16:960:665 - Syllabus

Pranav Tikkawar

December 9, 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

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*

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

"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
$\boxed{\textbf{ASK: If our } \mathcal{T} \textbf{ is a non convex set, does this still hold?}}$
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 $E(A) = E(B) = 0$, $Var(A) = Var(B) = 1$, $Cov(A,B) = 0$

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

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)
$X_t = Z_t + \Theta Z_{t-1}$

- $E(X_t) = 0$
- $Var(X_t) = Var(Z_t) + \Theta^2 Var(Z_{t-1}) = (1+\Theta^2)\sigma^2$
- $Cov(X_t, X_s) = E(X_t X_s) = E[(Z_t + \Theta Z_{t-1})(Z_s + \Theta Z_{s-1})] = \Theta\sigma^2$ 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
- $\mathrm{Cov}(S_r, S_t) = E(S_r S_t) = r\sigma^2$ 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 $k(h)$. □

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.

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

$\boxed{\textbf{Where does this formula on the RHS come from?}}$

- 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 $|\operatorname{Cov}(X, Y)| \leq 4C_1 C_2 \alpha(\mathcal{G}, \mathcal{H})$

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 $\mathcal{H} \subset \mathcal{F}$. $\boxed{\text{LOOK INTO TEXTBOOK ASSIGNMENTS}}$

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

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

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

Proof.

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

□

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

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* $n \geq 1$
Assume $E|X_0|^p < \infty$ *for some* $p > 2$
Then $|\gamma_x(k)| = |\operatorname{Cov}(X_0, X_k)| \leq 8||X_0||_p^2 \alpha_k^{1-\frac{2}{p}}$

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

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

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

Thus $|E(XY) - E(X)E(Y)| \leq 4CC_1\alpha(X,Y) + \frac{||X||_p^p}{C_1^{p-1}}$.
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
$|E(XY) - E(X)E(Y)| \leq 8||X||_p||Y||_q[\alpha(X,Y)]^{1-\frac{1}{p}-\frac{1}{q}}$

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)**.** $L^2(\Omega, \mathcal{F}, \mathbb{P}) = \{X : \Omega \to \mathbb{R} | X \text{ is measurable and } E(X^2) < \infty\}$
$\langle X, Y \rangle = E(XY) = \int_\Omega X(\omega)Y(\omega)d\mathbb{P}(\omega)$
$\langle X, X \rangle = E(X^2) = 0 \implies X = 0$ 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: $BX_t = X_{t-1}$, $B^2 X_t = X_{t-2}$, …, $B^k X_t = X_{t-k}$
- 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 $\hat{phi}(z)$ 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 $\{X_1, \ldots, X_k\}$.

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

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

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

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

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

1.9 10/2/2025 Lecture 9

Proposition 6 (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.*

Proof. Need to show

- Find $WN\{Z_t\}$
- 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 $\theta_q \neq 0$

Linear prediction problem: Predict X_{k+1} using $X_1, \ldots, X_k$

$$\underline{\hat{\phi}_k} = \arg \min_{\phi_1,\ldots,\phi_k} E[(X_{k+1} - \hat{X}_{k+1})^2]$$

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

$\underline{\hat{\phi}_k} = \Gamma_k^{-1} \underline{\gamma(k)}$ where $\underline{\gamma(k)} = [\gamma(1), \ldots, \gamma(k)]'$
$E[(X_{k+1} - \underline{\hat{\phi}_k}' \underline{X_k})^2] = \gamma(0) - \underline{\hat{\gamma(k)}}' \Gamma_k^{-1} \underline{\gamma(k)} = \nu_k$
Note that if we want

$$X_t = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$$

$$Z_t = \sum_{j=0}^{\infty} \psi_j X_{t-j}$$

$$Z_{t-1}, \ldots, Z_{t-q} \in \overline{sp}\{X_{t-1}, X_{t-2}, \ldots\}$$

Proof:
Define $\mathcal{M}_t = \overline{sp}\{X_t, X_{t-1}, \ldots\}$ and $Z_t = X_t - \mathcal{P}_{\mathcal{M}_{t-1}}(X_t)$ Notice that $Z_t \in \mathcal{M}_t$ and $Z_t \perp \mathcal{M}_{t-1}$

- $E[Z_s Z_t] = 0$ if $s \neq t$ $s > t$.

$$
\begin{aligned}
Z_t &\in \mathcal{M}_t \subset \mathcal{M}_{s-1} \\
Z_s &:= X_s - \mathcal{P}_{\mathcal{M}_{s-1}}(X_s) \\
Z_s \perp \mathcal{M}_{s-1} &\implies Z_s \perp Z_t \\
X_t &= Z_t + \phi_1 Z_{t-1} + \ldots + \phi_q Z_{t-q} \text{ for some } \phi_1, \ldots, \phi_q \\
Z_t &= \sum_{j=0}^{\infty} \psi_j X_{t-j} \text{ for some } \psi_j \\
Z_{t-1}, \ldots, Z_{t-q} &\in \overline{sp}\{X_{t-1}, X_{t-2}, \ldots\}
\end{aligned}
$$

  and $\mathcal{P}_{\overline{sp}\{X_{t-1},X_{t-2},\ldots\}}(X_t) \to \mathcal{P}_{\mathcal{M}_t} X_t$ in L^2 as $k \to \infty$.

- $||Z_t|| = \lim_{n\to\infty} ||X_t - \mathcal{P}_{\overline{sp}\{X_{t-1},\ldots,X_{t-n}\}}(X_t)||$
  $= \lim_{n\to\infty} E[(X_{t-1} - \mathcal{P}_{\overline{sp}\{X_{t-2},\ldots,X_{t-n-1}\}}(X_t))] = ||Z_{t-1}||$
  Denote $\sigma^2 = ||Z_t||^2$ then $\{Z_t\} \sim WN(0, \sigma^2)$

- $\mathcal{M}_{t-1} = \overline{sp}\{X_{t-1}, X_{t-2}, \ldots\} = \overline{sp}\{Z_{t-1}, \ldots Z_{t-q}, \ldots X_{t-q-1}, X_{t-q-2}, \ldots\}$
  Also $\overline{sp}\{Z_{t-1}, \ldots, Z_{t-q}\} \perp \overline{sp}\{X_{t-q-1}, X_{t-q-2}, \ldots\}$
  So $\mathcal{M}_{t-1} = \overline{sp}\{Z_{t-1}, \ldots, Z_{t-q}\} \oplus \mathcal{M}_{t-q-1}$

$$
\begin{aligned}
\mathcal{P}_{\mathcal{M}_{t-1}}(X_t) &= \mathcal{P}_{\overline{sp}\{Z_{t-1},\ldots,Z_{t-q}\}}(X_t) + \mathcal{P}_{\mathcal{M}_{t-q-1}}(X_t) \\
&= \frac{\langle X_t, Z_{t-1}\rangle}{\langle Z_{t-1}, Z_{t-1}\rangle} Z_{t-1} + \ldots + \frac{\langle X_t, Z_{t-q}\rangle}{\langle Z_{t-q}, Z_{t-q}\rangle} Z_{t-q} + \mathcal{P}_{\mathcal{M}_{t-q-1}}(X_t)
\end{aligned}
$$

  Note that $\gamma_x(h) = 0$ for $|h| > q \implies \mathcal{P}_{\mathcal{M}_{t-q-1}}(X_t) = 0$

$$
\implies X_t = Z_t + \sum_{j=1}^{q} \frac{\langle X_t, Z_{t-j}\rangle}{\sigma^2} Z_{t-j}
$$

$\square$

Proposition 7 (5.2.1 Durbin Levinson Algorithm)**.** *Suppose $\{X_t\}$ is a zero-mean stationary process with autocovariance function $\gamma(\cdot)$. Such that $\gamma(0) > 0$ and $\gamma(h) \to 0$ as $h \to \infty$.*
Then $\phi_{11} = \rho(1)$ and $\nu_1 = \gamma(0)(1 - \rho(1)^2)$
$\phi_{nn} = [\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j}\gamma(n-j)]\nu_{n-1}^{-1}$ for $n \geq 2$

$$\begin{bmatrix} \phi_{n1} \\ \phi_{n2} \\ \vdots \\ \phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \phi_{n-1,1} \\ \phi_{n-1,2} \\ \vdots \\ \phi_{n-1,n-1} \end{bmatrix} - \phi_{nn} \begin{bmatrix} \phi_{n-1,n-1} \\ \phi_{n-1,n-2} \\ \vdots \\ \phi_{n-1,1} \end{bmatrix} \textit{ for } n \geq 2$$

$\nu_n = \nu_{n-1}(1 - \phi_{nn}^2)$
Given ϕ_{n-1} and ν_{n-1} How do we get ϕ_n and ν_n?
The complexity is $O(n^2)$ instead of $O(n^3)$.

Proof. (1) $\phi_{11} = \rho(1)$ and $\nu_1 = \gamma(0)[1 - \rho(1)^2]$
Suppose we have $\underline{\phi_{n-1}}$ and ν_{n-1}, we want to find $\underline{\phi_n}$ and ν_n.
Want: $\mathcal{P}_{\overline{sp}\{X_1,\ldots,X_n\}}(X_{n+1})$

Let $w_1 = X_1 - \mathcal{P}_{\overline{sp}\{X_2,\ldots,X_n\}}(X_1)$ then $\begin{cases} w_1 \perp \overline{sp}\{X_n, \ldots, X_{n-2}\} \\ \overline{sp}\{X_1, \ldots, X_n\} = \overline{sp}\{w_1\} \oplus \overline{sp}\{X_2, \ldots, X_n\} \end{cases}$

$\mathcal{P}_{\overline{sp}\{X_n \ldots X_2\}}(X_{n+1}) = \underline{\phi_{n-1}}' \begin{bmatrix} X_n \\ X_{n-1} \\ \vdots \\ X_2 \end{bmatrix}$ Then want $\frac{\langle X_{n+1}, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1$

$\mathcal{P}_{\overline{sp}\{X_n,\ldots,X_{n-2}\}}(X_1) = \underline{\phi_{n-1}}' \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$

$w_1 = X_1 - \underline{\phi_{n-1}}' \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \quad ||w_1||^2 = \nu_{n-1} \quad \langle X_{n+1}, w_1 \rangle = \gamma(n) - \underline{\phi_{n-1}}' \begin{bmatrix} \gamma(n-1) \\ \gamma(n-2) \\ \vdots \\ \gamma(1) \end{bmatrix}$

$\mathcal{P}_{\overline{sp}_{w_1}}(X_{n+1}) = \frac{\langle X_{n+1}, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 = \phi_{nn} w_1$ $\underline{\phi_n}' \begin{bmatrix} X_n \\ X_{n-1} \\ \vdots \\ X_1 \end{bmatrix} - \phi_{nn} \begin{bmatrix} \phi_{n-1,n-1} \\ \phi_{n-1,n-2} \\ \vdots \\ \phi_{n-1,1} \end{bmatrix}$ $\nu_n = \nu_{n-1} - \phi_{nn}^2 \nu_{n-1} = (1 - \phi_{nn}^2)\nu_{n-1}$ □

$\boxed{\textit{PLEASE REVIEW THIS WHAT THE HELL IS THIS}}$

1.10 10/7/2025 Lecture 10

Remark. Review of last class:

- Linear Predictor: Suppose $Y \in \mathbb{R}$, $E[Y] = 0$, $\mathbf{X} \in \mathbb{R}^d$, $E[\mathbf{X}] = \mathbf{0}$.

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

The BLP is given by $\hat{X} = \phi_n' X_n$ where $\phi_n = \begin{bmatrix} \phi_{n1} \\ \phi_{n2} \\ \vdots \\ \phi_{nn} \end{bmatrix}$ and $X_n = \begin{bmatrix} X_n \\ X_{n-1} \\ \vdots \\ X_1 \end{bmatrix}$

Suppose $\{X_t\}$ follows causal AR(p) process: $\phi(B)X_t = Z_t$ where $\{Z_t\} \sim WN(0, \sigma^2)$

Predict X_t based on $X_{t-1}, \ldots, X_{t-p}$ Then use $\phi_p = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}$

Definition (Partial Autocorrelation Function)**.** LLet $\{X_t\}$ be a mean-zero stationary process. Its partial autocorrelation function (PACF) $\alpha(\cdot)$ is a function on positive integers defined as follows:

- $\alpha(1) = \rho(1)$
- $\alpha(k) = \text{Cor}[X_{k+1} - \mathcal{P}_{\overline{sp}\{X_k,\ldots,X_2\}}(X_{k+1}), X_1 - \mathcal{P}_{\overline{sp}\{X_2,\ldots,X_k\}}(X_1)]$ for $k \geq 2$

Example (3.4.1)**.** AR(1) process: $X_t = \phi X_{t-1} + Z_t$

$$\begin{aligned} \alpha(1) &= \rho(1) = \phi \\ \alpha(2) &= \text{Cor}[X_3 - \mathcal{P}_{\overline{sp}\{X_2\}}(X_3), X_1 - \mathcal{P}_{\overline{sp}\{X_2\}}(X_1)] \\ &= \text{Cor}[X_3 - \phi X_2, X_1 - \phi X_2] \\ &= \text{Cor}[Z_3, X_1 - \phi X_2] = 0 \\ \alpha(k) &= 0 \text{ for } k > 1 \end{aligned}$$

Example (3.4.2)**.** $MA(1)$ process: $X_t = Z_t + \theta Z_{t-1}$ where $\{Z_t\} \sim WN(0, \sigma^2)$

$$\begin{aligned} \alpha(1) &= \rho(1) = \frac{\theta}{1+\theta^2} \\ \alpha(2) &= \text{Cor}[X_3 - \mathcal{P}_{\overline{sp}\{X_2\}}(X_3), X_1 - \mathcal{P}_{\overline{sp}\{X_2\}}(X_1)] \\ &= \text{Cor}[X_3 - \rho(1)X_2, X_1 - \rho(1)X_2] \\ &= \frac{-\theta^2}{1+\theta^2+\theta^4} \end{aligned}$$

Definition (Partial Least Squares)**.** Model 1: $y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \epsilon_i$ for $i = 1, \ldots, n$
Model 2: $y_i = \beta_0 + \sum_{j=1}^{k} \beta_j z_{ij} + \epsilon_i$ for $i = 1, \ldots, n$
PLS procedure: Set LSE $\hat{\beta}_{(p-1)}$ based on Model 2

- regress $\underline{x_p}$ on $\underline{x_1}, \ldots, \underline{x_{p-1}}$ set the LSE to $\hat{\gamma}$ ie $\underline{x_p} = \sum_{j=1}^{p-1} \gamma_j \underline{x_j} + \underline{z_p}$
- Regress $\underline{y}$ on $\underline{z_p}$, minimize $||\underline{y} - c\underline{z}_{p-1}||$ $\hat{c} = \frac{\langle y, z_p \rangle}{\langle z_p, z_p \rangle}$
- get the LSE $\hat{\beta}_p$ BAsed on Model 1

- See that $\hat{c} = \hat{\beta}_p$

Remark. Prediction probkem 1:
Predict X_k using $X_2, \ldots, X_{k-1}$
Prediction problem 2:
Predict X_k using $X_{k-1}, \ldots, X_2, X_1$
Question how much does X_1 help in predicting X_k given $X_2, \ldots, X_{k-1}$
use $\frac{\langle X_1 - \mathcal{P}_{\overline{sp}\{X_2,\ldots,X_{k-1}\}}(X_1), X_{k+1} - \mathcal{P}_{\overline{sp}\{X_2,\ldots,X_{k-1}\}}(X_{k+1})\rangle}{||X_1 - \mathcal{P}_{\overline{sp}\{X_2,\ldots,X_{k-1}\}}(X_1)||^2}$ $w_1 = X_1 - \mathcal{P}_{\overline{sp}\{X_2,\ldots,X_{k-1}\}}(X_1)$
$w_1 = X_1 - \phi_{k-1,1}X_2 - \ldots - \phi_{k-1,k-1}X_k$

$$\begin{aligned}
\mathcal{P}_{\overline{sp}\{X_k \ldots X_1\}}(X_{k+1}) &= \mathcal{P}_{\overline{sp}\{X_k \ldots X_2\}}(X_k) + \mathcal{P}_{\overline{sp}\{w_1\}}(X_{k+1}) \text{ where } w_1 = X_1 - \mathcal{P}_{\overline{sp}\{X_2,\ldots,X_{k-1}\}}(X_1) \\
&= \phi_{k-1,1}X_k + \ldots + \phi_{k-1,k-1}X_2 + \hat{c}\left(X_1 - \phi_{k-1,1}X_2 - \ldots - \phi_{k-1,k-1}X_k\right) \\
&= \hat{c}X_1 + \ldots & = \phi_{k,k}X_1 + \ldots
\end{aligned}$$

Thus $\alpha(k) = \phi_{kk}$

Example. AR(p) process: $\phi(B)X_t = Z_t$

$$\begin{aligned}
\alpha(1) &= \rho(1) = \phi_1 \\
\alpha(p) &= \phi_p \\
\alpha(k) &= 0 \text{ for } k > p
\end{aligned}$$

Table 1: Theoretical behaviour of the ACF and PACF for common linear time series models

Model	ACF behaviour	PACF behaviour	Identification rule (sample)
AR(p)	Tails off (exponential or damped sinusoid depending on roots).	Cuts off after lag p (i.e. $\alpha(k) \approx 0$ for $k > p$).	If sample ACF decays and sample PACF shows a clear cutoff at lag p, prefer AR(p).
MA(q)	Cuts off after lag q (theoretical autocorrelations $\rho_k = 0$ for $k > q$).	Tails off (no finite cutoff).	If sample ACF has significant spikes up to lag q then ≈ 0 afterwards, prefer MA(q).
ARMA(p, q)	Tails off (mixture-shaped decay from both AR and MA parts).	Tails off (mixture-shaped).	If both sample ACF and PACF tail off (no short cutoff), try ARMA(p, q) and select (p, q) by AIC/BIC.

1.11 10/9/2025 Lecture 11

Remark (Estimating The PACF)**.** Assume $\{X_t\}$ is a mean-zero stationary process .
Predict X_{k+1} based on $X_k, \ldots, X_1$
Then the coeeficient of X_1 is $\alpha(k)$
Method 1 $\underline{\phi}_k = \Gamma_k^{-1} \underline{\gamma}_{k+1}$ where $\underline{\gamma}_k = [\gamma(1), \ldots, \gamma(k)]'$ and $\Gamma_k = [\gamma(i-j)]_{i,j=1}^k$
Plug in the sample autocovariance function $\hat{\gamma}(h)$ to get an estimated $p\hat{h}i_k$ and take $\hat{\alpha}(k) = \hat{\phi}_{kk}$
Method 2 $\min_{\phi_1,\ldots,\phi_k} \sum_{t=k+1}^n (X_t - \mathcal{P}_{\overline{sp}\{X_{k+1},\ldots,X_1\}}(X_t))^2$
$= \sum_{t=k+1}^n (X_t - \phi_{k1}X_{t-1} - \ldots - \phi_{kk}X_{t-k})^2$
Then take $\hat{\alpha}(k) = \hat{\phi}_k$

Remark (ACF and PACF Estimation for ARMA(p,q))**.** Assume $\{X_t\}$ is a mean-zero stationary process and a causal and invertable ARMA(p,q) process.
How to calculate the ACF?

- AR(1): $X_t = \phi X_{t-1} + Z_t$
  $\gamma(0) = \phi^2\gamma(0) + \sigma^2 \implies \gamma(0) = \frac{\sigma^2}{1-\phi^2}$ note that it needs to be causal so $|\phi| < 1$.
  $\gamma(h) = \phi\gamma(h-1)$ for $h \geq 1$
  $\implies \gamma(h) = \phi^{|h|}\gamma(0)$

- AR(2): $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$
  covariance with X_{t-1} is $\gamma(1) = \phi_1\gamma(0) + \phi_2\gamma(1) \implies \rho(1) = \phi_1 + \phi_2\rho(1) \implies \rho(1) = \frac{\phi_1}{1-\phi_2}$
  $\gamma(h) = \phi_1\gamma(h-1) + \phi_2\gamma(h-2)$ for $h \geq 2 \implies \rho(h) = \phi_1\rho(h-1) + \phi_2\rho(h-2)$ for $h \geq 2$

- General ARMA(p,q):
  Method 1: Write $X_t = \frac{\theta(B)Z_t}{\phi(B)} = \psi(B)Z_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$
  $\gamma(0) = \psi_0^2\sigma^2 + \psi_1^2\sigma^2 + \ldots = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2$
  $\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j\psi_{j+|h|}$ for $h \geq 1$

$\psi(z) = \frac{\theta(z)}{\phi(z)} \implies \phi(z)\psi(z) = \theta(z)$
$(1 - \phi_1 z - \ldots - \phi_p z^p)(\psi_0 + \psi_1 z + \psi_2 z^2 + \ldots) = 1 + \theta_1 z + \ldots + \theta_q z^q$
Match the coefficients of z^j for $j = 0, 1, 2, \ldots$ to get $\psi_0, \psi_1, \ldots$

- $\psi_0 = 1$
- $\psi_1 - \phi_1\psi_0 = \theta_1 \implies \psi_1 = \phi_1 + \theta_1$
- $\psi_2 - \phi_1\psi_1 - \phi_2\psi_0 = \theta_2 \implies \psi_2 = \phi_1\psi_1 + \phi_2 + \theta_2$

Method 2: Example ARMA(2,2) $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}$
Covariance with X_t: $\gamma(0) = \phi_1\gamma(1) + \phi_2\gamma(2) + \sigma^2(1 + \theta_1\psi_1 + \theta_2\psi_2)$
Covariance with X_{t-1}: $\gamma(1) = \phi_1\gamma(0) + \phi_2\gamma(1) + \sigma^2(\theta_1 + \theta_2\psi_1)$
Covariance with X_{t-2}: $\gamma(2) = \phi_1\gamma(1) + \phi_2\gamma(0) + \sigma^2\theta_2$
Now we can solve for $\gamma(0), \gamma(1), \gamma(2)$
Covariance with X_{t-h} for $h \geq 3$: $\gamma(h) = \phi_1\gamma(h-1) + \phi_2\gamma(h-2)$
Method 2.1: Solve the Difference Equation
Method 2.2: Get $\gamma(h)$ recursively
In general ARMA(p, q): make a system of $p + 1$ equations.

Remark (Statistical infrence)**.** Ask the questions: How do I know the order of the model? How do I estimate the parameters? How do I check if the model is good?

Remark (Spectral Representaions of Staitionary Processes)**.** New Chapter

Definition (Complex Random Variable)**.** $X = Re(X) + iIm(X)$ where $Re(X), Im(X)$ are real random variables.
or $X = X_1 + iX_2$ where X_1, X_2 are real random variables on the same probability space.
Properties
$E[X] = E[X_1] + iE[X_2]$
Suppose $Y = Y_1 + iY_2$ is another complex random variable.
$\text{Cov}(X, Y) = E[(X - E[X])\overline{(Y - E[Y])}]$
$\text{Cov}(X, Y) = \overline{\text{Cov}(Y, X)}$
Assume $E[X] = 0$ then $\text{Var}(X) = E[(X + iX_2)(X - iX_2)] = E[X_1^2 + X_2^2] \geq 0$
To get Properties of X_1, X_2 from X use Variance and Second moment
$L_c^2(\Omega, \mathcal{F}, P) = \{X : X = X_1 + iX_2, X_1, X_2 \in L^2(\Omega, \mathcal{F}, P)\}$

Definition (Stationary Process for Complex Random Variables)**.** $\{X_t\}$ is a complex-valued process.
It is a complex valed staition ary process if $E[|X_t|^2] < \infty$ and $E[X_t]$ does depend on t and $E[X_{t+h}\overline{X_t}]$ does not depend on t.
Write $X_t = X_{t1} + iX_{t2}$ where X_{t1}, X_{t2} are real-valued processes.

$$\begin{aligned}
E[X_{t+h}\overline{X_t}] &= E[(X_{t+h,1} + iX_{t+h,2})(X_{t1} - iX_{t2})] \\
&= E[X_{t+h,1}X_{t1}] + E[X_{t+h,2}X_{t2}] + i(E[X_{t+h,2}X_{t1}] - E[X_{t+h,1}X_{t2}])
\end{aligned}$$

This means the sum of the auto and cross covariance functions do not depend on t.
If a process is complex-valued stationary process does it imply that the real and imaginary parts are stat

1.12 10/14/2025 Lecture 12

Remark. Review of last class:

- $X = X_1 + iX_2$ where X_1, X_2 are real random variables

- $E[X] = E[X_1] + iE[X_2]$
- $\langle X, Y\rangle = E[X\overline{Y}]$
- $\mathrm{Cov}(X,Y) = E[(X - E[X])\overline{(Y - E[Y])}] = \langle X, Y\rangle - \langle E[X], E[Y]\rangle$
- Complex L^2 space
- Complex-valued stationary process:
  - $E[|X_t|^2] < \infty$
  - $E[X_t]$ does not depend on t
  - $E[X_{t+h}\overline{X_t}]$ does not depend on t
- the autocovariance of a complex-valued stationary process is $\gamma(h) = E[X_{t+h}\overline{X_t}] = \langle X_{t+h}, X_t\rangle - \langle E[X_{t+h}], E[X_t]\rangle$.
- $\gamma(h) = \overline{\gamma(-h)}$ it is hermitian

Theorem 10 (4.1.1)**.** *A function $k(\cdot)$ defined on the integers is an autocovariance function of a complex valued stationary process if and only if it is non-negative definite.*
IE $\sum_{j,k=1}^{n} a_j k(j-k)\overline{a_k} \geq 0$ for all $n \geq 1$ and $a_1, \dots, a_n \in \mathbb{C}$
This also implies that the acf is also hermitian

Proof. $\implies$: $\boxed{\text{Left As excerise}}$
$\Longleftarrow$: NND when $n = 1$ $a_1 = 1$ implies $k(0) \geq 0$
NND when $n = 2$ take $a_1 = 1$, $k(0) + k(0) \cdot |a_2|^2 + k(-1)\overline{a_2} + k(1)a_2 \geq 0$
it is always real $\implies$ $k(1) = -k(-1)$
for an arbitrary n take $a_j = 0$ for $j \neq 1, n$
$a_1 = 1$, $k(1-n) = -k(n-1)$
Suppose $k(\cdot)$ is the acf of $X_t = Y_t + iZ_t$ where Y_t, Z_t are real-valued stationary processes with mean 0
Let $\underline{X} = (X_1, \dots, X_n)'$ and $\underline{X} = \underline{Y} + i\underline{Z}$
$\mathrm{Cov}(\underline{X}) = E[\underline{X}\underline{X}^*] = E[(\underline{Y} + i\underline{Z})(\underline{Y}' - i\underline{Z}')]$
$= \Sigma_{yy} + \Sigma_{zz} + i(\Sigma_{zy} - \Sigma_{yz})$

$\mathrm{Cov}(\underline{X}) = \begin{bmatrix} k(0) & k(-1) & k(-2) & \dots & k(1-n) \\ k(1) & k(0) & k(-1) & \dots & k(2-n) \\ k(2) & k(1) & k(0) & \dots & k(3-n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ k(n-1) & k(n-2) & k(n-3) & \dots & k(0) \end{bmatrix} = K_1 + iK_2$ where K_1 and K_2 are real matrices.
write $\underline{a} = (a_1, \dots, a_n)' = \underline{b} + i\underline{c}$

$$\sum_{j,k=1}^{n} a_j k(j-k)\overline{a_k} = \underline{\mathrm{a}}'\underline{\mathrm{K}\overline{\mathrm{a}}}$$

$$\begin{aligned} &= (\underline{\mathrm{b}} + i\underline{\mathrm{c}})'(K_1 + iK_2)(\underline{\mathrm{b}} - i\underline{\mathrm{c}}) \\ &= \underline{\mathrm{b}}'K_1\underline{\mathrm{b}} + \underline{\mathrm{c}}'K_1\underline{\mathrm{c}} + b'K_2c - c'K_2b \\ \text{same as } &= (b_1' - c')\begin{bmatrix} k_1 & k_2' \\ k_2 & k_1 \end{bmatrix}\begin{bmatrix} b \\ c \end{bmatrix} \geq 0 \end{aligned}$$

Take $\Sigma_{yy} = \Sigma_{zz} = \frac{1}{2}K_1$ and $\Sigma_{zy} = \frac{1}{2}K_2$, $\Sigma_{yz} = \frac{1}{2}K_2' = -\frac{1}{2}K_2$
$K_1' - iK_2' = K_1 + iK_2$ thus $K_2' = -K_2$
Construct $\begin{bmatrix} \underline{\mathrm{Y}} \\ \underline{\mathrm{Z}} \end{bmatrix} \sim N(0, \frac{1}{2}\begin{bmatrix} K_1 & K_2' \\ K_2 & K_1 \end{bmatrix})$ □

Example. $X_t = \sum_{j=1}^{n} A_j e^{it\lambda_j}$

- $-\pi < \lambda_1 < \lambda_2 < \ldots < \lambda_n = \pi$
- A_j are complex-valued random variables
- $E[A_j] = 0$, $E[A_j\overline{A_k}] = 0$ for $j \neq k, E[|A_j|^2] = \sigma_j^2 < \infty$

Write $A_j = C_j + iD_j$ then

$$\begin{aligned} (C_j + iD_j)e^{it\lambda_j} &= (C_j + iD_j)(\cos(t\lambda_j) + i\sin(t\lambda_j)) \\ &= (C_j\cos(t\lambda_j) - D_j\sin(t\lambda_j)) + i(C_j\sin(t\lambda_j) + D_j\cos(t\lambda_j)) \end{aligned}$$

We can see that this is a complex-valued stationary process, but if I want the X to be real then we require the conition that $C_j\sin(t\lambda_j) + D_j\cos(t\lambda_j) = 0$

Want X_t to be real valued $\begin{cases} \lambda_j = -\lambda_{n-j} \text{ for } j = 1, \ldots, n-1 \\ A(\lambda_j) = \overline{A(\lambda_{n-j})} \text{ for } j = 1, \ldots, n-1 \\ A_n \text{ is real} \end{cases}$

This

1.13 10/16/2025 Lecture 13

Remark. Review of what we are looking at

$$X_t = \sum_{j=1}^{n} A_j e^{it\lambda_j}$$

where $\begin{cases} -\pi < \lambda_1 < \lambda_2 < \ldots < \lambda_n = \pi \\ A_j \text{ are uncorrelated complex-valued random variables} \\ E[A_j] = 0, E[A_j\overline{A_k}] = 0 \text{ for } j \neq k, E[|A_j|^2] = \sigma_j^2 < \infty \end{cases}$

When is X_t real-valued? $\begin{cases} \lambda_j = -\lambda_{n-j} & 1 \leq j \leq n-1 \\ A_j = \overline{A_{n-j}} & 1 \leq j \leq n-1 \\ A_n \text{ is real} \end{cases}$ Is X_t stationary?

$$E(X_{t+h}\overline{X_t}) = E\left[\left(\sum_{j=1}^{n} A_j e^{i(t+h)\lambda_j}\right)\left(\sum_{k=1}^{n} \overline{A_k} e^{-it\lambda_k}\right)\right] = E\left[A_1\overline{A_1}e^{ih\lambda_1} + A_2\overline{A_2}e^{ih\lambda_2} + \ldots + A_n\overline{A_n}e^{ih\lambda_n}\right]$$

$$= \sum_{j=1}^{n} \sigma_j^2 e^{ih\lambda_j}$$

This does not depend on t so it is stationary.
$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\nu} dF(\nu)$ where $F(\nu) = \sum_{j:\lambda_j \leq \nu} \sigma_j^2$ This is the Reimann-Stieltjes integral This is a step function with jumps of size σ_j^2 at λ_j for $j = 1, \ldots, n$
View $F(\nu)$ as a measure on $(-\pi, \pi]$ which assigns point measures $m(\nu) = \sigma_j^2$
The function $e^{ih\nu}$ takes on values $e^{ih\lambda_j}$ at the points λ_j with value σ_j^2
Every mean-zero stationary process $\{X_t\}$ has a representation

$$X_t = \int_{-\pi}^{\pi} e^{it\nu} dZ(\nu)$$

$\boxed{\text{If we have a continous path, everywhere differentialble, how is the different from the Riemann-Stieltjes}}$
where $Z(\nu)$ is a complex-valued process with the following properties:

- $E[dZ(\nu)] = 0$
- $E[|dZ(\nu)|^2] = dF(\nu)$ where $F(\nu)$ is a non-decreasing function on $(-\pi, \pi]$ with $F(-\pi) = 0$ and $F(\pi) = \gamma(0)$
- $E[dZ(\nu)\overline{dZ(\lambda)}] = 0$ for $\nu \neq \lambda$

Correspondingly $\gamma_X(h) = \int_{-\pi}^{\pi} e^{ih\nu} dF(\nu)$ **Riemann integral**: $\int_a^b g(x)dx$ where $g(x)$ is a function on $[a, b]$
$\lim_{\max|a_j - a_{j-1}| \to 0} \sum_{j=1}^{n} g(a_j)(a_j - a_{j-1})$
$\lim_{\max|\lambda_j - \lambda_{j-1}| \to 0} \sum_{j=1}^{n} e^{ih\lambda_j} Z(\lambda_j) - Z(\lambda_{j-1})$

$F(\cdot)$ is called the spectral distribution function of $\{X_t\}$
F is increasing and caddlag, $F(-\pi) = 0$, $F(\pi) = \gamma(0)$

Theorem 11 (4.3.1 Herglotz Theorem)**.** *A complex Values fn $\gamma(\cdot)$ defined on the integers is NND if and only if*

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\nu} dF(\nu) \quad h \in \mathbb{Z}$$

Where $F(\cdot)$ is a distribution function supported on $(-\pi, \pi]$ with $F(-\pi) = 0$ and $F(\pi) \leq \infty$
Note: Corresponding Bocher's for the characteristic function of a random variable
If $F(\cdot)$ is absolutely Continuous w.r.t the Lebesgue measure, (ie they are related but not the same) ie $F(\lambda) = \int_{-\pi}^{\lambda} f(\nu)d\nu$ then $f(\lambda)$ is called the spectral density function of $\{X_t\}$

Proof. $\Longleftarrow$

$$\begin{aligned}
\sum_{j,k=1}^{n} a_j \gamma(j-k)\overline{a_k} &= \sum_{j,k=1}^{n} a_j \int_{-\pi}^{\pi} e^{i(j-k)\nu} dF(\nu)\overline{a_k} \\
&= \int_{-\pi}^{\pi} \sum_{j,k=1}^{n} a_j e^{i(j)\nu} e^{-ik\nu} \overline{a_k} dF(\nu) \\
&= \int_{-\pi}^{\pi} \left(\sum_{j=1}^{n} a_j e^{ij\nu} \right) \left(\sum_{k=1}^{n} \overline{a_k e^{ik\nu}} \right) dF(\nu) \\
&= \int_{-\pi}^{\pi} \left| \sum_{j=1}^{n} a_j e^{ij\nu} \right|^2 dF(\nu) \geq 0
\end{aligned}$$

□

1.14 10/21/2025 Lecture 14

Remark (Spectral Representation of Complex Stationary Process)**.** Every zero-mean stationary process has a representation:

$$X_t = \int_{-\pi}^{\pi} e^{it\nu} dZ(\nu)$$

Correspondingly $\gamma_X(h) = \int_{-\pi}^{\pi} e^{ih\nu} dF(\nu)$ F is cadlag, increasing, $F(-\pi) = 0$, $F(\pi) = \gamma(0)$ where $Z(\nu)$ is a complex-valued process with the following properties:
If $F(\cdot)$ is absolutely continuous w.r.t the Lebesgue measure, ie $F(\lambda) = \int_{-\pi}^{\lambda} f(\nu)d\nu$ then $\gamma_X(h) = \int_{-\pi}^{\pi} e^{ih\nu} f(\nu)d\nu$ where $f(\nu)$ is called the spectral density function of $\{X_t\}$
F is the spectral distribution function of $\{X_t\}$ f is the spectral density function of $\{X_t\}$

Theorem 12 (4.3.1 Herglotz)**.** *A complex valued fn $\gamma(\cdot)$ defined on $\mathbb{Z}$ is NND if and only if*

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\nu} dF(\nu) \quad h \in \mathbb{Z}$$

Where $F(\cdot)$ is a distribution function supported on $(-\pi,\pi]$ with $F(-\pi)=0$ and $F(\pi)\le\infty$

Proof. $\implies$: Define $f_n(\nu)=\frac{1}{2\pi N}\sum_{j,k=1}^{N}\gamma(j-k)e^{-ij\nu}e^{ik\nu}$

$$\begin{aligned}f_n(\nu)&=\frac{1}{2\pi N}\sum_{j,k=1}^{N}\gamma(j-k)e^{-ij\nu}e^{ik\nu}\\&=\frac{1}{2\pi N}\sum_{m=1-N}^{N-1}(1-\frac{|m|}{N})\gamma(m)e^{im\nu}\end{aligned}$$

$\boxed{\text{Fejer Kernel}}$ Then $f_n(\nu)\ge 0$ and $\int_{-\pi}^{\pi}f_n(\nu)d\nu=\gamma(0)$
Next Define $F_N(\lambda)=\int_{-\pi}^{\lambda}f_n(\nu)d\nu$ Then $F_N(\lambda)$ is a distribution function on $(-\pi,\pi]$ with $F_N(-\pi)=0$ and $F_N(\pi)=\gamma(0)$
And

$$\begin{aligned}\int_{-\pi}^{\pi}e^{ih\nu}dF_N(\nu)&=\int_{-\pi}^{\pi}e^{ih\nu}f_n(\nu)d\nu\\&=\begin{cases}\gamma(h)(1-\frac{|h|}{N}) & |h|<N\\ 0 & |h|\ge N\end{cases}\end{aligned}$$

By Helly's Theorem: there exists a subsequence $\{N_k\}$ s.t. $F_{N_k}(\cdot)\to F(\cdot)$ and $\lim_{k\to\infty}\int_{-\pi}^{\pi}e^{ih\nu}dF_{N_k}(\nu)=\int_{-\pi}^{\pi}e^{ih\nu}dF(\nu)$

$$\lim_{k\to\infty}\int_{-\pi}^{\pi}e^{ih\nu}dF_{N_k}(\nu)=\lim_{k\to\infty}\gamma(h)(1-\frac{|h|}{N_k})=\gamma(h)$$

□

Remark. F as the limit might have a point mass at $-\pi$ so an additial step transfer it to π
It can be proven that $F_N(\cdot)\to F(\dot{})$
This also ensure the spectral distribution function is unique ie if $\exists$ another distribution function $G(\cdot)$ s.t $\gamma(h)=\int_{-\pi}^{\pi}e^{ih\nu}dG(\nu)$ then $F(\cdot)=G(\cdot)$ at all continuity points of both F and G

Theorem 13 (4.3.2). *Suppose $K(\cdot)$ is a function defined on $\mathbb{Z}$ s.t. $\sum_{h=-\infty}^{\infty}|K(h)|<\infty$ then*

$$f(\lambda):=\frac{1}{2\pi}\sum_{h=-\infty}^{\infty}K(h)e^{-ih\lambda}$$

is well -defined and $K(n)=\int_{-\pi}^{\pi}e^{in\lambda}f(\lambda)d\lambda$ for all $n\in\mathbb{Z}$

Corollary. Suppose $\gamma(\cdot)$ is a function defined on $\mathbb{Z}$ and it $\sum_{h=-\infty}^{\infty}|\gamma(h)| < \infty$ then $\gamma(\cdot)$ is the acf of a stationary process iff $f(\lambda) := \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\gamma(h)e^{-ih\lambda} \geq 0$ for all $\lambda \in [-\pi, \pi]$

Proof. $\Longleftarrow$: f is a density, can define $F(\lambda) = \int_{-\pi}^{\lambda} f(\nu)d\nu$ which is a distribution function on $(-\pi, \pi]$
$\implies f_N(\lambda) = \frac{1}{2\pi}\sum_{m=-1-N}^{N-1}(1 - \frac{|m|}{N})\gamma(m)e^{-im\lambda} \geq 0$ for all $\lambda \in [-\pi, \pi]$ ◻

Example (4.3.1)**.** $K(h) = \begin{cases} 1 & h = 0 \\ \rho & h = \pm 1 \quad \text{where } |\rho| < \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$
We can look at the spectral density function:

$$\begin{aligned} f(\lambda) &= \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} K(h)e^{-ih\lambda} \\ &= \frac{1}{2\pi}(1 + \rho e^{-i\lambda} + \rho e^{i\lambda}) \\ &= \frac{1}{2\pi}(1 + 2\rho\cos(\lambda)) \end{aligned}$$

Remark. If the acf is real, then the spectral density function is also real-valued.
$f(\lambda) = \frac{1}{2\pi}\left[1 + \sum_{n=1}^{\infty} 2\gamma(n)\cos(n\lambda)\right]$
In general the spectral distribution function is stne=netruc in the sense $F_X(\lambda) = F_X(\pi^-) - F_X(-\lambda^-)$
$\boxed{\text{HW}}$

1.15 10/23/2025 Lecture 15

Remark. If $\gamma(\cdot)$ is a complex balued acg then there exists a unique spectral distribution F s.t. $\gamma(h) = \int_{-\pi}^{\pi} e^{ih\nu}dF(\nu)$
If $F(\cdot)$ is absolutely continuous w.r.t the Lebesgue measure ie $F(\lambda) = \int_{-\pi}^{\lambda} f(\nu)d\nu$ then $\gamma(h) = \int_{-\pi}^{\pi} e^{ih\nu}f(\nu)d\nu$ where $f(\nu)$ is called the spectral density function of $\{X_t\}$
If $\sum_{h=-\infty}^{\infty}|\gamma(h)| < \infty$ then he spectral density eists and is given by $f(\lambda) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\gamma(h)e^{-ih\lambda}$
Consider a $WN(0, \sigma^2)$, the acf is $\gamma(h) = \begin{cases} \sigma^2 & h = 0 \\ 0 & h \neq 0 \end{cases}$
The spectral density function is $f(\lambda) = \frac{1}{2\pi}\sigma^2$ for $\lambda \in [-\pi, \pi]$

Theorem 14 (4.4.1)**.** *If $\{Y_t\}$ is zero-mean complex valued stationary process with the spectral denisty $F_Y(\cdot)$ and if $\{X_t\}$ is defined by $X_t = \sum_{j=-\infty}^{\infty}\psi_j Y_{t-j}$ where $\sum_{j=-\infty}^{\infty}|\psi_j| < \infty$ then $\{X_t\}$ is staitionary with acf $\gamma_X(h) = \sum_{j,k=-\infty}^{\infty}\psi_j\overline{\psi_k}\gamma_Y(h - j + k)$ AAnd the spectral distribution: $F_X(\lambda) = \int_{-\pi}^{\lambda}|\sum_{j=-\infty}^{\infty}\psi_j e^{-ij\nu}|^2 dF_Y(\nu)$*

This a Radon-Nikodym derivative

Proof.

$$\begin{aligned}
\gamma_X(h) &= \sum_{j,k=-\infty}^{\infty} \psi_j \overline{\psi_k} \int_{-\pi}^{\pi} e^{i(h-j+k)\nu} dF_Y(\nu) \\
&= \int_{-\pi}^{\pi} e^{ih\nu} \left[\sum_{j,k=-\infty}^{\infty} \psi_j e^{-ij\nu} \overline{\psi_k e^{-ik\nu}} \right] dF_Y(\nu) \\
&= \int_{-\pi}^{\pi} e^{ih\nu} | \sum_{j=-\infty}^{\infty} \psi_j e^{-ij\nu} |^2 dF_Y(\nu) \\
&= \int_{-\pi}^{\lambda} e^{ih\nu} dF_X(\nu) \quad \text{ where } F_X(\lambda) = \int_{-\pi}^{\lambda} | \sum_{j=-\infty}^{\infty} \psi_j e^{-ij\nu} |^2 dF_Y(\nu)
\end{aligned}$$

□

Theorem 15 (3.1.3)**.** *Consider ARMA(p,q) $\phi(B)X_t = \theta(B)Z_t$ If $\phi(z) \neq 0$ for $|z| = 1$ then there is a unique stationary solution given by $X_t = \sum_{j=\infty}^{\infty} \psi_j Z_{t-j}$ where $\psi(z) = \frac{\theta(z)}{\phi(z)} = \sum_{j=-\infty}^{\infty} \psi_j z^j$ for $|z| \leq 1$*
note:
$\phi(z) = (1 - .5z)(1 - 2z)$
$1/\phi(z) = 1/(1 - .5z) \cdot 1/(1 - 2z)$
$\frac{1}{1-2z} = \frac{1}{z} \cdot \frac{1}{\frac{1}{z} - 2}$

Theorem 16 (4.4.2)**.** *Consider the ARMA(p,q) $\phi(B)X_t = \theta(B)Z_t$ Assume that ϕ, θ have no common zeros and that $\phi(z) \neq 0$ for $|z| = 1$ then $\{X_t\}$ has a spectral density function given by*

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \cdot \frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2} \quad \lambda \in [-\pi, \pi]$$

since $dF_Z(\lambda) = \frac{\sigma^2}{2\pi} d\nu$

Proof. Define $U_t = \phi(B)X_t = \theta(B)Z_t$
$f_U(\lambda) = |\theta(e^{-i\lambda})|^2 f_Z(\lambda) = \frac{\sigma^2}{2\pi} |\theta(e^{-i\lambda})|^2$
And thus $\implies: f_X(\lambda) = \frac{\sigma^2}{2\pi} \cdot \frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2}$

□

Example. $AR(1) : X_t = \phi X_{t-1} + Z_t$ where $|\phi| \neq 1$
$f_X(\lambda) = \frac{\sigma^2}{2\pi} \cdot \frac{1}{|1-\phi e^{-i\lambda}|^2} = \frac{\sigma^2}{2\pi} \cdot \frac{1}{1-2\phi\cos(\lambda)+\phi^2}$
Suppose $|\phi| > 1$ We want to have $(1 - \frac{1}{\phi}B)X_t = Z_t^*$?.

$$(1-\frac{1}{\phi}B)(1-\phi B)X_t = (1-\frac{1}{\phi}B)Z_t$$
$$(1-\frac{1}{\phi}B)X_t = \frac{(1-\frac{1}{\phi}B)}{(1-\phi B)}Z_t = Z_t^*$$

Thus by Theorem 4.4.2 the spectral density function is $f_{Z^*}(\lambda) = \left|\frac{1-\frac{1}{\phi}e^{-i\lambda}}{1-\phi e^{-i\lambda}}\right| \frac{\sigma^2}{2\pi}$ We can verify that this is a constant by

$$\begin{aligned}|e^{i\lambda}e^{-i\lambda} - \frac{1}{\phi}e^{i\lambda}|^2 &= |e^{i\lambda} - \frac{1}{\phi}|^2 \\ &= \frac{1}{\phi^2}|-\phi e^{-i\lambda}+1|^2 \quad \text{Which cancels with the denominator}\end{aligned}$$

Thus $f_{Z^*}(\lambda) = \frac{\sigma^2}{2\pi\phi^2}$ for $\lambda \in [-\pi, \pi]$
IE $Z_t^* \sim WN(0, \frac{\sigma^2}{\phi^2})$

Example. $MA(1) : X_t = Z_t + \theta Z_{t-1}$ where $Z_t \sim WN(0, \sigma^2)$, $|\theta| > 1$
$X_t = (1+\theta B)Z_t = (1+\frac{1}{\theta}B)Z_t^*$
$\implies Z_t^* = \frac{(1+\theta B)}{(1+\frac{1}{\theta}B)}Z_t$
Thus $Z_t^* \sim WN(0, \sigma^2\theta^2)$

Remark. Suppose $\{X_t\}$ iscausal and invertable. LEt $\mathcal{H}_t = \overline{sp}\{X_t, X_{t-1}, \ldots\}$.

$$X_t - \mathcal{P}_{\mathcal{H}_{t-1}}X_t = Z_t$$

Since
Causality implies that $\mathcal{H}_{t-1} \subset \overline{sp}\{Z_{t-1}, Z_{t-2}, \ldots\} \implies Z_t \perp \mathcal{H}_{t-1}$
Invertability implies that $Z_{t-1}, Z_{t-2}, \ldots \in \mathcal{H}_{t-1}$

1.16 10/28/2025 Lecture 16

Remark. Suppose $\phi(B)X_t = \theta(B)Z_t$ ϕ, θ have no common zeros and $\phi(z) \neq 0$ for $|z| = 1$
The spectral of X_t is given by $f_X(\lambda) = \frac{\sigma^2}{2\pi} \cdot \frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2}$

Example (AR(1) Spectral density)**.** $X_t = \phi X_{t-1} + Z_t$ where $|\phi| > 1$ then $X_t = \frac{1}{\phi}X_{t-1} + Z_t^*$ where $Z_t^* \sim WN(0, \frac{\sigma^2}{\phi^2})$
And $Z_t^* = \frac{(1-\frac{1}{\phi}B)}{(1-\phi B)}Z_t$

Example (MA(1) Spectral density)**.** $X_t = Z_t + \theta Z_{t-1}$ where $|\theta| > 1$ then $X_t = (1 + \frac{1}{\theta}B)Z_t^*$ where $Z_t^* \sim WN(0, \sigma^2\theta^2)$
And $Z_t^* = \frac{(1+\theta B)}{(1+\frac{1}{\theta}B)} Z_t$

Definition. Suppose $\phi(B)X_t = \theta(B)Z_t$ is causal and invertable $\mathcal{M}_t = \overline{sp}\{X_t, X_{t-1}, \dots\}$
The BLP $\mathcal{P}_{\mathcal{M}_{t-1}} X_t$ what is $X_t - \mathcal{P}_{\mathcal{M}_{t-1}} X_t$?
Causal - $Z_t \perp M_{t-1}$
Invertable - $Z_{t-1}, Z_{t-2}, \dots \in \mathcal{M}_{t-1}$
Thus $X_t - \mathcal{P}_{\mathcal{M}_{t-1}} X_t = Z_t$ Consequently $||X_t - \mathcal{P}_{\mathcal{M}_{t-1}} X_t||^2 = E[|Z_t|^2] = \sigma^2$
Thus for AR(1) with $|\phi| > 1$, the BLP error variance is σ^2/ϕ^2
For MA(1) with $|\theta| > 1$, the BLP error variance is $\sigma^2\theta^2$

Remark. How do you make an ARMA that is not causal or invertable into one that is?
More generally write $\phi(z) = \prod_{j=1}^m (1 - a_j^{-1}z)$ where $|a_j| > 1$ for $1 < j < r$ and $|a_j| < 1$ for $r + 1 < j < m$

$$(1 - a_1^{-1}z)(1 - a_2^{-1}z)\dots(1 - a_r^{-1}z)(1 - a_{r+1}^{-1}z)\dots(1 - a_m^{-1}z)$$

The left part is the $(1 - a_{r+1}z)/(1 - a_{r+1}^{-1})z$

$$\frac{1 - \overline{a_{r+1}}B}{1 + a_{r+1}^{-1}B}\phi(B)X_t = \frac{1 - \overline{a_{r+1}}B}{1 - a_{r+1}^{-1}B}\theta(B)Z_t$$

$a_{r+1} = ce^{i\theta}$ for $c < 1$
Can verify that $\frac{1-a_{r+1}B}{1-a_{r+1}^{-1}B}Z_t \sim WN(0, |a_{r+1}|^2\sigma^2)$ Then do the same thing for every $r+1 \le j \le m$ ie replace a_{r+1} with somehting that makes it causal and invertable
Thus

$$\hat{\phi}(z) = \frac{j = r+1}{p}\frac{1 - \overline{a_j}B}{1 - a_j^{-1}B}\phi(z)$$

Thus

$$Z_t^* = \prod_{j=r+1}^{p}\frac{1 - \overline{a_j}B}{1 - a_j^{-1}B}Z_t \sim WN(0, \sigma^2\prod_{j=r+1}^{p}|a_j|^2)$$

Also

$$\theta(z) = \prod_{j=1}^{q}(1 - b_j^{-1}z) \text{ where } |b_j| > 1 \text{ for } 1 \le j \le s \text{ and } |b_j| < 1 \text{ for } s + 1 \le j \le q$$

Then

$$\hat{\theta}(z) = \prod_{j=s+1}^{q} \frac{1-\overline{b_j}B}{1-b_j^{-1}B}\theta(z)$$

More Generally $\phi(B)X_t = \theta(B)Z_t$ When conveted

$$\text{Var}(Z_t^*) = \sigma^2 \frac{\prod_{j=r+1}^{p} |a_j|^2}{\prod_{j=s+1}^{q} |b_j|^2}$$

This is a theorem in the Book.
Whenever there is a root that is bad, we can flip it by takeing the conjugate reciprocal and adjusting the variance accordingly.

Proposition 8 (4.4.1)**.** *Assume ARMA as usual. and $\theta(z) \neq 0$ for $|z| < 1$ then $Z_t \in \overline{sp}\{X_t, X_{t-1}, \ldots\}$*

Proof. First: Factorize $\theta(z) = \theta^+(z)\theta^*(z)$ where $\theta^+(z) = \prod_{j=1}^{s}(1-b_j^{-1}z)$ with $|b_j| > 1$ and $\theta^*(z) = \prod_{j=s+1}^{q}(1-b_j^{-1}z)$ with $|b_j| = 1$
Define $Y_t = \theta^*(B)Z_t$ then $\phi(B)X_t = \theta^+(B)Y_t$
Using the earlier proof on invertability, $\implies Y_t \in \overline{sp}\{X_t, X_{t-1}, \ldots\}$
So it suffices to show that $Z_t \in \overline{sp}\{Y_t, Y_{t-1}, \ldots\}$
Define $U_t = Y_t - \mathcal{P}_{\overline{sp}\{Y_k, k \leq t-1\}} T_t$
Then by Prop 3.2.1:

$$\begin{aligned} Y_t &= U_t + \alpha_1 U_{t-1} + \alpha_2 U_{t-2} + \ldots + \alpha_{q-s} U_{t-(q-s)} \\ \implies f_Y(\lambda) &= \frac{\sigma^2}{2\pi}|\alpha(e^{-i\lambda})|^2 = \frac{\sigma^2}{2\pi}|\theta^*(e^{-i\lambda})|^2 \end{aligned}$$

Both $\alpha(z)$ and $\theta^*(z)$ are real polynomials of the same degree $q-s$
Factoring $\alpha(z) = \prod_{j=1}^{q-s}(1-c_j^{-1}z)$ And $\theta^*(z) = \prod_{j=s+1}^{q}(1-b_j^{-1}z)$
We see that they must have the same roots ie $\{c_j\} = \{b_j\}$ as multisets
And they are the same everywhere
Thus $\alpha(z) = \theta^*(z)$
Then using this we can show that $Z_t = U_t$ □

1.17 10/30/2025 Lecture 17

Proposition 9 (4.4.1 cont)**.** *Consider the ARMA(p,q) $\phi(B)X_t = \theta(B)Z_t$ where $\theta(z) \neq 0$ for $|z| < 1$*
Then $Z_t \in \overline{sp}\{X_t, X_{t-1}, \ldots\}$

Proof. WLOG: assume $\theta(z) \neq 0$ for $|z| \neq 1$
Define $Y_t = \theta(B)Z_t$ then $\phi(B)X_t = Y_t$
$\implies \overline{sp}\{Y_t, k \leq t\} \subset \overline{sp}\{X_t, k \leq t\}$
So it suffices to show that $Z_t \in \overline{sp}\{Y_t, Y_{t-1}, \dots\}$
Define $U_t = Y_t - \mathcal{P}_{\overline{sp}\{Y_k, k \leq t-1\}} Y_t$
Then by Prop 3.2.1:

$$Y_t = U_t + \alpha_1 U_{t-1} + \alpha_2 U_{t-2} + \dots + \alpha_q U_{t-q}$$
$$\implies f_Y(\lambda) = \frac{\sigma^2}{2\pi}|\alpha(e^{-i\lambda})|^2 = \frac{\sigma^2}{2\pi}|\theta(e^{-i\lambda})|^2$$

Both $\alpha(z)$ and $\theta(z)$ are real polynomials of the same degree q
And they have the same roots ie $\{c_j\} = \{b_j\}$ as multisets
And they are the same everywhere
Thus $\alpha(z) = \theta(z)$
And $\sigma^2 = \sigma_n^2$
$\implies$ that $(U_t, Y_t, Y_{t-1}, \dots, Y_{t-q})$ and $(Z_t, Y_t, Y_{t-1}, \dots, Y_{t-q})$ have the same covarience matrix
$\implies \mathcal{P}_{\overline{sp}\{Y_k, k \leq t-1\}} U_t = \mathcal{P}_{\overline{sp}\{Y_k, k \leq t-1\}} Z_t$
$\implies U_t = \lim_{n \to \infty} \mathcal{P}_{\overline{sp}\{Y_t, Y_{t-1}, \dots, Y_{t-n}\}} U_t = \lim_{n \to \infty} \mathcal{P}_{\overline{sp}\{Y_t, Y_{t-1}, \dots, Y_{t-n}\}} Z_t$
Bescuase $\sigma_n^2 = \sigma^2$ it holds that $Z_t = U_t$
If you show that two projections are equal with the same variance then the two random variables are equal. $\square$

Proposition 10 (4.4.3)**.** *Assume the ARMA(p,q)* $\phi(B)X_t = \theta(B)Z_t$ *where* $\phi(z) \neq 0$ *for* $|z| < 1$
If $\theta(z) = 0$ *for some* $|z| \leq 1$ *then* $Z_t \notin \overline{sp}\{X_t, X_{t-1}, \dots\}$

Definition (Wold Decomposition)**. Aside**: Crimea-Wold Device to prove the multivariate CLT
Setup: $X_t, t \in \mathbb{Z}$ is a zero-mean stationary process.
Define $\mathcal{M}_n = \overline{sp}\{X_t, t \leq n\}$, $\mathcal{M} = \overline{\bigcup_{n \in \mathbb{Z}} \mathcal{M}_n} = \overline{sp}\{X_t, t \in \mathbb{Z}\}$
Define $\mathcal{M}_{-\infty} = \bigcap_{n \in \mathbb{Z}} \mathcal{M}_n$
$\boxed{\text{Look into Durret and Billinsley}}$

Definition (Mean Squared Error)**.** $\sigma^2 = E[X_{n+1}^2 - \mathcal{P}_{\mathcal{M}_n} X_{n+1}]$ is the mean squared error of the BLP of X_{n+1} based on $\mathcal{M}_n$
AKA the one step ahead prediction error variance

Definition (Determinisitic Process)**.** A stationary process $\{X_t\}$ is deterministic if $\sigma^2 = 0$ ie $X_{n+1} \in \mathcal{M}_n$ for all $n \in \mathbb{Z}$
ie X_{n+1} can be perfectly predicted based on the infinite past $\{X_n, X_{n-1}, \dots\}$
Note that for all Process $\sum_{j=1}^n A_j e^{-i\lambda_j t}$ is deterministic for A_j uncorrelated mean zero random variables

Theorem 17 (3.7.1 Wold Decomposition Theorem)**.** *If $\sigma^2 > 0$ then X_t can we expressed as $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t$ where*

1. $\psi_0 = 1$, $\sum_{j=0}^{\infty} |\psi_j|^2 < \infty$
2. $Z_t \sim WN(0, \sigma^2)$
3. $Z_t \in \mathcal{M}_t, \forall t \in \mathbb{Z}$
4. $E[Z_t V_s] = 0, \forall t, s \in \mathbb{Z}$
5. $V_t \in \mathcal{M}_{-\infty}, \forall t \in \mathbb{Z}$
6. V_t *is deterministic stationary process*

Proof. Define $Z_t = X_t - \mathcal{P}_{\mathcal{M}_{t-1}} X_t$ Thus the $Z_t \sim WN(0, \sigma^2)$ and $Z_t \in \mathcal{M}_t$ which gives us number 2 and 3
Project X_t onto $\overline{sp}\{Z_t, Z_{t-1}, \dots, Z_{t-n}\}$

$$\mathcal{P}_{\overline{sp}\{Z_t, Z_{t-1}, \dots, Z_{t-n}\}} X_t =: \sum_{j=0}^{n} \psi_j Z_{t-j}$$

Where $\psi_j = \frac{\langle X_t, Z_{t-j} \rangle}{\langle Z_{t-j}, Z_{t-j} \rangle}$
In particular $\psi_0 = 1$ thus We get number 1
Define $V_t = X_t - \lim_{n\to\infty} \mathcal{P}_{\overline{sp}\{Z_t, Z_{t-1}, \dots, Z_{t-n}\}} X_t$, $V_t \in \mathcal{M}_t$
Then $E[Z_s V_t] = 0$ for all $s, t \in \mathbb{Z}$ which gives us number 4 $\square$

1.18 11/04/2025 Lecture 18

Remark (Wold Decomposition)**.** Recall the Linear Past:

$$\begin{aligned} \mathcal{M}_n &= \overline{sp}\{X_t, t \le n\} \\ \mathcal{M} &= \overline{\bigcup_{n\in\mathbb{Z}} \mathcal{M}_n} = \overline{sp}\{X_t, t \in \mathbb{Z}\} \\ \mathcal{M}_{-\infty} &= \bigcap_{n\in\mathbb{Z}} \mathcal{M}_n \end{aligned}$$

$\sigma^2 = E[X_{n+1} - \mathcal{P}_{\mathcal{M}_n} X_{n+1}]$ is the one step ahead prediction error variance

Definition (Determinisitic)**.** A stationary process $\{X_t\}$ is deterministic if $\sigma^2 = 0$

Theorem 18 (5.7.1 Wold Decomposition Theorem)**.** *If $\sigma^2 > 0$ then X_t can we expressed as $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t$ where*

1. $\psi_0 = 1$, $\sum_{j=0}^{\infty} |\psi_j|^2 < \infty$

2. $Z_t \sim WN(0, \sigma^2)$

3. $Z_t \in \mathcal{M}_t, \forall t \in \mathbb{Z}$

4. $E[Z_t V_s] = 0, \forall t, s \in \mathbb{Z}$

5. $V_t \in \mathcal{M}_{-\infty}, \forall t \in \mathbb{Z}$

6. V_t *is deterministic stationary process*

The sequences $\{\psi_j\}$, $\{Z_t\}$ *and* $\{V_t\}$ *are uniquely determined by the assumptions above.*

Proof. $Z_t = X_t - \mathcal{P}_{\mathcal{M}_{t-1}} X_t$
$\psi_j = \frac{\langle X_t, Z_{t-j} \rangle}{\langle Z_{t-j}, Z_{t-j} \rangle}$ $V_t = X_t - \sum_{j=0}^{\infty} \psi_j Z_{t-j}$
$\mathcal{M}_t = \mathcal{M}_{t-1} \oplus \overline{sp}\{Z_t\}, V_t \perp Z_t \implies V_t \in \mathcal{M}_{t-1}$
$\mathcal{M}_t = \mathcal{M}_{t-2} \oplus \overline{sp}\{Z_{t-1}, Z_t\}, V_t \perp Z_{t-1}, Z_t \implies V_t \in \mathcal{M}_{t-2}$
Continuing this way we get $V_t \in \mathcal{M}_{t-k}$ for all $k \geq 0$ thus $V_t \in \mathcal{M}_{-\infty}$
This shows properties 5.
Let $\mathcal{M}_t^V = \overline{sp}\{V_t, V_{t-1}, \ldots\}$
Need to show that $V_t \in \mathcal{M}_{t-1}^V$ and thus will show that $V_t \in \mathcal{M}_{-\infty}^V$
it suffices to show that $\mathcal{M}_{-\infty}^V = \mathcal{M}_{-\infty}$
We know $V_T \in \mathcal{M}_t \implies \mathcal{M}_t^V \subset \mathcal{M}_t \implies \mathcal{M}_{-\infty}^V \subset \mathcal{M}_{-\infty}$
Now we show the reverse inclusion.
$\mathcal{M}_t = \overline{sp}\{Z_k, V_k, k \leq t\} = \overline{sp}\{Z_k, k \leq t\} \oplus \mathcal{M}_t^V$
$\mathcal{M}_{t-k} \perp \overline{sp}\{Z_t \ldots Z_{t-k}\}$
$\implies \mathcal{M}_{-\infty} \perp \overline{sp}\{Z_t, Z_{t-1}, \ldots\} \implies \mathcal{M}_{-\infty} \subset \mathcal{M}_t^V$
Thus $\mathcal{M}_{-\infty} \subset \mathcal{M}_{-\infty}^V$

□

Proof of Uniqueness. If $X_t = \sum_{j=0}^{\infty} \eta_j W_{t-j} + G_t$ with the same properties as above.
(5) implies that $E[G_t] \in \mathcal{M}_{-\infty}$
(3) implies that $W_{t-1}, W_{t-2}, \ldots \in \mathcal{M}_{t-1}$

$$X_t = W_t + \sum_{j=1}^{\infty} \eta_j W_{t-j} + G_t \in \mathcal{M}_{t-1}$$

And $W_t \perp X_{t-k}$ for any $k \geq 1$
Thus $W_t = X_t - \mathcal{P}_{\mathcal{M}_{t-1}} X_t = Z_t$

□

Theorem 19 (Kolmogorov or Doob)**.** *Let* $\{Y_{1t}\}$ *and* $\{Y_{2t}\}$ *be two zero-mean mutually orthogonal stationary processes and let* $X_t = Y_{1t} + Y_{2t}$ *Suppose* F_1 *and* F_2 *are the spectral distribution functions of* $\{Y_{1t}\}$ *and* $\{Y_{2t}\}$ *respectively.*

Then $\{Y_{1t}, Y_{2t}, t \in \mathbb{Z}\} \subset \mathcal{M}^X := \overline{sp}\{X_k, k \in \mathbb{Z}\}$
IFF
F_1 *and* F_2 *are mutually singular ie* F_1 *corresponds to a measure that is on the space* m_1 *and* F_2 *corresponds to a measure that is on the space* m_2. *If there exists a measureable set* $A \subset (-\pi, \pi]$ *such that,* $m_1(A^c) = 0$ *and* $m_2(A) = 0$ *then* F_1 *and* F_2 *are mutually singular.*

Theorem 20 (Rudin). *IF* $\{c_n, n \leq 0\}$ *are s.t.* $0 < \sum_{n=-\infty}^{0} |c_n|^2 < \infty$ *then* $\sum_{n=-\infty}^{0} c_n e^{in\lambda} \in L^2(-\pi, \pi] := \{g : (-\pi, \pi] \to \mathbb{C} | \int_{-\pi}^{\pi} |g(\lambda)|^2 d\lambda < \infty\}$
and $\int_{-\pi}^{\pi} log|ce^{i\theta}|d\theta > -\infty$

Remark (Back to Wold). $\sigma^2 > 0$ let $\psi(e^{-i\lambda}) = \sum_{n=0}^{\infty} \psi_j e^{-in\lambda}$
By Theorem Rudin $\int_{-\pi}^{\pi} log|\psi(e^{-i\lambda})|d\lambda > -\infty \implies \psi(e^{-i\lambda})$ is non-zero almost everywhere on $(-\pi, \pi]$. Let $U_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ then U_t and V_t (deterministic part) are orthogonal.
Let $f_U(\lambda)$ and $f_V(\lambda)$ be the spectral density functions of U_t and V_t respectively.
Then $f_X(\lambda) = f_U(\lambda) + f_V(\lambda)$ and $f_U(\lambda)$ and $f_V(\lambda)$ are mutually singular.
Can show that $\frac{\sigma^2}{2\pi}|\psi(e^{-i\lambda})|^2$ is the spectral density function of U_t
$\implies F_u$ is absolutely continuous almost everywhere with non-zero, non-negative density function.
F_v is singular with respect to F_u and singular to the Lebesgue measure.
IE $\exists$ a set A with Lebesgue measure 0 such that $F_v(A^c) = 0$

1.19 11/06/2025 Lecture 19

Theorem 21 (5.6.1 Wold Decomposition). *If* $\sigma^2 > 0$ *then* X_t *can we expressed as* $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t$ *where*

1. $\psi_0 = 1$, $\sum_{j=0}^{\infty} |\psi_j|^2 < \infty$
2. $Z_t \sim WN(0, \sigma^2)$
3. $Z_t \in \mathcal{M}_t, \forall t \in \mathbb{Z}$
4. $E[Z_t V_s] = 0, \forall t, s \in \mathbb{Z}$
5. $V_t \in \mathcal{M}_{-\infty}, \forall t \in \mathbb{Z}$
6. V_t *is deterministic stationary process*

The sequences $\{\psi_j\}$, $\{Z_t\}$ *and* $\{V_t\}$ *are uniquely determined by the assumptions above.*

Remark.
1. The spectral density function of $U_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ is absolutely continuous wrt the Lebesgue measure ie $F_U(\lambda) = \int_{-\pi}^{\lambda} f_U(\nu)d\nu$ and we know it is given by $f_U(\lambda) = \frac{\sigma^2}{2\pi}|\psi(e^{-i\lambda})|^2$ where $\psi(e^{-i\lambda}) = \sum_{j=0}^{\infty} \psi_j e^{-ij\lambda}$. Furthermore $\psi(e^{-i\lambda}) \neq 0$ almost everywhere on $(-\pi, \pi]$.

2. F_U and F_V are mutually singular ie $\exists$ a set A with Lebesgue measure 0 such that $F_V(A^c) = 0$ and $F_U(A) = 0$

3. $F_X = F_U + F_V$ is the Lebesgue decomposition of F_X wrt the Lebesgue measure.

Let F_X be the spectral distribution function of X_t. Let $f_X(\theta)$ be its derivative, then we can write $F_X(\lambda) = \int_{-\pi}^{\lambda} f_X(\theta)d\theta + F_s(\lambda)F_c(\lambda) + F_s(\lambda)$ where F_s is also a distribution functions

1. $f_x(\theta) = 0$ on a set of positive measure

2. $f_x(\theta) > 0$ almost everywhere $\int_{-\pi}^{\pi} log f_X(\theta)d\theta = -\infty$

3. $f_x(\theta) > 0$ almost everywhere $\int_{-\pi}^{\pi} log f_X(\theta)d\theta > -\infty$

Theorem 22 (Kolmogorov Formula)**.** *$\sigma^2 > 0$ iff $\int_{-\pi}^{\pi} log f_X(\lambda)d\lambda > -\infty$ and $\sigma^2 = 2\pi \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} log f_X(\lambda)d\lambda\right)$*

Example.

$$(1 - 0.5B)X_t = (1 + 0.3B)Z_t \quad Z_t \sim WN(0, 1)$$

$f_x(\theta) = \frac{\sigma^2}{2\pi} \cdot \frac{|1+0.3e^{-i\theta}|^2}{|1-0.5e^{-i\theta}|^2}$

$$\log(f_x(\theta)) = \log\left(\frac{\sigma^2}{2\pi}\right) + 2\log|1 + 0.3e^{-i\theta}| - 2\log|1 - 0.5e^{-i\theta}|$$
$$\implies \int_{-\pi}^{\pi} \log(f_x(\theta))d\theta = 2\pi \log\left(\frac{\sigma^2}{2\pi}\right)$$

Remark. Why? are we doing this.
$X_t = \sum_{j=1}^{n} A_j e^{i\lambda_j t}$ where A_j are uncorrelated mean zero random variables.

Remark (Infrence For ARMA Models)**.** Not only we do we have coefficients ϕ and θ to estimate but we also have to determine the order of the model p and q.
Suppose we get data for ARMA(p,q): $\phi(B)(X_t - \mu) = \theta(B)Z_t$ where $Z_t \sim WN(0, \sigma^2)$
Assume p, q are known.
We estimate mean μ, by $\hat{\mu} = \bar{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t$
Suppose $\{X_1 \dots X_n\}$ is a realization of a sp.

$$\hat{\mu} = \bar{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t$$
$$\bar{X}_n \xrightarrow{a.s.} E[X_0|\mathcal{M}_{-\infty}] \text{ as } n \to \infty$$

When $\bar{X} \xrightarrow{a.s} \mu$ then we call the process ergodic in the mean.

1. IID sequence with finite mean is ergodic
2. Suppose Z_t iid
3. $X_t := g(Z_t, Z_{t-1}, \ldots)$ and g is measurable and $E[|X_t|] < \infty$ then X_t is ergodic.

1.20 11/11/2025 Lecture 20

Remark (Infrence for ARMA models)**.** We need to look at Strictly stationary processes
Suppose $\{X_t\}$ is strictly stationary process with mean μ and autocovariance function $\gamma_X(h)$
If $\{X_t\}$ is ergoic then $\bar{X}_n := \frac{1}{n}\sum_{t=1}^{n} X_t \xrightarrow{a.s.} \mu$ as $n \to \infty$
This can be thought of as the time average converging to the space average.
IF $\{Z_t\}$ is IID then it is ergodic.
If $X_t = g(Z_t, Z_{t-1}, \ldots)$ then X_t is ergodic and strictly stationary.
For the causal and invertable ARMA(p,q) $\phi(B)(X_t - \mu) = \theta(B)Z_t$ where $Z_t \sim WN(0, \sigma^2)$
Then $X_t = \mu + \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, so it is strictly stationary and ergodic.
In particular $\bar{X}_n \xrightarrow{a.s.} \mu$ as $n \to \infty$
Sample Autocovariance Function:

$$\begin{aligned}
\hat{\gamma}(h) &= \frac{1}{n}\sum_{t=h+1}^{n}(X_t - \bar{X}_n)(X_{t-h} - \bar{X}_n) \quad h \geq 0 \\
&= \frac{1}{n}\sum_{t=h+1}^{n} X_t X_{t-h} - \frac{n-h}{n}\bar{X}_n^2 - \bar{X}_n \cdot \frac{1}{n}\sum_{t=h+1}^{n}(X_t + X_{t-h}) \\
&= \frac{1}{n}\sum_{t=h+1}^{n} X_t X_{t-h} - \mu^2 - 2\mu^2 \quad \text{Note that } X_t X_{t-h} \text{ is ergodic}
\end{aligned}$$

Therefore $\frac{1}{n}\sum_{t=h+1}^{n} X_t X_{t-h} \xrightarrow{a.s.} E[X_t X_{t-h}] = \gamma(h) + \mu^2$
Thus $\hat{\gamma}(h) \xrightarrow{a.s.} \gamma(h)$ as $n \to \infty$

$$X_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$$

$$\text{If IID} \quad \sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \gamma_X(0))$$

$$\operatorname{Var}(\bar{X}_n) = \frac{1}{n}\gamma_X(0) \operatorname{Var}(n\bar{X}_n) = n\gamma_X(0) + 2\sum_{h=1}^{n-1}\left(1 - \frac{h}{n}\right)\gamma_X(h)$$

$$\operatorname{Var}(\sqrt{n}(\bar{X}_n - \mu)) = \sum_{-\infty}^{\infty}\gamma_X(h) \text{ if converges}$$

CLT for $\bar{X}_n$:

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N\left(0, \tau^2 := \sum_{h=-\infty}^{\infty}\gamma_X(h)\right)$$

IF $X_t = \mu + \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ where $Z_t \sim IID(0, \sigma^2)$ and $\sum_{j=0}^{\infty} |\psi_j| < \infty$ then the CLT holds and the spectral density function is of the form:

$$\begin{aligned}
f_X(\lambda) &= \frac{\sigma^2}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_X(h) e^{-ih\lambda} \\
f_X(0) &= \frac{\sigma^2}{2\pi} \left(\sum_{j=0}^{\infty} \psi_j \right)^2 = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_X(h) \\
\tau^2 &:= \sum_{h=-\infty}^{\infty} \gamma_X(h) = 2\pi f_X(0)
\end{aligned}$$

For $\hat{\gamma}_X(h)$ we need to have $E[X_t^4] < \infty$. and for $\hat{\rho}_X(h)$ we need $E[X_t^2] < \infty$
Book has information of these CLTs
In particular of $\{X_t\}$ is IID then

$$\sqrt{n} \begin{bmatrix} \hat{\rho}(1) \\ \hat{\rho}(2) \\ \vdots \\ \hat{\rho}(m) \end{bmatrix} \xrightarrow{d} N(0, I_m)$$

Remark. $X_t = \phi_0 + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$
where $Z_t \sim IID(0, \sigma^2)$
$\phi_p \neq 0, \theta_q \neq 0$
$\phi(z)\theta(z) \neq 0$ for $|z| = 1$

$$\mu = \frac{\phi_0}{1 - \phi_1 - \ldots - \phi_p}$$

Center the data: $X_t = X_t - \bar{X}_t$ and fit the model $\phi(B)X_t = \theta(B)Z_t$
Parameter Space: Let $\beta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q) \in \mathbb{R}^{p+q}$
$\mathbb{C} = \{\beta \in \mathbb{R}^{p+q} | \phi(z) \neq 0, \theta(z) \neq 0 \text{ for } |z| \leq 1\}$
Consider AR(p): $X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + Z_t$

$$\begin{aligned}
X_{p+1} &= \phi_1 X_p + \ldots + \phi_p X_1 + Z_{p+1} \\
X_{p+2} &= \phi_1 X_{p+1} + \ldots + \phi_p X_2 + Z_{p+2} \\
&\vdots \\
X_n &= \phi_1 X_{n-1} + \ldots + \phi_p X_{n-p} + Z_n
\end{aligned}$$

Now we can do least squares to estimate $\phi_1, \ldots, \phi_p$

$$\begin{bmatrix} X_{p+1} \\ X_{p+2} \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} X_p & X_{p-1} & \dots & X_1 \\ X_{p+1} & X_p & \dots & X_2 \\ \vdots & \vdots & \ddots & \vdots \\ X_{n-1} & X_{n-2} & \dots & X_{n-p} \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} + \begin{bmatrix} Z_{p+1} \\ Z_{p+2} \\ \vdots \\ Z_n \end{bmatrix}$$

LSE of $\phi = (\phi_1, \dots, \phi_p)$ is given by:

$$\hat{\phi}_{LSE} = (X^T X)^{-1} X^T Y$$

Where X is the design matrix and Y is the response vector.

Remark (Yule Walker Estimator)**.** Also consider the Yule-Walker Estimator for AR(p)

$$\begin{aligned} \gamma(1) &= \phi_1\gamma(0) + \phi_2\gamma(-1) + \ldots + \phi_p\gamma((p-1)) \\ \gamma(2) &= \phi_1\gamma(1) + \phi_2\gamma(0) + \ldots + \phi_p\gamma((p-2)) \\ &\vdots \\ \gamma(p) &= \phi_1\gamma(p-1) + \phi_2\gamma(p-2) + \ldots + \phi_p\gamma(0) \end{aligned}$$

$$\begin{aligned} \hat{\Gamma}_p\phi &= \hat{\gamma}_p \\ \hat{\phi}_{YW} &= \hat{\Gamma}_p^{-1}\hat{\gamma}_p \end{aligned}$$

1.21 11/13/2025 Lecture 21

Remark. $\phi(B)X_t = \theta(B)Z_t$ where $Z_t \sim IID(0, \sigma^2)$
$\beta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q) \in \mathbb{R}^{p+q}$
$\mathbb{C} = \{\beta \in \mathbb{R}^{p+q} | \phi(z) \neq 0, \theta(z) \neq 0 \text{ for } |z| \leq 1\}$
Consider AR(p): $X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + Z_t$
Yule Walker Estimator:

$$\hat{\phi}_{YW} = \hat{\Gamma}_p^{-1}\hat{\gamma}_p$$

Where $\hat{\Gamma}_p = [\hat{\gamma}(i-j)]_{i,j=1}^p$ and $\hat{\gamma}_p = [\hat{\gamma}(1), \dots, \hat{\gamma}(p)]^T$
LSE Estimator:

$$\hat{\phi}_{LSE} = (X^T X)^{-1} X^T Y$$

Where X is the design matrix and Y is the response vector.
Define $Y = [X_{p+1}, X_{p+2}, \dots, X_n]^T$ and X is the matrix with rows $[X_{t-1}, X_{t-2}, \dots, X_{t-p}]$ for

$t = p+1, \dots, n$
Let Y^* and X^* be the time starting at 0 values.
$\phi^* = (X^{*T}X^*)^{-1}X^{*T}Y^*$
Then $\hat{\phi}_{LSE} - \phi = \phi^* - \phi + o_p(1/\sqrt{n})$
Note: $\frac{1}{n}X^{*T}X^* \xrightarrow{a.s.} \Gamma_p$ as $n \to \infty$
$\implies (\frac{1}{n}X^{*T}X^*)^{-1} \xrightarrow{a.s.} \Gamma_p^{-1}$ as $n \to \infty$

Proposition 11. *IF $\{X_t\}$ is a causal AR(p) with $Z_T \sim IID(0, \sigma^2)$ then*

$$\sqrt{n}(\phi^* - \phi) \xrightarrow{d} N(0, \sigma^2\Gamma_p^{-1}) \text{ as } n \to \infty$$

This also does not depend on σ^2

Proof.

$$\begin{aligned}
\phi^* &= (X^{*T}X^*)^{-1}X^{*T}Y^* \\
&= (X^{*T}X^*)^{-1}(X^{*T}X^*\phi + X^{*T}Z) \\
&= \phi + (X^{*T}X^*)^{-1}X^{*T}Z \\
\implies \phi^* - \phi &= (\frac{1}{n}X^{*T}X^*)^{-1} \cdot \frac{1}{\sqrt{n}}X^{*T}Z \cdot \frac{1}{\sqrt{n}} \\
\sqrt{n}(\phi^* - \phi) &= (\frac{1}{n}X^{*T}X^*)^{-1} \cdot \frac{1}{\sqrt{n}}X^{*T}Z
\end{aligned}$$

Will do CLT on $\frac{1}{\sqrt{n}}X^{*T}Z$
Will apply Slutsky's Theorem
$X^{*T}Z = \sum_{t=1}^n Z_t \begin{bmatrix} X_{t-1} \\ X_{t-2} \\ \vdots \\ X_{t-p} \end{bmatrix}$
Let $\mathcal{F}_t = \sigma(X_{1-p}, \dots, X_t)$
Then $\mathcal{F}_t \subset \mathcal{F}_{t+1}$
And $S_t := \sum_{k=1}^t Z_k \begin{bmatrix} X_{k-1} \\ X_{k-2} \\ \vdots \\ X_{k-p} \end{bmatrix} \in \mathcal{F}_t$
Note that

$$\begin{aligned}
E[S_t|\mathcal{F}_{t-1}] &= E[S_{t-1} + Z_tX_t|X_{1-p}, \dots, X_{t-1}] \\
&= S_{t-1}
\end{aligned}$$

by the previous staements, this shows that $\{S_t, \mathcal{F}_t\}$ is a martingale.
IE increasing sequence of sigma algebras and $E[|S_t|] < \infty$ and $E[S_t|\mathcal{F}_{t-1}] = S_{t-1}$

Apply Martingale CLT:

$$\frac{1}{n}\sum_{t=1}^{n} E\left[Z_t X_{t-i} X_{t-j} | \mathcal{F}_{t-1}\right] = \frac{1}{n}\sum_{t=1}^{n} X_{t-i}X_{t-j}E[Z_t^2|\mathcal{F}_{t-1}]$$
$$= \sigma^2 \cdot \frac{1}{n}\sum_{t=1}^{n} X_{t-i}X_{t-j} \xrightarrow{a.s./p} \sigma^2\gamma_X(i-j) \text{ as } n \to \infty$$

Additionally, the Lindeburg condition holds since

$$\frac{1}{n}\sum_{t=1}^{n} E\left[Z_t^2||X_t||^2 \cdot I\left(||X_t||^2|Z_t|^2 > \epsilon n\right)|\mathcal{F}_{t-1}\right] \xrightarrow{a.s.} 0 \text{ as } n \to \infty$$

By the MCLT we have $\frac{1}{\sqrt{n}}S_n \xrightarrow{d} N(0, \sigma^2\Gamma_p)$ as $n \to \infty$
By Slutsky's Theorem we have:

$$\sqrt{n}(\phi^* - \phi) = (\frac{1}{n}X^{*T}X^*)^{-1} \cdot \frac{1}{\sqrt{n}}X^{*T}Z \xrightarrow{d} N(0, \sigma^2\Gamma_p^{-1}) \text{ as } n \to \infty$$

□

Remark. The LSE and Yule-Walker estimators are asymptotically equivalent ie

$$\sqrt{n}(\hat{\phi}_{LSE} - \hat{\phi}_{YW}) \xrightarrow{p} 0 \text{ as } n \to \infty$$

1.22 11/18/2025 Lecture 22

Remark. AR(p) $\phi(B)X_t = Z_t$ where $Z_t \sim IID(0, \sigma^2)$
Let $\hat{\phi}$ be either the LSE or Yule-Walker estimator.
Then $\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{d} N(0, \sigma^2\Gamma_p^{-1})$ as $n \to \infty$
Remark: sure fits an $AR(m)$ with $m > p$ and get the estimator $\hat{\phi}_m$
Then we can say that $\hat{\phi}_m \xrightarrow{p} \phi_m$ as $n \to \infty$.
And $\sqrt{n}(\hat{\phi}_m - \phi_m) \xrightarrow{d} N(0, \sigma^2\Gamma_m^{-1})$ as $n \to \infty$
If $m > p$ then the (m, m) entry of Γ_m^{-1} is 1

Remark (PACF)**.** For any statuionary process $\{X_t\}$ solve

$$\min_{c_1,\ldots,c_p} E\left[\left(X_t - \sum_{j=1}^{p} c_j X_{t-j}\right)^2\right]$$

then $\hat{C}_p$ is the lag-p partial autocorrelation coefficient ie the coefficient of X_{t-p} in the above minimization problem.

$$c^* = \Gamma_p^{-1}\gamma_p$$

Estimator $\hat{\phi}_{p,p}$ is the last element of $\hat{\Gamma}_p^{-1}\hat{\gamma}_p$
Now back to the AR(p) model if we fit AR(m) to the data then $\hat{\phi}_{m,m}$ satisfy: $\sqrt{n}\hat{\phi}_{m,m} \xrightarrow{d} N(0,1)$ as $n \to \infty$
I applied timer seireis: plot data, plot acf, plot pacf

Remark. ARMA(p,q): $\phi(B)X_t = \theta(B)Z_t$ where $Z_t \sim IID(0, \sigma^2)$

$$\begin{aligned}\mathbb{C} = \{\beta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q) :\\ \phi(z)\theta(z) \neq 0 \text{ for } |z| \leq 1;\\ \phi(z)\theta(z) \neq 0 \text{ for } |z| \leq 1;\\ \phi_p\theta_q \neq 0;\\ \phi(z) \text{ and } \theta(z) \text{ have no common zeros}\}\end{aligned}$$

We cannot use LSE or Yule-Walker to estimate parameters.
Instead we use Maximum Likelihood Estimation (MLE) with the prior that $Z_t \sim N(0, \sigma^2)$
Then X_t is a Gaussian process.
We can also consider the Quasi-Maximum Likelihood Estimation (QMLE) where we still use the Gaussian likelihood even if Z_t is not Gaussian.

Definition (Gaussian likelihood)**.** Let $\Gamma_n(\beta)$ be the $n \times n$ autocovariance matrix of $\mathbf{X}_n = (X_1, \ldots, X_n)^T$ where X_t is the ARMA(p,q) process with parameter $\beta \in \mathbb{C}$.
Then the Gaussian likelihood function is given by:

$$L(\mathbf{X}, \beta) := (2\pi)^{-\frac{n}{2}} |\Gamma_n(\beta)|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{X}_n^T \Gamma_n(\beta)^{-1}\mathbf{X}_n\right)$$

The QMLE finds the $\hat{\beta}$ and $\hat{\sigma}^2$ that maximize the Gaussian likelihood function.
Can write $\Gamma_n(\beta, \sigma^2) = \sigma^2 \cdot G_n(\beta)$ where $G_n(\beta)$ is the correlation matrix of $\mathbf{X}_n$.
then

$$L(\mathbf{X}_n, \beta, \sigma^2) := (2\pi\sigma^2)^{-\frac{n}{2}} |G_n(\beta)|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}\mathbf{X}_n^T G_n(\beta)^{-1}\mathbf{X}_n\right)$$

We can maximize wrt σ^2 first to get $\hat{\sigma}^2(\beta) = \frac{1}{n}\mathbf{X}_n^T G_n(\beta)^{-1}\mathbf{X}_n$
Then we can maximize the profile likelihood:

$$l(\mathbf{X}_n, \beta) = -\frac{n}{2}\log\left[\frac{1}{n}\mathbf{X}_n^T G_n(\beta)^{-1}\mathbf{X}_n\right] - \frac{1}{2}\log|G_n(\beta)|$$

Remark (How to find QMLE)**.** minimize

$$\min_{\beta \in \mathbb{C}} \frac{1}{n}\log|G_n(\beta)| + \log\left[\frac{1}{n}\mathbf{X}_n^T G_n(\beta)^{-1}\mathbf{X}_n\right]$$

And also take $\hat{\sigma}^2 = \frac{1}{n}\mathbf{X}_n^T G_n(\hat{\beta})^{-1}\mathbf{X}_n$
Method 1:
Predict X_1 using no data, predict X_2 using X_1, predict X_3 using X_1, X_2 and so on.
let v_{n-1} be the predictor of the next step using $n-1$ observations, i.e., $v_{n-1} = E[X_n \mid X_1, \ldots, X_{n-1}]/\sigma^2$.

$$\begin{bmatrix} X_1 - \hat{X}_1 \\ X_2 - \hat{X}_2 \\ \vdots \\ X_n - \hat{X}_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ -\phi_{11} & 1 & 0 & \ldots & 0 \\ -\phi_{21} & -\phi_{22} & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\phi_{n-1,1} & -\phi_{n-1,2} & -\phi_{n-1,3} & \ldots & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

$$\hat{X}_{j+1} = \hat{\phi}_j \begin{bmatrix} X_j \\ X_{j-1} \\ \vdots \\ X_1 \end{bmatrix}$$

1.23 11/20/2025 Lecture 23

Remark. ARMA(p,q): $\phi(B)X_t = \theta(B)Z_t$ where $Z_t \sim IID(0, \sigma^2)$
Find the BLP of X_{n+1} using $X_1, \ldots, X_n$
Ler $v_n = E[X_{n+1} - \hat{X}_{n+1}]^2$ be the mean squared error of the predictor and $r_j = \frac{v_j}{\sigma^2}$

- $X_{j+1} - \hat{X}_{j+1}$ is independent across j

- The joint densities of $\mathbf{X}_n$ and $[X_{n+1} - \hat{X}_{n+1}]$ are the same

$L(X_n, \beta, \sigma^2) = \prod_{j=1}^n \frac{1}{2\pi\sqrt{\sigma^2 r_{j-1}}} \exp\left(-\frac{(X_j - \hat{X}_j)^2}{2\sigma^2 r_{j-1}}\right)$
$L(X_n, \beta, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \left(\prod_{j=0}^{n-1} r_j\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^n \frac{(X_j - \hat{X}_j)^2}{r_{j-1}}\right)$
Maximize wrt σ^2 first to get $\hat{\sigma}^2(\beta) = \frac{1}{n}\sum_{j=1}^n \frac{(X_j - \hat{X}_j)^2}{r_{j-1}}$
Then we can maximize the log likelihood: (Let $S_n(\beta) = \sum_{j=1}^n \frac{(X_j - \hat{X}_j)^2}{r_{j-1}}$)

$$l(X_n, \beta) = -\frac{n}{2}\log\left[\frac{1}{n}S_n(\beta)\right] - \frac{1}{2}\sum_{j=0}^{n-1}\log r_j$$

To find MLE we minimize:

$$\min_{\beta \in \mathbb{C}} \frac{1}{n}\sum_{j=1}^n \frac{(X_j - \hat{X}_j)^2}{r_{j-1}} + \frac{1}{n}\sum_{j=0}^{n-1}\log r_j$$

And we can instead take the minimization problem:

$$\min_{\beta \in \mathbb{C}} S_n(\beta) = \sum_{j=1}^{n} \frac{(X_j - \hat{X}_j)^2}{r_{j-1}}$$

Remark. Method 2 Periodogram Method
Suppose we have data $X_1, \dots, X_n$ from a mean zero stationary process.

Definition (10.1.2)**.** The periodogram function is defined as:

$$\omega_j = \frac{2\pi j}{n}$$
$$I(\omega_j) = \frac{1}{n} \left| \sum_{t=1}^{n} X_t e^{-i\omega_j t} \right|^2 \quad -\pi < \omega_j \leq \pi, j = 1, 2, \dots, n$$

The set $F_n := \{j \in \mathbb{Z} : -\pi < \omega_j \leq \pi, j = 1, 2, \dots, n\}$ are called the Fourier frequencies. $|F_n| = n$

Proposition 12 (10.1.2)**.** *For each $j \in F_n$*

$$I(\omega_j) = \sum_{k=-(n-1)}^{n-1} \hat{\gamma}(k) e^{-ik\omega_j}$$

where $\hat{\gamma}(k)$ is the sample autocovariance function. We can see that the periodogram is a good estimator for the spectral density function since:

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-ik\lambda}$$
$$I(\omega_j) = \sum_{k=-(n-1)}^{n-1} \hat{\gamma}(k) e^{-ik\omega_j}$$

$I(\omega_j)$ corresponds to an estimation of the Spectral Density Function at the Fourier frequencies.
$E[I(\omega_j)] = 2\pi f(\omega_j)$ as $n \to \infty$
Let $\alpha(\omega_j) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t \cos(\omega_j t)$ and $\beta(\omega_j) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t \sin(\omega_j t)$
Then $I(\omega_j) = \alpha(\omega_j)^2 + \beta(\omega_j)^2$
Note that $\alpha(\omega_j)$ and $\beta(\omega_j)$ are asymptotically independent
Let $G(\lambda, \beta) = \frac{1}{2\pi} \left| \frac{\theta(e^{-i\lambda})}{\phi(e^{-i\lambda})} \right|^2 = \frac{1}{\sigma^2} f(\lambda, \beta)$
Then $\alpha(\omega_j) \approx N(0, \frac{\sigma^2}{2} g(\omega_j, \beta)$ We then reach to the whittle likelihood function:

$$\min_{\beta \in \mathbb{C}} \sum_{j \in F_n} \left[\frac{I(\omega_j)}{2\pi g(\omega_j, \beta)} + \log g(\omega_j, \beta) \right]$$

1.24 11/25/2025 Lecture 24

Theorem 23. *Let $\{X_1, \dots, X_n\}$ be a relization of a ttritly stationary ARMA(p,q) model with $Z_t \sim IID(0, \sigma^2)$ and $\beta \in \mathbb{C}$ be the parameters of the model.*
Then let $\hat{\beta}$ be the MLE, LSE, or Whittle Estimator then

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V(\beta)) \text{ as } n \to \infty$$

Where $V(\beta) = \sigma^2 \begin{bmatrix} E[U_t U_t^T] & E[U_t V_t^T] \\ E[V_t U_t^T] & E[V_t V_t^T]^{-1} \end{bmatrix}$ *With* $U_t = (U_{t-1}, \dots, U_{t-p})^T$ *and* $V_t = (V_{t-1}, \dots, V_{t-q})^T$ *and* $\phi(\beta)U_t = Z_t$ *and* $\theta(\beta)V_t = Z_t$ *with the same* Z_t *for both equations.*
Equivalently

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, W^{-1}(\beta))$$
$$W(\beta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} [\nabla \log g(\lambda, \beta)] [\nabla \log g(\lambda, \beta)]^T \, d\lambda$$
$$g(\lambda, \beta) = \frac{\sigma^2}{2\pi} \left| \frac{\theta(e^{-i\lambda}, \beta)}{\phi(e^{-i\lambda}, \beta)} \right|^2$$

Remark. Suppose $\{X_t\}$ is a SS process with mean μ

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \tau^2) \text{ as } n \to \infty$$

Where $\tau^2 = \sum_{h=-\infty}^{\infty} \gamma_X(h)$
$\hat{\tau} = \sum_{k=-h_n}^{h_n} \hat{\gamma}_X(k)$ where $h_n \to \infty$ and $\frac{h_n}{n} \to 0$ as $n \to \infty$
Then

Remark (batch estimator)**.** Suppose $n = bm$ choose $m_n \to \infty$ and $\frac{m_n}{n} \to 0$ as $n \to \infty$ let

$$\begin{aligned} X_1, \dots, X_m \\ X_{m+1}, \dots, X_{2m} \\ \vdots \\ X_{(b-1)m+1}, \dots, X_{bm} \end{aligned}$$

Let $\bar{X}_k = \frac{1}{m} \sum_{t=(k-1)m+1}^{km} X_t$ for $k = 1, 2, \dots, b$
Then

$$\sqrt{m}(\bar{X}_k - \bar{X}) \xrightarrow{d} N(0, \tau^2) \text{ as } m \to \infty$$
$$\hat{\tau}^2 = \frac{1}{b} \sum_{k=1}^{b} m(\bar{X}_k - \bar{X})^2$$

Remark (Spectral Approach)**.**

$$\tau^2 = \sum_{h=-\infty}^{\infty} \gamma_X(h) = 2\pi f_X(0)$$
$$f_X(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_X(h) e^{-ih\lambda}$$

The periodogram:

$$I(\omega_j) = \frac{1}{n} \left| \sum_{t=1}^{n} X_t e^{-i\omega_j t} \right|^2 \quad \text{where } \omega_j = \frac{2\pi j}{n}, j = 1, 2, \dots, n \quad -\pi < \omega_j \leq \pi$$

$E[I(\omega_j)] \approx 2\pi f_X(\omega_j)$ as $n \to \infty$
$\frac{1}{2\pi 2B_n+1} \sum_{j:|\omega_j| \leq B_n} I(\omega_j) \approx \hat{\tau}_3^2$ smoothing the periodogram

Remark (multivariate Time Series)**.** $\{\underline{X_t} : t \in \mathbb{Z}\}$ where $\underline{X_t} = (X_{t,1}, X_{t,2}, \dots, X_{t,m})^T$ is an m-dimensional vector.

$$E[\underline{X_t}^2] \leq \infty \text{ for all } t \in \mathbb{Z}$$
$$\mu_{ti} = E[X_{t,i}] \quad i = 1, 2, \dots, m$$
$$\gamma_{ij}(t+h, t) = Cov(X_{t+h,i}, X_{t,j}) = E[(X_{t+h,i} - \mu_{t+h,i})(X_{t,j} - \mu_{t,j})]$$
$$\Gamma(t+h, t) = [\gamma_{ij}(t+h, t)]_{i,j=1}^{m} \text{ is the autocovariance/crosscovariance matrix}$$

Definition. The time series $\{\underline{X_t}\}$ is said to be staitionary if

- $E[\underline{X_t}] = \underline{\mu_0}$ for all $t \in \mathbb{Z}$ Time Invariant Mean
- $\Gamma(t+h, t) = \Gamma(h, 0) = \Gamma(h)$ for all $t, h \in \mathbb{Z}$ Time Invariant Covariance

Also consider $R(h) = [\rho_{ij}(h)]_{i,j=1}^{m}$ where $\rho_{ij}(h) = \frac{\gamma_{ij}(h)}{\sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}}$ is the autocorrelation/crosscorrelation matrix.

- $\gamma_{ii}(0)$ is the autocovariance fn of the univariate sereis $\{X_{t,i}\}$
- $|\gamma_{ij}(h)| \leq \sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}$ by Cauchy-Schwarz Inequality
- $\Gamma(h) = [\Gamma(-h)]^T$ for all $h \in \mathbb{Z}$
- $sum_{j,k=1}^{n} a_j^T \Gamma(j-k) a_k \geq 0$ for all $n \in \mathbb{N}$ and $a_1, \dots, a_n \in \mathbb{R}^m$ (non-negative definite)
- Strict Staitionarity is defined the same

1.25 12/02/2025 Lecture 25

Definition (11.3.1 Multivariate ARMA(p,q) Models)**.** $\{\underline{X_t}, t \in \mathbb{Z}\}$ is an m-variate ARMA(p,q) process if
$\{\underline{X_t}\}$ is the stationary solution of the difference equations:

$$X_t - \Phi_1 X_{t-1} - \ldots - \Phi_p X_{t-p} = Z_t + \Theta_1 Z_{t-1} + \ldots + \Theta_q Z_{t-q}$$
$$\Phi(B)\underline{X_t} = \Theta(B)\underline{Z_t}$$

Where Φ_i and Θ_j are $m \times m$ coefficient matrices for $i = 1, \ldots, p$ and $j = 1, \ldots, q$
And $\{\underline{Z_t}\} \sim WN(0, \Sigma_Z)$ is a multivariate white noise process with $E[Z_t] = 0$ and $Cov(Z_t, Z_s)\Sigma_Z$ for all $t \in \mathbb{Z}$
$\Phi(z) = I_m - \Phi_1 z - \ldots - \Phi_p z^p$ and $\Theta(z) = I_m + \Theta_1 z + \ldots + \Theta_q z^q$
Consider the example: for an m = 2 variate ARMA(2,2) process:

$$\underline{X_t} = \begin{bmatrix} 0.3 & 0.1 \\ 0.1 & 0.5 \end{bmatrix} \underline{X_{t-1}} + \begin{bmatrix} 0.1 & 0.05 \\ 0.03 & 0.1 \end{bmatrix} \underline{X_{t-2}} + \underline{Z_t} + \begin{bmatrix} 0.5 & 0.1 \\ 0.1 & 0.6 \end{bmatrix} + \underline{Z_{t-1}} + \begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.3 \end{bmatrix} \underline{Z_{t-2}}$$

This is a system of question with the cross terms.

Example (MV AR(1))**.** $X_t = \Phi X_{t-1} + Z_t$ where $Z_t \sim WN(0, \Sigma_Z)$
$\Phi = UDV'$ $P(\Phi) = \max(|\lambda_i| : |\lambda I - \Phi| = 0)$
$\Phi^k \to 0$ as $k \to \infty$ iff $P(\Phi) < 1$
Equivalently $|\Phi(z)| = |I - \Phi z|$ for $|z| \leq 1$

Theorem 24 (11.3.1)**.** *If* $\det \Phi(z) \neq 0$ *for* $|z| \leq 1$ *then* $\{\underline{X_t}\}$ *has the unique stationary solution of the form*

$$X_t = \sum_{j=0}^{\infty} \Psi_j Z_{t-j}$$

where $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j = \Phi^{-1}(z)\Theta(z)$ *for* $|z| \leq 1$
We can do a fast inversion of $\Phi(z)$ *by taking a ton of determinants*
Similarly we can see for invertability if $\det \Theta(z) \neq 0$ *for* $|z| \leq 1$ *then* X_t *is invertible in the sense that* $Z_t = \sum_{j=0}^{\infty} \Pi_j X_{t-j}$ *where* $\sum_{j=0}^{\infty} ||\Pi_j|| < \infty$ *and* $\Pi(z) = \sum_{j=0}^{\infty} \Pi_j z^j = \Theta^{-1}(z)\Phi(z)$ *for* $|z| \leq 1$

Example. If $Z_t \sim IIDN(0, \Sigma_Z)$ then $\{X_t\}$ is a Gaussian process.
If $\{X_t\}$ is gaussian with autocovariance function $\Gamma_X(h)$ and Y_t is another gaussian process with autocovariance function $\Gamma_Y(h)$ and $\Gamma_X(h) = \Gamma_Y(h)$ for all $h \in \mathbb{Z}$ then $\{X_t\}$ and $\{Y_t\}$ have the same distribution and same coeeficients
Note we cannot have the same "no same roots" condition as the univariate case since the determinant of a matrix polynomial can have multiple roots even if the matrix polynomial does not.

Remark (Estimation)**.** Mean: $\underline{\mu} = E[\underline{X_t}]$ can be estimated by $\bar{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t$
Autocovariance: $\Gamma(h) = Cov(X_{t+h}, X_t)$ can be estimated by

$$\hat{\Gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(X_{t+|h|} - \bar{X}_n)(X_t - \bar{X}_n)^T$$

For $VAR(p)$ model: $X_t = \Phi_1 X_{t-1} + \ldots + \Phi_p X_{t-p} + Z_t$ where $Z_t \sim WN(0, \Sigma_Z)$
We can use LSE or Yule-Walker to estimate the parameters.
For VARMA(p,q) model we can use MLE or QMLE to estimate the parameters.

1.26 12/04/2025 Lecture 26

Definition (Factor Models)**.** $X_i = AF_i + e_i$ where

- X_i is the observed p-variate time series
- F_i is the unobserved r-variate factor time series with $r < p$
- A is the $p \times r$ loading matrix
- e_i is the p-variate noise vector

$X_{ij} = \sum_{k=1}^{r} a_{jk} f_{ik} + e_{ij}$ for $j = 1, 2, \ldots, p$

Example (Model 1)**.** Assume f_t are iid normal with mean 0 and covariance I_r
Assume e_t are iid normal with mean 0 and covariance Σ_e where Σ_e is diagonal.
F_t and e_t are independent for all t
Assume $A'\sigma A$ is a diagonal matrix with distinct eigenvalues.
Then

$$\begin{aligned}\text{Cov}(X_t) &= A\,\text{Cov}(F_t)A^T + \text{Cov}(e_t)\\ &= AA^T + \Sigma_e\end{aligned}$$

Thus $X_T = AQQ'f_t + e_t$ for any orthogonal matrix Q since $QQ' = I_r$
$Q'A'\Sigma_e AQ$ is also diagonal with distinct eigenvalues.
Also m is fixed
This is the classical factor model

Example (Model 2: High Dimensional Factor Model)**.** M is large, $m \to \infty$ as $n \to \infty$
r is fixed
Assume f_t are iid normal with mean 0 and covariance I_r
Assume e_t are iid normal with mean 0 and covariance Σ_e for a general covariance matrix Σ_e
$A = UDV'$

1.27 12/09/2025 Lecture 27

Remark. consider $X_i = U\Lambda f_i + Z_i$ with r factors where

- X_i is the observed p-variate time series
- f_i is the unobserved r-variate factor time series with $r < p$
- U is the $p \times r$ loading matrix with orthonormal columns
- Λ is a $r \times r$ diagonal matrix with positive diagonal entries
- Z_i is the p-variate noise vector with covariance $\sigma^2 I_p$

Metric between teo suubspaces.
The let U and $\hat{U}$ be two $p \times r$ orthonormal matrices.
Then the distance between the column spaces of U and $\hat{U}$ is defined as:

$$d(U, \hat{U}) = ||UU' - \hat{U}\hat{U}'||_2$$

Theorem 25 (Wedins theorem). *Suppose $X = UDV' + Z$ and $\hat{X} = \hat{U}\hat{D}\hat{V}' + \hat{Z}$ where*

- *U and $\hat{U}$ are $p \times r$ orthonormal matrices*
- *V and $\hat{V}$ are $n \times r$ orthonormal matrices*
- *D and $\hat{D}$ are $r \times r$ diagonal matrices with positive diagonal entries*
- *Z and $\hat{Z}$ are noise matrices*

Then

$$||U_\perp' \hat{U}||_2 \leq \frac{Z||Z||}{\sigma_r}$$

Remark (Operator Norm: of Gaussian RM). Sudokov-Fermque Inequality:
Let $\{X\}$ and $\{Y\}$ be mean zero Gaussian random vectors.
Assume $E(X_t - X_s)^2 \leq E(Y_t - Y_s)^2$ for all s, t in the index set T.
Then

$$E\left[\sup_{t \in T} X_t\right] \leq E\left[\sup_{t \in T} Y_t\right]$$

Theorem 26 (7.3.1). *Let A be $d_1 \times d_2$ matrix with entries $A_{ij} \sim N(0, \sigma^2)$ iid. such that $E[\alpha' vec(A)] \leq ||\alpha||$ for all $\alpha \in \mathbb{R}^{d_1 d_2}$*
Then

$$E[||A||] \leq (\sqrt{d_1} + \sqrt{d_2})$$

Proof. $\square$