Characteristics of Time Series

A time series $\{X_1, X_2, X_3, \ldots\}$ is a sequence of random variables

We will talk about how to model $\{X_t\}$, $t=1,2,\ldots$ and then how to use the models to forecast.

Concepts for modeling time series:

1. White Noise: uncorrelated random variables, w_t, with mean 0 and finite variance σ_w^2. $w_t \sim wn(0, \sigma_w^2)$

If w_t are independent and identically distributed (iid), we write $w_t \sim iid(0, \sigma_w^2)$

If w_t are Gaussian white noise, then $w_t \sim N(0, \sigma_w^2)$

2. Moving Averages: the white noise can be reduced by taking average

Suppose $y_t = a + bx_t + w_t$, $w_t \sim N(0, \sigma_w^2)$

If σ_w^2 is large, the relation between y_t and x_t will be hard to notice. Consider $\dfrac{y_{t-1}+y_t+y_{t+1}}{3} = a + b\dfrac{x_{t-1}+x_t+x_{t+1}}{3} + \dfrac{w_{t-1}+w_t+w_{t+1}}{3}$

The variance of $v_t = \frac{1}{3}(w_{t-1}+w_t+w_{t+1})$ is $Var(v_t) = \frac{1}{9}(\sigma_w^2+\sigma_w^2+\sigma_w^2)$
$= \frac{1}{3}\sigma_w^2$

A linear combination of values in a time series, e.g. $v_t = \frac{1}{3}(w_{t-1}+w_t+w_{t+1})$ is called a filtered series.

3. Autoregressions: A linear regression model with x_t as the response and its past values as the predictors, e.g. $x_t = x_{t-1} - 0.9x_{t-2} + w_t$, $w_t \sim wn(0, \sigma_w^2)$.

Usually we label the observed/generated values starting from $x_1, x_2, x_3, \ldots$ and setting $x_0 = x_{-1} = x_{-2} = \cdots = 0$

The first few generated values are removed to avoid startup problems.

4. Random Walk with Drift : $X_t = \delta + X_{t-1} + W_t$, $X_0 = 0$

W_t is white noise. The constant δ is called drift.

When $\delta = 0$, $X_t = X_{t-1} + W_t$ is called a random walk.

Note that $X_t = \delta t + \sum_{j=1}^{t} W_j \Rightarrow EX_t = \delta t$

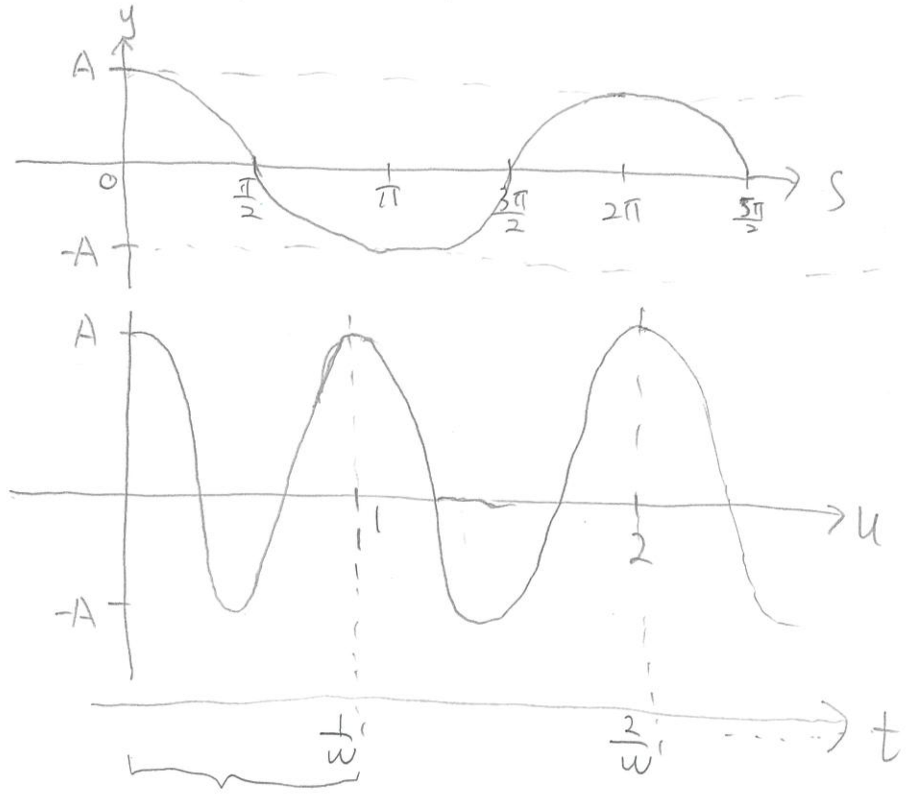
5. Periodic variation and Signal in Noise

Many realistic models for generating time series assume consistent periodic variation. Consider $X_t = A\cos(2\pi\omega t + \phi) + W_t$, —— (1)
where A is the amplitude, ω is the frequency of oscillation, and ϕ is a phase shift.

Consider $y = A\cos(s)$

$y = A\cos(2\pi u)$

$y = A\cos(2\pi\omega t)$



One cycle every $\frac{1}{\omega}$ time points

Back to (1), the first term $A\cos(2\pi\omega t + \phi)$ is regarded as the signal. The ratio A/σ_w is called the signal-to-noise ratio (SNR), here $\sigma_w^2 = Var(W_t)$.

Let $\mu_{X_t} = E(X_t)$ ($= \mu_t$ if no possible confusion)

$\boxed{\text{Definition 1.2}}$ Autocovariance function

$$\gamma_X(s,t) = Cov(X_s, X_t) = E[(X_s - \mu_s)(X_t - \mu_t)] \quad (= \gamma(s,t))$$

Note that $\gamma_X(t,t) = Var(X_t)$

For white noise W_t, $\gamma_W(s,t) = Cov(W_s, W_t) = \begin{cases} \sigma_W^2 & s = t \\ 0 & s \neq t \end{cases}$

$\boxed{\text{Property 1.1}}$ If $U = \sum_{j=1}^{m} a_j X_j$ and $V = \sum_{k=1}^{r} b_k Y_k$, a_j, b_k are constan[ts]

then $Cov(U, V) = \sum_{j=1}^{m} \sum_{k=1}^{r} a_j b_k Cov(X_j, Y_k)$

For moving average $V_t = \frac{1}{3}(W_{t-1} + W_t + W_{t+1})$

$$\gamma_V(s,t) = Cov(V_s, V_t) = \frac{1}{9} Cov(W_{s-1} + W_s + W_{s+1}, W_{t-1} + W_t + W_{t+1})$$

Clearly, for $|s-t| > 2$, $\gamma_V(s,t) = 0$

for $|s-t| = 2$, since $\gamma_V(s,t) = \gamma_V(t,s)$, we only need to consider $s = t+2$

$$\gamma_V(s,t) = \frac{1}{9} Cov(W_{t+1} + W_{t+2} + W_{t+3}, W_{t-1} + W_t + W_{t+1})$$
$$= \frac{1}{9} Cov(W_{t+1}, W_{t+1}) = \frac{1}{9}\sigma_W^2$$

for $|s-t| = 1$, consider $s = t+1$

$$\gamma_V(s,t) = \frac{1}{9} Cov(W_t + W_{t+1} + W_{t+2}, W_{t-1} + W_t + W_{t+1})$$
$$= \frac{1}{9}\left(Cov(W_t, W_t) + Cov(W_{t+1}, W_{t+1})\right) = \frac{2}{9}\sigma_W^2$$

for $s = t$, $\gamma_V(s,t) = \frac{1}{3}\sigma_W^2$

$$\therefore \gamma_V(s,t) = \begin{cases} \sigma_W^2/3 & s = t \\ 2\sigma_W^2/9 & |s-t| = 1 \\ \sigma_W^2/9 & |s-t| = 2 \\ 0 & |s-t| > 2 \end{cases}$$

For random walk $X_t = \sum_{j=1}^{t} W_j$, WLOG, for $s < t$

$$\gamma_X(s,t) = Cov(X_s, X_t) = Cov(X_s, X_s + \sum_{j=s+1}^{t} W_j) = Cov(\sum_{j=1}^{s} W_j, \sum_{j=1}^{s} W_j) = s\sigma_W^2$$

In general, $\gamma_X(s,t) = \min\{s,t\}\sigma_W^2$

Definition 1.3 Autocorrelation function (ACF)

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}} = Corr(X_s, X_t),$$ the correlation between X_s and X_t

Note that $-1 \le \rho(s,t) \le 1 \quad \forall s,t$

Similarly, we can define, for two series X_t and y_t,

Cross-covariance function: $\gamma_{xy}(s,t) = Cov(X_s, y_t) = E[(X_s - \mu_{xs})(y_t - \mu_{yt})]$

Cross-correlation function (CCF): $\rho_{xy}(s,t) = \dfrac{\gamma_{xy}(s,t)}{\sqrt{\gamma_x(s,s)\ \gamma_y(t,t)}}$

Definition 1.6 A time series $\{X_t\} = \{X_1, X_2, \dots\}$ is <u>strictly stationary</u>

if $P(X_{t_1} \le c_1, \dots, X_{t_k} \le c_k) = P(X_{t_1+h} \le c_1, \dots, X_{t_k+h} \le c_k)$

for all $k = 1, 2, \dots$, all time points $t_1, t_2, \dots, t_k$,

all numbers $c_1, c_2, \dots, c_k$, and all time shifts $h = 0, \pm 1, \pm 2, \dots$

That is, the joint distribution of any subset of $\{X_t\}$ does not depend on time t

---

Strictly stationary is hard to verify. It is more common to talk about weakly stationary

Definition 1.7 A time series $\{X_t\}$ is <u>weakly stationary</u> if

(i) $E(X_t) = \mu_{xt} = \mu_x$ for all t

(ii) $\gamma(s,t)$ depends on s and t only through their difference $|s-t|$

We will use the term <u>stationary</u> to mean weakly stationary

For stationary time series $\{X_t\}$, we can define

autocovariance function $\gamma(h) = Cov(X_{t+h}, X_t) = E[(X_{t+h} - \mu)(X_t - \mu)]$

autocorrelation function $\rho(h) = \dfrac{\gamma(t+h, t)}{\sqrt{\gamma(t+h, t+h)\gamma(t,t)}} = \dfrac{\gamma(h)}{\gamma(0)}$

To check if a time series $\{X_t\}$ is (weakly) stationary, we need to prove that $E(X_t) = \mu$ and $\gamma_x(t, t+h) = \gamma(h)$ do not depend on time t

For white noise w_t, $E(w_t) = 0$

$$\gamma_w(t, t+h) = cov(w_t, w_{t+h}) = \begin{cases} \sigma_w^2 & h=0 \\ 0 & h \neq 0 \end{cases}$$

$\therefore$ White noise is (weakly) stationary.

If $w_t \sim N(0, \sigma_w^2)$ iid, it is also strictly stationary.

For $v_t = \frac{1}{3}(w_{t-1} + w_t + w_{t+1})$, $E(v_t) = 0$

$$\gamma_v(t, t+h) = \begin{cases} \frac{1}{3}\sigma_w^2 & \text{if } h=0 \\ \frac{2}{9}\sigma_w^2 & \text{if } h=\pm 1 \\ \frac{1}{9}\sigma_w^2 & \text{if } h=\pm 2 \\ 0 & \text{if } |h|>2 \end{cases} \quad \therefore \text{ stationary}$$

The autocorrelation function is $\rho_v(h) = \dfrac{\gamma(h)}{\gamma(0)} = \begin{cases} 1 & h=0 \\ \frac{2}{3} & h=\pm 1 \\ \frac{1}{3} & h=\pm 2 \\ 0 & |h|>2 \end{cases}$

For random walk $x_t = \sum_{j=1}^{t} w_j$,

$\gamma_x(t, t+h) = \min\{t, t+h\}\sigma_w^2$ depends on t

$\therefore$ It is not weakly stationary

Some properties of $\gamma(h)$ for stationary $\{X_t\}$

1. For any $n \geq 1$ and constants $a_1, \dots, a_n$

$$\sum_{j=1}^{n}\sum_{k=1}^{n} a_j a_k \gamma(j-k) = Var(a_1X_1 + \dots + a_nX_n) \geqslant 0$$

2. $|\gamma(h)| \leq \gamma(0)$ (By Cauchy-Schwarz inequality)
3. $\gamma(h) = \gamma(-h)$

$\boxed{\text{Definition 1.10}}$ $\{X_t\}$ and $\{y_t\}$ are said to be <u>jointly stationary</u> if they are each stationary, and the cross-covariance function

$$\gamma_{xy}(t+h, t) = cov(X_{t+h}, y_t) = \gamma_{xy}(h)$$

is a function only of lag h.

Note that $\gamma_{xy}(h) \neq \gamma_{xy}(-h)$ in general as $Cov(X_{t+h}, y_t) \neq Cov(X_t, y_{t+h})$

Definition 1.11 Cross-correlation function (CCF) of jointly stationary $\{X_t\}$ and $\{y_t\}$ is defined as

$$\rho_{xy}(h) = \frac{\gamma_{xy}(h)}{\sqrt{\gamma_x(0)\,\gamma_y(0)}}$$

Example 1.23 Consider $X_t = W_t + W_{t-1}$ $\quad y_t = W_t - W_{t-1}$

$E(X_t) = E(y_t) = 0$

$$\gamma_x(t+h, t) = Cov(X_{t+h}, X_t) = Cov(W_{t+h} + W_{t+h-1}, W_t + W_{t-1}) = \begin{cases} 2\sigma_w^2 & h=0 \\ \sigma_w^2 & |h|=1 \\ 0 & |h|>1 \end{cases}$$

$$\gamma_y(t+h, t) = Cov(W_{t+h} - W_{t+h-1}, W_t - W_{t-1}) = \begin{cases} 2\sigma_w^2 & h=0 \\ -\sigma_w^2 & |h|=1 \\ 0 & |h|>1 \end{cases}$$

$\therefore \{X_t\}$ and $\{y_t\}$ are stationary

$$\gamma_{xy}(t+h, t) = Cov(X_{t+h}, y_t) = Cov(W_{t+h} + W_{t+h-1}, W_t - W_{t-1}) = \begin{cases} 0 & \text{if } h=0 \\ \sigma_w^2 & \text{if } h=1 \\ -\sigma_w^2 & \text{if } h=-1 \\ 0 & \text{if } |h|>1 \end{cases}$$

does not depend on t $\Rightarrow$ jointly stationary

$$\rho_{xy}(h) = \frac{\gamma_{xy}(h)}{\sqrt{\gamma_x(0)\,\gamma_y(0)}} = \begin{cases} 0 & \text{if } h=0 \\ 1/2 & \text{if } h=1 \\ -1/2 & \text{if } h=-1 \\ 0 & \text{if } |h|>1 \end{cases}$$

Consider $\quad y_t = A X_{t-\ell} + W_t$

the series X_t is said to lead y_t for $\ell > 0$, and is said to lag y_t for $\ell < 0$. Assuming the noise W_t is uncorrelated with $\{X_t\}$

The cross-covariance function is

$$\gamma_{yx}(h) = Cov(y_{t+h}, X_t) = Cov(A X_{t+h-\ell} + W_{t+h}, X_t)$$
$$= A\, Cov(X_{t+h-\ell}, X_t) = A\gamma_x(h-\ell)$$

Definition 1.12 A <u>linear process</u>, X_t, is of the form

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j}, \qquad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty$$

If $\psi_j = 0$ for $j < 0$, ie.

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j w_{t-j},$$ X_t is called <u>causal linear process</u>

Definition 1.13 $\{X_t\}$ is a <u>Gaussian process</u> if $\vec{X} = (X_{t_1}, X_{t_2}, \dots, X_{t_n})^T$ for any n, any $t_1 < t_2 < \dots < t_n$, have a multivariate normal distribution.

If $\{X_t\}$ is Gaussian, then $\{X_t\}$ is a causal linear process with $w_t \sim N(0, \sigma_w^2)$ iid.

Estimation of Correlation

For stationary time series $\{X_t\}$, consider $\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$, we have $E(\bar{X}) = \frac{1}{n}\sum_{t=1}^{n} E(X_t) = \mu$. $\therefore \bar{X}$ is an unbiased estimator.

We can estimate $\gamma(h)$ by the <u>sample autocovariance function</u>

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(X_{t+h} - \bar{X})(X_t - \bar{X}) \approx E\left((X_{t+h} - \mu)(X_t - \mu)\right)$$

Note that

1. Dividing by n instead of $n-h$ to ensure that the sample covariance matrix $\left(\widehat{Cov}(X_j, X_k)\right)_{1\le j,k\le n} = \left(\hat{\gamma}(j-k)\right)_{1\le j,k\le n}$ is non-negative definite. No such guarantee if we divide by $n-h$.
2. Neither dividing by n nor $n-h$ yields an unbiased estimator of $\gamma(h)$.

We can then definite the sample autocorrelation function by

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

Property 1.2 If X_t is iid with finite fourth moment, then as $n \to \infty$,

$$\hat{\rho}_X(h) \xrightarrow{D} N\left(0, \frac{1}{\sqrt{n}}\right)$$ for $h = 1, 2, \dots, H$, where H is fixed

∴ For white noise sequence, approximately 95% of $\hat{\rho}(h)$ should be within $\pm 2/\sqrt{n}$. ∴ The plotted ACFs of residuals are helpful to check if an assumed model is correct.

Similarly, we can estimate $\gamma_{xy}(h) = \gamma_{xy}(t+h,t) = E[(X_{t+h}-\mu)(y_t-\mu)]$

by $\hat{\gamma}_{xy}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(X_{t+h}-\bar{X})(y_t-\bar{y})$ (sample cross-covariance)

and $\rho_{xy}(h)$ by $\hat{\rho}_{xy}(h) = \dfrac{\hat{\gamma}_{xy}(h)}{\sqrt{\hat{\gamma}_x(0)\,\hat{\gamma}_y(0)}}$ (sample cross-correlation)

$\boxed{\text{Property 1.3}}$ As $n\to\infty$ $\hat{\rho}_{xy}(h) \xrightarrow{D} N(0, \frac{1}{\sqrt{n}})$ if at least one of the processes is <u>independent white noise</u>

Vector Time Series

It is common that a time series X_t depends on not just its past values, but also some other time series $X_{t2}, X_{t3}, \dots, X_{tp}$.

For example, we can have

$$X_{t1} = 0.2 + 0.3X_{t-1,1} + 0.1X_{t-1,2} + w_{t1}$$

and

$$X_{t2} = -0.1 + 0.7X_{t-1,1} - 0.3X_{t-1,2} + w_{t2}$$

We can rewrite into matrix-vector form:

$$\begin{pmatrix} X_{t1} \\ X_{t2} \end{pmatrix} = \begin{pmatrix} 0.2 \\ -0.1 \end{pmatrix} + \begin{pmatrix} 0.3 & 0.1 \\ 0.7 & -0.3 \end{pmatrix}\begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \end{pmatrix} + \begin{pmatrix} w_{t1} \\ w_{t2} \end{pmatrix}$$

$$\Rightarrow \vec{X}_t = \vec{a} + \vec{A}\vec{X}_{t-1} + \vec{w}_t$$

In general, consider a vector time series $\vec{X}_t = (X_{t1}, X_{t2}, \dots, X_{tp})^T$.

Assume $\vec{X}_t$ is stationary, it means

$$\vec{\mu} = E(\vec{X}_t) \quad \text{does not depend on } t$$
$$= (\mu_1, \mu_2, \dots, \mu_p)^T$$

and the $p\times p$ autocovariance matrix

$$\vec{\Gamma}(h) = E[(\vec{X}_{t+h}-\vec{\mu})(\vec{X}_t-\vec{\mu})^T] \quad \text{depends on the lag } h \text{ only}$$

Note that $\vec{\Gamma}(h)_{ij} = E[(X_{t+h,i}-\mu_i)(X_{tj}-\mu_j)] = \gamma_{ij}(h) = \gamma_{ij}(-h) \Rightarrow \vec{\Gamma}(-h) = \vec{\Gamma}(h)^T$

By using the sample autocovariance $\hat{\gamma}(h)$ and sample cross-covariance $\hat{\gamma}_{xy}(h)$, we can compute the sample autocovariance matrix of X_t

$$\hat{\Gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h} (\vec{X}_{t+h} - \bar{x})(\vec{X}_t - \bar{x})^T , \quad \text{where } \bar{x} = \frac{1}{n}\sum_{t=1}^{n} X_t$$

We can check that $\hat{\Gamma}(-h) = \hat{\Gamma}(h)^T$

An observed series may be indexed by more than time alone. It may be indexed also by, for example, location.

Example 1.30 the soil surface temperature X at location (s_1, s_2) is written as $X_{\vec{s}} = X_{s_1, s_2}$

Since $\vec{s}$ is a vecter, the lag in each dimension can also be different. For example, it makes sense to say X_{s_1,s_2} is highly correlated with $X_{s_1+h_1, s_2+h_2}$. In general, for $\vec{s} = (s_1, s_2, \ldots, s_r)$ and $\vec{h} = (h_1, \ldots, h_r)$, we define the autocovariance function of a stationary multidimensional process, $X_{\vec{s}}$, as

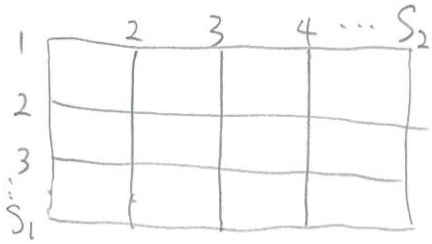
$$\gamma(\vec{h}) = E\left[(X_{\vec{s}+\vec{h}} - \mu)(X_{\vec{s}} - \mu)\right], \text{ where}$$

$\mu = E(X_{\vec{s}})$ does not depend on the index $\vec{s}$.

We can estimate μ by taking the average of all observed $X_{\vec{s}}$.

Suppose we have X_{s_1,s_2} for $s_1 = 1, 2, \ldots, S_1$, $s_2 = 1, 2, \ldots, S_2$ then we estimate μ by

$$\bar{x} = \frac{1}{S_1 S_2} \sum_{s_1=1}^{S_1} \sum_{s_2=1}^{S_2} X_{s_1,s_2}$$



In general, $\bar{x} = (S_1 S_2 \cdots S_r)^{-1} \sum_{s_1} \cdots \sum_{s_r} X_{s_1,\ldots,s_r}$

Similarly, we can estimate $\gamma(\vec{h})$ by taking average of $(X_{\vec{s}+\vec{h}} - \bar{x})(X_{\vec{s}} - \bar{x})$ with all possible $\vec{s}$, i.e.

$$\hat{\gamma}(\vec{h}) = (S_1 S_2 \cdots S_r)^{-1} \sum_{s_1}\sum_{s_2}\cdots\sum_{s_r} (X_{\vec{s}+\vec{h}} - \bar{x})(X_{\vec{s}} - \bar{x}),$$

the range of summation for each argument is $1 \le s_i \le S_i - h_i$, $i = 1, \ldots, r$.