{cases} 1 & s = t \\ 0 & s \neq t \end{cases}$ for all $s, t \in \mathcal{T}$

Definition (Complete Orthonormal Set)**.** An orthonormal set $\{e_t : t \in \mathcal{T}\}$ in a H.S $\mathcal{H}$ is said to be complete if $\overline{sp}\{e_t, t \in \mathcal{T}\} = \mathcal{H}$

Definition (Seperability)**.** The HS is separable if it has a finite or countable infinite complete orthonormal set.

Example (Separable HS)**.** 1. $\mathbb{R}^d$
2. $L^2(\Omega, \mathcal{F}, \mathbb{P})$

Theorem 5 (2.4.2)**.** *If $\mathcal{H}$ is a separable H.S and $\mathcal{H} = \overline{sp}\{e_t : t \in \mathcal{T}\}$ where $\{e_t : t \in \mathcal{T}\}$ is an orthonormal set then*

- *The set of all finite linear combinations of $\{e_t : t \in \mathcal{T}\}$ is dense in $\mathcal{H}$. ie for every $x \in \mathcal{H}$ and $\epsilon > 0$ there exists $y = \sum_{j=1}^{n} a_j e_{t_j}$ such that $||x - y|| < \epsilon$*
- *$x = \sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$ for each $x \in \mathcal{H}$ ie $||x - \sum_{i=1}^{n} \langle x, e_i \rangle e_i|| \to 0$ as $n \to \infty$*
- *$||x||^2 = \sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2$ for each $x \in \mathcal{H}$ (Parseval's Identity)*
- *$\langle x, y \rangle = \sum_{i=1}^{\infty} \langle x, e_i \rangle \langle y, e_i \rangle$ for all $x, y \in \mathcal{H}$*
- *$x = 0 \iff \langle x, e_i \rangle = 0$ for all $i \geq 1$*

1.6 9/23/2025 Lecture 6

Definition (ARMA models: ARMA(p, q))**.** Let $\{Z_t\} \sim WN(0, \sigma^2)$. The process $\{X_t, t \in \mathbb{Z}\}$ is said to be an $ARMA(p, q)$ process if

- $\{X_t\}$ is stationary for all $t \in \mathbb{Z}$
- $X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$ for all $t \in \mathbb{Z}$ where $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ are real constants with $\phi_p, \theta_q \neq 0$.

Remark. There are a few special cases of the $ARMA(p, q)$ model:

- When $q = 0$ we can write the model as $X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t$ and call it an $AR(p)$ model.
- When $p = 0$ we can write the model as $X_t = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$ and call it a $MA(q)$ model.
- When $p = 0$ and $q = 0$ we have $X_t = Z_t$ and call it a white noise model.
- $\{X_t\}$ is defined relative to the white noise process $\{Z_t\}$.
- Staitionarity is a critial requirement for the $ARMA(p, q)$ model.
- AR polynomial: $\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p$
- MA polynomial: $\theta(z) = 1 + \theta_1 z + \ldots + \theta_q z^q$
- Backshift operator: $BX_t = X_{t-1}$, $B^2 X_t = X_{t-2}$, $\ldots$, $B^k X_t = X_{t-k}$
- AR(p) model: $\phi(B) X_t = Z_t$
- MA(q) model: $X_t = \theta(B) Z_t$

- ARMA(p,q) model: $\phi(B)X_t = \theta(B)Z_t$
- More general model with a mean: $\{X_t + \mu : t \in \mathbb{Z}\}$
- Can also be characterized by $X_t = \phi_0 + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$ where $\phi_0 = \mu(1 - \phi_1 - \ldots - \phi_p)$

Example (Staitionary solution to AR(1))**.**

$$
\begin{aligned}
X_t &= \phi X_{t-1} + Z_t \\
&= Z_t + \phi(Z_{t-1} + \phi X_{t-2}) = Z_t + \phi Z_{t-1} + \phi^2 X_{t-2} \\
&= Z_t + \phi Z_{t-1} + \phi^2(Z_{t-2} + \phi X_{t-3}) = Z_t + \phi Z_{t-1} + \phi^2 Z_{t-2} + \phi^3 X_{t-3} \\
&\vdots \\
&= Z_t + \phi Z_{t-1} + \phi^2 Z_{t-2} + \ldots + \phi^k Z_{t-k} + \phi^{k+1} X_{t-(k+1)}
\end{aligned}
$$

If $|\phi| < 1$ then $\phi^{k+1} X_{t-(k+1)} \to 0$ as $k \to \infty$
Thus the stationary solution is $X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$
If $|\phi| \geq 1$ then there is no stationary solution since we can see that $X_{t+1} = \phi X_t + Z_{t+1} \iff X_t = -\frac{1}{\phi} Z_{t+1} + \frac{1}{\phi} X_{t+1}$

$$
\begin{aligned}
X_t &= \phi^{-1} X_{t+1} - \phi^{-1} Z_{t+1} \\
&= \phi^{-1}(\phi^{-1} X_{t+2} - \phi^{-1} Z_{t+2}) - \phi^{-1} Z_{t+1} \\
&= \phi^{-2} X_{t+2} - \phi^{-1} Z_{t+1} - \phi^{-2} Z_{t+2} \\
&\vdots \\
&= \phi^{-k} X_{t+k} - \sum_{j=1}^{k} \phi^{-j} Z_{t+j} \\
&= -\sum_{j=1}^{\infty} \phi^{-j} Z_{t+j}
\end{aligned}
$$

$\boxed{\text{We will see later that why this is the unique stationary solution when } |\phi| < 1}$

Remark. Uniqueness of stationary solution to AR(1):

- If $X_t = \phi X_{t-1} + Z_t$, where $|\phi| > 1$ then we can rewrite this as $X_t = \phi^{\star} X_{t-1} + Z_t^{\star}$ with $\phi^{\star} < 1$ and $Z_t^{\star} \sim WN(0, \sigma^2)$ $\boxed{\text{Homework problem}}$

Definition (3.1.3: Causality)**.** An ARMA(p,q) process $\phi(B)X_t = \theta(B)Z_t$ is said to be causal if ther exists a sequence of constants $\{\psi_j\}$ such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ for all $t \in \mathbb{Z}$.

Proposition 3 (3.1.1)**.** *If $\{X_t, t \in \mathbb{Z}\}$ is a sequence of rv st.* $\sup_t E|X_t| < \infty$ *and if* $\{\psi_j\}_{j\geq 0}$ *is a sequence of numbers s.t* $\sum_{j=0}^{\infty} |\psi_j| < \infty$ *then the series* $\psi(B)X_t = \left(\sum_{j=0}^{\infty} \psi_j B^j\right) X_t = \sum_{j=0}^{\infty} \psi_j X_{t-j}$ *converges absolutely w.p 1*
If in addition $\sup_t E(X_t^2) < \infty$ *then the series converges in* L^2 *to the same limit.*

Proof.

- Consider $\sum_{j=0}^{\infty} |\psi_j||X_{t-j}|$, which always exists (may be infinite)
- Monotone Convergence Theorem implies $E\left(\sum_{j=0}^{\infty} |\psi_j||X_{t-j}|\right) = \sum_{j=0}^{\infty} |\psi_j| E|X_{t-j}| \leq (\sup_t E|X_t|) \sum_{j=0}^{\infty} |\psi_j| < \infty \implies \sum_{j=0}^{\infty} |\psi_j||X_{t-j}| < \infty$ w.p 1
- $\implies \sum_{j=0}^{\infty} \psi_j X_{t-j}$ converges absolutely w.p 1, call the limit W_t.
- Verify $\sum_{j=0}^{n} \psi_j X_{t-j}$ is a Cauchy sequence in L^2: We do this by showing $|| \sum_{j=n}^{m} \psi_j X_{t-j}||_2 \to 0$ as $n, m \to \infty$.
- So it converges in L^2 to some limit S_t.
- $E(S_t - W_t)^2 = E[\liminf_n (S - \sum_{j=0}^{n} \psi_j X_{t-j})^2]$ by Fatou's Lemma
  $\leq \liminf_n E(S - \sum_{j=0}^{n} \psi_j X_{t-j})^2 = 0$
  $\implies S_t = W_t$ a.s. since the second moment is 0.

□

1.7 9/25/2025 Lecture 7

Remark (Review)**.** Review of last week:

- ARMA(p, q) process: $\phi(B)X_t = \theta(B)Z_t$ where $\{Z_t\} \sim WN(0, \sigma^2)$
- MA(q) process: $X_t = \theta(B)Z_t$

Proposition 4 (3.1.2)**.** *If* $\{X_t\}$ *is a staitionary process with autocovariance function* $\gamma_x(\cdot)$ *and if* $\{\psi_j\}_{j\geq 0}$, $\sum_{j=0} |\psi_j| < \infty$, *define* $Y_t = \sum_{j=0}^{\infty} \psi_j X_{t-j}$ *(converges absolutely, w.p 1). Then* Y_t *is also staitionary with autocovariance function* $\gamma_y(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_x(h + j - k)$ *where* $\psi_j = 0$ *for* $j < 0$.

Proof. We need to show that $E(Y_t)$ is constant and $\text{Cov}(Y_{t+h}, Y_t)$ depends only on h.

$$
\begin{aligned}
E(Y_t) &= E\left(\sum_{j=0}^{\infty}\psi_j X_{t-j}\right) = \sum_{j=0}^{\infty}\psi_j E(X_{t-j}) = \mu_x \sum_{j=0}^{\infty}\psi_j \text{ (constant)} \\
\operatorname{Cov}(Y_{t+h}, Y_t) &= E[(Y_{t+h} - E(Y_{t+h}))(Y_t - E(Y_t))] \\
&= E\left[\left(\sum_{j=0}^{\infty}\psi_j (X_{t+h-j} - \mu_x)\right)\left(\sum_{k=0}^{\infty}\psi_k (X_{t-k} - \mu_x)\right)\right] \\
&= E\left[\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\psi_j\psi_k (X_{t+h-j} - \mu_x)(X_{t-k} - \mu_x)\right] \\
&= \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\psi_j\psi_k E[(X_{t+h-j} - \mu_x)(X_{t-k} - \mu_x)] \\
&= \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\psi_j\psi_k \gamma_x(h+j-k) \\
&= \sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\psi_j\psi_k \gamma_x(h+j-k) \text{ where } \psi_j = 0 \text{ for } j < 0
\end{aligned}
$$

□

Remark. Let $\alpha(B) = \sum_{j=0}^{\infty}\alpha_j B^j$ and $\beta(B) = \sum_{j=0}^{\infty}\beta_j B^j$ $\sum_{j=0}^{\infty}|\alpha_j| < \infty$ $\sum_{j=0}^{\infty}|\beta_j| < \infty$. Then the product $\psi(B) = \alpha(B)\beta(B) = \sum_{j=0}^{\infty}\psi_j B^j$ then $\sum_{j=0}^{\infty}|\psi_j| < \infty$

Theorem 6 (3.1.1.a). *If $\phi(z)$ and $\theta(z)$ have no common zeros, if $phi(z) \neq 0$ for $|z| = 1$ and if $\{Z_t\} \sim WN(0, \sigma^2)$ then exists a unique stationary solution given by*

$$
X_t = \sum_{j=0}^{\infty}\psi_j Z_{t-j} \text{ where } \psi(z) = \frac{\theta(z)}{\phi(z)} = \sum_{j=0}^{\infty}\psi_j z^j
$$

and $\sum_{j=0}^{\infty}|\psi_j| < \infty$. so that X_t is well-defined and causal.

Proof. (i) Find Solution
If $\phi(z) \neq 0$ for $|z| = 1$ then $\exists\epsilon > 0$ such that

$$
\begin{aligned}
\frac{1}{\phi(z)} &:= \sum_{j=0}^{\infty}\zeta_j z^j =: \zeta(z), |z| \leq 1 + \epsilon \\
&\implies |\zeta_j| \leq (1 + \epsilon/2)^{-j} \text{ for some } K > 0
\end{aligned}
$$

Consider $\frac{1}{1-z} = \sum_{j=0}^{\infty} z^j$ for $|z| < 1$
Consider $\frac{1}{1-0.5z} = \sum_{j=0}^{\infty}(0.5z)^j$ for $|z| < 2$
$\phi(z) = \prod_{j=1}^{p}(1 - w_j z)$, ie each of the roots are $\frac{1}{w_j}$.
Then $\frac{1}{\phi(z)} = \prod_{j=1}^{p}\frac{1}{1-w_j z}$

$\implies \frac{1}{\phi(z)} = \sum_{j=0}^{\infty} \zeta_j z^j$ for $|z| < \min_{1 \le j \le p} |w_j|^{-1}$
We know that $\forall j, |w_j| < 1$ and then if we take $\epsilon = \min_{1 \le j \le p} |w_j|^{-1} - 1 > 0$ then we are done.
(ii) Find Stationary Solution
Define $X_t = \frac{\theta(B)}{\phi(B)} Z_t$ which is staitionary
$\phi(B)X_t = \theta(B)Z_t$
(iii) Uniqueness of Stationary Solution
Suppose $\{W_t\}$ is another stationary solution to $\phi(B)W_t = \theta(B)Z_t$

$$\begin{aligned} \phi(B)W_t &= \theta(B)Z_t \\ \zeta(B[\phi(B)W_t]) &= \zeta(B[\theta(B)Z_t]) \\ \implies W_t &= \zeta(B)[\theta(B)Z_t] = \frac{\theta(B)}{\phi(B)} Z_t = X_t \end{aligned}$$

□

Theorem 7 (3.1.1.b). *Assume $\phi(z)$ and $\theta(z)$ have no common zeros. If there exists a stationary solution which is also causal then $\phi(z) \neq 0$ for $|z| \le 1$.*

1.8 9/30/2025 Lecture 8

Remark (Review). Prior class review:

- ARMA(p, q) process: $\phi(B)X_t = \theta(B)Z_t$ where $\{Z_t\} \sim WN(0, \sigma^2)$

$\phi(z)$ and $\theta(z)$ have no common zeros.

Theorem 8 (3.1.1.a & .b). *(a) If $\phi(z) \neq 0$ for all $|z| \le 1$ then there exists a unique stationary solution given by*

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \text{ where } \psi(z) = \frac{\theta(z)}{\phi(z)} = \sum_{j=0}^{\infty} \psi_j z^j$$

and they satisfy $\sum_{j=0}^{\infty} |\psi_j| < \infty$.
(b) If there exists a stationary solution which is also causal then $\phi(z) \neq 0$ for all $|z| \le 1$.

Remark. Not proving

- If $\phi(z) \neq 0$ for all $|z| = 1$ then there a unique stationary solution.
- If $\phi(z) = 0$ for some $|z| = 1$ then there is no stationary solution.
- If $\phi(z) \neq 0$ for all $|z| = 1$ and $\{X_t\}$ is the unique staitionary solution then one can find $p\hat{h}i(z)$ and $WN\{Z_t^*\}$ st $\hat{\phi}(z)X_t = \phi(B)Z_t^*$ and $\hat{\phi}(z) \neq 0$ for all $|z| \le 1$.
- Only Focus on Causal and Invertable ARMA models

Definition (3.1.4)**.** Suppose $\{X_t\}$ is a staitionary solution of $\phi(B)X_t = \theta(B)Z_t$, it is said to be invertible if $\exists \pi_j$ such that $\sum_{j=0}^{\infty} |\pi_j| < \infty$ and $Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$ for all $t \in \mathbb{Z}$.

Theorem 9 (3.1.2)**.** *Suppose X_t is the unique stationary solution of $\phi(B)X_t = \theta(B)Z_t$, then it is invertible iff $\theta(z) \neq 0$ for all $|z| \leq 1$.*
When the condition holds $\{\pi_j\}$ are determined by $\pi(z) = \frac{\phi(z)}{\theta(z)} = \sum_{j=0}^{\infty} \pi_j z^j$.

Remark. IF the definition of invertability is relaxted to:

$$Z_t \in \overline{sp}\{X_t, X_{t-1}, \ldots\}$$

then the condition relaxted to $\theta(z) \neq 0$ for all $|z| < 1$

Definition (3.2.1)**.** Suppose $\{Z_t\} \sim WN(0, \sigma^2)$, we say $\{X_t\}$ is an infinite order moving average denoted by $MA(\infty)$ if

$$\exists \{\psi_j\} \text{ such that } \sum_{j=0}^{\infty} |\psi_j| < \infty \text{ and } X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

May relax condition to $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ then take X_t as the L^2 limit.
Sometimes MA(∞) is called the linear process.
This is related to the Wold Decomposition Theorem.

Proposition 5 (3.2.1)**.** *If $\{X_t\}$ is a zero-mean stationary process with autocovariance function $\gamma_x(\cdot)$ such that $\gamma_x(h) = 0$ for $|h| > q$ and $\gamma_x(q) \neq 0$ then $\{X_t\}$ is an $MA(q)$ process.*
IE: $\exists WN\{Z_t\}$ s.t. $X_t = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$ where $\theta_q \neq 0$.

Proof.
- Find the WN $\{Z_t\}$
- Show that $X_t = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$ for some $\theta_1, \ldots, \theta_q$ with $\theta_q \neq 0$

□

Definition (Linear Predictor)**.** Suppose $Y \in \mathbb{R}$, $E[Y] = 0$, $\mathbf{X} \in \mathbb{R}^d$, $E[\mathbf{X}] = \mathbf{0}$.

$$\text{Cov}(\begin{bmatrix} Y \\ \mathbf{X} \end{bmatrix}) = \begin{bmatrix} \sigma_Y^2 & \sigma_{\mathbf{YX}}' \\ \sigma_{\mathbf{YX}} & \Sigma_X \end{bmatrix}$$

A linear predictor takes the form $C^T X$ where $C \in \mathbb{R}^d$.
The best linear predictor (BLP) of Y based on $\mathbf{X}$ is the linear predictor $\hat{Y} = C^T \mathbf{X}$ that minimizes the mean squared error $\min_{C \in \mathbb{R}^d} E[(Y - C^T \mathbf{X})^2]$.

$$\begin{aligned} E[(Y - C^T \mathbf{X})^2] &= E[Y^2] - 2C^T E[Y\mathbf{X}] + C^T E[\mathbf{X}\mathbf{X}^T] C \\ &= \sigma_Y^2 - 2C^T \sigma_{\mathbf{YX}} + C^T \Sigma_X C \end{aligned}$$

The best solution is given taking the partial derivative and setting it to 0:

$$\begin{aligned} \frac{\partial}{\partial C} E[(Y - C^T \mathbf{X})^2] &= -2\sigma_{\mathbf{YX}} + 2\Sigma_X C = 0 \\ &\implies \hat{C} = \Sigma_X^{-1} \sigma_{\mathbf{YX}} \\ &\implies \hat{Y} = \hat{C}^T \mathbf{X} = \sigma_{\mathbf{YX}}' \Sigma_X^{-1} \mathbf{X} \end{aligned}$$

$$E[(Y - \hat{Y})^2] = \sigma_Y^2 - \sigma_{\mathbf{YX}}' \Sigma_X^{-1} \sigma_{\mathbf{YX}}$$

Remark. $\{X_t\}$ is a mean-zero stationary process.
Want to predict X_{k+1} based on $\{X_1, \dots, X_k\}$.

$$\min_{\phi_1,\dots,\phi_k} E[(X_{k+1} - \hat{X}_{k+1})^2]$$

Where $\hat{X}_{k+1} = \sum_{j=1}^{k} \phi_j X_{k+1-j}$

$$Gamma_{k+1} = \mathrm{Cov}(\begin{bmatrix} X_{k+1} \\ X_k \\ \vdots \\ X_1 \end{bmatrix}) = \begin{bmatrix} \gamma(0) & \gamma(\mathbf{k})' \\ \gamma(\mathbf{k}) & \Gamma_k \end{bmatrix}$$

Where $\mathbf{gamma}(\mathbf{k}) = [\gamma(1), \dots, \gamma(k)]'$ and $\Gamma_k = [\gamma(i-j)]_{i,j=1}^{k}$

$$\begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_k \end{bmatrix} = \Gamma_k^{-1} \gamma(\mathbf{k})$$