DS-GA 3001.008 Modelling time series data

2. AR(I)MA

Instructor: Cristina Savin

NYU, CNS & CDS

Recap: basic statistics of a time series

Time series $\{X_t\}$

Recap: basic statistics of a time series

Time series $\{X_t\}$

Mean $\boxed{\mu_X(t) = \mathbb{E}(X_t)}$

Recap: basic statistics of a time series

Time series $\{X_t\}$

Mean $\mu_X(t) = \mathbb{E}(X_t)$

Covariance $R_X(t, u) = \mathrm{cov}(X_t, X_u)$

Recap: basic statistics of a time series

Time series $\{X_t\}$

Mean $\mu_X(t) = \mathbb{E}(X_t)$

Covariance $R_X(t,u) = \mathrm{cov}(X_t, X_u)$

Auto-Correlation Function (ACF) $\rho_X(t,u) = \frac{R_X(t,u)}{\sqrt{R_X(t,t), R_X(u,u)}}$

Recap: basic statistics of a time series

Time series $\{X_t\}$

Mean $\mu_X(t) = \mathbb{E}(X_t)$

Covariance $R_X(t,u) = \mathrm{cov}(X_t, X_u)$

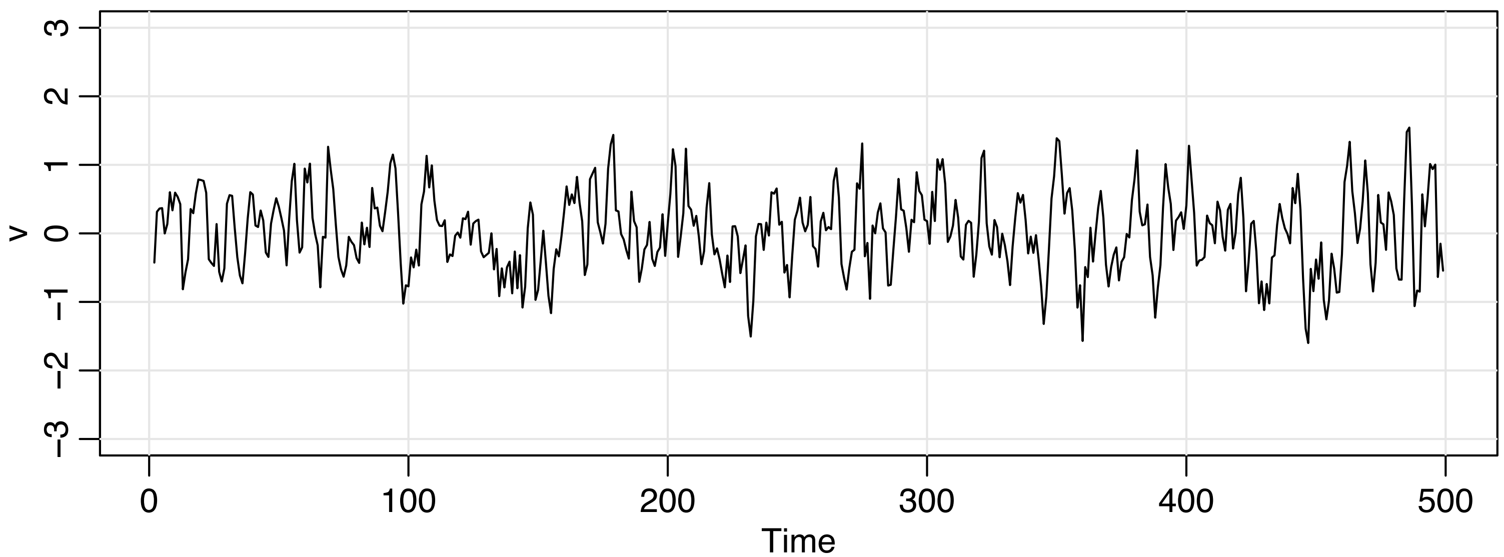
Auto-Correlation Function (ACF) $\rho_X(t,u) = \frac{R_X(t,u)}{\sqrt{R_X(t,t), R_X(u,u)}}$

Cross-Covariance $R_{X,Y}(t,u) = \mathrm{cov}\left(X_t, Y_u\right)$

Recap: basic statistics of a time series

Time series	$\{X_t\}$
Mean	$\mu_X(t) = \mathbb{E}(X_t)$
Covariance	$R_X(t,u) = \mathrm{cov}(X_t, X_u)$
Auto-Correlation Function (ACF)	$\rho_X(t,u) = \frac{R_X(t,u)}{\sqrt{R_X(t,t), R_X(u,u)}}$
Cross-Covariance	$R_{X,Y}(t,u) = \mathrm{cov}\left(X_t, Y_u\right)$
Cross-Correlation Function (CCF)	$\rho_{X,Y}(t,u) = \frac{R_{X,Y}(t,u)}{\sqrt{R_{X,Y}(t,u), R_{X,Y}(u,u)}}$

White noise $W_t \sim \mathcal{N}(0, \sigma^2)$ i.i.d.



Trivially, white noise has

$$\mu_W(t) = 0$$

$$R_W(t,u) = \begin{cases} \sigma^2, t = u \\ 0, t \neq u \end{cases}$$

Moving averages, e.g. MA(1)

$$\boxed{X_t = W_t + \lambda W_{t-1}}$$

Where $\{W_t\}$ is white noise

Moments:

Moving averages, e.g. MA(1)

$$X_t = W_t + \lambda W_{t-1}$$

Where $\{W_t\}$ is white noise

Moments:

$$\mu_X = 0$$

Moving averages, e.g. MA(1)

$$X_t = W_t + \lambda W_{t-1}$$

Where $\{W_t\}$ is white noise

Moments:

$$\mu_X = 0$$

$$R_X(t, t+h) = \begin{cases} \sigma^2\left(1+\lambda^2\right), & h = 0 \\ \sigma^2\lambda, & |h| = 1 \\ 0, & \text{otherwise} \end{cases}$$

Moving averages, e.g. MA(1)

$$X_t = W_t + \lambda W_{t-1}$$

Where $\{W_t\}$ is white noise

Moments:

$$\mu_X = 0$$

$$R_X(t, t+h) = \begin{cases} \sigma^2 \left(1+\lambda^2\right), & h = 0 \\ \sigma^2 \lambda, & |h| = 1 \\ 0, & \text{otherwise} \end{cases}$$

(weakly) stationary

Moving averages, e.g. MA(1)

$$X_t = W_t + \lambda W_{t-1}$$

Where $\{W_t\}$ is white noise

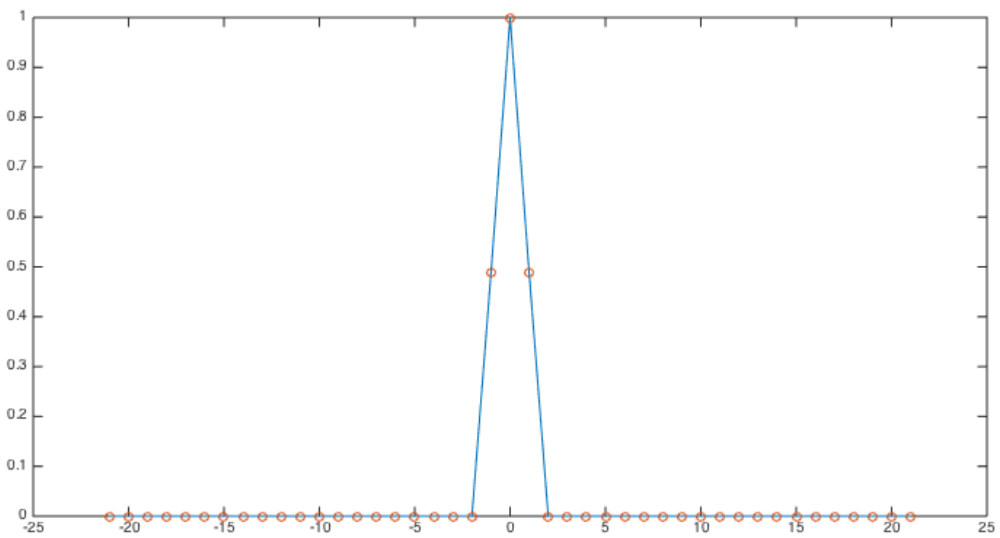
Moments:

$$\mu_X = 0$$

$$R_X(t, t+h) = \begin{cases} \sigma^2 \left(1 + \lambda^2\right), & h = 0 \\ \sigma^2 \lambda, & |h| = 1 \\ 0, & \text{otherwise} \end{cases}$$

(weakly) stationary

ACF



Autoregressive process AR(1)

$$X_t = \lambda X_{t-1} + W_t$$

Where $\{W_t\}$ is white noise and $|\lambda| < 1$

Autoregressive process AR(1)

$$X_t = \lambda X_{t-1} + W_t$$

Where $\{W_t\}$ is white noise and $|\lambda| < 1$

By expanding the recursion we get: $X_t = W_t + \lambda W_{t-1} + \lambda^2 W_{t-2} + \cdots$

Autoregressive process AR(1)

$$X_t = \lambda X_{t-1} + W_t$$

Where $\{W_t\}$ is white noise and $|\lambda| < 1$

By expanding the recursion we get: $X_t = W_t + \lambda W_{t-1} + \lambda^2 W_{t-2} + \cdots$

$$\mu_X = \mathbb{E}\left[\sum_{h=0}^{\infty} \lambda^h W_{t-h}\right] = 0$$

Autoregressive process AR(1)

$$X_t = \lambda X_{t-1} + W_t$$

Where $\{W_t\}$ is white noise and $|\lambda| < 1$

By expanding the recursion we get: $X_t = W_t + \lambda W_{t-1} + \lambda^2 W_{t-2} + \cdots$

$$\mu_X = \mathbb{E}\left[\sum_{h=0}^{\infty} \lambda^h W_{t-h}\right] = 0$$

$$\mathbb{E}\left[X_t^2\right] = \mathbb{E}\left[\sum_h \lambda^{2h} W_{t-h}^2\right] = \sigma^2 \sum \lambda^{2h} = \frac{\sigma^2}{1-\lambda^2}$$

Autoregressive process AR(1)

$$X_t = \lambda X_{t-1} + W_t$$

Where $\{W_t\}$ is white noise and $|\lambda| < 1$

By expanding the recursion we get: $X_t = W_t + \lambda W_{t-1} + \lambda^2 W_{t-2} + \cdots$

$$\mu_X = \mathbb{E}\left[\sum_{h=0}^{\infty} \lambda^h W_{t-h}\right] = 0$$

$$\mathbb{E}\left[X_t^2\right] = \mathbb{E}\left[\sum_h \lambda^{2h} W_{t-h}^2\right] = \sigma^2 \sum \lambda^{2h} = \frac{\sigma^2}{1-\lambda^2}$$

For now, assume $h > 0$

Autoregressive process AR(1)

$$X_t = \lambda X_{t-1} + W_t$$

Where $\{W_t\}$ is white noise and $|\lambda| < 1$

By expanding the recursion we get: $X_t = W_t + \lambda W_{t-1} + \lambda^2 W_{t-2} + \cdots$

$$\mu_X = \mathbb{E}\left[\sum_{h=0}^{\infty} \lambda^h W_{t-h}\right] = 0$$

$$\mathbb{E}\left[X_t^2\right] = \mathbb{E}\left[\sum_h \lambda^{2h} W_{t-h}^2\right] = \sigma^2 \sum \lambda^{2h} = \frac{\sigma^2}{1-\lambda^2}$$

For now, assume $h > 0$

$$R_x(h) = \mathrm{cov}(X_t, X_{t+h}) = \mathrm{cov}(X_t, \lambda X_{t+h-1} + W_{t+h})$$

Autoregressive process AR(1)

$$X_t = \lambda X_{t-1} + W_t$$

Where $\{W_t\}$ is white noise and $|\lambda| < 1$

By expanding the recursion we get: $X_t = W_t + \lambda W_{t-1} + \lambda^2 W_{t-2} + \cdots$

$$\mu_X = \mathbb{E}\left[\sum_{h=0}^{\infty} \lambda^h W_{t-h}\right] = 0$$

$$\mathbb{E}\left[X_t^2\right] = \mathbb{E}\left[\sum_h \lambda^{2h} W_{t-h}^2\right] = \sigma^2 \sum \lambda^{2h} = \frac{\sigma^2}{1-\lambda^2}$$

For now, assume $h > 0$

$$\begin{aligned} R_x(h) = \operatorname{cov}(X_t, X_{t+h}) &= \operatorname{cov}(X_t, \lambda X_{t+h-1} + W_{t+h}) \\ &= \lambda \operatorname{cov}(X_t, X_{t+h-1}) \end{aligned}$$

Autoregressive process AR(1)

$$X_t = \lambda X_{t-1} + W_t$$

Where $\{W_t\}$ is white noise and $|\lambda| < 1$

By expanding the recursion we get: $X_t = W_t + \lambda W_{t-1} + \lambda^2 W_{t-2} + \cdots$

$$\mu_X = \mathbb{E}\left[\sum_{h=0}^{\infty} \lambda^h W_{t-h}\right] = 0$$

$$\mathbb{E}\left[X_t^2\right] = \mathbb{E}\left[\sum_h \lambda^{2h} W_{t-h}^2\right] = \sigma^2 \sum \lambda^{2h} = \frac{\sigma^2}{1-\lambda^2}$$

For now, assume $h > 0$

$$\begin{aligned} R_x(h) = \operatorname{cov}(X_t, X_{t+h}) &= \operatorname{cov}(X_t, \lambda X_{t+h-1} + W_{t+h}) \\ &= \lambda \operatorname{cov}(X_t, X_{t+h-1}) \\ &= \lambda^h \operatorname{cov}(X_t, X_t) \end{aligned}$$

Autoregressive process AR(1)

$$X_t = \lambda X_{t-1} + W_t$$

Where $\{W_t\}$ is white noise and $|\lambda| < 1$

By expanding the recursion we get: $X_t = W_t + \lambda W_{t-1} + \lambda^2 W_{t-2} + \cdots$

$$\mu_X = \mathbb{E}\left[\sum_{h=0}^{\infty} \lambda^h W_{t-h}\right] = 0$$

$$\mathbb{E}\left[X_t^2\right] = \mathbb{E}\left[\sum_h \lambda^{2h} W_{t-h}^2\right] = \sigma^2 \sum \lambda^{2h} = \frac{\sigma^2}{1-\lambda^2}$$

For now, assume $h > 0$

$$\begin{aligned} R_x(h) = \operatorname{cov}(X_t, X_{t+h}) &= \operatorname{cov}(X_t, \lambda X_{t+h-1} + W_{t+h}) \\ &= \lambda \operatorname{cov}(X_t, X_{t+h-1}) \\ &= \lambda^h \operatorname{cov}(X_t, X_t) \\ &= \sigma^2 \frac{\lambda^{|h|}}{1-\lambda^2} \end{aligned}$$

Autoregressive process AR(1)

$$X_t = \lambda X_{t-1} + W_t$$

Where $\{W_t\}$ is white noise and $|\lambda| < 1$

By expanding the recursion we get: $X_t = W_t + \lambda W_{t-1} + \lambda^2 W_{t-2} + \cdots$

$$\mu_X = \mathbb{E}\left[\sum_{h=0}^{\infty} \lambda^h W_{t-h}\right] = 0$$

$$\mathbb{E}\left[X_t^2\right] = \mathbb{E}\left[\sum_h \lambda^{2h} W_{t-h}^2\right] = \sigma^2 \sum \lambda^{2h} = \frac{\sigma^2}{1-\lambda^2}$$

For now, assume $h > 0$

$$\begin{aligned} R_x(h) = \operatorname{cov}(X_t, X_{t+h}) &= \operatorname{cov}(X_t, \lambda X_{t+h-1} + W_{t+h}) \\ &= \lambda \operatorname{cov}(X_t, X_{t+h-1}) \\ &= \lambda^h \operatorname{cov}(X_t, X_t) \\ &= \sigma^2 \frac{\lambda^{|h|}}{1-\lambda^2} \end{aligned}$$

(weakly) stationary

Autoregressive process AR(1)

$$X_t = \lambda X_{t-1} + W_t$$

Where $\{W_t\}$ is white noise and $|\lambda| < 1$

By expanding the recursion we get: $X_t = W_t + \lambda W_{t-1} + \lambda^2 W_{t-2} + \cdots$

$$\mu_X = \mathbb{E}\left[\sum_{h=0}^{\infty} \lambda^h W_{t-h}\right] = 0$$

$$\mathbb{E}\left[X_t^2\right] = \mathbb{E}\left[\sum_h \lambda^{2h} W_{t-h}^2\right] = \sigma^2 \sum \lambda^{2h} = \frac{\sigma^2}{1-\lambda^2}$$

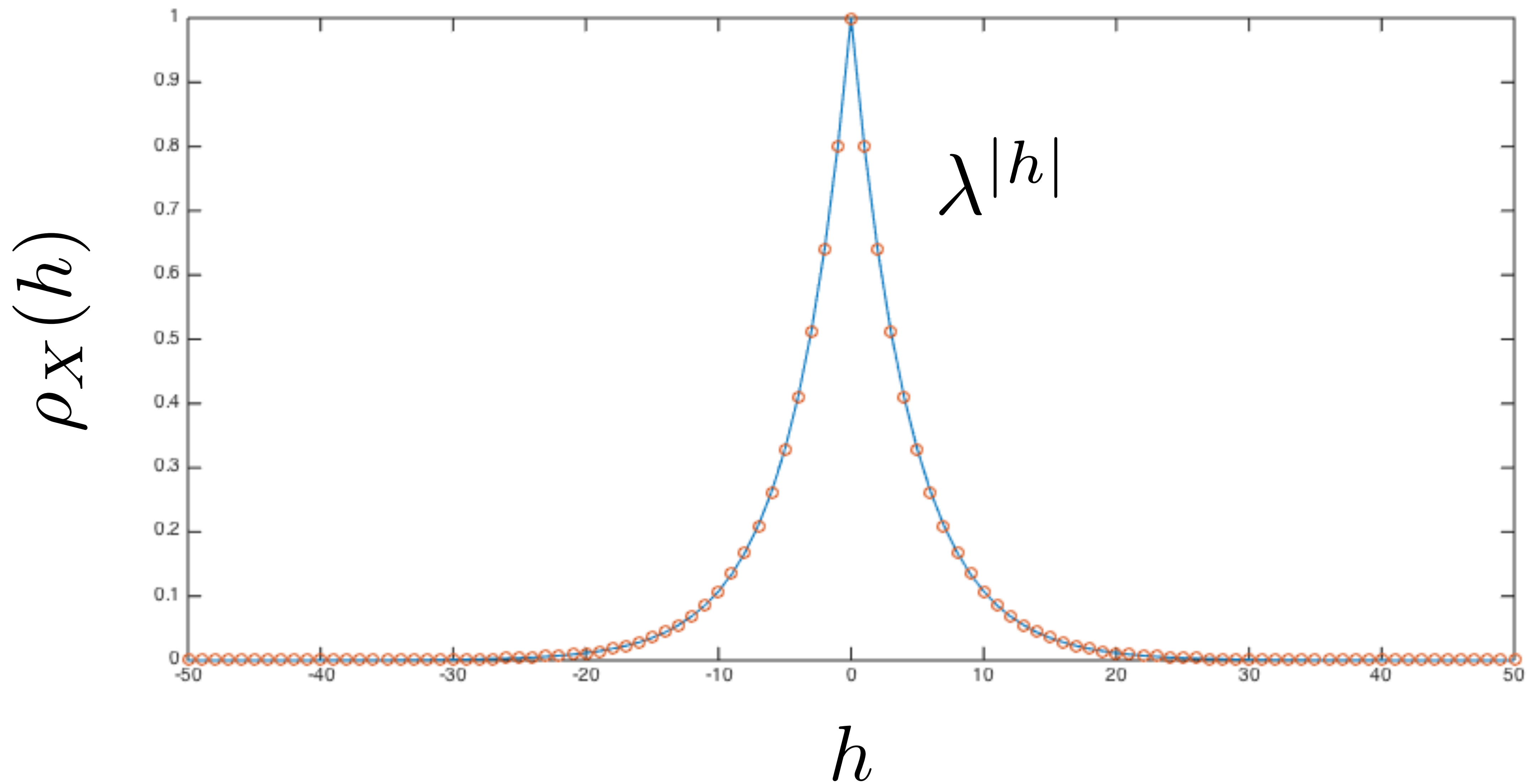
For now, assume $h > 0$

$$\begin{aligned} R_x(h) = \operatorname{cov}(X_t, X_{t+h}) &= \operatorname{cov}(X_t, \lambda X_{t+h-1} + W_{t+h}) \\ &= \lambda \operatorname{cov}(X_t, X_{t+h-1}) \\ &= \lambda^h \operatorname{cov}(X_t, X_t) \\ &= \sigma^2 \frac{\lambda^{|h|}}{1-\lambda^2} \end{aligned}$$

(weakly) stationary

*Check other direction at home

AR(1) ACF



$\rho_X(h)$
$\lambda^{|h|}$
h
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-50
-40
-30
-20
-10
0
10
20
30
40
50

AR and **MA** are special instances of a **linear process**

AR and **MA** are special instances of a **linear process**

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

AR and **MA** are special instances of a **linear process**

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

$$\mu_X = \mu$$

AR and **MA** are special instances of a **linear process**

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

$$\mu_X = \mu$$

$$R_X(h) = \sigma^2 \sum_k \psi_k \psi_{k+h}$$

AR and **MA** are special instances of a **linear process**

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

$$\mu_X = \mu$$

$$R_X(h) = \sigma^2 \sum_k \psi_k \psi_{k+h}$$

***Useful: Cov. of linear combinations**

$$U = \sum_i a_i X_i$$

$$V = \sum_i b_i Y_i$$

$$\mathrm{cov}(V, U) = \sum_{i,j} a_i b_j \mathrm{cov}(X_i, Y_j)$$

AR and **MA** are special instances of a **linear process**

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

***Useful: Cov. of linear combinations**

$$U = \sum_i a_i X_i$$
$$V = \sum_i b_i Y_i$$
$$\text{cov}(V, U) = \sum_{i,j} a_i b_j \text{cov}(X_i, Y_j)$$

$$\mu_X = \mu$$

$$R_X(h) = \sigma^2 \sum_k \psi_k \psi_{k+h}$$

Special cases:

$\mu = 0$

AR and **MA** are special instances of a **linear process**

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

$$\mu_X = \mu$$

$$R_X(h) = \sigma^2 \sum_k \psi_k \psi_{k+h}$$

***Useful: Cov. of linear combinations**

$$U = \sum_i a_i X_i$$
$$V = \sum_i b_i Y_i$$
$$\mathrm{cov}(V, U) = \sum_{i,j} a_i b_j \mathrm{cov}(X_i, Y_j)$$

Special cases:

$\mu = 0$

White noise $\psi_k = \begin{cases} 1 & \text{if } k = 0 \text{ ,} \\ 0 & \text{otherwise.} \end{cases}$

MA(1) $\psi_k = \begin{cases} 1 & \text{if } k = 0 \text{ ,} \\ \lambda & \text{if } k = 1 \text{ ,} \\ 0 & \text{otherwise.} \end{cases}$

AR(1) $\psi_k = \begin{cases} \lambda^k & \text{if } k \geq 0 \text{ ,} \\ 0 & \text{otherwise.} \end{cases}$

AR and **MA** are special instances of a **linear process**

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

$$\mu_X = \mu$$

$$R_X(h) = \sigma^2 \sum_k \psi_k \psi_{k+h}$$

***Useful: Cov. of linear combinations**

$$U = \sum_i a_i X_i$$
$$V = \sum_i b_i Y_i$$
$$\mathrm{cov}(V, U) = \sum_{i,j} a_i b_j \mathrm{cov}(X_i, Y_j)$$

Special cases:

$\mu = 0$

White noise $\psi_k = \begin{cases} 1 & \text{if } k = 0\ , \\ 0 & \text{otherwise.} \end{cases}$

MA(1) $\psi_k = \begin{cases} 1 & \text{if } k = 0\ , \\ \lambda & \text{if } k = 1\ , \\ 0 & \text{otherwise.} \end{cases}$

AR(1) $\psi_k = \begin{cases} \lambda^k & \text{if } k \geq 0\ , \\ 0 & \text{otherwise.} \end{cases}$

How about the random walk?

$$X_t = \sum_{0 \leq k \leq t} W_{t-k}$$

$$\neq \sum_k \psi_k W_{t-k}$$

Prediction

Suppose $\{X_t\}$ is a linear process

Prediction

Suppose $\{X_t\}$ is a linear process

How can we use observed data to predict what happens next?

Prediction

Suppose $\{X_t\}$ is a linear process

How can we use observed data to predict what happens next?

How does the prediction depend on ACF?

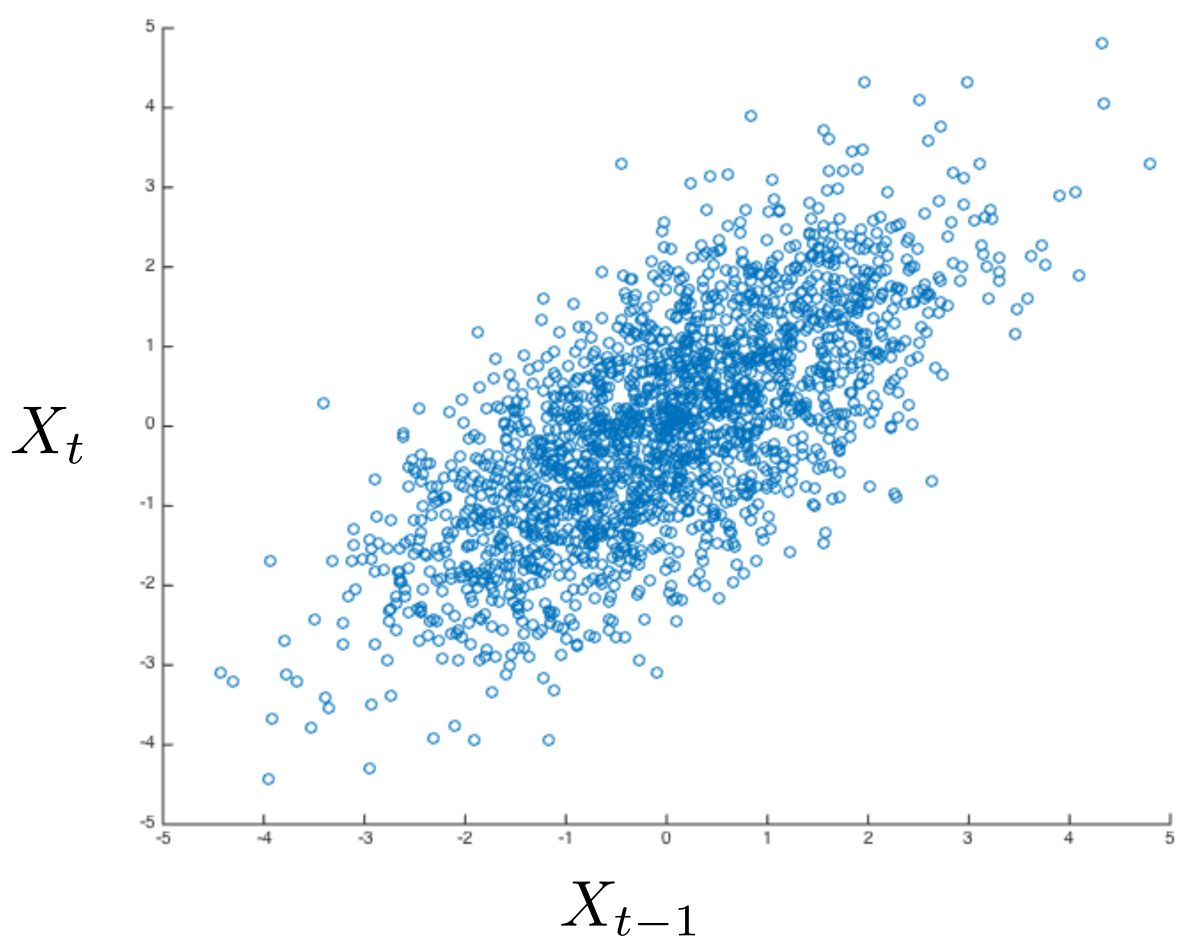
Prediction

Suppose $\{X_t\}$ is a linear process

How can we use observed data to predict what happens next?

How does the prediction depend on ACF?

example: AR(1)



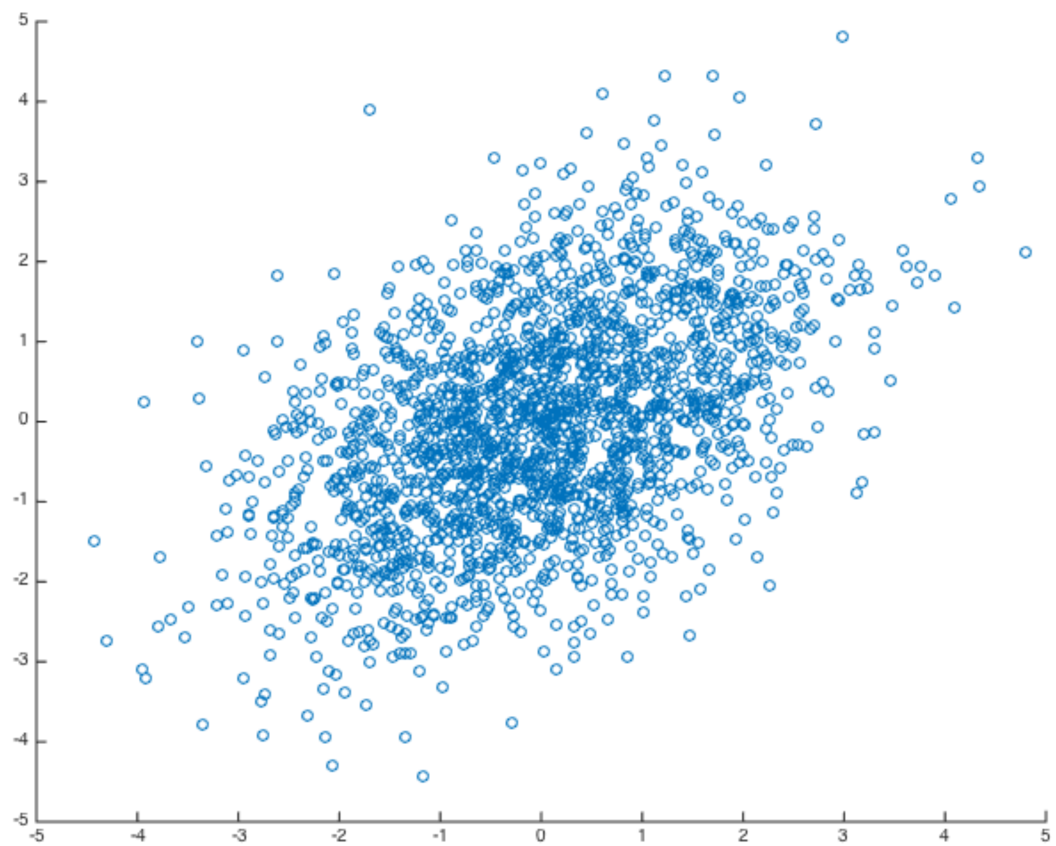
Prediction

Suppose $\{X_t\}$ is a linear process

How can we use observed data to predict what happens next?

How does the prediction depend on ACF?

example: AR(1)



X_t

X_{t-2}

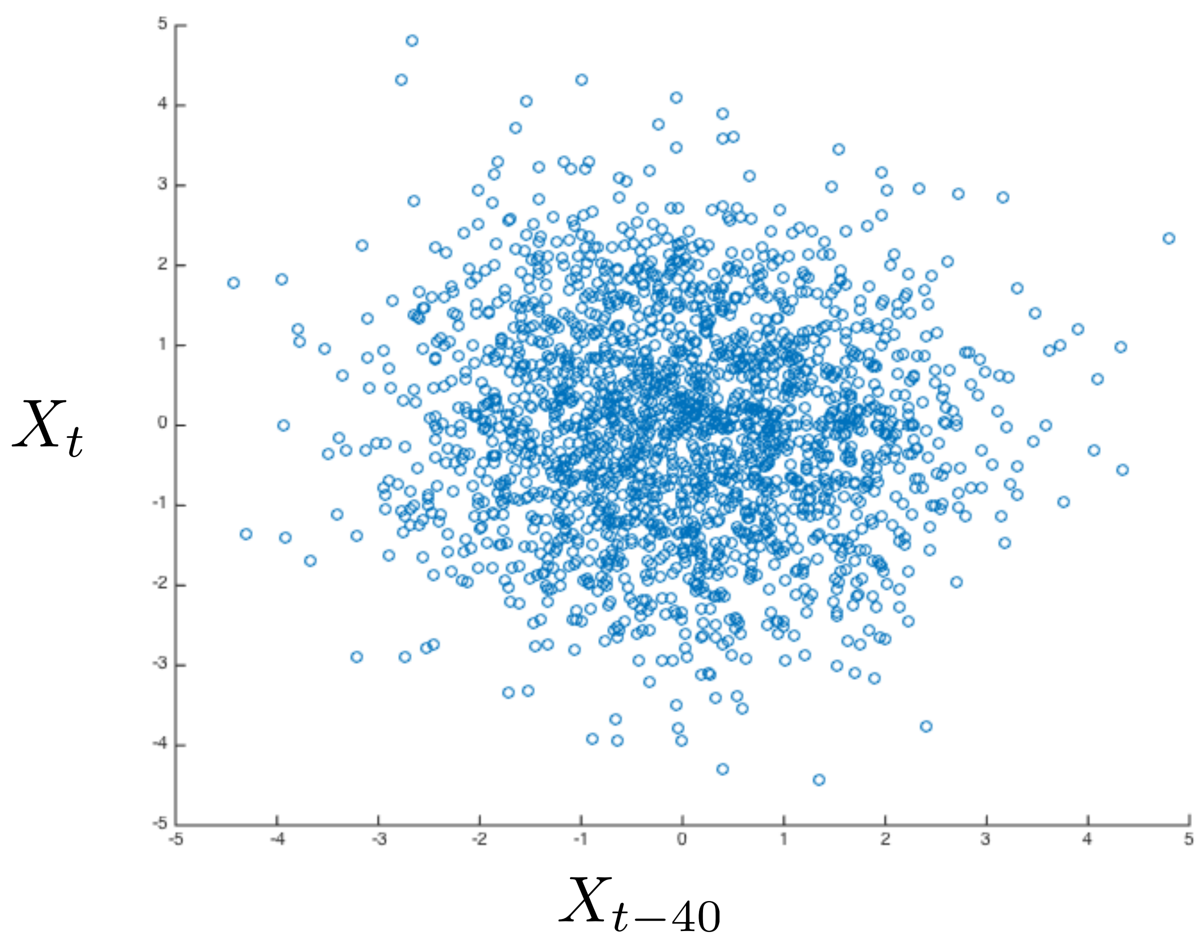
Prediction

Suppose $\{X_t\}$ is a linear process

How can we use observed data to predict what happens next?

How does the prediction depend on ACF?

example: AR(1)



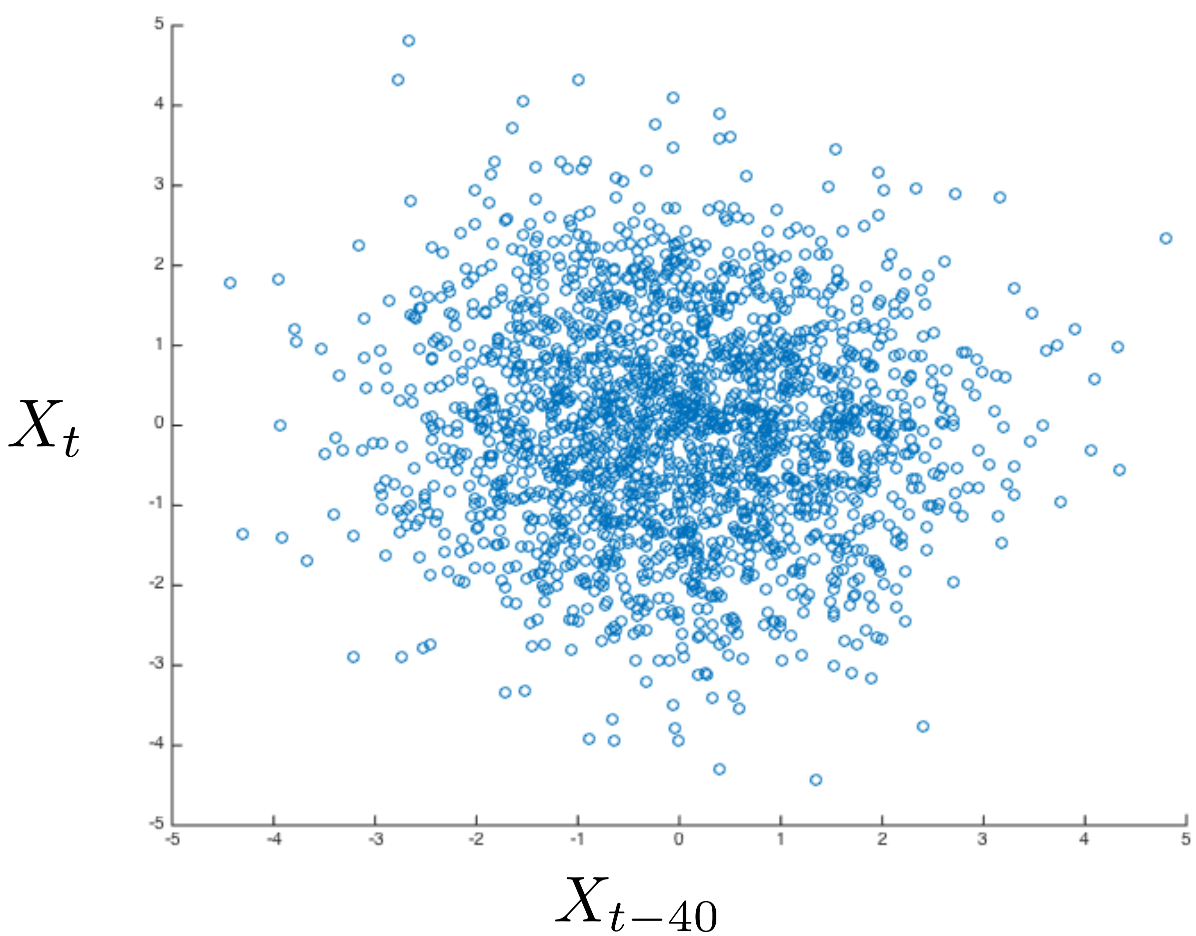
Prediction

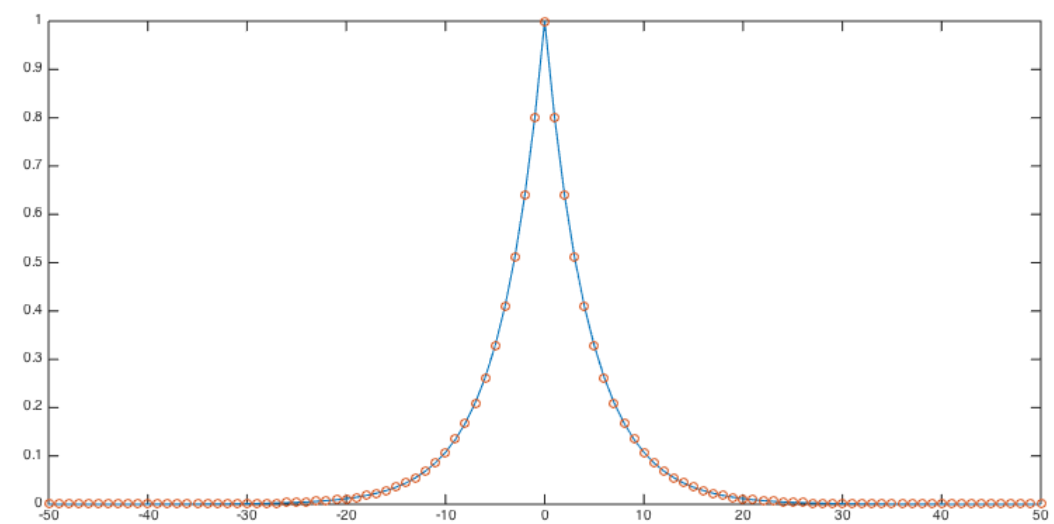
Suppose $\{X_t\}$ is a linear process

How can we use observed data to predict what happens next?

How does the prediction depend on ACF?

example: AR(1)





ACF determines
linear predictability

Least squares and ACF

Least squares and ACF

Least squares estimation reminder

Least squares and ACF

Least squares estimation reminder

$$\hat{f} = \mathrm{argmin}_f (Y - f)^2$$

Least squares and ACF

Least squares estimation reminder

$$\hat{f} = \mathrm{argmin}_f (Y - f)^2$$

$$\hat{f} = \mathbb{E}[Y]$$

Least squares and ACF

Least squares estimation reminder

$$\hat{f} = \mathrm{argmin}_f (Y - f)^2$$

$$\hat{f} = \mathbb{E}[Y]$$

With MSE $\mathrm{var}[Y]$

Least squares and ACF

Least squares estimation reminder

$$\hat{f} = \mathrm{argmin}_f (Y - f)^2$$

$$\hat{f} = \mathbb{E}[Y|X]$$

With MSE $\mathrm{var}[Y|X]$

Least squares and ACF

Least squares estimation reminder

$$\hat{f} = \text{argmin}_f (Y - f)^2$$

$$\hat{f} = \mathbb{E}[Y|X]$$

With MSE $\text{var}[Y|X]$

We can compute a least square estimate of X_{t+h} given X_t

Least squares and ACF

Least squares estimation reminder

$$\hat{f} = \mathrm{argmin}_f (Y - f)^2$$

$$\hat{f} = \mathbb{E}[Y|X]$$

With MSE $\mathrm{var}[Y|X]$

We can compute a least square estimate of X_{t+h} given X_t

If everything is Gaussian, then conditional expectations are easy!

Least squares and ACF

Least squares estimation reminder

$$\hat{f} = \mathrm{argmin}_f (Y - f)^2$$

$$\hat{f} = \mathbb{E}[Y|X]$$

With MSE $\mathrm{var}[Y|X]$

We can compute a least square estimate of X_{t+h} given X_t

If everything is Gaussian, then conditional expectations are easy!

If $\{X_t\}$ is jointly gaussian

Least squares and ACF

Least squares estimation reminder

$$\hat{f} = \mathrm{argmin}_f (Y - f)^2$$

$$\hat{f} = \mathbb{E}[Y|X]$$

With MSE $\mathrm{var}[Y|X]$

We can compute a least square estimate of X_{t+h} given X_t

If everything is Gaussian, then conditional expectations are easy!

If $\{X_t\}$ is jointly gaussian

$$f_X(x) = \frac{1}{2\pi^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

Least squares and ACF

Least squares estimation reminder

$$\hat{f} = \text{argmin}_f (Y - f)^2$$

$$\hat{f} = \mathbb{E}[Y|X]$$

With MSE $\text{var}[Y|X]$

We can compute a least square estimate of X_{t+h} given X_t

If everything is Gaussian, then conditional expectations are easy!

If $\{X_t\}$ is jointly gaussian

$$f_X(x) = \frac{1}{2\pi^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

The pair (X_t, X_{t+h}) is also jointly gaussian, with covariance

Least squares and ACF

Least squares estimation reminder

$$\hat{f} = \mathrm{argmin}_f (Y - f)^2$$

$$\hat{f} = \mathbb{E}[Y|X]$$

With MSE $\mathrm{var}[Y|X]$

We can compute a least square estimate of X_{t+h} given X_t

If everything is Gaussian, then conditional expectations are easy!

If $\{X_t\}$ is jointly gaussian

$$f_X(x) = \frac{1}{2\pi^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

The pair (X_t, X_{t+h}) is also jointly gaussian, with covariance

$$\begin{pmatrix} \sigma_t^2 & \rho(t, t+h)\sigma_t\sigma_{t+h} \\ \rho(t, t+h)\sigma_t\sigma_{t+h} & \sigma_{t+h}^2 \end{pmatrix}$$

Least squares and ACF

Least squares estimation reminder

$$\hat{f} = \mathrm{argmin}_f (Y - f)^2$$

$$\hat{f} = \mathbb{E}[Y|X]$$

With MSE $\mathrm{var}[Y|X]$

We can compute a least square estimate of X_{t+h} given X_t

If everything is Gaussian, then conditional expectations are easy!

If $\{X_t\}$ is jointly gaussian

$$f_X(x) = \frac{1}{2\pi^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

The pair (X_t, X_{t+h}) is also jointly gaussian, with covariance

$$\begin{pmatrix} \sigma_t^2 & \rho(t, t+h)\sigma_t\sigma_{t+h} \\ \rho(t, t+h)\sigma_t\sigma_{t+h} & \sigma_{t+h}^2 \end{pmatrix}$$

$$X_{t+h}|X_t = x_t$$

Least squares and ACF

Least squares estimation reminder

$$\hat{f} = \mathrm{argmin}_f (Y - f)^2$$

$$\hat{f} = \mathbb{E}[Y|X]$$

With MSE $\mathrm{var}[Y|X]$

We can compute a least square estimate of X_{t+h} given X_t

If everything is Gaussian, then conditional expectations are easy!

If $\{X_t\}$ is jointly gaussian

$$f_X(x) = \frac{1}{2\pi^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

The pair (X_t, X_{t+h}) is also jointly gaussian, with covariance

$$\begin{pmatrix} \sigma_t^2 & \rho(t, t+h)\sigma_t\sigma_{t+h} \\ \rho(t, t+h)\sigma_t\sigma_{t+h} & \sigma_{t+h}^2 \end{pmatrix}$$

$$X_{t+h}|X_t = x_t$$

$$\mathcal{N}\left(\mu_{t+h} + \frac{\sigma_{t+h}\rho(t, t+h)(x_t - \mu_t)}{\sigma_t}, \sigma^2(1 - \rho(t, t+h)^2)\right)$$

Least squares and ACF

Least squares estimation reminder

$$\hat{f} = \mathrm{argmin}_f (Y - f)^2$$
$$\hat{f} = \mathbb{E}[Y|X]$$

With MSE $\mathrm{var}[Y|X]$

We can compute a least square estimate of X_{t+h} given X_t

If everything is Gaussian, then conditional expectations are easy!

If $\{X_t\}$ is jointly gaussian

$$f_X(x) = \frac{1}{2\pi^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

The pair (X_t, X_{t+h}) is also jointly gaussian, with covariance

$$\begin{pmatrix} \sigma_t^2 & \rho(t,t+h)\sigma_t\sigma_{t+h} \\ \rho(t,t+h)\sigma_t\sigma_{t+h} & \sigma_{t+h}^2 \end{pmatrix}$$

$$X_{t+h}|X_t = x_t$$

$$\mathcal{N}\left(\mu_{t+h} + \frac{\sigma_{t+h}\rho(t,t+h)(x_t-\mu_t)}{\sigma_t}, \sigma^2(1-\rho(t,t+h)^2)\right)$$

*Gaussian identities .pdf

For a gaussian stationary process, the optimal predictor for $X_{t+h}|X_t = x_t$

takes the form:

$$f(x_t) = \mathbf{E}\left(X_{t+h}|X_t = x_t\right) = \mu + \rho_X(h)(x_t - \mu)$$

With MSE

$$\mathbf{E}\left(|X_{t+h} - f(x_t)|^2 \ , |X_t = x_t\right) = \sigma^2(1 - \rho_X(h)^2)$$

For a gaussian stationary process, the optimal predictor for $X_{t+h}|X_t = x_t$

takes the form:

$$f(x_t) = \mathbf{E}\left(X_{t+h}|X_t = x_t\right) = \mu + \rho_X(h)(x_t - \mu)$$

Linear in x_t

With MSE

$$\mathbf{E}\left(|X_{t+h} - f(x_t)|^2 \ , |X_t = x_t\right) = \sigma^2(1 - \rho_X(h)^2)$$

For a gaussian stationary process, the optimal predictor for $X_{t+h}|X_t = x_t$

takes the form:

$$f(x_t) = \mathbf{E}\left(X_{t+h}|X_t = x_t\right) = \mu + \rho_X(h)(x_t - \mu)$$

Linear in x_t

With MSE

$$\mathbf{E}\left(|X_{t+h} - f(x_t)|^2 \,, |X_t = x_t\right) = \sigma^2(1 - \rho_X(h)^2)$$

The higher the autocorrelation coeff.
the better the prediction

For a gaussian stationary process, the optimal predictor for $X_{t+h}|X_t = x_t$

takes the form: $$f(x_t) = \mathbf{E}\left(X_{t+h}|X_t = x_t\right) = \mu + \rho_X(h)(x_t - \mu)$$ Linear in x_t

With MSE $$\mathbf{E}\left(|X_{t+h} - f(x_t)|^2\ ,|X_t = x_t\right) = \sigma^2(1 - \rho_X(h)^2)$$

The higher the autocorrelation coeff.
the better the prediction

For more complicated processes, the best **linear** predictor

For a gaussian stationary process, the optimal predictor for $X_{t+h}|X_t = x_t$

takes the form:

$$f(x_t) = \mathbf{E}\left(X_{t+h}|X_t = x_t\right) = \mu + \rho_X(h)(x_t - \mu)$$

Linear in x_t

With MSE

$$\mathbf{E}\left(|X_{t+h} - f(x_t)|^2\ ,|X_t = x_t\right) = \sigma^2(1 - \rho_X(h)^2)$$

The higher the autocorrelation coeff.
the better the prediction

For more complicated processes, the best **linear** predictor

$$\mathbf{E}\left(|X_{t+h} - \alpha - \beta X_t|^2\right) = E(\alpha, \beta)$$

For a gaussian stationary process, the optimal predictor for $X_{t+h}|X_t = x_t$

takes the form:

$$f(x_t) = \mathbf{E}\left(X_{t+h}|X_t = x_t\right) = \mu + \rho_X(h)(x_t - \mu)$$

Linear in x_t

With MSE

$$\mathbf{E}\left(|X_{t+h} - f(x_t)|^2 \ , |X_t = x_t\right) = \sigma^2(1 - \rho_X(h)^2)$$

The higher the autocorrelation coeff.
the better the prediction

For more complicated processes, the best **linear** predictor

$$\mathbf{E}\left(|X_{t+h} - \alpha - \beta X_t|^2\right) = E(\alpha, \beta)$$

$$\alpha = \mu(1 - \rho_X(h)) \ , \ \beta = \rho_X(h)$$

For a gaussian stationary process, the optimal predictor for $X_{t+h}|X_t = x_t$

takes the form:

$$f(x_t) = \mathbf{E}\left(X_{t+h}|X_t = x_t\right) = \mu + \rho_X(h)(x_t - \mu)$$

Linear in x_t

With MSE

$$\mathbf{E}\left(|X_{t+h} - f(x_t)|^2 \ , |X_t = x_t\right) = \sigma^2(1 - \rho_X(h)^2)$$

The higher the autocorrelation coeff.
the better the prediction

For more complicated processes, the best **linear** predictor

$$\mathbf{E}\left(|X_{t+h} - \alpha - \beta X_t|^2\right) = E(\alpha, \beta)$$

$$\alpha = \mu(1 - \rho_X(h)) \ , \ \beta = \rho_X(h)$$

$$MSE = \sigma^2(1 - \rho_X(h)^2)$$

For a gaussian stationary process, the optimal predictor for $X_{t+h}|X_t = x_t$

takes the form:

$$f(x_t) = \mathbf{E}\left(X_{t+h}|X_t = x_t\right) = \mu + \rho_X(h)(x_t - \mu)$$

Linear in x_t

With MSE

$$\mathbf{E}\left(|X_{t+h} - f(x_t)|^2 \;, |X_t = x_t\right) = \sigma^2(1 - \rho_X(h)^2)$$

The higher the autocorrelation coeff.
the better the prediction

For more complicated processes, the best **linear** predictor

$$\mathbf{E}\left(|X_{t+h} - \alpha - \beta X_t|^2\right) = E(\alpha, \beta)$$

minimum->
derivatives zero
(check at home)

$$\alpha = \mu(1 - \rho_X(h)) \;, \; \beta = \rho_X(h)$$

$$MSE = \sigma^2(1 - \rho_X(h)^2)$$

For a gaussian stationary process, the optimal predictor for $X_{t+h}|X_t = x_t$

takes the form:

$$f(x_t) = \mathbf{E}\left(X_{t+h}|X_t = x_t\right) = \mu + \rho_X(h)(x_t - \mu)$$

Linear in x_t

With MSE

$$\mathbf{E}\left(|X_{t+h} - f(x_t)|^2 \;, |X_t = x_t\right) = \sigma^2(1 - \rho_X(h)^2)$$

The higher the autocorrelation coeff.
the better the prediction

For more complicated processes, the best **linear** predictor

$$\mathbf{E}\left(|X_{t+h} - \alpha - \beta X_t|^2\right) = E(\alpha, \beta)$$

minimum->
derivatives zero
(check at home)

$$\alpha = \mu(1 - \rho_X(h)) \;,\; \beta = \rho_X(h)$$

$$MSE = \sigma^2(1 - \rho_X(h)^2)$$

Optimal **linear** predictor

$$f(x_t) = \mu + \rho_X(h)(x_t - \mu)$$

For a gaussian stationary process, the optimal predictor for $X_{t+h}|X_t = x_t$

takes the form:

$$f(x_t) = \mathbf{E}\left(X_{t+h}|X_t = x_t\right) = \mu + \rho_X(h)(x_t - \mu)$$

Linear in x_t

With MSE

$$\mathbf{E}\left(|X_{t+h} - f(x_t)|^2 \ , |X_t = x_t\right) = \sigma^2(1 - \rho_X(h)^2)$$

The higher the autocorrelation coeff.
the better the prediction

For more complicated processes, the best **linear** predictor

$$\mathbf{E}\left(|X_{t+h} - \alpha - \beta X_t|^2\right) = E(\alpha, \beta)$$

minimum->
derivatives zero
(check at home)

$$\alpha = \mu(1 - \rho_X(h)) \ , \ \beta = \rho_X(h)$$

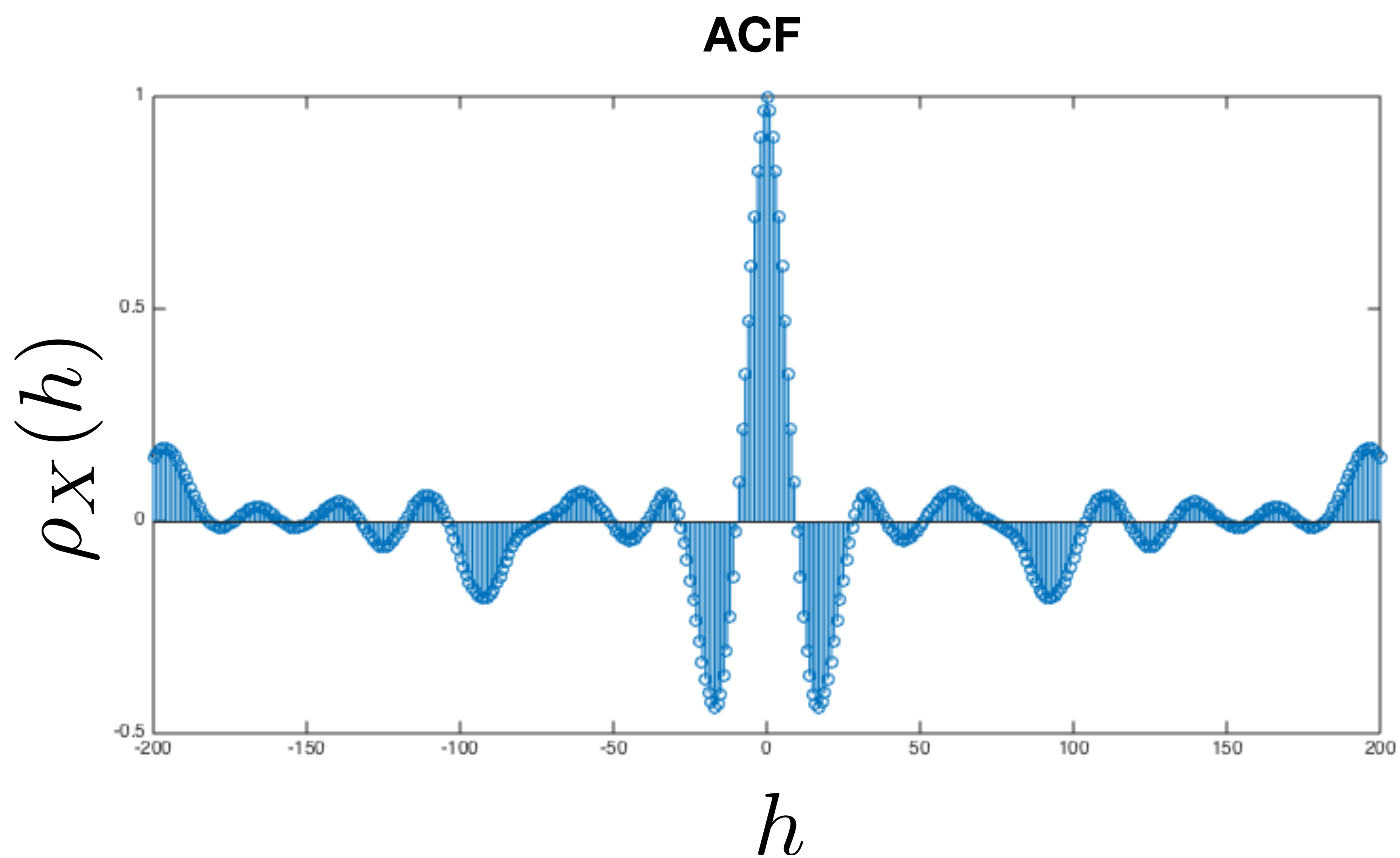
$$MSE = \sigma^2(1 - \rho_X(h)^2)$$

Optimal **linear** predictor

$$f(x_t) = \mu + \rho_X(h)(x_t - \mu)$$

The optimal predictor
if stationary gaussian

How do we use this to model real data?



Real life looks a bit more complicated than a simple AR(1)

Can we combine the basic idea of simple linear processes to get more **expressive** power, while keeping math nice and simple?

Go back to AR(1), rewrite using backshift operator

Rewrite equation

$$X_t - \lambda X_{t-1} = W_t$$

$$(1 - \lambda B)X_t = W_t$$

$$P(B)X_t = W_t$$

Using backshift /translation operator **B**

*properties on the board

Go back to AR(1), rewrite using backshift operator

Rewrite equation

$$X_t - \lambda X_{t-1} = W_t$$
$$(1 - \lambda B)X_t = W_t$$
$$P(B)X_t = W_t$$

Using backshift /translation operator **B**
*properties on the board

We can write the differentiation

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t.$$
$$B^2 X_t = BBX_t = BX_{t-1} = X_{t-2},$$
$$B^k X_t = X_{t-k}.$$

Go back to AR(1), rewrite using backshift operator

Rewrite equation

$$\begin{aligned} X_t - \lambda X_{t-1} &= W_t \\ (1 - \lambda B) X_t &= W_t \\ P(B) X_t &= W_t \end{aligned}$$

Using backshift
/translation operator **B**
*properties on the board

We can write the differentiation

$$\nabla X_t = X_t - X_{t-1} = (1 - B) X_t.$$

$$B^2 X_t = BBX_t = BX_{t-1} = X_{t-2},$$

$$B^k X_t = X_{t-k}.$$

$$X_t = \sum_{k=0} \lambda^k W_{t-k}$$

Go back to AR(1), rewrite using backshift operator

Rewrite equation

$$X_t - \lambda X_{t-1} = W_t$$
$$(1 - \lambda B)X_t = W_t$$
$$P(B)X_t = W_t$$

Using backshift /translation operator **B**
*properties on the board

We can write the differentiation

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t.$$
$$B^2 X_t = BBX_t = BX_{t-1} = X_{t-2},$$
$$B^k X_t = X_{t-k}.$$

$$X_t = \sum_{k=0} \lambda^k W_{t-k} \quad = \sum_{k=0} \lambda^k B^k W_t$$

Go back to AR(1), rewrite using backshift operator

Rewrite equation

$$X_t - \lambda X_{t-1} = W_t$$
$$(1 - \lambda B)X_t = W_t$$
$$P(B)X_t = W_t$$

Using backshift
/translation operator **B**
*properties on the board

We can write the differentiation

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t.$$
$$B^2 X_t = BBX_t = BX_{t-1} = X_{t-2},$$
$$B^k X_t = X_{t-k}.$$

$$X_t = \sum_{k=0} \lambda^k W_{t-k} = \underbrace{\boxed{\sum_{k=0} \lambda^k B^k}}_{Q(B)} W_t$$

$P(B) = 1 - \lambda B$ *and* $Q(B) = \sum_{k \geq 0} \lambda^k B^k$ *are related by*

$$P(B)Q(B) = 1 \ , \quad \textit{or } Q(B) = P(B)^{-1} \ .$$

$P(B) = 1 - \lambda B$ *and* $Q(B) = \sum_{k \geq 0} \lambda^k B^k$ *are related by*

$$P(B)Q(B) = 1 \ , \quad \textit{or } Q(B) = P(B)^{-1} \ .$$

Since $P(B)X_t = W_t$

$P(B) = 1 - \lambda B$ and $Q(B) = \sum_{k \geq 0} \lambda^k B^k$ are related by

$$P(B)Q(B) = 1 \ , \quad \textit{or } Q(B) = P(B)^{-1} \ .$$

Since $P(B)X_t = W_t$

we have

$P(B) = 1 - \lambda B$ *and* $Q(B) = \sum_{k \geq 0} \lambda^k B^k$ *are related by*

$$P(B)Q(B) = 1 \ , \quad \textit{or } Q(B) = P(B)^{-1} \ .$$

Since $P(B)X_t = W_t$

we have $$X_t = P(B)^{-1} W_t$$

$P(B) = 1 - \lambda B$ *and* $Q(B) = \sum_{k \geq 0} \lambda^k B^k$ *are related by*

$$P(B)Q(B) = 1 \ , \quad \textit{or } Q(B) = P(B)^{-1} \ .$$

Since $P(B)X_t = W_t$

we have
$$\begin{aligned} X_t &= P(B)^{-1} W_t \\ &= Q(B) W_t \end{aligned}$$

$P(B) = 1 - \lambda B$ *and* $Q(B) = \sum_{k \geq 0} \lambda^k B^k$ *are related by*

$$P(B)Q(B) = 1 \ , \quad \text{or } Q(B) = P(B)^{-1} \ .$$

Since $P(B)X_t = W_t$

we have
$$\begin{aligned} X_t &= P(B)^{-1} W_t \\ &= Q(B) W_t \end{aligned}$$

This is weird…
WHY BOTHER?

$P(B) = 1 - \lambda B$ *and* $Q(B) = \sum_{k \geq 0} \lambda^k B^k$ *are related by*

$$P(B)Q(B) = 1 \ , \quad \textit{or } Q(B) = P(B)^{-1} \ .$$

Since $P(B)X_t = W_t$

we have
$$\begin{aligned} X_t &= P(B)^{-1} W_t \\ &= Q(B) W_t \end{aligned}$$

This is weird…
WHY BOTHER?

Operators **P** and **Q** behave like normal polynomials

$P(B) = 1 - \lambda B$ *and* $Q(B) = \sum_{k \geq 0} \lambda^k B^k$ *are related by*

$$P(B)Q(B) = 1 \ , \quad \text{or } Q(B) = P(B)^{-1} \ .$$

Since $P(B)X_t = W_t$

we have
$$\begin{aligned} X_t &= P(B)^{-1} W_t \\ &= Q(B) W_t \end{aligned}$$

This is weird…
WHY BOTHER?

Operators **P** and **Q** behave like normal polynomials

$$\frac{1}{1 - \lambda z} = \sum_{k \geq 0} \lambda^k z^k \ , \ |\lambda| < 1, |z| \leq 1$$

$$X_t = \lambda X_{t-1} + W_t$$

What happens when $|\lambda| > 1$?

$$X_t = \lambda X_{t-1} + W_t$$

What happens when |λ| > 1 ?

$Q(B)W_t = \sum_{k \geq 0} \lambda^k B^k W_t$ does not converge

$$X_t = \lambda X_{t-1} + W_t$$

What happens when $|\lambda| > 1$?

$Q(B)W_t = \sum_{k\geq 0} \lambda^k B^k W_t$ does not converge

But we can rewrite everything

(essentially flipping time axis)

$$X_t = \lambda X_{t-1} + W_t$$

What happens when $|\lambda| > 1$?

$Q(B)W_t = \sum_{k \geq 0} \lambda^k B^k W_t$ does not converge

But we can rewrite everything

(essentially flipping time axis)

$$\frac{1}{\lambda} X_t = \frac{\lambda}{\lambda} X_{t-1} + \frac{1}{\lambda} W_t$$

$$X_t = \lambda X_{t-1} + W_t$$

What happens when $|\lambda| > 1$?

$Q(B)W_t = \sum_{k\geq 0} \lambda^k B^k W_t$ does not converge

But we can rewrite everything

(essentially flipping time axis)

$$\frac{1}{\lambda}X_t = \frac{\lambda}{\lambda}X_{t-1} + \frac{1}{\lambda}W_t$$

$$X_{t-1} = \lambda^{-1}X_t - \lambda^{-1}W_t$$

$$X_t = \lambda X_{t-1} + W_t$$

What happens when $|\lambda| > 1$?

$Q(B)W_t = \sum_{k \geq 0} \lambda^k B^k W_t$ does not converge

But we can rewrite everything

(essentially flipping time axis)

$$\frac{1}{\lambda} X_t = \frac{\lambda}{\lambda} X_{t-1} + \frac{1}{\lambda} W_t$$

$$X_{t-1} = \lambda^{-1} X_t - \lambda^{-1} W_t$$

Anti-causal : future determines the past

$$X_t = \lambda X_{t-1} + W_t$$

What happens when |λ| > 1 ?

$Q(B)W_t = \sum_{k\geq 0} \lambda^k B^k W_t$ does not converge

But we can rewrite everything (essentially flipping time axis)

$$\frac{1}{\lambda}X_t = \frac{\lambda}{\lambda}X_{t-1} + \frac{1}{\lambda}W_t$$

$$X_{t-1} = \lambda^{-1}X_t - \lambda^{-1}W_t$$

Anti-causal : future determines the past

$$X_t = -\sum_{k=1}^{\infty} \lambda^{-k} W_{t+k}$$

Revisiting MA(1)

$$X_t = W_t + \theta W_{t-1} = (1 + \theta B)W_t = P(B)W_t$$

$|\theta| < 1,$

Revisiting MA(1)

$$X_t = W_t + \theta W_{t-1} = (1 + \theta B) W_t = P(B) W_t$$

$|\theta| < 1$,

$$\begin{aligned} P(B)^{-1} X_t &= W_t \\ \frac{1}{1+\theta B} X_t &= W_t \\ \left(1 - \theta B + \theta^2 B^2 - \theta^3 B^3 + \dots\right) X_t &= W_t \\ \sum_{k \geq 0} (-\theta)^k X_{t-k} &= W_t \ , \end{aligned}$$

Revisiting MA(1)

$$X_t = W_t + \theta W_{t-1} = (1 + \theta B) W_t = P(B) W_t$$

$|\theta| < 1,$

$$\begin{aligned} P(B)^{-1} X_t &= W_t \\ \frac{1}{1 + \theta B} X_t &= W_t \\ \left(1 - \theta B + \theta^2 B^2 - \theta^3 B^3 + \dots\right) X_t &= W_t \\ \sum_{k \geq 0} (-\theta)^k X_{t-k} &= W_t \ , \end{aligned}$$

essentially, we have inverted the roles of X and W

DEF: Invertible Process

A linear process $\{X_t\}$ is **invertible** if there exist $\psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \dots$ with $\sum_k |\psi_k| < \infty$ and

$$\psi(B) X_t = W_t \; .$$

DEF: Invertible Process

A linear process $\{X_t\}$ is **invertible** if there exist $\psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \ldots$ with $\sum_k |\psi_k| < \infty$ and

$$\psi(B) X_t = W_t \ .$$

AR(1)

$$X_t - \lambda X_{t-1} = (1 - \lambda B) X_t = W_t$$

Causal (wrt $\{W_t\}$) iff $|\lambda| < 1$.

Always invertible (wrt $\{W_t\}$).

DEF: Invertible Process

A linear process $\{X_t\}$ is **invertible** if there exist $\psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \dots$ with $\sum_k |\psi_k| < \infty$ and

$$\psi(B) X_t = W_t \ .$$

AR(1)

$$X_t - \lambda X_{t-1} = (1 - \lambda B) X_t = W_t$$

Causal (wrt $\{W_t\}$) iff $|\lambda| < 1$.

Always invertible (wrt $\{W_t\}$).

MA(1)

$$X_t = W_t + \theta W_{t-1} = (1 + \theta B) W_t$$

Always causal (wrt $\{W_t\}$).

Invertible (wrt $\{W_t\}$) iff $|\theta| < 1$.

Increasing complexity: AR(p)

An AR(p) process $\{X_t\}$ is a stationary process satisfying

$$X_t - \lambda_1 X_{t-1} - \ldots \lambda_p X_{t-p} = W_t \ ,$$

Where $\{W_t\}$ is white noise

$$\lambda_p \neq 0$$

Increasing complexity: AR(p)

An AR(p) process $\{X_t\}$ is a stationary process satisfying

$$X_t - \lambda_1 X_{t-1} - \ldots \lambda_p X_{t-p} = W_t \ ,$$

Where $\{W_t\}$ is white noise

$$\lambda_p \neq 0$$

$$P(B) = 1 - \lambda_1 B - \lambda_2 B^2 - \ldots \lambda_p B^p$$

Increasing complexity: AR(p)

An AR(p) process $\{X_t\}$ is a stationary process satisfying

$$X_t - \lambda_1 X_{t-1} - \ldots \lambda_p X_{t-p} = W_t \ ,$$

Where $\{W_t\}$ is white noise

$$\lambda_p \neq 0$$

$$P(B) = 1 - \lambda_1 B - \lambda_2 B^2 - \ldots \lambda_p B^p$$

Constraints on polynomial P(B)

$|z_k^*| \neq 1$ for all (complex) roots of P(B)

Increasing complexity: AR(p)

An AR(p) process $\{X_t\}$ is a stationary process satisfying

$$X_t - \lambda_1 X_{t-1} - \ldots \lambda_p X_{t-p} = W_t \ ,$$

Where $\{W_t\}$ is white noise

$$\lambda_p \neq 0$$

$$P(B) = 1 - \lambda_1 B - \lambda_2 B^2 - \ldots \lambda_p B^p$$

Constraints on polynomial P(B)

$|z_k^*| \neq 1$ for all (complex) roots of P(B)

Polynomials refresher

A polynomial of order n has n complex roots

If coeff. are real valued- pairs of conjugate roots

Stationarity and causality

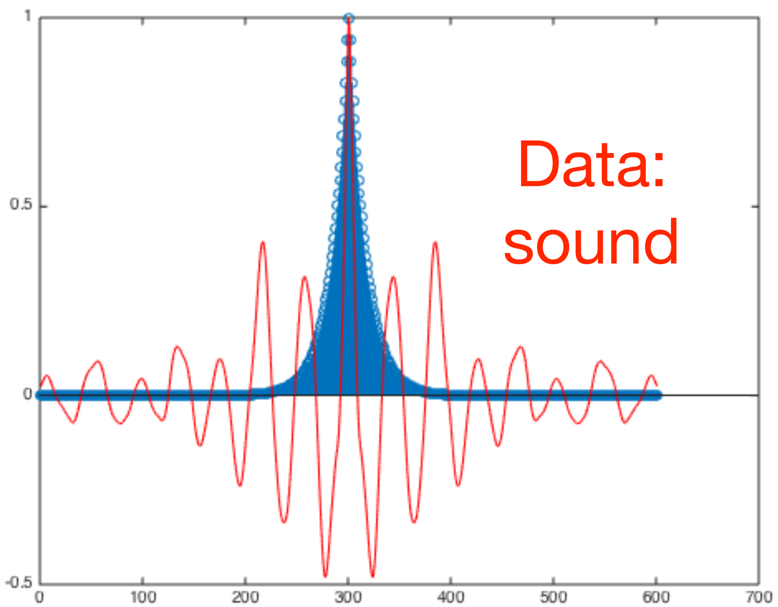
Theorem

1. *The equation* $P(B)X_t = W_t$ *has a unique stationary solution if and only if*
$$P(z) = 0 \Rightarrow |z| \neq 1 \ .$$
We call this unique solution an AR(p) process.
2. *Moreover, this process is causal if and only if*
$$P(z) = 0 \Rightarrow |z| > 1 \ .$$

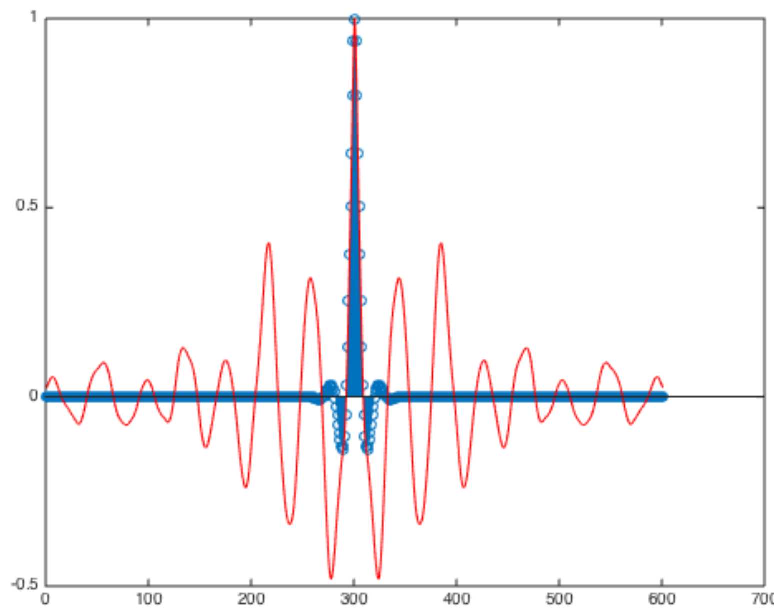
Once we have the polynomial form, recursion can be solved by traditional linear diff. eq. methods.

Increasing complexity: AR(p)

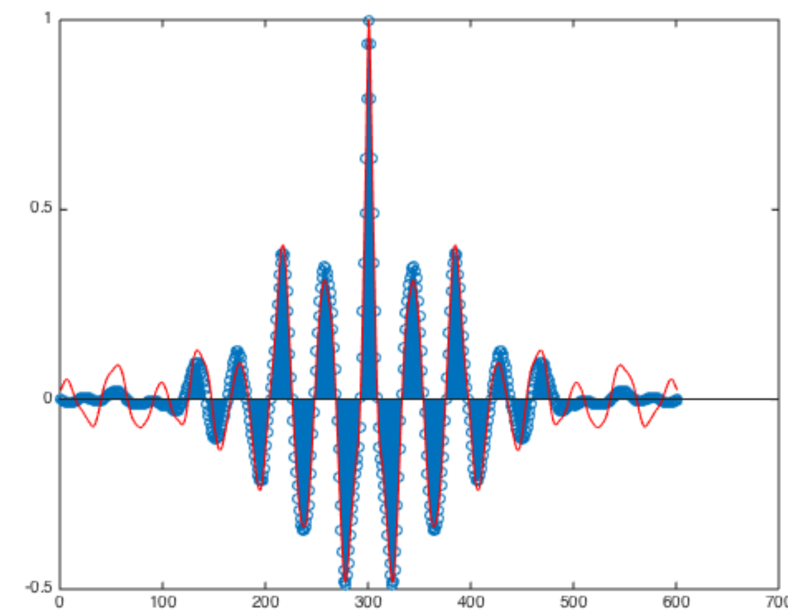
AR(1)



AR(4)



AR(16)



Increased expressive power

Increasing complexity: MA(q)

The moving average model of order q, or MA(q), is defined as

$$X_t = W_t + \theta_1 W_{t-1} + \theta_2 W_{t-2} + \dots \theta_q W_{t-q} \ ,$$

Where $\{W_t\}$ is white noise

$\lambda_q \neq 0$

Increasing complexity: MA(q)

The moving average model of order q, or MA(q), is defined as

$$X_t = W_t + \theta_1 W_{t-1} + \theta_2 W_{t-2} + \dots \theta_q W_{t-q} \ ,$$

Where $\{W_t\}$ is white noise

$\lambda_q \neq 0$

$$X_t = \theta(B) W_t$$

Increasing complexity: MA(q)

The moving average model of order q, or MA(q), is defined as

$$X_t = W_t + \theta_1 W_{t-1} + \theta_2 W_{t-2} + \dots \theta_q W_{t-q} \ ,$$

Where $\{W_t\}$ is white noise

$$\lambda_q \neq 0$$

$$X_t = \theta(B) W_t$$

$$\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$$

Putting it all together: ARMA

An ARMA(p,q) process $\{X_t\}$ is a stationary process that satisfies

$$X_t - \lambda_1 X_{t-1} - \dots \lambda_p X_{t-p} = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q} \ ,$$

Where $\{W_t\}$ is white noise

$$\lambda_p \neq 0$$

$$\lambda_q \neq 0$$

*no **common** roots

Check : causal and invertible

Putting it all together: ARMA

An ARMA(p,q) process $\{X_t\}$ is a stationary process that satisfies

$$X_t - \lambda_1 X_{t-1} - \dots \lambda_p X_{t-p} = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q} \ ,$$

Where $\{W_t\}$ is white noise

$$\lambda_p \neq 0$$

$$\lambda_q \neq 0$$

*no **common** roots

Example 3.8

$$x_t = .4x_{t-1} + .45x_{t-2} + w_t + w_{t-1} + .25w_{t-2},$$

Check : causal and invertible

Putting it all together: ARMA

An ARMA(p,q) process $\{X_t\}$ is a stationary process that satisfies

$$X_t - \lambda_1 X_{t-1} - \ldots \lambda_p X_{t-p} = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q} \ ,$$

Where $\{W_t\}$ is white noise

$$\lambda_p \neq 0$$

$$\lambda_q \neq 0$$

*no **common** roots

Example 3.8

$$x_t = .4x_{t-1} + .45x_{t-2} + w_t + w_{t-1} + .25w_{t-2},$$

$$(1 - .4B - .45B^2)x_t = (1 + B + .25B^2)w_t$$

Check : causal and invertible

Putting it all together: ARMA

An ARMA(p,q) process $\{X_t\}$ is a stationary process that satisfies

$$X_t - \lambda_1 X_{t-1} - \dots \lambda_p X_{t-p} = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q} \ ,$$

Where $\{W_t\}$ is white noise

$\lambda_p \neq 0$

$\lambda_q \neq 0$

*no **common** roots

Example 3.8

$$x_t = .4x_{t-1} + .45x_{t-2} + w_t + w_{t-1} + .25w_{t-2},$$

$$(1 - .4B - .45B^2)x_t = (1 + B + .25B^2)w_t$$

$$\theta(B) = (1 + B + .25B^2) = (1 + .5B)^2$$

Check : causal and invertible

Putting it all together: ARMA

An ARMA(p,q) process $\{X_t\}$ is a stationary process that satisfies

$$X_t - \lambda_1 X_{t-1} - \ldots \lambda_p X_{t-p} = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q} \ ,$$

Where $\{W_t\}$ is white noise

$$\lambda_p \neq 0$$

$$\lambda_q \neq 0$$

*no **common** roots

Example 3.8

$$x_t = .4x_{t-1} + .45x_{t-2} + w_t + w_{t-1} + .25w_{t-2},$$

$$(1 - .4B - .45B^2)x_t = (1 + B + .25B^2)w_t$$

$$\theta(B) = (1 + B + .25B^2) = (1 + .5B)^2$$

$$\phi(B) = 1 - .4B - .45B^2 = (1 + .5B)(1 - .9B)$$

Check : causal and invertible

Putting it all together: ARMA

An ARMA(p,q) process $\{X_t\}$ is a stationary process that satisfies

$$X_t - \lambda_1 X_{t-1} - \ldots \lambda_p X_{t-p} = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q} \ ,$$

Where $\{W_t\}$ is white noise

$$\lambda_p \neq 0$$

$$\lambda_q \neq 0$$

*no **common** roots

Example 3.8

$$x_t = .4x_{t-1} + .45x_{t-2} + w_t + w_{t-1} + .25w_{t-2},$$

$$(1 - .4B - .45B^2)x_t = (1 + B + .25B^2)w_t$$

$$\theta(B) = (1 + B + .25B^2) = (1 + .5B)^2$$

$$\phi(B) = 1 - .4B - .45B^2 = (1 + .5B)(1 - .9B)$$

Simplified is actually ARMA(1,1)

Check : causal and invertible

Putting it all together: ARMA

An ARMA(p,q) process $\{X_t\}$ is a stationary process that satisfies

$$X_t - \lambda_1 X_{t-1} - \ldots \lambda_p X_{t-p} = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q} \ ,$$

Where $\{W_t\}$ is white noise

$$\lambda_p \neq 0$$
$$\lambda_q \neq 0$$

*no **common** roots

Example 3.8

$$x_t = .4x_{t-1} + .45x_{t-2} + w_t + w_{t-1} + .25w_{t-2},$$

$$(1 - .4B - .45B^2)x_t = (1 + B + .25B^2)w_t$$

$$\theta(B) = (1 + B + .25B^2) = (1 + .5B)^2$$

$$\phi(B) = 1 - .4B - .45B^2 = (1 + .5B)(1 - .9B)$$

Simplified is actually ARMA(1,1)

$$x_t = .9x_{t-1} + .5w_{t-1} + w_t$$

Check : causal and invertible

Putting it all together: ARMA

An ARMA(p,q) process $\{X_t\}$ is a stationary process that satisfies

$$X_t - \lambda_1 X_{t-1} - \ldots \lambda_p X_{t-p} = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q} \ ,$$

Where $\{W_t\}$ is white noise

$$\lambda_p \neq 0$$
$$\lambda_q \neq 0$$

*no **common** roots

Example 3.8

Putting it all together: ARMA

An ARMA(p,q) process $\{X_t\}$ is a stationary process that satisfies

$$X_t - \lambda_1 X_{t-1} - \dots \lambda_p X_{t-p} = W_t + \theta_1 W_{t-1} + \dots + \theta_q W_{t-q} \ ,$$

Where $\{W_t\}$ is white noise

$\lambda_p \neq 0$

$\lambda_q \neq 0$

*no **common** roots

Example 3.8

Special cases

AR(p) = ARMA(p, 0), ie $\theta(B) = 1$.

MA(q) = ARMA(0,q), ie $P(B) = 1$.

Putting it all together: ARMA

An ARMA(p,q) process $\{X_t\}$ is a stationary process that satisfies

$$X_t - \lambda_1 X_{t-1} - \dots \lambda_p X_{t-p} = W_t + \theta_1 W_{t-1} + \dots + \theta_q W_{t-q} \ ,$$

Where $\{W_t\}$ is white noise

$\lambda_p \neq 0$

$\lambda_q \neq 0$

*no **common** roots

Example 3.8

Special cases

AR(p) $=$ ARMA(p, 0), ie $\theta(B) = 1$.

MA(q) $=$ ARMA(0,q), ie $P(B) = 1$.

Has p+q parameters

Putting it all together: ARMA

An ARMA(p,q) process $\{X_t\}$ is a stationary process that satisfies

$$X_t - \lambda_1 X_{t-1} - \dots \lambda_p X_{t-p} = W_t + \theta_1 W_{t-1} + \dots + \theta_q W_{t-q} \ ,$$

Where $\{W_t\}$ is white noise

$$\lambda_p \neq 0$$
$$\lambda_q \neq 0$$

*no **common** roots

Example 3.8

Special cases

AR(p) $=$ ARMA(p, 0), ie $\theta(B) = 1$.

MA(q) $=$ ARMA(0,q), ie $P(B) = 1$.

Has p+q parameters

For any stationary process with autocovariance R and any $k > 0$, there is an ARMA process $\{X_t\}$ such that

$$R_X(h) = R(h) \ , h \leq k \ .$$

The wonderful world of ARMA polynomials

$$P(B)X_t = \theta(B)W_t$$

Where P(B) has degree p and
Q(B) has degree q

The wonderful world of ARMA polynomials

$$P(B)X_t = \theta(B)W_t$$

Where P(B) has degree p and Q(B) has degree q

We can think an ARMA model as concatenating two models:

$$Y_t = \theta(B)W_t \ , \text{ and } \ P(B)X_t = Y_t \ .$$

The wonderful world of ARMA polynomials

$$P(B)X_t = \theta(B)W_t$$

Where P(B) has degree p and
Q(B) has degree q

We can think an ARMA model as concatenating two models:

$$Y_t = \theta(B)W_t \ , \ \text{and} \ \ P(B)X_t = Y_t \ .$$

Theorem

- *If P and θ have no common factors, a stationary solution to $P(B)X_t = \theta(B)W_t$ exists iff the roots of $P(z)$ avoid the unit circle: $P(z) = 0 \Rightarrow |z| \neq 1$. This is called an ARMA(p,q) process.*
- *This process is* **causal** *iff the roots of $P(z)$ are* **outside** *the unit circle: $P(z) = 0 \Rightarrow |z| > 1$.*
- *This process is* **invertible** *iff the roots of $\theta(B)$ are* **outside** *the unit circle: $\theta(z) = 0 \Rightarrow |z| > 1$.*

Autocovariance of ARMA processes

$$X_t - \lambda_1 X_{t-1} - \ldots \lambda_p X_{t-p} = \theta_0 W_t + \cdots + \theta_q W_{t-q}$$

Autocovariance of ARMA processes

$$X_t - \lambda_1 X_{t-1} - \ldots \lambda_p X_{t-p} = \theta_0 W_t + \cdots + \theta_q W_{t-q}$$

Left side: $\mathrm{cov}(X_t - \lambda_1 X_{t-1} - \cdots - \lambda_p X_{t-p}, X_{t-h}) =$

Autocovariance of ARMA processes

$$X_t - \lambda_1 X_{t-1} - \ldots \lambda_p X_{t-p} = \theta_0 W_t + \cdots + \theta_q W_{t-q}$$

Left side: $$\mathrm{cov}(X_t - \lambda_1 X_{t-1} - \cdots - \lambda_p X_{t-p}, X_{t-h}) =$$

$$\mathrm{cov}(X_t, X_{t-h}) - \lambda_1 \mathrm{cov}(X_{t-1}, X_{t-h}) - \cdots - \lambda_p \mathrm{cov}(X_{t-p}, X_{t-h})$$

Autocovariance of ARMA processes

$$X_t - \lambda_1 X_{t-1} - \dots \lambda_p X_{t-p} = \theta_0 W_t + \cdots + \theta_q W_{t-q}$$

Left side: $$\text{cov}(X_t - \lambda_1 X_{t-1} - \cdots - \lambda_p X_{t-p}, X_{t-h}) =$$

$$\text{cov}(X_t, X_{t-h}) - \lambda_1 \text{cov}(X_{t-1}, X_{t-h}) - \cdots - \lambda_p \text{cov}(X_{t-p}, X_{t-h})$$

Right side: $$\theta_0 \text{cov}(W_t, X_{t-h}) + \cdots + \theta_q \text{cov}(W_{t-q}, X_{t-h})$$

Autocovariance of ARMA processes

$$X_t - \lambda_1 X_{t-1} - \ldots \lambda_p X_{t-p} = \theta_0 W_t + \cdots + \theta_q W_{t-q}$$

Left side: $$\mathrm{cov}(X_t - \lambda_1 X_{t-1} - \cdots - \lambda_p X_{t-p}, X_{t-h}) =$$

$$\mathrm{cov}(X_t, X_{t-h}) - \lambda_1 \mathrm{cov}(X_{t-1}, X_{t-h}) - \cdots - \lambda_p \mathrm{cov}(X_{t-p}, X_{t-h})$$

Right side: $$\theta_0 \mathrm{cov}(W_t, X_{t-h}) + \cdots + \theta_q \mathrm{cov}(W_{t-q}, X_{t-h})$$

The auto-covariance satisfies a homogeneous recurrence

Autocovariance of ARMA processes

$$X_t - \lambda_1 X_{t-1} - \dots \lambda_p X_{t-p} = \theta_0 W_t + \cdots + \theta_q W_{t-q}$$

Left side: $$\mathrm{cov}(X_t - \lambda_1 X_{t-1} - \cdots - \lambda_p X_{t-p}, X_{t-h}) =$$

$$\mathrm{cov}(X_t, X_{t-h}) - \lambda_1 \mathrm{cov}(X_{t-1}, X_{t-h}) - \cdots - \lambda_p \mathrm{cov}(X_{t-p}, X_{t-h})$$

Right side: $$\theta_0 \mathrm{cov}(W_t, X_{t-h}) + \cdots + \theta_q \mathrm{cov}(W_{t-q}, X_{t-h})$$

The auto-covariance satisfies a homogeneous recurrence

$$R_X(h) - \lambda_1 R_X(h-1) - \cdots - \lambda_p R_X(h-p) = \sigma^2 \sum_{k=0}^{q-h} \theta_{k+h} \psi_k \ .$$

Autocovariance of ARMA processes

$$X_t - \lambda_1 X_{t-1} - \ldots \lambda_p X_{t-p} = \theta_0 W_t + \cdots + \theta_q W_{t-q}$$

Left side: $$\mathrm{cov}(X_t - \lambda_1 X_{t-1} - \cdots - \lambda_p X_{t-p}, X_{t-h}) =$$

$$\mathrm{cov}(X_t, X_{t-h}) - \lambda_1 \mathrm{cov}(X_{t-1}, X_{t-h}) - \cdots - \lambda_p \mathrm{cov}(X_{t-p}, X_{t-h})$$

Right side: $$\theta_0 \mathrm{cov}(W_t, X_{t-h}) + \cdots + \theta_q \mathrm{cov}(W_{t-q}, X_{t-h})$$

The auto-covariance satisfies a homogeneous recurrence

$$R_X(h) - \lambda_1 R_X(h-1) - \cdots - \lambda_p R_X(h-p) = \sigma^2 \sum_{k=0}^{q-h} \theta_{k+h} \psi_k \ .$$

Polynomial form defines the moments of the process!

So the autocorrelation $R_X(h)$ also satisfies an homogeneous recurrence:

$$R_X(h) - \lambda_1 R_X(h-1) - \cdots - \lambda_p R_X(h-p) = 0 \ , \ \text{for } h > q \ ,$$

with initial conditions given by

$$R_X(h) - \lambda_1 R_X(h-1) - \cdots - \lambda_p R_X(h-p) = \sigma^2 \sum_{k=0}^{q-h} \theta_{k+h} \psi_k \ , \ (h = 0 \ldots, q-1$$

What do we do with this?

So the autocorrelation $R_X(h)$ also satisfies an homogeneous recurrence:

$$R_X(h) - \lambda_1 R_X(h-1) - \cdots - \lambda_p R_X(h-p) = 0 \ , \ \text{for } h > q \ ,$$

with initial conditions given by

$$R_X(h) - \lambda_1 R_X(h-1) - \cdots - \lambda_p R_X(h-p) = \sigma^2 \sum_{k=0}^{q-h} \theta_{k+h}\psi_k \ , \ (h = 0 \ldots, q-1$$

What do we do with this?

Linear homogeneous eq of order p

$$a_0 X_t + a_1 X_{t-1} + \cdots + a_p X_{t-p} = 0 \ .$$

$$(a_0 + a_1 B + \cdots + a_p B^p) X_t = 0 \ .$$

$$a(B) X_t = 0 \ , \ \text{with } a(z) = a_0 + a_1 z + \cdots + a_p z^p$$

So the autocorrelation $R_X(h)$ also satisfies an homogeneous recurrence:

$$R_X(h) - \lambda_1 R_X(h-1) - \cdots - \lambda_p R_X(h-p) = 0 \ , \ \text{for } h > q \ ,$$

with initial conditions given by

$$R_X(h) - \lambda_1 R_X(h-1) - \cdots - \lambda_p R_X(h-p) = \sigma^2 \sum_{k=0}^{q-h} \theta_{k+h} \psi_k \ , \ (h = 0 \ldots, q-1$$

What do we do with this?

Linear homogeneous eq of order p

$$a_0 X_t + a_1 X_{t-1} + \cdots + a_p X_{t-p} = 0 \ .$$

$$(a_0 + a_1 B + \cdots + a_p B^p) X_t = 0 \ .$$

$$a(B) X_t = 0 \ , \ \text{with } a(z) = a_0 + a_1 z + \cdots + a_p z^p$$

with characteristic polynomial $a(z) = a_p(z - z_1)(z - z_2) \ldots (z - z_p)$

Big picture

Big picture

*Note: proof tsa4.pdf, page 91

Big picture

Goal: solve $a_0 X_t + \cdots + a_p X_{t-p} = 0$

With initial conditions $X_1, \ldots, X_p$.

*Note: proof tsa4.pdf, page 91

Big picture

Goal: solve $a_0 X_t + \cdots + a_p X_{t-p} = 0$

With initial conditions $X_1, \ldots, X_p$.

Equivalent to finding the roots of polynomial

$$(z - z_1)^{m_1} \ldots (z - z_k)^{m_k} = 0$$

*Note: proof tsa4.pdf, page 91

Big picture

Goal: solve $a_0 X_t + \cdots + a_p X_{t-p} = 0$

With initial conditions $X_1, \ldots, X_p$.

Equivalent to finding the roots of polynomial

$$(z - z_1)^{m_1} \ldots (z - z_k)^{m_k} = 0$$

General solution of the form*

$$X_t = c_1(t) z_1^{-t} + \cdots + c_k(t) z_k^{-t}$$

Where $c_i(t)$ polynomials of order m_i

*Note: proof tsa4.pdf, page 91

Big picture

Goal: solve $a_0 X_t + \cdots + a_p X_{t-p} = 0$

With initial conditions $X_1, \ldots, X_p$.

Equivalent to finding the roots of polynomial

$$(z - z_1)^{m_1} \ldots (z - z_k)^{m_k} = 0$$

General solution of the form* $X_t = c_1(t) z_1^{-t} + \cdots + c_k(t) z_k^{-t}$

Where $c_i(t)$ polynomials of order m_i

Coefficients adjusted from initial conditions

*Note: proof tsa4.pdf, page 91

Estimating ARMA parameters from data

2 methods: **Maximum likelihood**

Method of moments

Estimating ARMA parameters from data

2 methods: **Maximum likelihood**

Method of moments

Assume: model known, and zero mean (preprocessing)

Estimating ARMA parameters from data

2 methods: **Maximum likelihood**

Method of moments

Assume: model known, and zero mean (preprocessing)

Gaussian process $P(B)X_t = \theta(B)W_t$ Where $\{W_t\}$ is iid gaussian noise

0 mean, σ^2 variance

Estimating ARMA parameters from data

2 methods: **Maximum likelihood**

Method of moments

Assume: model known, and zero mean (preprocessing)

Gaussian process $P(B)X_t = \theta(B)W_t$ Where $\{W_t\}$ is iid gaussian noise

0 mean, σ^2 variance

Find parameters $\lambda_i, \; \Theta_j \; \sigma^2$

that maximize likelihood

$$\mathcal{L}(\lambda, \theta, \sigma^2) = f_{\lambda,\theta,\sigma^2}(x_1, \ldots, x_n)$$

Estimating ARMA parameters from data

2 methods: **Maximum likelihood**

Method of moments

Assume: model known, and zero mean (preprocessing)

Gaussian process $P(B)X_t = \theta(B)W_t$ Where $\{W_t\}$ is iid gaussian noise

0 mean, σ^2 variance

Find parameters $\lambda_i, \ \Theta_j \ \sigma^2$

that maximize likelihood

$$\mathcal{L}(\lambda, \theta, \sigma^2) = f_{\lambda,\theta,\sigma^2}(x_1, \ldots, x_n)$$

Jointly gaussian

$$\mathcal{L}(\lambda, \theta, \sigma^2) = (2\pi)^{-n/2} |\Gamma_n|^{-1/2} \exp\left(-\frac{1}{2}\mathbf{x}^T \Gamma_n^{-1} \mathbf{x}\right)$$

Where we've collated the data in vector $\mathbf{x} = (x_1, \ldots, x_n)$

Maximum likelihood

we can show

Maximum likelihood

Pros: efficient, works well even if model assumptions not 100% right

we can show

Maximum likelihood

Pros: efficient, works well even if model assumptions not 100% right

Cons: unpleasant optimisation, need good initialization

we can show

Maximum likelihood

Pros: efficient, works well even if model assumptions not 100% right

Cons: unpleasant optimisation, need good initialization

We need to find a cheap initial guess

we can show

Maximum likelihood

Pros: efficient, works well even if model assumptions not 100% right

Cons: unpleasant optimisation, need good initialization

We need to find a cheap initial guess

Yule Walker equations

we can show

Maximum likelihood

Pros: efficient, works well even if model assumptions not 100% right

Cons: unpleasant optimisation, need good initialization

We need to find a cheap initial guess

Yule Walker equations

In the AR(p) case, we can show the forecasting coefficients

$$X_{t+1}^{t} = \phi_{t,1} X_t + \cdots + \phi_{t,t} X_1$$

correspond exactly to the model parameters λ_i, $i = 1, \ldots, p$.

So we can regress X_t onto $X_{t-1}, \ldots, X_{t-p}$ to estimate λ_i.

Maximum likelihood

Pros: efficient, works well even if model assumptions not 100% right

Cons: unpleasant optimisation, need good initialization

We need to find a cheap initial guess

Yule Walker equations

In the AR(p) case, we can show the forecasting coefficients

$$X_{t+1}^{t} = \phi_{t,1} X_t + \cdots + \phi_{t,t} X_1$$

correspond exactly to the model parameters λ_i, $i = 1, \ldots, p$.

So we can regress X_t onto $X_{t-1}, \ldots, X_{t-p}$ to estimate λ_i.

If $\{X_t\}$ is a casual AR(p) model $P(B)X_t = W_t$, it results that

$$\mathbf{E}\left(X_{t-i}\left(X_t - \sum_{j=1}^{p} \lambda_j X_{t-j}\right)\right) = \mathbf{E}\left(X_{t-i} W_t\right) \ , \ (i = 0, \ldots, p) \Leftrightarrow$$

$$\boxed{R_X(0) - \lambda^T R_p = \sigma^2 \ \text{, and } R_p = \Gamma_p \lambda} \ .$$

where $\Gamma_p = \{\gamma(k-j)\}^{p}_{j,k=1}$ is a $p \times p$ matrix, $\phi = (\phi_1, \ldots, \phi_p)'$ is a $p \times 1$ vector, and $\gamma_p = (\gamma(1), \ldots, \gamma(p))'$ is a $p \times 1$ vector.

Method of moments

Identities linking **parameters** to moments (in our case cov),
use empirical estimates,
adjust parameters to match

Yule Walker equations, with empirical estimates

$$\hat{\lambda}^T \hat{R}_p = \hat{R}_X(0) - \hat{\sigma}^2 \text{ , and } \hat{\Gamma}_p \hat{\lambda} = \hat{R}_p$$

Efficient implementation using the **Durbin-Levinson** algorithm
(you'll implement this during the lab)

What do we do about the mean? ARIMA

Integrated models for non stationary data

If linear time dependence $\mu_t = \beta_0 + \beta_1 t$

What do we do about the mean? ARIMA

Integrated models for non stationary data

Trend stationary processes: varying mean + stationary process

If linear time dependence $\mu_t = \beta_0 + \beta_1 t$

What do we do about the mean? ARIMA

Integrated models for non stationary data

Trend stationary processes: varying mean + stationary process

$$x_t = \mu_t + y_t,$$

If linear time dependence $\mu_t = \beta_0 + \beta_1 t$

What do we do about the mean? ARIMA

Integrated models for non stationary data

Trend stationary processes: varying mean + stationary process

$$x_t = \mu_t + y_t,$$

If linear time dependence $\mu_t = \beta_0 + \beta_1 t$

$$\nabla x_t = x_t - x_{t-1} = \beta_1 + y_t - y_{t-1} = \beta_1 + \nabla y_t.$$

What do we do about the mean? ARIMA

Integrated models for non stationary data

Trend stationary processes: varying mean + stationary process

$$x_t = \mu_t + y_t,$$

If linear time dependence $\mu_t = \beta_0 + \beta_1 t$

$$\nabla x_t = x_t - x_{t-1} = \beta_1 + y_t - y_{t-1} = \beta_1 + \nabla y_t.$$

In general, it may take several differentiations to get there
(d-th order polynomial dependence on time)

What do we do about the mean? ARIMA

Integrated models for non stationary data

Trend stationary processes: varying mean + stationary process

$$x_t = \mu_t + y_t,$$

If linear time dependence $\mu_t = \beta_0 + \beta_1 t$

$$\nabla x_t = x_t - x_{t-1} = \beta_1 + y_t - y_{t-1} = \beta_1 + \nabla y_t.$$

In general, it may take several differentiations to get there
(d-th order polynomial dependence on time)

A process is **ARIMA** (p,**d**,q) if **d**-th difference

$\nabla^d x_t = (1 - B)^d x_t$ is ARMA (p,q)

What do we do about the mean? ARIMA

Integrated models for non stationary data

Trend stationary processes: varying mean + stationary process

$$x_t = \mu_t + y_t,$$

If linear time dependence $\mu_t = \beta_0 + \beta_1 t$

$$\nabla x_t = x_t - x_{t-1} = \beta_1 + y_t - y_{t-1} = \beta_1 + \nabla y_t.$$

In general, it may take several differentiations to get there (d-th order polynomial dependence on time)

A process is **ARIMA** (p,**d**,q) if **d**-th difference

$$\nabla^d x_t = (1 - B)^d x_t$$ is ARMA (p,q)

$$\phi(B)(1 - B)^d x_t = \theta(B) w_t$$

What do we do about the mean? ARIMA

Integrated models for non stationary data

Trend stationary processes: varying mean + stationary process

$$x_t = \mu_t + y_t,$$

If linear time dependence $\mu_t = \beta_0 + \beta_1 t$

$$\nabla x_t = x_t - x_{t-1} = \beta_1 + y_t - y_{t-1} = \beta_1 + \nabla y_t.$$

In general, it may take several differentiations to get there
(d-th order polynomial dependence on time)

A process is **ARIMA** (p,**d**,q) if **d**-th difference

$$\nabla^d x_t = (1 - B)^d x_t$$ is ARMA (p,q)

$$\phi(B)(1 - B)^d x_t = \theta(B) w_t$$

These are tricky to use,
motivate state space models

Lab:

1. Prediction
2. Polynomials - AR/AM/ARMA
3. Method of moments

Next week: Kalman filtering

Homework 1: posted on Thu, due Febr.12