Raya Feldman Spring 2023

WEEK 1: Examples of Time Series. Autocorrelation function. Stationarity.

Some common abbreviations used in these notes:

[BD]: P. J. Brockwell and R. A. Davis, Introduction to Time Series and Forecasting, 2016

TS: time series

IID or iid: independent identically distributed

WN: white noise

r.v.: random variable

(fi-di): finite-dimensional (distributions)
fdds: finite-dimensional (distributions)

b/c: because

. . . .

0. Introduction to Time Series. (Based on §§1.1-1.2, [BD].)

Times series is a data obtained from observations collected sequentially over time. Such data is extremely common and times series forecasting is performed in nearly every organization that works with quantifiable data:
- Retails stores use it to forecast sales.
- Energy companies use it to forecast reserves, production, demand, and prices.
- Educational institutions use it to forecast enrollment.
- International financial organizations such as the World Bank and International Monetary Fund use it to forecast inflation and economic activity.
- Transportation companies use time series forecasting to forecast future travel.
- Banks and lending institutions use it to forecast new homes purchases.
- Venture capital firms use it to forecast market potential/evaluate business plans.
- Meteorologists use it to predict precipitation, temperatures, etc.

In many other classes you dealt with predicting data where time is not a factor and where the order of measurements in time does not matter. These are typically classed cross-sectional data. In contrast, this course deals with a different type of data: time series.

A comment on time scales. With modern technology, many times series are recorded on very frequent times scales. Although data might be available at a very frequent scale, for the purpose of forecasting it is not always preferable to use this scale. In considering choice of time scale, one must consider the scale of the required forecasts and the level of noise in data. For example, if the goal is to forecast next-day sales at a grocery store, using minute-by-minute sales data are likely to be less useful for forecasting than using daily aggregates. Minute-by-minute series may contain many sources of noise (e.g., variation by peak and non-peak shopping hours) that degrade forecasting power, and these noise errors, when the data is aggregated to a coarser level, are likely to average out.

Some examples of short data series:
annual data for FSU– approximately 30 observations; for Israel – less than 100, etc.

Examples on slides:

• Population of USA: Denote t = year: 1790, 1800, ... 1990 and denote X_t = population of USA at year t, 21 observations.

• International airline data. Represents monthly totals of international airline passengers from Jan. 49 to Dec. 1960. 144 observations.

here: t = month: Jan 1949, Feb. 1949, ..., Dec. 1960
X_t = total number of passengers taking flights on the airline in a month t.

This is an example of a *non-stationary seasonal* time series:
• Low in winter, high in summer; • Upward trend • Variability increases with time

---

Important points from the slides:

A time series $\{X_t\}$ is a series of observations taken sequentially over time: x_t is an observation recorded at a specific time $t.x_t$ is also called a realization.

For each fixed time t the observation x_t is one realization of the r.v. X_t. Time series is a collection of dependent r.v.'s $X_{t_1}, \dots, X_{t_n}, \dots$.

Characteristics of times series data: observations are dependent, become available at equally spaced time points and are time-ordered.

The purposes of time series analysis are to **model** and **to predict or forecast** future values of a series based on the history of that series.

A very short review of random variables and vectors.

Time series analysis relies strongly on use of covariance and correlation functions.

....

Take a step backwards: how do we describe a r.v. or a random vector?

• Most important characteristics of one r.v. X:
–c.d.f. $F_X(x) := P(X \le x)$, (density: $f_X(x) = F'_X(x)$, $p_X(x) = p_k = P(X = k)\}$).
–mean $\mu = EX$ (average value, expectation) and
–variance $\sigma^2 = Var(X) = E(X - EX)^2 \equiv E(X^2) - (EX)^2 = E(X^2) - \mu^2$ (spread).

Example: $X \sim N(\mu, \sigma^2)$ Gaussian, then

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Spread is 99.7% around μ: 3σ rule

• Let us now take two variables, a random vector (X, Y):
for each, we have: $X \sim F_X, EX, VarX$, $Y \sim F_Y, EY, VarY$.

If X and Y are independent, then
$E(XY) = E(X)E(Y)$ and $Var(X + Y) = Var(X) + Var(Y)$.
Why? Because

$$\begin{aligned} Var(X+Y) &= E[X+Y-E(X+Y)]^2 = E[(X-EX)+(Y-EY)]^2 \\ &= E(X-EX)^2 + E(Y-EY)^2 + 2E[(X-EX)(Y-EY)] \\ &= Var(X) + Var(Y) + 2Cov(X,Y), \text{ where} \end{aligned}$$

$$Cov(X,Y) \stackrel{def}{=} E[(X-EX)(Y-EY)] \equiv E(XY) - E(X)E(Y) \equiv Cov(Y,X).$$

• When X and Y are independent,
$E[(X-EX)(Y-EY)] = E[(X-EX)]\cdot E[(Y-EY)] = 0$ and $Var(X+Y) = Var(X) + Var(Y)$.

• When X and Y are NOT independent, the product rule does not necessarily hold and we have

$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$ and

$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var X_i + 2\sum_{i<j} Cov(X_i, X_j)$.

Things are obviously simpler when $Cov(X_i, X_j) = 0$ so that we make the following definition:

Definition: If $Cov(X,Y) = 0$, then X and Y are uncorrelated.

• All independent variables are uncorrelated. In general, the opposite is not true.

. . . .

Some review of covariance/correlation discussed in sections:

• Example to illustrate that it is possible to have X and Y dependent but uncorrelated:

Let $X \sim U(-1,1)$ be uniform on $(-1,1)$, i.e., $f_X(x) = 1/2$ when $-1 < x < 1$ and zero otherwise. Let $Y = X^2$. Clearly, X and Y are dependent (non linear quadratic dependence).
We show that X and Y are uncorrelated:

$Cov(X,Y) = E(XY) - E(X)E(Y) = E(X^3) - 0 \cdot EY = \int_{-1}^{1} (x^3)(1/2)dx = 0.$

(Note that the same result will hold for any symmetric bounded r.v.)

Despite this, covariance is used as a rough guide to mutual dependence. A more useful tool is

Definition. The correlation:

$$\rho(X,Y) \equiv \rho_{XY} \equiv Cor(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}.$$

Properties:

• X and Y uncorrelated implies $\rho(X,Y) = 0$.
• $|\rho(X,Y)| \leq 1$ (Cauchy-Schwartz inequality).
• $\rho(aX+b, cY+d) = sign(ac)\rho(X,Y)$, i.e., correlation is essentially unchanged under the change of location and scale which is not true of covariance:

$Cov(aX+b, cY+d) = (ac)Cov(X,Y)$.

• $\rho(X, cX+d) = sign(c)\rho(X,X) = sign(c)$.
Thus, if Y is a linear function of X, the absolute value of the $Cor(X,Y)$ achieves its maximum value of 1.

---

1. Some descriptive techniques. (Based on [BD] §1.3 and §1.4)

1.1 Definitions.

To describe a time series $\{X_t, t = 1, 2, \ldots\}$ we define

(i) **The finite-dimensional distributions**: (fi-di) d.f. is the joint d.f. for the vector $(X_{t_1}, \ldots, X_{t_n})$:

$$F_{t_1 \ldots t_n}(x_1, \ldots, x_n) = P(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n), \forall n \geq 1, \forall t_1 < t_2 < \ldots < t_n.$$

(ii) **First- and Second-order moments.**

• **Mean**: $\mu_X(t) = EX_t$

- **Variance:** $\sigma_X^2(t) = E(X_t - \mu_X(t))^2 \equiv EX_t^2 - \mu_X(t)^2$
- **Autocovariance function (acvf):**

$$\gamma_X(t,s) = Cov(X_t, X_s) = E[(X_t - \mu_X(t))(X_s - \mu_X(s))] \equiv E(X_t X_s) - \mu_X(t)\mu_X(s)$$

(Note: this is an infinite matrix).

- **Autocorrelation function (acf):**

$$\rho_X(t,s) = Cor(X_t, X_s) = \frac{Cov(X_t, X_s)}{\sqrt{Var(X_t)Var(X_s)}} = \frac{\gamma_X(t,s)}{\sigma_X(t)\sigma_X(s)}$$

Properties of the process which are determined by the first- and second- order moments are called second-order properties.

Example: stationarity. – will be defined soon!

1.2 **Meaning and Estimation of the Correlation function**

$Cor(X,Y)$ measures linear dependence between variables. It takes values between -1 and 1.

– If X and Y are independent $\rho = 0$, scatter diagram for observations of (X,Y) looks like a ball
– If X and Y are linearly dependent, points of observations of (X,Y) lie on a line.
– If $|\rho| \approx 1$, strong linear relationship.

Estimate of the correlation coefficient from data $(x_1, \ldots, x_n)$:

$$\hat{\rho} = \Big(\frac{1}{n-1}\Big)\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{s_X^2 s_Y^2}}, \; s_X^2 \equiv \hat{\sigma}_X^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2, \; \bar{x} = \frac{1}{n}\sum_{i=1}^n x_i.$$

Note: if observations in a sample are i.i.d., the order of summation is unimportant.

. . . .

In time series: instead of vector (X,Y), we consider two variables (X_t, X_{t+k}),
instead of $\rho(X,Y)$ we consider $\rho_X(t, t+k) = \rho(X_t, X_{t+k})$, called acf.

The prefix *auto* is to convey the notion of self-correlation (both variables come from the same TS), correlation of the series with itself. It is used to assess numerically the dependence between values X_t and X_{t+k} and tell us about smoothness of the time series:

$\rho_X(t, t+1) \approx 1$ means strong positive dependence between the adjacent observations.
$\rho_X(t, t+k) > 0$ large for $k > 1$ means strong positive dependence in the series.
Such series has a smooth graph: $\rho = 1$ means that X_{t+1} increases with increase of X_t.

$\rho_X(t, t+1) \approx 0$, weak correlation, means that adjacent values X_t and X_{t+1} are almost independent. Thus, the series does not look smooth. It looks choppy (wild, rough).

1.3 Two "extreme" examples.

1.3.1 Example: White Noise WN $(0, \sigma^2)$.

Consider series $\{Z_t\}$, where Z_t are mean zero uncorrelated r.v.'s:

$$E(Z_t) = 0, \gamma_Z(s,t) = E(Z_t Z_s) = \sigma_Z^2, \text{ if s = t , and 0, if } s \neq t \qquad (*)$$

WN is often refered to as the white noise or "shocks".

If $\{Z_t\}$ are i.i.d. (not a requirement for WN), we write $\{Z_t\} \sim IID(0, \sigma_Z^2)$.

If, in addition, $\{Z_t\}$ is Gaussian, we call it Gaussian WN.

In Gaussian case, Z_t's are i.i.d r.v.s $\mathcal{N}(0, \sigma_Z^2)$ and take values between $(-2\sigma_Z, 2\sigma_Z)$ w.p. 95%, independently of its neighbors.

1.3.2 Example. Apply smoothing operation $X_t = \frac{1}{3}(Z_{t-1} + Z_t + Z_{t+1})$ to Gaussian WN of Example 1.3.1. Calculate acvf:

$EX_t = \frac{1}{3}E(Z_{t-1} + Z_t + Z_{t+1}) = 0$ (b/c $EZ_t = 0$.)

$$\begin{aligned}\gamma_X(t,s) &= E[(X_t - EX_t)(X_s - EX_s)] = E[X_t X_s] = \tfrac{1}{9}E[(Z_{t-1} + Z_t + Z_{t+1})(Z_{s-1} + Z_s + Z_{s+1})] \\ &= \tfrac{1}{9}E[(Z_{t-1}Z_{s-1} + Z_t Z_s + Z_{t+1}Z_{s+1}) \\ &\quad +(Z_{t-1}Z_s + Z_t Z_{s+1}) + (Z_t Z_{s-1} + Z_{t+1}Z_s) \\ &\quad +(Z_{t-1}Z_{s+1} + Z_{t+1}Z_{s-1})].\end{aligned}$$

Recall that $EZ_t Z_s = 0$ when $s \neq t$. We get:

$$\gamma_X(t,s) = \sigma_Z^2 \begin{cases} 3/9 & \text{if } t = s, & \text{(from the first row)} \\ 2/9 & \text{if } t-1 = s \text{ or } t = s-1, \text{ i.e. } |t-s| = 1 & \text{(from the second row)} \\ 1/9 & \text{if } t-1 = s+1 \text{ or } t+1 = s-1, \text{ i.e., } |t-s| = 2 & \text{(from the third row)} \\ 0 & \text{if } |t-s| > 2. \end{cases}$$

Autocorrelation function is a normalized autocovariance function.

2. Stationarity

2.1 Stationarity and Strict Stationarity (Based on §1.4 and 2.1 of [BD])

Intuitively stationarity means that the graphs over two equal-length time intervals of a realization of the TS should exhibit similar statistical characteristics. For example, the proportion of ordinates not exceeding a given level x should be roughly the same for both intervals.

On a graph:
• no trend • no seasonality • no change of variability • no apparent sharp changes of behavior

Two approaches to stationarity:

Definition 2.1.1. A time series is said to be **strictly stationary** if the joint probability distribution associated with the n r.v.'s $X_{t_1}, \ldots, X_{t_n}$ for any set of times $t_1, \ldots, t_n$, is the same as that associated with the n r.v.'s $X_{t_1+k}, \ldots, X_{t_n+k}$ for any integer k.

Definition 2.1.2 A time series is said to be (weakly, second-order) **stationary** if
(a) $E|X_t|^2 < \infty$ (b) $EX_t = \mu$ (c) $\gamma_X(t,s) = \gamma_X(t+r, s+r) = \gamma_X(t-s)$ for all $t, s, r \in T$.

Since index set $T = \{1, 2, \ldots\}$, the time difference $t-s$ is an integer and typically denoted by $k = 1, 2, \ldots$. The function $\gamma_X(k)$ is often referred to as the value of the autocov. f'n (acvf) at *lag k*.

2.1.3 $\rho_X(k) = \frac{\gamma_X(k)}{\gamma_X(0)}$ is the acf (the autocorr. f'n) for a stationary process.

Note: $\gamma_X(k) = \gamma_X(-k)$, $\rho_X(k) = \rho_X(-k)$, $\gamma_X(0) = \sigma_X^2$.

2.1.4 Example. Apply smoothing operation to a WN series: $X_t = \frac{1}{3}(Z_{t-1} + Z_t + Z_{t+1})$. Autocov. f'n found in Example 1.3.2 shows that X_t is stationary:

$$\gamma_X(t,s) = \sigma_Z^2 \begin{cases} 3/9 = 1/3 & \text{if } t-s = 0, \\ 2/9 & \text{if } |t-s| = 1 \\ 1/9 & \text{if } |t-s| = 2 \\ 0 & \text{if } |t-s| > 2 \end{cases} \equiv \gamma_X(t-s).$$

2.1.5 Example. Let $\{Z_t\}$ be I.I.D. WN, that is, Z_t are i.i.d. r.v.s with $EZ_t = 0$ and $Var Z_t = \sigma^2$.
We write: $\{Z_t\} \sim IID(0, \sigma^2)$.

Let $X_t = Z_1 + \ldots + Z_t$. Such a process is called a random walk (RW).

Clearly, IID WN is strictly stationary. We show: **RW is NOT stationary.**

Derivation. $EX_t = 0$. *Take* $s = t + k$, $k \geq 1$.
$$\begin{aligned} \gamma_X(t,s) &= E[X_t X_{t+k}] = E[(Z_1 + \ldots + Z_t)(Z_1 + \ldots + Z_t + Z_{t+1} + \ldots + Z_{t+k})] \\ &= E[(Z_1 + \ldots + Z_t)^2] + E[(Z_1 + \ldots + Z_t)(Z_{t+1} + \ldots + Z_{t+k})] \\ &= t \cdot \sigma^2 + 0 = t\sigma^2, \end{aligned}$$
i.e. $\gamma_X(t, t+k)$ *depends on* t, $\{X_t\}$ *is non-stationary.*

2.1.6 Short Summary. In general:
• strict stationarity + finite second moment imply stationarity
• stationarity does not imply strict stationarity
Fact: For Gaussian TS *stationarity* = *strict stationarity.*

3. Moving Average models, MA(q). (Based on §1.4 of [BD])

3.1 White Noise or shocks (WN). IID Noise (IID).

See also Example 1.3.1 of the lecture notes and graphs on ppts for week 1.

$\{Z_t\} \sim WN(0, \sigma_Z^2)$ is called the **white noise or "shocks"** if

- $EZ_t = 0$, $EZ_t^2 = \sigma_Z^2$, and $\gamma_Z(k) = \rho_Z(k) = 0, k > 0$ (uncorrelated).

If $\{Z_t\}$ are iid (not a requirement for WN), we write $\{Z_t\} \sim IID(0, \sigma_Z^2)$.

3.2 Moving average of order one, MA(1).

$X_t = Z_t + \theta_1 Z_{t-1}$ where $Z_t \sim WN(0, \sigma_Z^2)$, $|\theta_1| < 1$.

In this model the previous shock Z_{t-1} is still influencing the current value X_t.

Calculate acvf: $EX_t = 0$ b/c $EZ_t = 0$;

$$\gamma_X(k) = Cov(X_t, X_{t+k}) = E(Z_t + \theta_1 Z_{t-1})(Z_{t+k} + \theta_1 Z_{t+k-1}) =$$
$$= E[(Z_t Z_{t+k} + \theta_1^2 Z_{t-1} Z_{t+k-1}) + \theta_1(Z_t Z_{t+k-1} + Z_{t-1} Z_{t+k})] = \sigma_Z^2 \cdot \begin{cases} (1+\theta_1^2) & \text{if } k = 0, \\ \theta_1 & \text{if } k = 1, -1, \\ 0 & \text{if } |k| > 1. \end{cases}$$

Thus, X is stationary with the following second-order moments:
$\mu_X = 0$, $\sigma_X^2 = \gamma_X(0) = \sigma_Z^2(1+\theta_1^2)$, $\rho_X(0) = 1$, $\rho_X(1) = \theta_1/(1+\theta_1^2)$, $\rho_X(k) = 0, k > 1$.

Note: since $|\theta_1| < 1$ it follows that $\rho_X(1) < .5$.

Question: How does $\rho_X(k)$ look like?