DS-GA 3001.001
Probabilistic time series analysis
Lecture 2
AR(I)MA

Instructor: Cristina Savin

Quick recap

stochastic process

$$\{X_1, X_2, \ldots, X_t \ldots\}$$

$$\mathrm{P}(X_1 \leq x_1, \ldots, X_t \leq x_t \ldots)$$

Examples of stochastic process

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

$v_t = \frac{1}{3}\left(w_{t-1} + w_t + w_{t+1}\right)$ filtered white noise **Moving Average**

$x_t = x_{t-1} - 0.9x_{t-2} + w_t$ **Auto-Regressive process**

ARIMA models provide a general treatment for studying such processes and their generalizations

Quick recap

Basic statistical properties

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

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

ACF $$\rho_X(t,u) = \frac{R_X(t,u)}{\sqrt{R_X(t,t), R_X(u,u)}}$$

Causality, stationarity

$$\{X_t, ..., X_{t+K}\}$$
$$\{X_{t+h}, ..., X_{t+h+K}\}$$

Identically distributed subsets
for all t,h,K

jointly gaussian -> strongly stationary, 2 moments, linear prediction

Quick recap

Cross-Covariance

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

Cross-Correlation Function (ACF)

$$\rho_{X,Y}(t,u) = \frac{R_{X,Y}(t,u)}{\sqrt{R_X(t,t), R_Y(u,u)}}$$

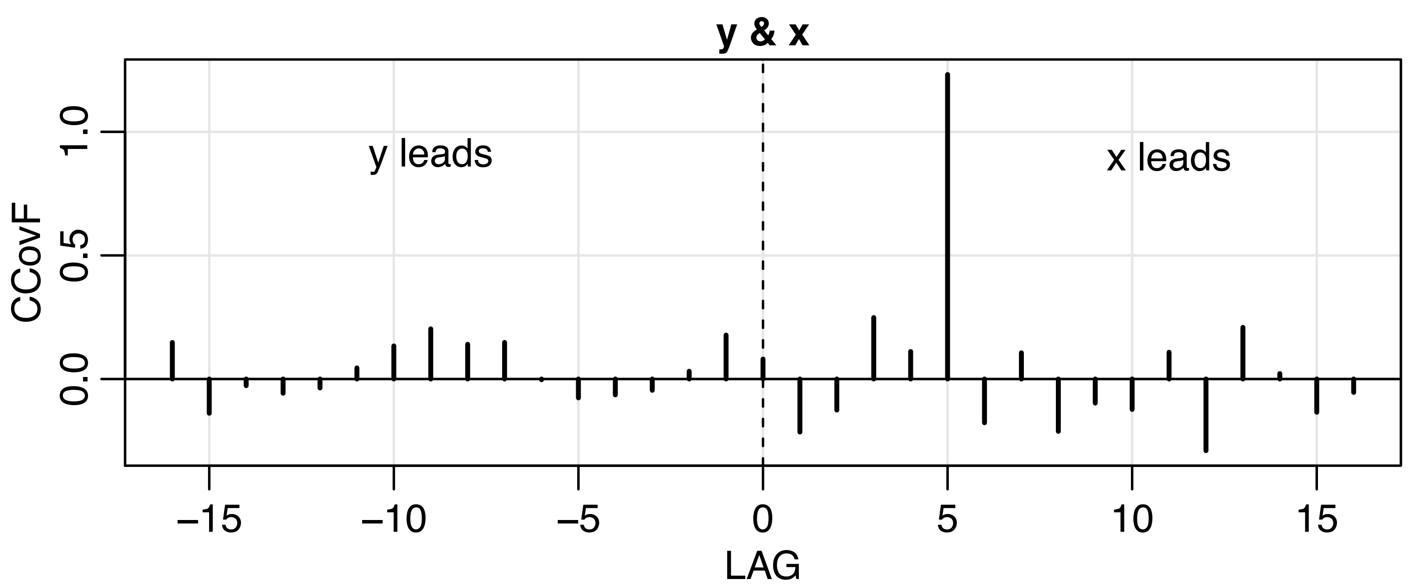
stationarity

E.g. $x_t = w_t + w_{t-1}$ and $y_t = w_t - w_{t-1}$,

$$\rho_{xy}(h) = \begin{cases} 0 & h = 0, \\ 1/2 & h = 1, \\ -1/2 & h = -1, \\ 0 & |h| \geq 2. \end{cases}$$

Lead-lag $y_t = Ax_{t-\ell} + w_t$

$$\begin{aligned}\gamma_{yx}(h) &= \operatorname{cov}(y_{t+h}, x_t)\\ &= \operatorname{cov}(Ax_{t+h-\ell} + w_{t+h}\,, x_t)\\ &= \operatorname{cov}(Ax_{t+h-\ell}\,, x_t)\\ &= A\gamma_x(h-\ell).\end{aligned}$$



Notes on empirical estimation (board)
Tsa4 - pg27

Moving averages, e.g. MA(1)

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

Where $\{W_t\}$ is white noise

Moments:

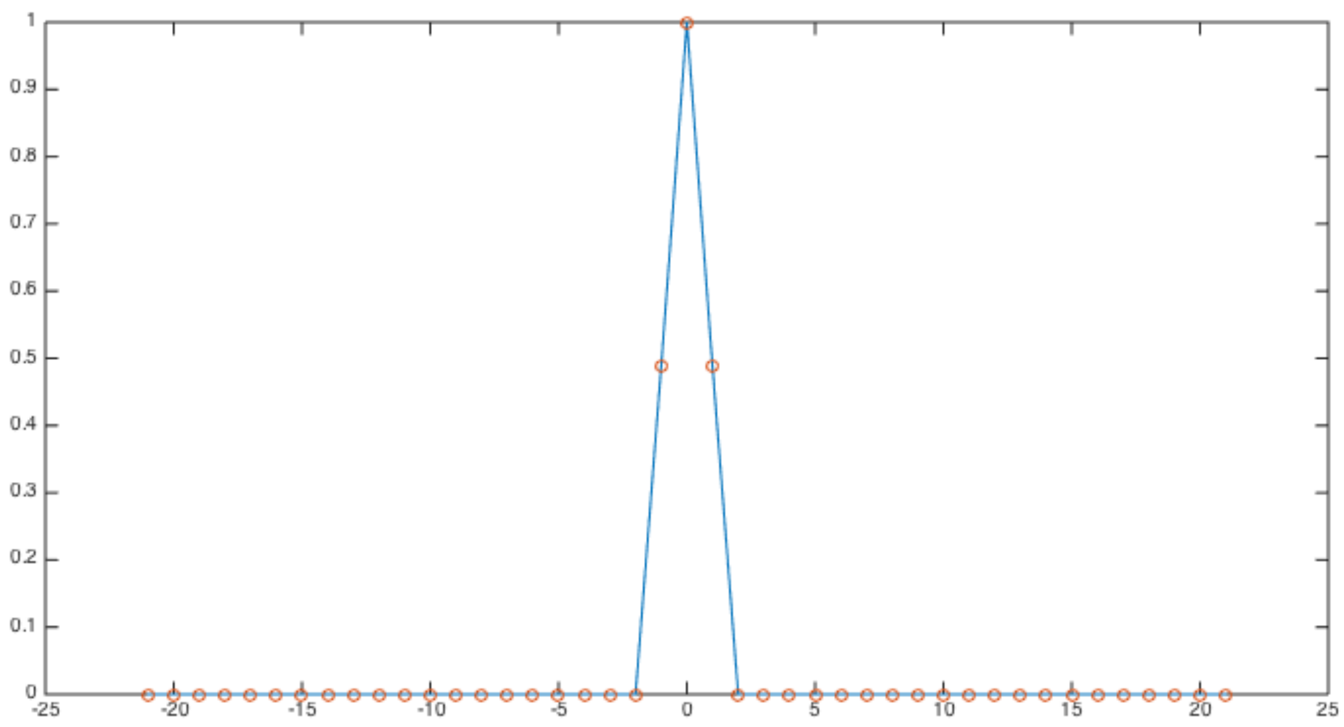
$$\mu_X = 0$$

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

stationary

parameters not unique

MA(1) ACF



Increasing complexity: MA(q)

Definition *The* **moving average model** *of order q, or* **MA**(q) *model, is defined to be*

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q},$$

where $w_t \sim wn(0, \sigma_w^2)$, and $\theta_1, \theta_2, \ldots, \theta_q$ ($\theta_q \neq 0$) are parameters.

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: $X_t = W_t + \lambda W_{t-1} + \lambda^2 W_{t-2} + \cdots$

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

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

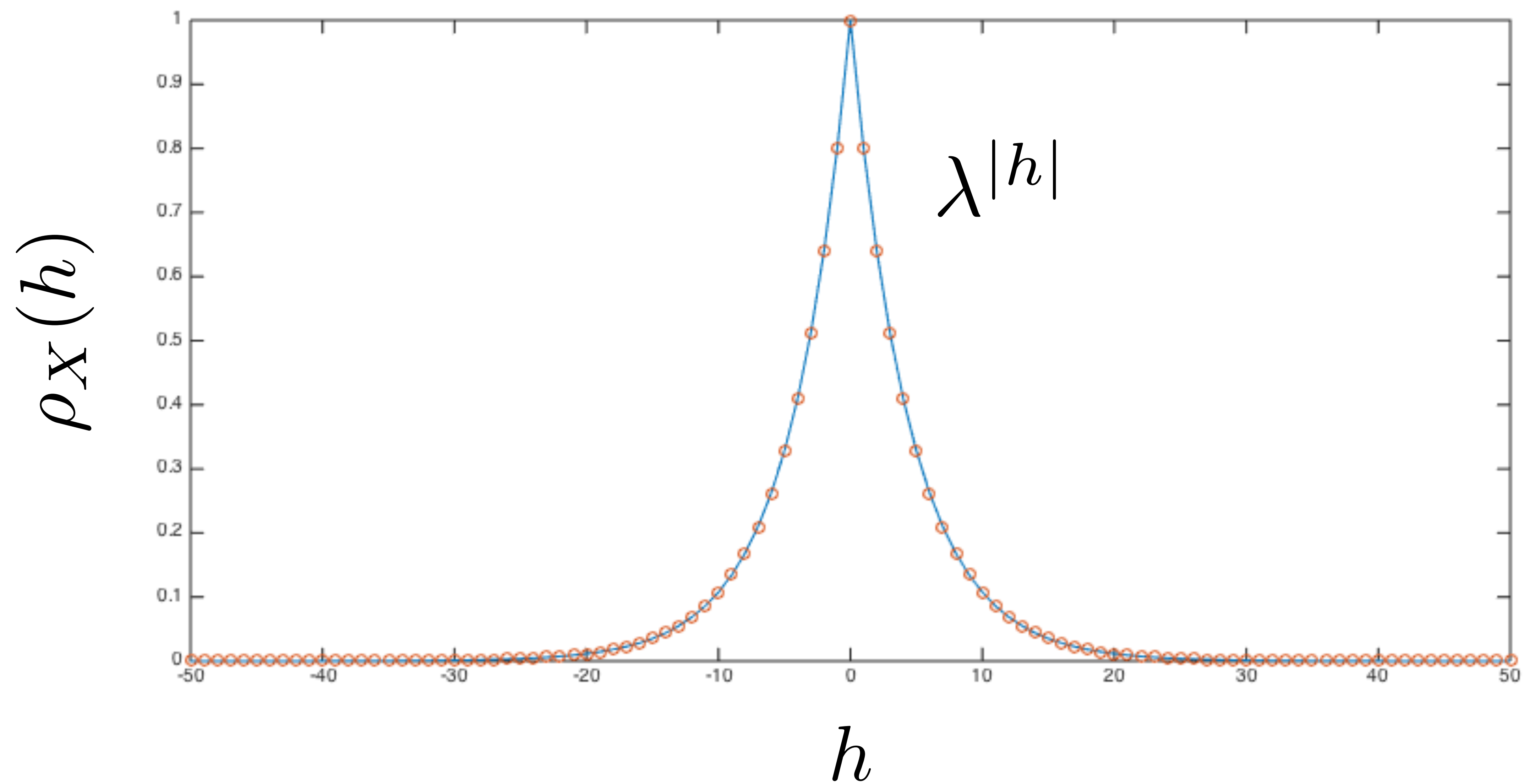
For now, assume $h > 0$

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

stationary

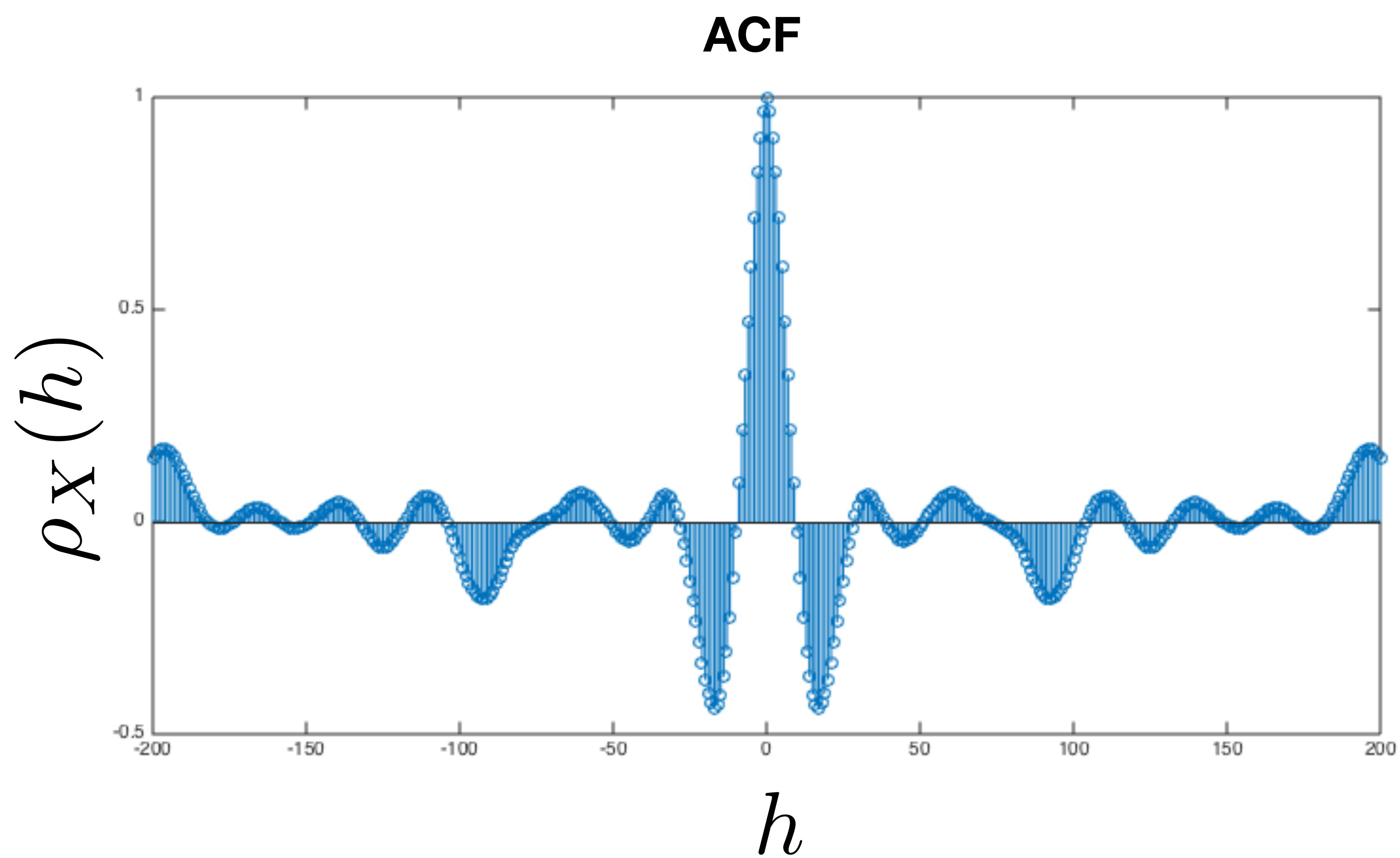
*Check other direction at home

AR(1) ACF



Note: explosive processes

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?

Increasing complexity: AR(p)

Definition *An* **autoregressive model** *of order p, abbreviated* **AR(**p**)**, *is of the form*

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t,$$

where x_t is stationary, $w_t \sim wn(0, \sigma_w^2)$, and $\phi_1, \phi_2, \ldots, \phi_p$ are constants ($\phi_p \neq 0$). The mean of x_t in (3.1) *is zero. If the mean, μ, of x_t is not zero, replace x_t by $x_t - \mu$ in* (3.1),

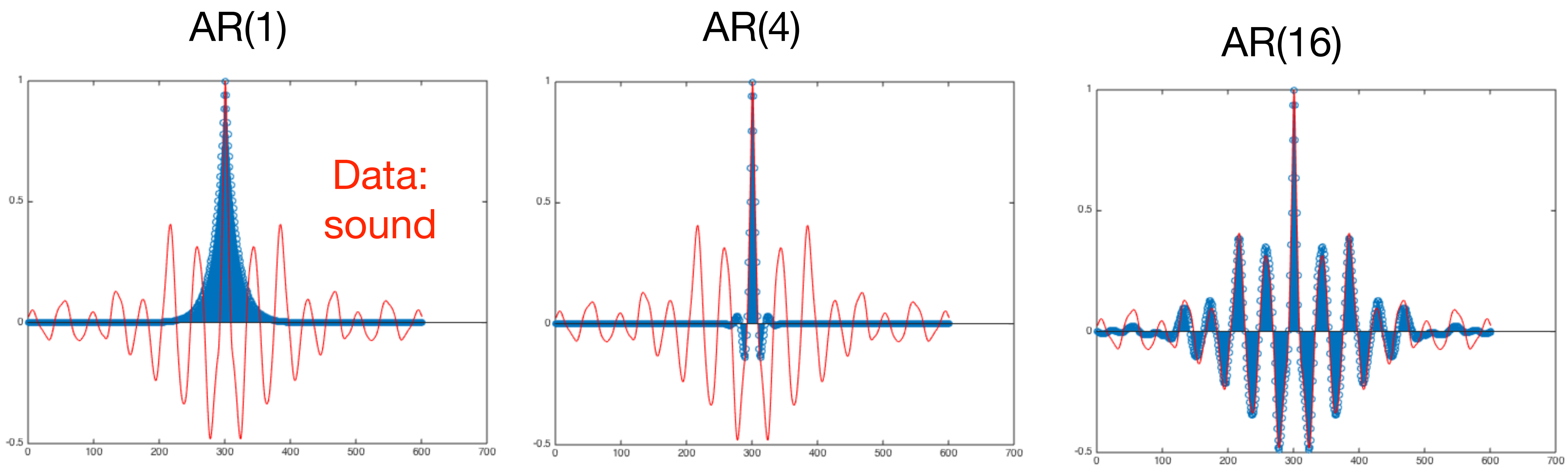
$$x_t - \mu = \phi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) + \cdots + \phi_p(x_{t-p} - \mu) + w_t,$$

or write

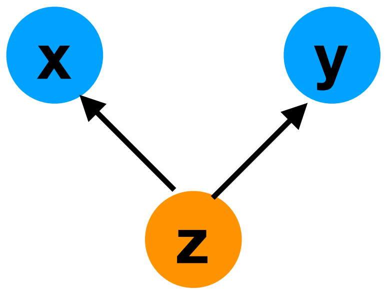
$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t,$$

where $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$.

Increasing complexity: AR(p)



Partial correlations



$$\rho_{XY|Z} = \text{corr}\{X - \hat{X},\, Y - \hat{Y}\}.$$

Definition 3.9 *The* **partial autocorrelation function (PACF)** *of a stationary process, x_t, denoted ϕ_{hh}, for $h = 1, 2, \ldots,$ is*

$$\phi_{11} = \text{corr}(x_{t+1}, x_t) = \rho(1) \tag{3.55}$$

and

$$\phi_{hh} = \text{corr}(x_{t+h} - \hat{x}_{t+h},\; x_t - \hat{x}_t), \quad h \geq 2. \tag{3.56}$$

The reason for using a double subscript will become evident in the next section. The PACF, ϕ_{hh}, is the correlation between x_{t+h} and x_t with the linear dependence of $\{x_{t+1}, \ldots, x_{t+h-1}\}$ on each, removed. If the process x_t is Gaussian, then $\phi_{hh} = \text{corr}(x_{t+h}, x_t \mid x_{t+1}, \ldots, x_{t+h-1})$; that is, ϕ_{hh} is the correlation coefficient between x_{t+h} and x_t in the bivariate distribution of (x_{t+h}, x_t) conditional on $\{x_{t+1}, \ldots, x_{t+h-1}\}$.

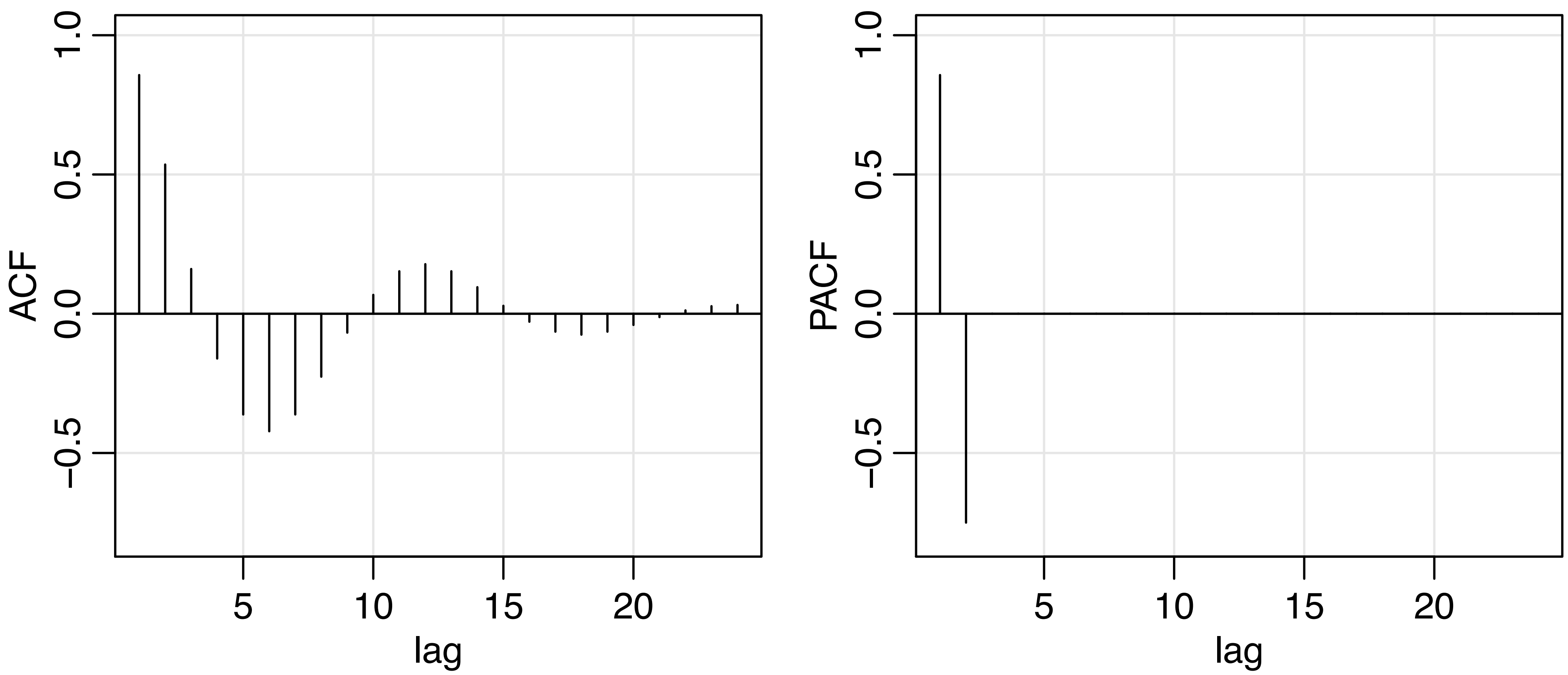


Fig. 3.5. *The ACF and PACF of an AR(2) model with* $\phi_1 = 1.5$ *and* $\phi_2 = -.75$.

AR and **MA** are special instances of **linear processes**

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

$$\mu_X = \mu$$

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

***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?

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

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

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

$$\lambda_p \neq 0$$

$$\lambda_q \neq 0$$

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 of the mean

$$\mu_t = \beta_0 + \beta_1 t$$

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

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

How do we use such models to do prediction?

Gaussian conditioning, reminder

let the vector $\mathbf{z} = [\mathbf{x}^T \mathbf{y}^T]^T$ be normally distributed according to:

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}^T & \mathbf{B} \end{bmatrix}\right) \tag{5a}$$

where $\mathbf{C}$ is the (non-symmetric) cross-covariance matrix between $\mathbf{x}$ and $\mathbf{y}$ which has as many rows as the size of $\mathbf{x}$ and as many columns as the size of $\mathbf{y}$. then the marginal distributions are:

$$\mathbf{x} \sim \mathcal{N}(\mathbf{a}, \mathbf{A}) \tag{5b}$$

$$\mathbf{y} \sim \mathcal{N}(\mathbf{b}, \mathbf{B}) \tag{5c}$$

and the conditional distributions are:

$$\mathbf{x}|\mathbf{y} \sim \mathcal{N}\left(\mathbf{a} + \mathbf{C}\mathbf{B}^{-1}(\mathbf{y} - \mathbf{b}), \mathbf{A} - \mathbf{C}\mathbf{B}^{-1}\mathbf{C}^T\right) \tag{5d}$$

$$\mathbf{y}|\mathbf{x} \sim \mathcal{N}\left(\mathbf{b} + \mathbf{C}^T\mathbf{A}^{-1}(\mathbf{x} - \mathbf{a}), \mathbf{B} - \mathbf{C}^T\mathbf{A}^{-1}\mathbf{C}\right) \tag{5e}$$

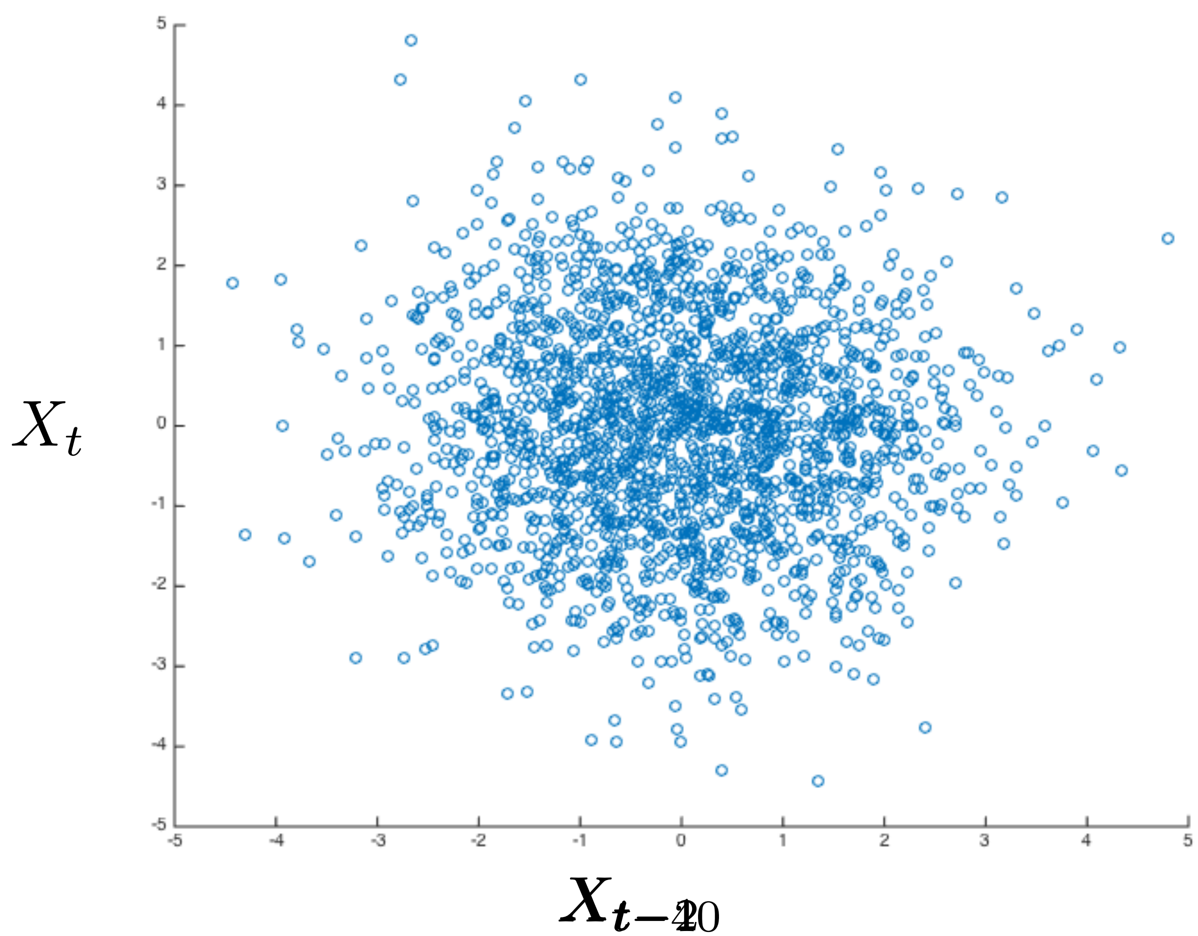
Prediction

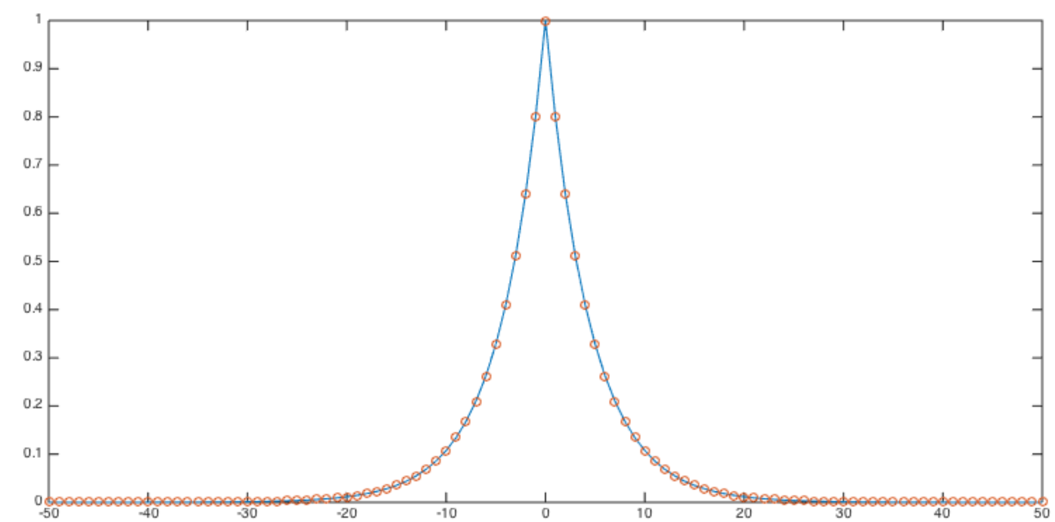
Suppose $\{X_t\}$ is a linear process

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 estimation reminder

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

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

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

$$\mathbf{x}|\mathbf{y} \sim \mathcal{N}\left(\mathbf{a} + \mathbf{C}\mathbf{B}^{-1}(\mathbf{y} - \mathbf{b}), \mathbf{A} - \mathbf{C}\mathbf{B}^{-1}\mathbf{C}^T\right)$$

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

Since everything is Gaussian, 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:

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

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, tsa4 theorem B3)

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

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

Optimal **linear** predictor

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

The optimal predictor
if stationary gaussian

Projection theorem

Best Linear Prediction for Stationary Processes

Given data $x_1, \ldots, x_n$*, the best linear predictor,* $x_{n+m}^n = \alpha_0 + \sum_{k=1}^n \alpha_k x_k$*, of* x_{n+m}*, for* $m \geq 1$*, is found by solving*

$$\mathrm{E}\left[\left(x_{n+m} - x_{n+m}^n\right) x_k\right] = 0, \quad k = 0, 1, \ldots, n,$$

where $x_0 = 1$*, for* $\alpha_0, \alpha_1, \ldots \alpha_n$.

Durbin-Levinson algorithm

It all boils down to computing the ACF
How can we do this in the most general case?

The backshift operator

Definition *We define the* **backshift operator** *by*

$$Bx_t = x_{t-1}$$

and extend it to powers $B^2 x_t = B(Bx_t) = Bx_{t-1} = x_{t-2}$, *and so on. Thus,*

$$B^k x_t = x_{t-k}.$$

Inverse (forward-shift operator)

$$x_t = B^{-1}Bx_t = B^{-1}x_{t-1}$$

Finite differences:

$$\nabla x_t = (1 - B)x_t$$

$$\nabla^d = (1 - B)^d$$

*show on board

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 B powers:

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

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

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

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

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

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

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

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

Revisiting MA(1)

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

$|\theta| < 1,$

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

essentially, we have inverted the roles of X and W

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*
$$P(z) = 0 \Rightarrow |z| > 1 \ .$$

Roots of polynomial determine properties of the stochastic process

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

$$\lambda_p \neq 0$$

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

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

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

$\theta_q \neq 0$

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

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

Putting it all together: ARMA

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

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

Where $\{W_t\}$ is white noise

$\lambda_p \neq 0 \ \ \theta_q \neq 0$

The **autoregressive operator** *is defined to be*

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p.$$

The **moving average operator** *is*

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q.$$

$$\phi(B) x_t = \theta(B) w_t.$$

***NO COMMON ROOTS**

Example: minimal models

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

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

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

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

Simplified, this is actually ARMA(1,1)

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

Putting it all together: ARMA

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

$$\phi(B)x_t = \theta(B)w_t.$$

Where $\{W_t\}$ is white noise

*no **common** roots

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

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: $\theta(z) = 0 \Rightarrow |z| > 1$.*

The ACF of an AR(2) Process

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$$

$$\mathrm{E}(x_t x_{t-h}) = \phi_1 \mathrm{E}(x_{t-1} x_{t-h}) + \phi_2 \mathrm{E}(x_{t-2} x_{t-h}) + \mathrm{E}(w_t x_{t-h}).$$

$$\gamma(h) = \phi_1 \gamma(h-1) + \phi_2 \gamma(h-2), \quad h = 1, 2, \ldots .$$

Where we used $$\mathrm{E}(w_t x_{t-h}) = \mathrm{E}\Big(w_t \sum_{j=0}^{\infty} \psi_j w_{t-h-j}\Big) = 0.$$

This reduces to solving a difference eq

$$\rho(h) - \phi_1 \rho(h-1) - \phi_2 \rho(h-2) = 0, \quad h = 1, 2, \ldots .$$

A short intermezzo: difference equations

$$u_1 = \alpha u_0$$
$$u_2 = \alpha u_1 = \alpha^2 u_0$$
$$\vdots$$
$$u_n = \alpha u_{n-1} = \alpha^n u_0.$$

$$u_n = \alpha^n c = \left(z_0^{-1}\right)^n c.$$

$$(1 - \alpha B) u_n = 0.$$

$$u_n - \alpha_1 u_{n-1} - \alpha_2 u_{n-2} = 0,$$

$$\alpha(z) = 1 - \alpha_1 z - \alpha_2 z^2,$$

2 distinct roots

$$u_n = c_1 z_1^{-n} + c_2 z_2^{-n}$$

Same root

$$u_n = z_0^{-n}(c_1 + c_2 n).$$

In general:

$$u_n - \alpha_1 u_{n-1} - \cdots - \alpha_p u_{n-p} = 0, \quad \alpha_p \neq 0,$$

$$u_n = z_1^{-n} P_1(n) + z_2^{-n} P_2(n) + \cdots + z_r^{-n} P_r(n),$$

where $P_j(n)$, for $j = 1, 2, \ldots, r$, is a polynomial in n, of degree $m_j - 1$. Given p initial conditions $u_0, \ldots, u_{p-1}$, we can solve for the $P_j(n)$ explicitly.

*board, tsa4.pdf page 88

The ACF of an AR(2) Process

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$$

$$\mathrm{E}(x_t x_{t-h}) = \phi_1 \mathrm{E}(x_{t-1} x_{t-h}) + \phi_2 \mathrm{E}(x_{t-2} x_{t-h}) + \mathrm{E}(w_t x_{t-h}).$$

$$\gamma(h) = \phi_1 \gamma(h-1) + \phi_2 \gamma(h-2), \quad h = 1, 2, \ldots .$$

Where we used $$\mathrm{E}(w_t x_{t-h}) = \mathrm{E}\Big(w_t \sum_{j=0}^{\infty} \psi_j w_{t-h-j}\Big) = 0.$$

This reduces to solving a difference eq

$$\rho(h) - \phi_1 \rho(h-1) - \phi_2 \rho(h-2) = 0, \quad h = 1, 2, \ldots .$$

2 roots $$\rho(h) = c_1 z_1^{-h} + c_2 z_2^{-h},$$

Double root $$\rho(h) = z_0^{-h}(c_1 + c_2 h),$$

If complex conjugates

$$\rho(h) = a|z_1|^{-h} \cos(h\theta + b),$$

*check, tsa4.pdf example 3.10

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

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

Polynomial form defines the ACF 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?

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

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 -1

Coefficients adjusted from initial conditions

*Note: proof tsa4.pdf, page 91

How do we estimate the model parameters?

Estimating ARMA parameters from data

2 methods: **Maximum likelihood**

Method of moments

Assume: model known, and zero mean (preprocessing)

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

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

Cons: unpleasant optimization, 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 $\Gamma_p = \{\gamma(k-j)\}_{j,k=1}^{p}$ 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

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)

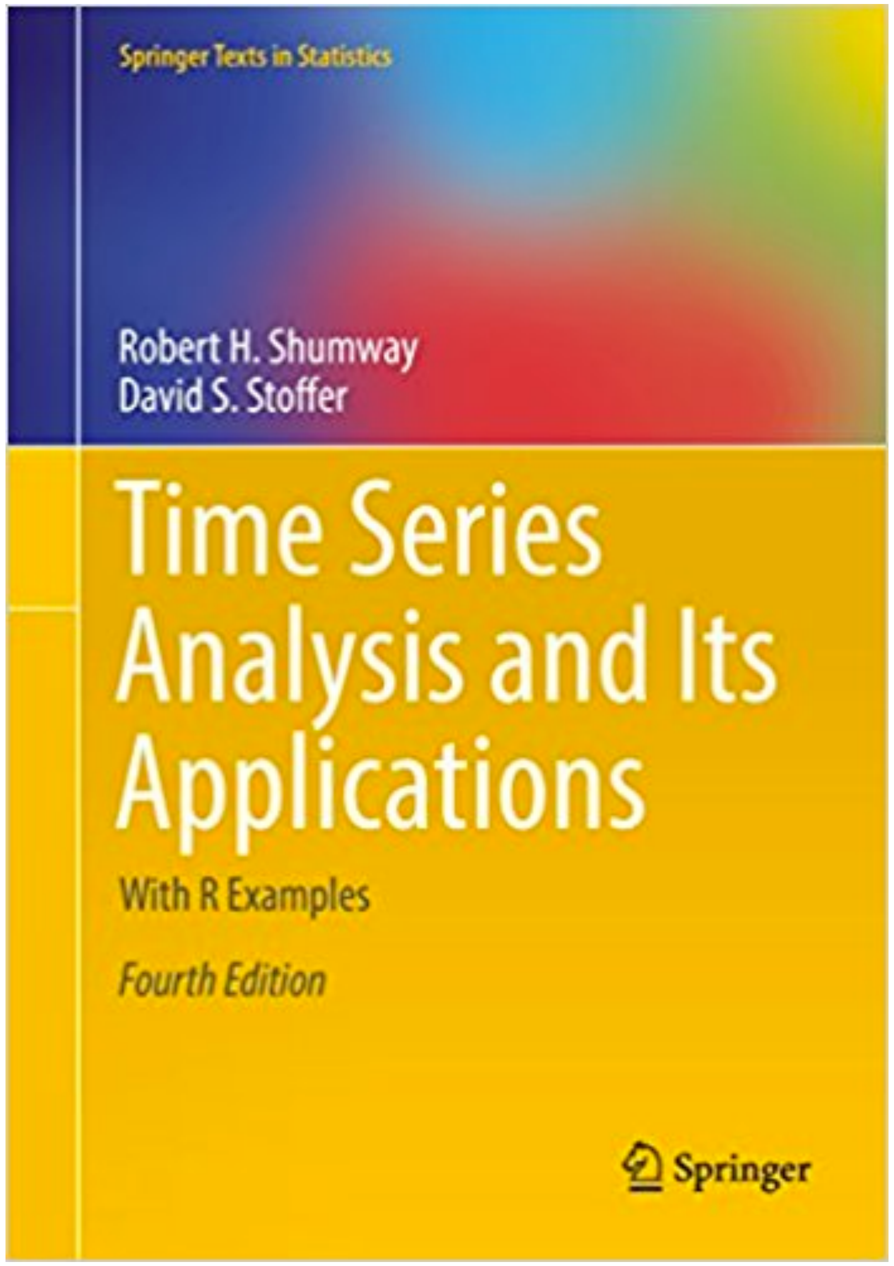
A process is **ARIMA** (p,**d**,q) if **d**-th difference

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

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

These are tricky to use, motivate state space models

Chapter 1 & 3



Homework 1: posted tonight, due Sept.27