01:XXX:XXX - Homework n

Pranav Tikkawar

March 5, 2025

1 2/19

$$dB_{f(t)} = \sqrt{f'(t)}dB_t$$

Loo into hornstin olbeck process Understand Stationatiy (strong and weak) Resad about ARMA and ARIMA what they mean, what they do, and differences

1.1 Reading 2/19-2/26

Definition (Ergodic Property with a constant limit)**.** also known as EPCL

$$\exists \mu \in \mathbb{R} \text{ s.t. } \mathbb{P}(\lim_{n\to\infty} \bar{x} = \mu) = 1$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Definition (L^2 - Ergodic propetty with a constant limit)**.** also known as L^2-EPCL

$$\exists \mu \in \mathbb{R} \text{ s.t. } \lim_{n\to\infty} \mathbb{E}((\bar{x}-\mu)^2) = 0$$

EPLC doesnt hold in a lack of stablilty, high variabilty of marginal distributions, and absobsing states.

Definition (Strict Stationarity)**.** $X_t \in \mathbb{R}(t \in \mathbb{Z})$ is said to be strictly stationary if

$$\forall k \in \mathbb{Z}, \forall m \in \mathbb{N}, \forall t_1, t_2, \ldots, t_m \in \mathbb{Z}$$
$$(X)_{t_1}, (X)_{t_2}, \ldots, (X)_{t_m} =_d (X)_{t_1+k}, (X)_{t_2+k}, \ldots, (X)_{t_m+k}$$

In other words, the colllection of distributions of the random variables $(X)_{t_1}, (X)_{t_2}, \ldots, (X)_{t_m}$ is the same as the collection of distributions of the random variables shifted over $(X)_{t_1+k}, (X)_{t_2+k}, \ldots, (X)_{t_m+k}$

Example (Moverage Average Process of order q)**.** $X_t = \sum_{j=0}^{q} \psi_j \epsilon_{t-1} (t \in \mathbb{Z})$ where $\epsilon_t \in \mathbb{R}$ iid, $\psi_j \in \mathbb{R}$ and $j = 0, 1, \ldots, q$.

Definition (Weak Stationarity)**.** Let $X_t \in \mathbb{R}(t \in \mathbb{Z})$ s.t. $\forall t \in \mathbb{Z}, \mathbb{E}(|X_t|) < \infty$
Then $\mu_t = \mathbb{E}(X_t)$ is called the expected balue funtion or mean function of the process.
IF $E(x_t^2) < \infty$ then $\gamma : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$ with $\gamma(s,t) = \text{Cov}(X_s, X_t)$ is called the autocovariance function of the process. (acf)
Also $\mathbb{E}(X_s - \mu_s)(X_t - \mu_t) = \text{Cov}(X_s, X_t)$
Additionally

$$\rho(s,t) = \text{corr}(X_s, X_t) = \frac{\text{Cov}(X_s, X_t)}{\sqrt{\text{Var}(X_s)\text{Var}(X_t)}}$$

is called the autocorrelation function of the process. (acf)
We can also say something is weakly stationary if

$$\mathbb{E}X_t^2 < \infty$$
$$\exists \mu \in \mathbb{R} \text{ s.t. } \mathbb{E}(X_t) = \mu$$
$$\exists \gamma : \mathbb{Z} \to \mathbb{R} \text{ s.t. } \forall s, t \in \mathbb{Z} : \text{Cov}(X_s, X_t) = \gamma(t-s)$$

Definition (k-step prediction mean squared error)**.** Let $\mathscr{F}_{\leq t} = \sigma(X_s, s \leq t)$
$\mathscr{X}_t = \{Y | Y \in L^2(\Omega), \mathscr{F}_{\leq t} - \text{measurable}\}$
Then the k-step prediction mean squared error is defined for $k \geq 1$ and $Y \in \mathscr{X}_t$ as

$$\sigma_k^2(Y) = \mathbb{E}((X_{t+k} - Y)^2$$

ASK to explain thm 2.1 (pg 23) and thm 2.2 (pg 24)

Definition (Orthogonal Complement)**.** Let $(H, \langle \cdot, \cdot \rangle)$ be a Hilbert space and $A \subseteq H$ be a closed subspace of H.
The orthogonal complement of A is defined as

$$A^{\perp} = \{x : x \in H \text{ s.t. } \forall y \in A : \langle x, y \rangle = 0\}$$

We also write $A \perp B \iff \forall x \in A, y \in B : \langle x, y \rangle = 0$
Note that $A^{\perp}$ is a closed linear subspace of H

Dont understand the implications of corollary 2.3 pg 26

Definition (Optimal Forecast)**.** Let $X_t \in \mathbb{R}(t \in \mathbb{Z})$ be a weakly stationary process.

$$\begin{aligned} \mathscr{F}_{\leq t} &= \sigma(X_s : s \leq t) \\ \mathscr{X}_t &= \{Y | Y \in L^2(\Omega), \mathscr{F}_{\leq t} - \text{measurable}\} \\ k &\in \mathbb{N}, \hat{X}_{t+k} \in \mathscr{X}_t \end{aligned}$$

Then $\hat{X}_{t+k}$ is the optimal forecast of X_{t+k} given the information up to time t (i.e. $\mathscr{F}_{\leq t}$)

$$\iff$$

$$\sigma_k^2(\hat{X}_{t+k}) = \inf_{Y \in \mathscr{X}_t} \sigma^2(Y)$$

Definition (Conditional Expectation)**.** A $\mathscr{G}$-measurable random variable Y is said to be the conditional expectation of X given $\mathscr{G}$ if

$$\begin{aligned} & \mathbb{E}(X|\mathscr{G}) = Y \\ \iff \forall A \in \mathscr{G} : \int_A X dP &= \int_A Y dP \end{aligned}$$

Note: $\mathbb{E}(X|\mathscr{G})$ is a $\mathscr{G}$-measurable random variable.

Corollary.

$$\begin{aligned} X_t \in \mathbb{R}(t \in \mathbb{Z} &\text{ is a weakly stationary process}) \\ &\hat{X}_{t+k} \in \mathscr{X}_t \\ \sigma_k^2(\hat{X}_{t+k}) = \inf_{Y \in \mathscr{X}_t} \sigma^2(Y) \iff &\hat{X}_{t+k} = \mathbb{E}(X_{t+k} | \mathscr{F}_{\leq t}) \end{aligned}$$

In other words the optimal forecast of X_{t+k} given the information up to time t is the conditional expectation of X_{t+k} given $\mathscr{F}_{\leq t}$

Definition (Infinte linear past of X_t)**.**

$$L_t^0 = \left\{ Y | Y = \sum_{j=1}^{k} a_j X_{t_j}, k \in \mathbb{N}, a_j \in \mathbb{R}, t_j \in \mathbb{Z}, t_j \leq t \right\}$$

$$L_t = \overline{L_t^0} = \left\{ Y | \exists Y \in L_t^0 (n \in \mathbb{N}) \text{ s.t. } \lim_{n\to\infty} ||Y - Y_t||^2_{L^2(\Omega)} = 0 \right\}$$

$$L_{-\infty} = \bigcap_{t=-\infty}^{\infty} L_t = \text{infinite linear past of } X_t$$

Definition (Optimal Linear Forcast of X_{t+k} given $\mathscr{F}_{\leq t}$)**.** Let $X_t \in \mathbb{R}(t \in \mathbb{Z})$ be a weakly stationary process with $k \geq 1$ and $\hat{X}_{t+k} \in L_t$
Then

$$\begin{aligned} \hat{X}_{t+k} = \text{optimal linear forecast of } X_{t+k} \text{ given } \mathscr{F}_{\leq t} \\ \Longleftrightarrow \\ \sigma_k^2(\hat{X}_{t+k}) = \inf_{Y \in \mathscr{X}_t} \sigma^2(Y) \end{aligned}$$

Definition (Deterministic Stochastic)**.** Let $X_t \in \mathbb{R}(t \in \mathbb{Z})$ be a weakly stationary process. It is called deterministic if $\sigma_k^2(X_{t+k}) = 0$
More generally:

$$\forall t \in \mathbb{Z} : \inf_{Y \in L_t} \mathbb{E}\left[(Z_{t+1} - Y)^2 \right] = 0$$

Theorem 1 (Wold Decomposition Theorem)**.** *Let $X_t \in \mathbb{R}(t \in \mathbb{Z})$ be a weakly stationary process.*
Then

$$\exists a_0, a_1, a_2, \ldots \ s.t. \ a_0 = 1, \sum_{j=1}^{\infty} a_j^2 < \infty$$

and

$$\begin{aligned} \exists \epsilon_t, \mu_t (t \in \mathbb{Z}) \ s.t. \ \forall s, t \in \mathbb{Z} : \\ \epsilon_t \in L_t, \mu_t \in L_{-\infty} \\ E(\epsilon_t) = 0, Cov(\epsilon_s, \epsilon_t) \sigma_\epsilon^2 \delta_{s,t} \leq \infty, Cov(\epsilon_s, \mu_t) = 0 \\ X_{t_{a.s., L^2(\Omega)}} = \mu_t + \sum_{j=0}^{\infty} a_j \epsilon_{t-j} (t \in \mathbb{Z}) \end{aligned}$$

pg(26/34) This is literally fourier series for time series.
https://math.stackexchange.com/questions/703246/i-have-trouble-understanding-the-proof-of-the-wold-decomposition-theorem

Definition (Purely Stochastic or Regular)**.** We say if X_t has wold decomposition with $\mu_t \equiv \mu \in \mathbb{R}, \sigma_\epsilon^2 > 0$
Then X_t is called purely stochastic or regular.

Definition. Let $a = \{a_j\}_{j\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$, $0 < \beta < \infty$
Then

$$||a||_{\ell^\beta} = \left(\sum_{j=1}^{\infty} |a_j|^\beta\right)^{\frac{1}{\beta}} < \infty$$

Definition (Future and Asymptotic events)**.**

$$\begin{aligned} \mathscr{F}_t(t \in \mathbb{Z}) = &\text{ sequence of } \sigma\text{-algebras on } \Omega \\ \mathscr{F}_{>t} = \sigma\left(\bigcup_{s=t+1}^{\infty} \mathscr{F}_s\right) &\text{ future events} \\ \mathscr{F}^\infty = \bigcap_{t=1}^{\infty} \mathscr{F}_{>t} &\text{ asymptotic events} \end{aligned}$$

Definition (Ergodic Processes)**.** Let $X_t(t \in \mathbb{Z})$ be $\mathscr{F}_t$-measurable.

$$\begin{gathered} X_t \text{ is an Ergodic process} \\ \iff \\ \forall B \in \mathscr{F}^\infty : \mathbb{P}^2(B) = \mathbb{P}(B) \\ \iff \\ \forall B \in \mathscr{F}^\infty : \mathbb{P}(B) \in \{0, 1\} \end{gathered}$$

Theorem 2 (Kolmogorov's 0-1 Law)**.**

$$X_t(t \in \mathbb{Z}) \textit{ iid} \implies X_t \textit{ is ergodic}$$

Theorem 3 (Birkhoff's Ergodic Theorem)**.**

$$\begin{gathered} X_t(t \in \mathbb{Z}) \textit{ strictly stationary, ergodic}, \mathbb{E}(|X_t|) < \infty \\ \implies \\ \mu = E(X_t) \in \mathbb{R} \textit{ and } \overline{x} \to \mu \textit{ a.s.} \end{gathered}$$

2 2/26

Definition (Fockker Plank)**.**

$$\frac{dp(t,x)}{dt} = \frac{d}{dx}(\mu(t,x)p(t,x)) + \frac{d^2}{dx^2}(\sigma(t,x)p(t,x))$$

where $p(t,x)$ is the probability density function of the process X_t

Definition ($B_{f(t)}$)**.** $dB_{f(t)} = \sqrt{f'(t)}dB_t$

$$X_t = \frac{B_{e^t}}{\sqrt{e^t}}$$
$$dX_t = \frac{dB_{e^t}}{e^{t/2}} + \frac{B_{e^t}}{d}e^{-t/2} + \{$$

Look through example of Random walk +1,-1 and the linear past of it.

2.1 Reading 2/26-3/5

Definition (A1-A4)**.**

$$\forall t \in \mathbb{Z} : \mu_t = \mathbb{E}(X_t) \in \mathbb{R} \tag{1}$$
$$\forall t \in \mathbb{Z} : \sigma_t^2 = \mathrm{Var}(X_t) < \infty \tag{2}$$
$$\exists \mu \in \mathbb{R} \text{ s.t. } \lim_{t\to\infty} \mu_t = \mu \tag{3}$$
$$\lim_{t\to\infty} \mathrm{cov}(\bar{x_n}, X_n) = 0 \tag{4}$$

Weak stationarity implies A1-A3

$$\lim_{k\to\infty} \gamma_X(k) = 0$$
$$\gamma_X(k) \text{ Cesaro summable with limit } 0$$
$$\lim_{k\to\infty} \mathrm{Cov}(\bar{x_n}, X_n) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n-1} \gamma_X(k) = 0A4$$

This is an implication from top to bottom for A4

Definition (Backshift/ Lag Operator)**.**

$$B : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$$
$$x = \{x_t\}_{t\in\mathbb{Z}} \to y = Bx$$
$$\text{with } y_t = x_{t-1}$$

This shifts the sequence to the left by 1.

Definition (Linear Process)**.** $X_t \in \mathbb{R}(t \in \mathbb{Z}), \mathbb{E}(X_t) = \mu \in \mathbb{R}$ is a linear cprocess in $L^2(\Omega)$ if

$$\exists \epsilon_t \in \mathbb{R}(t \in \mathbb{Z}) \text{ iid } \mathbb{E}(\epsilon_t) = 0, \sigma_\epsilon^2 = \text{var}(\epsilon_t) < \infty$$

and

$$\exists a_j \in \mathbb{R}(j \in \mathbb{Z}), \sum_{j=-\infty}^{\infty} a_j^2 < \infty$$

s.t.

$$\forall t \in \mathbb{Z} : \lim_{N \to \infty} ||X_t - X_{t,N}||_{L^2(\Omega)}^2 = 0$$

We then write

$$X_t = \mu + \sum_{j=-\infty}^{\infty} a_j \epsilon_{t-j} (t \in \mathbb{Z}) = \mu + \left(\sum_{j=-\infty}^{\infty} a_j B^j \right) \epsilon_t$$

If $a_j = 0$ for $j < 0$ then we call it a causal linear process.
By definition if X_t is a linear process then it is weakly and strictly stationary.

NOTE: I am skipping most of the rest of the specific assumtions in favor of Reading more Probability Measures for Time Series