Lecture Notes #1

Time series analysis (STAT 5140/4140)

1 Definitions and probability review

Definition *A time series is a collection of observations made sequentially, usually in time. But observations may be collected in other domain as well (space etc.)*

Example Daily temperature data, Hourly satellite image.

Remark

1. Time series (TS) data may be collected regular irregular time intervals.
2. Many variables can be observed (Multivariate time series.)
3. In this course we will discuss uni-variate time series models.
4. Time series can be deterministic or stochastic. We will study stochastic time series models. It can also be discrete or continuous.

Time series plot exhibits different features. See some TS plots and try to infer.

1. Non-Stationary.
2. Stationary.
3. Seasonality / Periodicity.
4. Trend.

1.1 Univariate random variables

Before we go into TS, let us review few probability and statistics topics.

Definition *For a discrete random variable X the probability mass function is defined as:*

$$P(x) = P(X = x)$$

Definition *For a random variable X the cumulative probability distribution function (CDF) defined as:*

$$F_X(x) = P(X \leq x)$$

Example If you toss a coin twice and let X be the number of head. Then range of X is $\{0, 1, 2\}$ and for each point the probability:

$$P(X=0) = \frac{1}{4}$$
$$P(X=1) = \frac{1}{2}$$
$$P(X=2) = \frac{1}{4}$$

Using the above probabilities CFD of X is:

$$F_X(x) = \begin{cases} 0, & \text{for } x < 0 \\ \frac{1}{4} & \text{for } 0 \leq x < 1 \\ \frac{3}{4} & \text{for } 1 \leq x < 2 \\ 1 & \text{for } x \geq 2 \end{cases}$$

Note that when you are asked to find the CDF of a random variable, you need to find the function for the entire real line.

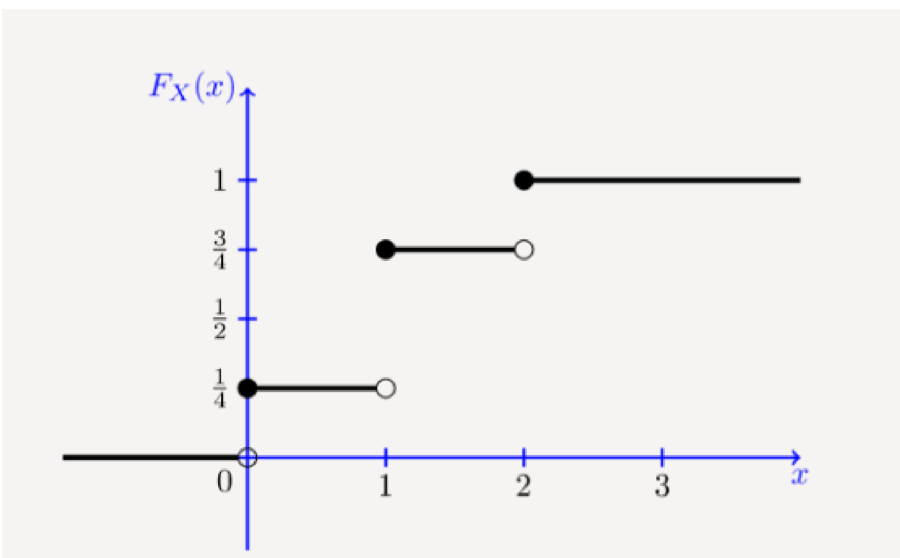

Figure 1: plot of the cdf

Remark
1. $F_X(x)$ is non decreasing.
2. $F_X(x)$ is right-continuous.
3. $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to+\infty} F_X(x) = 1$

Definition *When the random variable X is continuous then $f(x)$ is the probability density function (pdf) of X if:*

$$P(a < X < b) = \int_a^b f(x)dx$$

Remark
1. $f(x) \geq 0$ for all $x \in R$
2. $\int_{-\infty}^{+\infty} f(x)dx = 1$
3. $f(x) = \frac{d}{dx}F_X(x)$
4. $P(x < X < x + dx) \approx f(x)dx$

Definition: The n^{th} moment of a continuous random variable X is defined as:

$$\mu_n = \int_{-\infty}^{+\infty} x^n f(x)dx$$

In practice we are interested in (1) most likely value or expected value of X (ii) how far the outcomes are spread. So we are interested in first order and second order moment of a random variable.

$$\begin{aligned}
E(X) = \mu = \mu_1 &= \int_{-\infty}^{+\infty} xf(x)dx \\
\mu_2(x) &= \int_{-\infty}^{+\infty} x^2 f(x)dx \\
Var(X) &= E[X - E(X)]^2 \\
&= \mu_2 - \mu_1
\end{aligned}$$

Remark
1. $E[cX] = cE[X]$, for any constant c.
2. $E[X + Y] = E[X] + E[Y]$.
3. $E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx$.

4. $E[\sum_{i=1}^{k} c_i g_i(X)] = \sum_{i=1}^{k} c_i E[g_i(X)]$.

5. $E[X-c]^2$ is minimized when $c = E[X]$.

6. $Var(X+c) = Var(X)$

7. $Var(cX) = c^2 Var(X)$

Mean is the best prediction of the random variable in the minimum mean square error scene.

Example Consider a random variable $Y(t) = Sin(\omega t + \phi)$, where $\phi \sim Uniform[-\pi, \pi]$.

$$
\begin{aligned}
E[Y(t)] &= E[Sin(\omega t + \phi)] \\
&= \int_{-\pi}^{+\pi} Sin(\omega t + \phi) \frac{1}{2\pi} d\phi \\
&= \frac{1}{2\pi} \{Cos(\omega t + \phi)\Big|_{-\pi}^{\pi}\} \\
&= \frac{1}{2\pi} \{Cos(\omega t - \pi) - Cos(\omega t + \pi)\} \\
&= \frac{1}{2\pi} \{Cos(\omega t) - Cos(\omega t)\} \\
&= 0.
\end{aligned}
$$

1.2 Bi variate Random variable

Definition *For a random vector (X, Y) joint CDF is defined as*

$$P(X \le x \cap Y \le y) = F(x, y)$$

Joint PDF function, f(x,y), is a function such that:

$$P[(X, Y) \in A] = \int \int_A f(x, y) dx dy$$

for any $A \in R^2$

Remark Probability density function of (X, Y) has the following properties:

1. $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$

2. $f(x, y) \ge 0$ for $\forall (x, y) \in \mathbb{R}^2$

3. $\frac{\partial^2}{\partial x \partial y} F(x,y) = f(x,y)$

4. Marginal density of X is given by $f_X(x) = \int_{\mathbb{R}} f(x,y)dy$.

5. Marginal density of Y is given by $f_Y(y) = \int_{\mathbb{R}} f(x,y)dx$.

6. Conditional density of X given Y, $f(X|Y) = \frac{f(x,y)}{f_Y(y)}$.

7. Conditional density of Y given X, $f(Y|X) = \frac{f(x,y)}{f_X(x)}$.

8. Conditional expectation of X given Y, $E(X|Y) = \int_{\mathbb{R}} xf(x|Y=y)dx$.

9. Conditional expectation of Y given X, $E(Y|X) = \int_{\mathbb{R}} yf(y|X=x)dy$.

10. $E(X) = E[E(X|Y)]$.

11. $Var(X) = E[Var(X|Y)] + Var[E(X|Y)]$

12. Two events are independent iff $P(X \cap Y) = P(X)P(Y)$.

13. For independent events E$[XY] = E[X]E[Y]$

Linear relation between two random variables can be measured by covariance and correlation.

Definition *Co-variance of X and Y is given by:*

$$\begin{aligned} \sigma_{X,Y} = Cov(X,Y) &= E\Big[\{X - E(X)\}\{Y - E(Y)\}\Big] \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

Correlation of X and Y $\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$, where σ_X and σ_Y are standard deviation of X and Y respectively.

Remark 1. $|\rho_{X,Y}| \leq 1$.

2. If X and Y are independent $Cov(X,Y) = 0$

3. $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$.

4. $\{Cov(X,Y)\}^2 \leq Var(X)Var(Y)$

For multidimensional vector $(X_1, \ldots X_n)$ we can define a variance co-variance matrix as:

$$\Sigma = \begin{bmatrix} VarX_1 & Cov(X_1,X_2) & Cov(X_1,X_3) & \dots & Cov(X_1,X_n) \\ Cov(X_2,X_1) & Var(X_2) & Cov(X_2,X_3) & \dots & Cov(X_2,X_n) \\ \dots & \dots & \dots & \dots & \dots \\ Cov(X_n,X_1) & Cov(X_n,X_2) & Cov(X_n,X_3) & \dots & Var(X_n) \end{bmatrix}$$

Remark Σ is a symmetric, positive semi-definite matrix. That is $\Sigma^T = \Sigma$ and $\boldsymbol{\lambda}^T \Sigma \boldsymbol{\lambda} \geq 0$ for all non zero column vector $\boldsymbol{\lambda}$.

Remark Covariance is only a measure of linear relationship.Consider $X \sim N(0,1)$ and $Y = X^2$ then $Cov(X,Y) = 0$ but they are related. So zero covariance implies no linear relationship.

In most of the cases we would like to predict the value of the response variable Y using a value of X. Conditional expectation of Y given X is a function of x. This function is the "best" predictor of Y given X among all predictors that minimizes mean square errors.

Theorem *Let $\pi(x)$ be any predictor of Y given the value of X then $E[\{Y - E(Y|X)\}^2] \leq E[\{Y - \pi(x)\}^2]$*

Proof Let $E(Y|X = x) = \phi(x)$.Now:

$$\begin{aligned} E[\{Y - \pi(x)\}^2] &= E[\{(Y - \phi(x)) + (\phi(x) - \pi(x))\}^2] \\ &= E\Big[(Y - \phi(x))^2 + (\phi(x) - \pi(x))^2 + 2(Y - \phi(x))(\phi(x) - \pi(x))\Big] \\ &= E\Big[(Y - \phi(x))^2\Big] + E\Big[(\phi(x) - \pi(x))^2\Big] + 2E\Big[(Y - \phi(x))(\phi(x) - \pi(x))\Big] \\ &= E\Big[(Y - \phi(x))^2\Big] + E\Big[(\phi(x) - \pi(x))^2\Big] \quad \text{[See equation (1) below]} \end{aligned}$$

So, $E[\{Y - \pi(X)\}^2] = E\Big[(Y - \phi(x))^2\Big] + E\Big[(\phi(x) - \pi(x))^2\Big]$ and $E\Big[(\phi(x) - \pi(x))^2\Big] \geq 0$, combining we can write $E[\{Y - E(Y|X)\}^2] \leq E[\{Y - \pi(x)\}^2]$. This proves the theorem. But in real life computing conditional expectation in not possible. As in real life we only have data sets. We do not know anything about the joint or marginal distribution for sure. ∎

Product term in the expectation:

$$
\begin{aligned}
E\Big[(Y-\phi(x))(\phi(x)-\pi(x))\Big] &= \int_x\int_y [(y-\phi(x))(\phi(x)-\pi(x))\Big] f(x,y)dydx \\
&= \int_x\int_y [(y-\phi(x))(\phi(x)-\pi(x))\Big] f(y|x)f(x)dydx \\
&= \int_x \Big[\int_y (y-\phi(x))f(y|x)dy\Big](\phi(x)-\pi(x))f(x)dx \\
&= \int_x \Big[\int_y yf(y|x)dy - \int_y f(y|x)\phi(x)dy\Big](\phi(x)-\pi(x))f(x)dx \\
&= \int_x \Big[E(Y|X)-\phi(x)\int_y f(y|x)dy\Big](\phi(x)-\pi(x))f(x)dx \\
&= \int_x \Big[E(Y|X)-\phi(x)1](\phi(x)-\pi(x))f(x)dx \\
&= \int_x \Big[E(Y|X)-E(Y|X)](\phi(x)-\pi(x))f(x)dx \\
&= 0 \ \text{ as } \ \phi(x)=E[Y|X] \ \text{ and} \int_y f(y|x)=1. \qquad (1)
\end{aligned}
$$

Remark The above result in nice but it has limited application the practice, as $f(x,y)$ is not available in real life. But if we assume $(X,Y)\sim N(\boldsymbol{\mu},\Sigma)$ then the conditional expectation function $E(Y|X=x)$ is a linear function in x. So we will assume normality and work with a linear model for prediction.