Module 1.3
Basic Concepts in TSA

S. Lakshmivarahan

School of Computer Science
University of Oklahoma
Norman, OK, 73071
USA

Topics Covered

- Introduces the basic concepts related to
  1. Time Series
  2. Ensembles
  3. Stationarity - strong, weak
  4. Ergodicity
  5. Properties of auto-covariance / auto correlation functions

Stochastic Process

- Let $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ denote the set of natural integers
- A time series $x_t,\ t \in \mathbb{Z}$ is a sequence of scalar, real valued random variables defined on an appropriate Probability Space - $(\Omega, \mathcal{F}, P)$
- $X : \Omega \times \mathbb{Z} \to \mathbb{R}$ and $x_t(\omega, t)$ is denoted as $x_t(\omega), t \in \mathbb{Z}$ and $\omega \in \Omega$

Stochastic Process

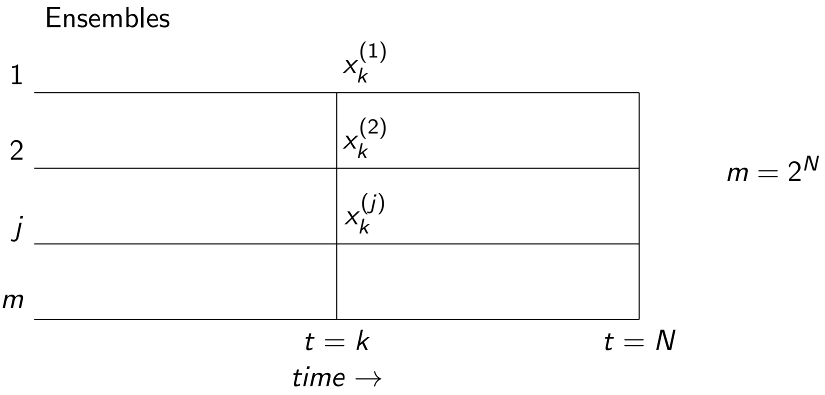
- For each $t, x_t(\cdot) : \Omega \to \mathbb{R}$ is a random variable
- For each $\omega \in \Omega, x_t(\omega)$ as a function from $\mathbb{Z} \to \mathbb{R}$ defines a sequence of random variables called a realization of the stochastic process
- The collection of $x_t(\omega)$ for all $\omega \in \Omega$ and $t \in \mathbb{Z}$ is the discrete time stochastic process of interest to us
- For a fixed t, the collection of random variables $x_t(\omega)$ are called the ensembles of realization of the process

Example: Ensemble and Time Series

- Let $\{x_1, x_2, \ldots, x_N\}$ be a finite sequence of independent, identically distributed random variables x_i where:

$$x_i = \begin{cases} +1 \text{ with probability } p = 1/2 \\ -1 \text{ with probability } p = 1/2 \end{cases}$$

- Clearly there are 2^N distinct sequences whose members are either $+1$ or -1



- The collection $x_k^{(j)}(\omega)$ of random variables for a fixed k and $1 \le j \le m$ is the ensemble of realization at time k
- $x_k^{(1)}(\omega)$ as k varies from 1 to N is a particular realization of the underlying stochastic process

- Complete characterization of $x_t(\omega)$ is given by joint distribution of

$$(x_{j_1}, x_{j_2}, \ldots, x_{j_k})$$

for $j_1 < j_2 < \ldots < j_k$ and $1 \leq k \leq N < \infty$

- If the joint probability distributions are invariant in time, then $x_t(\omega)$ is called a strictly stationary process
- Otherwise, it is called a non-stationary process

- In practice, we only have access to a single realization of the time series and may not learn much about the underlying probability space and information about joint distribution
- Hence, we need to settle down for a less ambitious program dealing only with sample <u>statistical moments</u>

First two moments of time series

- Let $\mu_t = \mathbf{E}(x_t)$ denote the mean of x_t
- Let $\gamma(t, s) = \mathbf{Cov}(x_t(\omega), x_s(\omega)) = \mathbf{E}\left[(x_t - \mathbf{E}(x_t))(x_s - \mathbf{E}(x_s))\right]$ denote the covariance between $x_t(\omega)$ and $x_s(\omega)$ for a fixed pair of times t and s
- In here, $\mathbf{E}(\cdot)$, the expectation operator is w.r. to the joint probability distribution of $x_t(\omega)$ and $x_s(\omega)$.

Second-Order Stationary Process

- The given TS x_t is said to be second-order /weakly/ covariance stationary if

$$\mathbf{E}(x_t^2) < \infty$$
$$\mathbf{E}(x_t) = \mu \text{ a constant, and}$$

- $\gamma(t, s) = \gamma(t + k, s + k)$ for any integer $k \geq 0$, that is, $\gamma(t, s)$ does not depend on t and s but only on the difference $|t - s|$

Weakly Stationary Process

- For a weakly stationary process with $t - s = k$, $\gamma(t,s) = \gamma(t+k,t) = \gamma(k,0) \equiv \gamma(k)$
- $\gamma(k)$ is auto-covariance at *lag* $k \geq 0$
- $\gamma(0) = \mathbf{Cov}(x_t, x_t) = \mathbf{Var}(x_t)$
- $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$ is called auto-correlation function of x_t
- Clearly, $\gamma(k)$ and $\rho(k)$ can be positive or negative

Properties of $\gamma(k)$ and $\rho(k)$

- Let $\gamma : \mathbb{Z} \to \mathbb{R}$ and $\rho : \mathbb{Z} \to \mathbb{R}$ be the auto-covariance and auto-correlation functions of a time series x_t
-
$$\begin{aligned}\gamma(k) &= \mathbf{Cov}(x_{t+k}, x_t) \\ &= \mathbf{Cov}(x_t, x_{t+k}) \\ &= \mathbf{Cov}(x_{t-k}, x_t) = \gamma(-k)\end{aligned}$$
- That is, $\gamma(k)$ is an <u>even</u> function
- Similarly $\rho(k) = \rho(-k)$ and auto-correlation is also an even function

Properties of $\rho(k)$

- $\rho(0) = 1$ by definition
- $|\rho(k)| \leq 1$
- To verify this claim: Let $x_i, i = 1, 2$ be two correlated random variables with ρ as their correlation coefficient
- Let μ_i and σ_i^2 be the mean and variance of x_i, for $i = 1, 2$
- Define $z_i = \frac{x_i - \mu_i}{\sigma_i}$ and $\rho = \mathbf{E}(z_1 z_2)$
- Clearly $\mathbf{E}(z_i) = 0$ and $\mathbf{Var}(z_i) = 1, i = 1, 2.$

Properties of $\rho(k)$

- Let $z = z_1 - z_2$. Then
  $0 \leq \mathbf{Var}(z) = \mathbf{E}(z^2) = \mathbf{E}(z_1^2) - 2\mathbf{E}(z_1, z_2) + \mathbf{E}(z_1^2) = 2 - 2\rho$
- Hence $\rho \leq 1$
- Similarly, if $z = z_1 + z_2$, by a similar arguments, it can be verified $\rho \geq -1$
- Hence, $|\rho| \leq 1$

- Let $\epsilon_t \sim \textbf{IID} N(0, \sigma^2)$ and $x_t = \mu + \epsilon_t$
- $\textbf{E}(x_t) = \mu$
- $\textbf{Var}(x_t) = \textbf{E}(x_t - \mu)^2 = \sigma^2$
-
$$\begin{aligned}\gamma(k) &= \textbf{Cov}(x_{t+k}, x_t) \\ &= \textbf{E}[(x_{t+k} - \mu)(x_t - \mu)] = \textbf{E}[\epsilon_{t+k}\epsilon_t] = 0 \ \text{ for } \ k \geq 1\end{aligned}$$
-
$$\rho(k) = \begin{cases} 1 \ \text{if} \quad k = 0 \\ 0 \ \text{for} \ \ k \neq 0 \end{cases}$$
- x_t is a weakly stationary process

Example 2 - Gaussian Random Walk

- Let $x_t \sim$ **IID** $N(0, \sigma^2)$
- Define $S_t = \sum_{k=1}^{t} x_k$ with $S_0 = 0$
- $(S_0, S_1, S_2, \ldots)$ is known as the Gaussian Random Walk
- $\mathbf{E}(S_t) = \mathbf{E}(\sum_{k=1}^{t} x_k) = \sum_{k=1}^{t} \mathbf{E}(x_k) = 0$
- $\mathbf{Var}(S_t) = \mathbf{Var}(\sum_{k=1}^{t} x_k) = \sum_{k=1}^{t} \mathbf{Var}(x_k) = \sigma^2 t$
- $$\begin{aligned} \gamma(t+k, t) &= \mathbf{Cov}\left(S_{t+k}, S_t\right) \\ &= \mathbf{Cov}\left(\sum_{i=1}^{t+k} x_i, \sum_{i=1}^{t} x_i\right) \\ &= \mathbf{Cov}\left(\sum_{i=1}^{t} x_i, \sum_{i=1}^{t} x_i\right) = \sigma^2 t \end{aligned}$$
- S_t is a non-stationary process

Example 3

- Let $x_t = a\cos(\theta t) + b\sin(\theta t)$, where a, b are two uncorrelated Gaussian random variables with zero mean and unit variance. That is: $a, b \sim N(0,1)$ - uncorrelated
- $\mathbf{E}(x_t) = 0$
- $$\begin{aligned}\gamma(k) &= \mathbf{Cov}(x_{t+k}, x_t) \\ &= \mathbf{E}\{\,[a\cos(\theta(t+k)) + b\sin(\theta(t+k))] \\ &\qquad [a\cos(\theta(t)) + b\sin(\theta(t))]\} \\ &= \cos\theta(t+k)\cos\theta t + \sin\theta(t+k)\sin\theta t = \cos\theta k\end{aligned}$$
- x_t is a second-order stationary process

Relation between Ensemble and Time Series Statistics

- In TSA, we are given only one realization of a discrete time stochastic process with no access to the underlying probability space $(\Omega, \mathcal{F}, P)$ over which the observed process is defined
- Accordingly, we don't have access to the probability measure and so the population moments - mean, variance, auto-covariance etc., <u>can not</u> be calculated as described.
- This <u>limits us</u> to computing all the required <u>sample moments</u> exclusively from the observed series as a function of time

Time Averages from the Data

- Let $\{x_t\}$ be a weakly stationary process
- Define $\overline{x}_n = \frac{1}{n}\sum_{t=1}^{n} x_t$ - time average
- $\overline{\gamma}(k) = \frac{1}{n}\sum_{t=k+1}^{n}(x_t - \overline{x}_n)(x_{t-k} - \overline{x}_n)$ - time average
- The quality of these estimates are better when n is large and k is smaller
- A guidline is: $n \geq 50$ and $0 \leq k \leq \frac{n}{4}$

Ergodicity in the Mean

- A weakly stationary process $\{x_t\}$ is said to be ergodic in the mean, if

$$\lim_{n\to\infty} \overline{x}_n = \mathbf{E}(x_t) = \mu$$

that is, $\overline{x}_n$ converges to μ in probability as $n \to \infty$:

$$\lim_{n\to\infty} \mathbf{Prob}\left[|\overline{x}_n - \mu| > \delta\right] < \epsilon,$$

where δ and ϵ are arbitrary positive real numbers. That is, $\overline{x}_n$ is a consistent estimate of μ

- Stated in words: the time average converges to the ensemble average if the process is ergodic in the mean

Ergodicity in the Second Moment

- A weakly stationary process $\{x_t\}$ is said to be ergodic in the second moment if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=k+1}^{n} (x_t - \mu)(x_{t-k} - \mu) = \gamma(k)$$

- That is, the sample auto-covariance on the l.h.s converges in probability to $\gamma(k)$ as $n \to \infty$. That is, l.h.s is a consistent estimate if $\gamma(k)$:
- Again, the time average on the l.h.s converges to the ensemble statistics

Absolute Summability

- Let $\{a_n\}$ be a real sequence. This sequence is said to be absolutely summable if

$$\sum_{n=1}^{\infty} |a_n| < \infty$$

- $a_n = \frac{1}{n}$. Then, $\sum_{n=1}^{k} \frac{1}{n} \approx \int_1^k \frac{1}{x} \, \mathrm{d}\, x = \log k \to \infty$ as $k \to \infty$. Hence, $1/n$ is not absolutely summable
- $a_n = \frac{1}{n^2}$. Then $\sum_{n=1}^{k} \frac{1}{n^2} \approx \int_1^k x^{-2} \, \mathrm{d}\, x = 1 - \frac{1}{k} < 1$. Hence $a_n = \frac{1}{n^2}$ is absolutely summable

A Sufficient Condition for Ergodicity

- A sufficient condition for $\{x_t\}$ to be ergodic in the first two moments is that the auto-covariance function must be absolutely summable

$$\sum_{k=0}^{\infty} |\gamma(k)| < \infty \tag{1}$$

- In the following, we will tacitly assume that $\gamma(k)$ is absolutely summable which will allow us to perform analysis based on sample moments with reasonably large samples

Stationarity does not imply Ergodicity

- Let $a \sim N(0, \sigma_1^2)$, $\epsilon_t \sim \mathbf{IID}N(0, \sigma_2^2)$ and let a and $\{\epsilon_t\}$ are also <u>mutually uncorrelated</u>, with $\sigma_i^2 > 0$ for $i = 1, 2$.
- Let $x_t^{(i)} = a^{(i)} + \epsilon_t$ to be a time series for a fixed i when $1 \leq i \leq n$, $a^{(i)}$ are $\mathbf{IID}N(0, \sigma_1^2)$, and $t \geq 1$.
- $\mathbf{E}(x_t^{(i)}) = \mathbf{E}(a^{(i)}) + \mathbf{E}(\epsilon_t) = 0$

Stationarity does not imply Ergodicity

- $\gamma(0) = \textbf{Var}(x_t) = \textbf{E}(a^{(i)} + \epsilon_t)^2 = \sigma_1^2 + \sigma_2^2$
- $$\begin{aligned}\gamma(k) &= \textbf{E}\left[x_t^{(i)} x_{t-k}^{(i)}\right] \\ &= \textbf{E}\left[\left(a^{(i)} + \epsilon_t^{(i)}\right)\left(a^{(i)} + \epsilon_{t-k}^{(i)}\right)\right] \\ &= \textbf{E}\left[\left(a^{(i)}\right)^2\right] = \sigma_1^2\end{aligned}$$
- Since $\sum_{k=0}^{\infty} |\gamma(k)| = \infty$, $\{x_t^{(i)}\}$ is <u>not ergodic</u>

Stationarity does not imply Ergodicity

- But $\frac{1}{n}\sum_{t=1}^{n} x_t^{(i)} = \frac{1}{n}\sum_{t=1}^{n}(a^{(i)} + \epsilon_t) = a^{(i)} + \frac{1}{n}\sum_{t=1}^{n}\epsilon_t$
- By Central limit theorem, $\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\epsilon_t = 0$
- Hence the time average of any realization $x_t^{(i)}$ in the limit as $n \to \infty$ is $a^{(i)}$. Hence, $\{x_t^{(i)}\}$ is <u>stationary</u>

Positive/Non-Negative Definite Functions

- Let $f : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$ be a real valued function in two variables with $f(i, j)$ defined for $1 \leq i, j \leq n$
- This function f is said to be non-negative definite if

$$\sum_{i,j=1}^{n} a_i f(i, j) a_j \geq 0$$

for all real vectors $\mathbf{a} = (a_1, a_2, \ldots, a_n)^T \in \mathbb{R}^n$ and for all $n \geq 1$

- If the inequality $\geq$ is replaced by $>$, then f is called a positive definite function
- If the inequality $\geq$ is replaced by $\leq$, then f is said to be non-positive definite

Auto-Covariance Function is Non-Negative Definite

- Let $\{x_t\}$ be a weakly stationary process with $\gamma(k)$ as its auto-correlation function
- Let $\mathbf{a} = (a_1, a_2, \dots a_n)^T \in \mathbb{R}^n$
- Let $z_i = x_i - \mathbf{E}(x_t), 1 \leq i \leq n$ and $\mathbf{z} = (z_1, z_2, \dots z_n)^T \in \mathbb{R}^n$
- Then

$$0 \leq \mathbf{Var}(a^t z) = \mathbf{E}(a^t z)^2 = \mathbf{E}\left[a^t z z^t a\right] = a^T \Gamma a = \sum_{i,j=1}^{n} a_i \Gamma(i,j) a_j,$$

  where $\Gamma = \Gamma(i,j) = \mathbf{E}(z_i z_j)$ is the $n \times n$ covariance matrix
- Since $\Gamma(i,j) = \gamma(|i-j|)$ by definition the auto-covariance function is non-negative definite