Time Series Analysis

Foundational Concepts

Hossein Moradi Rekabdarkolaee

South Dakota State University

email: *hossein.moradirekabdarkolaee@sdstate.edu*.

Office: CAME Building, *Room: 240*.

Time Series and Stochastic Processes

- A random variable is a measurable function from a sample space $\mathbb{E}$ to a measurable space $\boldsymbol{\Omega}$ (set of outcomes).
- Stochastic process (random process) refers to a series of events where each event through random occurrence has an inbuilt pattern.
- **Stochastic processes** are widely used as mathematical models of systems and phenomena that appear to vary in a random manner.
- Stochastic Processes can come in zero-dimensional, one-dimensional, all the way to infinite dimensional processes form.

Time Series and Stochastic Processes

- A sequence of random variables defined at fixed sampling intervals is sometimes referred to as a discrete-time stochastic process, though the shorter name time series model is often preferred.

- In other words, a time series is a one-dimensional stochastic process.

- What are the different components of Time Series.

Time Series and Stochastic Processes

- A systematic change in a time series that does not appear to be periodic is known as a *trend*.
- A repeating pattern within each year is known as *seasonal variation*, although the term is applied more generally to repeating patterns within any fixed period.
- Sometimes, there are *cycles* in a time series that do not correspond to some fixed natural period.
- There are other sources of variation that are "irregular" fluctuations.
- In order to define a stochastic process, we need to specify its mean, variance, and covariance functions.

Means, Variances, and Covariances

- Suppose we represent the *time series* by $\{Y_t : t \in T\}$ with T be an index set.
- The sequence of random variables $\{Y_t : t \in T\}$ is a *stochastic process* if Y_t is a random variable for all $t \in T$.
- The mean function of a random variable $\{Y_t : t \in \mathbb{Z}\}$ is defined by $\mu_t = E(Y_t)$ for $t \in \mathbb{Z}$.
- In general, μ_t can be different at each time point t.
- The *Autocovariance function* is defined by

$$\begin{aligned} \gamma_{t,s} = Cov(Y_t, Y_s) &= E[(Y_t - \mu_t)(Y_s - \mu_s)] \\ &= E(Y_t Y_s) - \mu_t \mu_s, \ t, \ s \in \mathbb{Z} \end{aligned}$$

Means, Variances, and Covariances

- From the autocovariance function, we have

$$\gamma_{t,t} = Var(Y_t) = Cov(Y_t, Y_s), \text{ when } s = t,$$

- The *Autocorrelation function*

$$\rho_{t,s} = Corr(Y_t, Y_s) = \frac{Cov(Y_t, Y_s)}{\sqrt{Var(Y_t)Var(Y_s)}} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}}, \qquad t, s \in \mathbb{Z}$$

- Since $Corr(Y_t, Y_s) = Corr(Y_s, Y_t)$, thus $\rho_{t,s} = \rho_{s,t}$.
- Thus we have

$$\begin{cases} \gamma_{t,t} = Var(Y_t) & \rho_{t,t} = 1, \\ \gamma_{t,s} = \gamma_{s,t} & \rho_{t,s} = \rho_{s,t}, \\ |\rho_{t,s}| \leq \sqrt{\rho_{t,t}\rho_{s,s}} & |\rho_{t,s}| \leq 1, \end{cases}$$

Means, Variances, and Covariances

- If we have constant $c_1, c_2, \ldots, c_m$ and $d_1, d_2, \ldots, d_n$. Then, for time points $t_1, t_2, \ldots, t_m$ and $s_1, s_2, \ldots, s_n$, we have

$$Cov\left[\sum_{i=1}^{m} c_i Y_{t_i}, \sum_{j=1}^{n} d_j Y_{s_j}\right] = \sum_{i=1}^{m}\sum_{j=1}^{n} c_i d_j Cov\left[Y_{t_j}, Y_{s_j}\right].$$

- A special case for this result is the following

$$Var\left(\sum_{i=1}^{n} c_i Y_{t_i}\right) = \sum_{i=1}^{n} c_i^2 Var(Y_{t_i}) + 2\sum_{i=2}^{n}\sum_{j=1}^{i-1} c_i c_j Cov\left[Y_{t_i}, Y_{t_j}\right].$$

- Let us start to see how these different components would work for various time series.

Random Walk

- Suppose, we have an independent identically distributed sequence of random variables $e_1, e_2, \ldots$ each with mean 0 and variance σ^2.
- Let us construct the time series as follows

$$\begin{cases} y_1 = e_1, \\ y_2 = e_1 + e_2, \\ \ldots \\ y_t = e_1 + e_2 + \ldots + e_t, \end{cases}$$

- The other way to write this time series is

$$y_t = y_{t-1} + e_t,$$

Random Walk

- What is the mean function?

$$\mu_t = E(y_t) = E\left(\sum_{i=1}^{t} e_i\right) = \sum_{i=1}^{t} E(e_i) = 0.$$

- Since the mean function does not depend on t, therefore, for all t, this process has a mean 0.
- The variance of the process?

$$Var(y_t) = Var\left(\sum_{i=1}^{t} e_i\right) = \sum_{i=1}^{t} Var(e_i) = t\sigma^2$$

- Thus, the variance of this process will linearly with time.

Random Walk

- Suppose, $1 \leq t \leq s$. The autocovariance function is

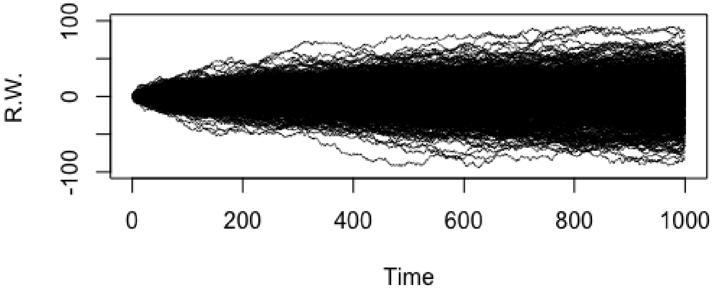
$$\begin{aligned}
\gamma_{t,s} &= Cov(y_t, y_s) \\
&= Cov(e_1 + \ldots + e_t, e_1 + \ldots + e_t + e_{t+1} + \ldots + e_s) \\
&= \sum_{j=1}^{s} \sum_{i=1}^{t} Cov(e_i, e_j) = t\sigma^2.
\end{aligned}$$

- Therefore the autocovariance function for all time point t and s is $\gamma_{t,s} = t\sigma^2$ for $1 \leq t \leq s$.
- Thus, the autocorrelation function for the random walk, for $1 \leq t \leq s$, is

$$\rho_{t,s} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}} = \frac{t\sigma^2}{\sqrt{t\sigma^2 s\sigma^2}} = \sqrt{\frac{t}{s}}.$$

Random Walk

- What does this mean for the correlation?
- What is the correlation between steps 1 and 2? How about steps 49 and 50? How about steps 1 and 100?



Moving Average Process

- Suppose $\{y_t\}$ is defined as

$$y_t = \frac{e_t + e_{t-1}}{2},$$

- where e's are assumed to be independent identically distributed with mean 0 and variance σ^2.
- Again for this process, we have to specify the mean, variance, and covariance function.
- The mean function is

$$\mu_t = E\left(\frac{e_t + e_{t-1}}{2}\right) = \frac{E(e_t) + E(e_{t-1})}{2} = 0.$$

Moving Average Process

- The Variance function is

$$Var(y_t) = Var\left(\frac{e_t + e_{t-1}}{2}\right) = \frac{Var(e_t) + Var(e_{t-1})}{4} = \frac{2\sigma^2}{4}.$$

- The Covariance function for time t and $t-1$ is

$$\begin{aligned}
Cov(y_t, y_{t-1}) &= Cov\left(\frac{e_t + e_{t-1}}{2}, \frac{e_{t-1} + e_{t-2}}{2}\right) \\
&= \frac{Cov(e_t, e_{t-1}) + Cov(e_t, e_{t-2})}{4} \\
&+ \frac{Cov(e_{t-1}, e_{t-1}) + Cov(e_{t-1}, e_{t-2})}{4} \\
&= \frac{Cov(e_{t-1}, e_{t-1})}{4} = \frac{\sigma^2}{4}.
\end{aligned}$$

Moving Average Process

- The covariance function for time t and $t-2$ is

$$
\begin{aligned}
Cov(y_t, y_{t-2}) &= Cov\left(\frac{e_t + e_{t-1}}{2}, \frac{e_{t-2} + e_{t-3}}{2}\right) \\
&= \frac{Cov(e_t, e_{t-2}) + Cov(e_t, e_{t-3})}{4} \\
&+ \frac{Cov(e_{t-1}, e_{t-2}) + Cov(e_{t-1}, e_{t-3})}{4} = 0.
\end{aligned}
$$

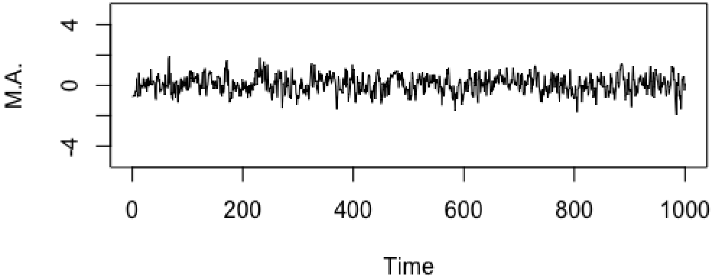
- Therefore, the covariance function can be written as

$$
\gamma_{t,s} = \begin{cases} 0.5\sigma^2 & t = s;\ i.e.\ |t-s| = 0, \\ 0.25\sigma^2 & |t-s| = 1, \\ 0 & |t-s| > 1, \end{cases}
$$

Moving Average Process

- The autocorrelation function can be written as

$$\rho_{t,s} = \begin{cases} 1 & t = s;\ i.e.\ |t-s| = 0, \\ 0.5 & |t-s| = 1, \\ 0 & |t-s| > 1, \end{cases}$$



Stationarity

- In order to make inferences about the structure of a stochastic process, we have to make some simplifying assumptions.
- One of these assumptions is about the laws (read probability laws) that govern the process's behavior.
- If we assume that the behavior of a process does not change over time, then we call that a **stationary process**.
- There are two types of stationarity
  - Strictly Stationary,
  - Weakly Stationary.

Stationary Processes

- A stochastic processes is *strictly stationary* (or *strongly stationary*) if for all choices of time points $t_1, \cdots, t_n$ and all choices of time lag k, the joint distributions of $Y_{t_1}, \cdots, Y_{t_n}$ and $Y_{t_1-k}, \cdots, Y_{t_n-k}$ are the same.
- The definition says that k dimensional distributions don't change with a shift in time.
  - For $n = 1$ this means the distribution of Y_t is the same for all values of t.
  - This results that all the existing moments of the Y_t's are the same including mean and variance.
  - That is provided the first two moments are finite,

$$\begin{aligned} \mu(t) &= \mu \\ \sigma^2(t) &= \sigma^2 \end{aligned}$$

Stationary Processes

- For $n = 2$ this means the joint distribution of Y_t, Y_s, depends only on the time difference $s - t = \tau$.
- That implies acv.f depends only on $t - s$, and may be written as $\gamma(\tau)$, where

$$\begin{aligned} \gamma_{t,s} &= \gamma_\tau = Cov(Y_t, Y_s) = Cov(Y_{t+\tau}, Y_t) \\ &= Cov(Y_\tau, Y_0) = \gamma_{0,|t-s|} \end{aligned}$$

- The acf for a strictly stationary time series reduces to

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$$

- Thus, we have

$$\begin{cases} \gamma_0 = var(Y_t) & \rho_0 = 1 \\ \gamma_k = \gamma_{-k} & \rho_k = \rho_{-k} \\ |\gamma_k| \leq |\gamma_0| & |\rho_k| \leq 1 \end{cases}$$

Second Order Stationarity

- A stochastic process $\{Y_t\}$ is said to be weakly (or second-order) stationary if
  - The mean function is constant over time i.e. $\mu_t = \mu$, and
  - The covariance function between two-time points is a function of their difference in time i.e. $\gamma_{t,t-k} = \gamma_{0,k}$ for all time t and lag k.
- The definition of a weakly stationary process implies that the variance of the process is constant.
- In general, is it possible that a weakly stationary process to be a strongly stationary process?
- Is any strongly stationary process a weakly stationary process?

Weak Stationarity

- For a stationary process $\{Y_t\}$, write the *autocovariance function* and the *autocorrelation function (AFC)* as

$$\gamma_k = Cov(Y_t, Y_{t-k}), \qquad \rho_k = corr(Y_t, Y_{t-k})$$

- Note that the mean function $E(Y_t) = \mu$ and the variance function $V(Y_t) = Var(Y_t) = \gamma_0$ are constant and do not depend on t.
- Also, the autocovariance function γ_k and the autocorrelation function ρ_k depend only on the lag k but not on the time t.
- In order to show that a process $\{Y_t\}$ is stationary:
  - we have to show that $\mu_t = E(Y_t) = \mu$ is constant for all t and that $\gamma_{t,t-k} = Cov(Y_t, Y_{t-k})$ depends only on the lag k, not on time t.